TABLE OF READINGS

I. Core Readings

No.	Title	Author(s)	Organization / Source	Date	Notes
1	*Who Benefits, Who Gets Left Behind, and What We Do About It?*		California Emerging Technology Fund	—	AI Equity Primer
1.1	*An AI Equity Primer*		California Emerging Technology Fund	—	Supporting reference
2	*CETF Audit Committee Policy Materials: Risks of AI for Nonprofit Organizations*		California Emerging Technology Fund, Audit Committee	April 2026	Ongoing guidance
	Co-Intelligence: How to Live and Work With AI	Ethan Mollick	—	April 2, 2024	Book
3	*Use of Artificial Intelligence in School2Home*		California Emerging Technology School2Home		
4	*California School Boards Association Policies Including Artificial Intelligence Content*		California School Boards Association	—	GAMUT Policy
5	*An AI Bubble Will Not Trigger a Financial Crisis*	Lenny Mendonca; Martin Neil Baily	Project Syndicate, Elite Plus Magazine	March 9, 2026	Opinion and analysis
6	*You Already Have Everything You Need: The CENIC Artificial Intelligence Resource*		Corporation for Education Network Initiatives in California	—	Artificial Intelligence Resource
7	*AI-Enabled Fiber-Optic Geodesy: CENIC Networking and Services Enable Cutting-Edge Seismology*		Corporation for Education Network Initiatives in California	February 4, 2026	Case study
8	*CENIC and San Diego Supercomputer Center Create Sustainable Agriculture for California's Future*		Corporation for Education Network Initiatives in California	December 4, 2025	Case study
9	*CENIC Members Enable Data-Driven Agriculture: Optimizing Harvests in a Changing Environment*		Corporation for Education Network Initiatives in California	September 24, 2025	Case study
10	*CENIC and Iron Horse Vineyards Celebrate One-Year Anniversary of Innovation*		Corporation for Education Network Initiatives in California	December 9, 2025	Case study
11	*Data Center Boom Risks the Health of Already Vulnerable Communities*	Cecilia Marrinan	Tech Policy Press	June 12, 2025	Policy analysis

II. Additional Readings

	The AI Risk We Need to Focus On	Lenny Mendonca; Martin Neil Baily	Project Syndicate, Elite Plus Magazine	April 24, 2026	
	FSB Examines Al Adoption Across the Financial Sector	Norton Rose Fulbright	Financial Stability Board - **LEXOLOGY**	October 17 2025	
	The California Report on Frontier AI Policy	Jennifer Tour Chayes, University of California, Berkeley Mariano-Florentino Cuéllar, Carnegie Endowment for International Peace Li Fei-Fei, Stanford University	The California Report on Frontier AI Policy	June 17,2025	
	Practice What You Teach: Academic Integrity and AI	*Randy Kolset* *Mike Lawrence*	EdTech Digest	October 10, 2025	
	Maximize AI ROI: Invest In The People Who Use It	*Maria Flynn*	Forbes q	March 27, 2026	
	Data Centers and California *Electricity Policy*	Milton Marks Commission on California State Government Organization and Economy	Little Hoover Commission LHC	March 2026	

III. Background Documents

No.	Title	Author(s)	Organization / Source	Date
B-1	*Harnessing Artificial Intelligence for the Common Good*	Lenny Mendonca	Project Syndicate, Elite Plus Magazine	July 8, 2024
B-2	*Sovereign Artificial Intelligence: Building Ecosystems for Strategic Resilience and Impact*	Ali Ustun; Arnaud Tournesac; Luca Bennici; Justin De Niese; Newfel Drahmoune; Ruben Schaubroeck; Kaavini Takkar; Melanie Krawina	McKinsey and Company	March 2026
B-3	*Stanford Artificial Intelligence Experts Predict What Will Happen in 2026*	Shana Lynch	Stanford Institute for Human-Centered Artificial Intelligence	December 15, 2025
B-4	*Artificial Intelligence and Education Summit Addresses Learning and Equity*	Shana Lynch	Stanford Institute for Human-Centered Artificial Intelligence	February 19, 2026
B-5	*Stanford Scholars Train Artificial Intelligence to Better Augment Human Creativity*	Nikki Goth Itoi	Stanford Human-Centered Artificial Intelligence	March 10, 2026
B-6	*As Artificial Intelligence Reshapes Work, What Should Workers Do Next?*		Stanford Institute for Economic Policy Research	March 13, 2026
B-7	*The Telco Reinvention: How Artificial Intelligence Can Fuel Value Creation*	Luca Fiandro; Nemanja Vucevic; Lorraine Salazar; Shu Chern Lim	McKinsey and Company	March 2026

ARTIFICIAL INTELLIGENCE EQUITY – PANEL DISCUSSION
Who Benefits, Who Gets Left Behind, and What Should Be Done

I. Digital Divide 2.0: Beyond Internet Access

- Access to Artificial Intelligence (AI) tools, digital literacy, and computing resources
- Unequal availability of premium AI tools compared with free or limited versions
- Risk of underserved communities becoming consumers rather than creators of AI

Question: *Are existing inequalities being perpetuated at a higher technological level?*

II. Data Bias and Representation

- AI systems trained on data that underrepresents specific populations
- Impacts on healthcare, hiring, lending, and law enforcement
- Feedback loops that reinforce structural inequality

Example: Executive recruitment and selection criteria.

III. Automation and Job Displacement

- Disproportionate impacts on low-income workers
- Impacts on gig economy workers
- Displacement risks for administrative and service-sector roles
- Limited access to reskilling and workforce transition infrastructure in underserved communities

Question: *Will AI widen the economic gap more rapidly than public policy can respond?*

IV. Education Gaps in the Age of AI

- Unequal access to AI-enhanced learning tools
- Schools in underserved communities lagging in AI integration
- Limited training, resources, and access for teachers

Question: *How large will the future opportunity gap be between schools with AI tools and those without?*

V. Language, Culture, and Inclusion

- AI performance concentrated in dominant languages, particularly English
- Underrepresentation of languages, dialects, cultural context, and minority perspectives
- Risk of cultural homogenization in AI systems

Question: *Who determines the default worldview embedded in AI systems?*

VI. Healthcare Inequities and Artificial Intelligence

- AI-driven diagnostics trained on non-diverse populations
- Unequal access to AI-enabled healthcare tools, including telehealth and predictive analytics
- Risk of widening disparities in health outcomes

VII. Financial Systems and Algorithmic Decision-Making

- Use of AI in credit scoring, insurance underwriting, and fraud detection
- Potential for algorithmic redlining
- Limited transparency in automated financial decisions

VIII. Governance, Regulation, and Accountability

- Accountability when AI systems harm marginalized communities
- Inclusivity and accessibility of regulatory frameworks and rulemaking processes
- Role of community participation in AI policy development

IX. Access to AI Infrastructure and Capital

- Concentration of AI startups and innovation in high-wealth regions
- Limited funding for founders from underserved communities
- Barriers to entry, including computing resources, talent, and professional networks

X. Trust, Adoption, and Historical Context

- Community skepticism rooted in histories of surveillance and discrimination
- Risk that non-adoption leads to deeper exclusion

XI. Solutions-Focused Discussion

- Public-private partnerships to expand equitable AI access
- AI literacy and workforce programs in underserved communities
- Development of inclusive datasets and participatory design practices

Discussion Questions

- *What is one way AI is widening inequality today that is often overlooked?*
- *What does equitable AI look like in practice?*
- *Who should be accountable when AI harms underserved communities?*
- *If no action is taken, how wide will the gap be in five to ten years?*

Artificial intelligence offers nonprofits powerful ways to streamline operations, expand services, and improve decision-making, but it also introduces meaningful risks that can affect data security, equity, legal compliance, and stakeholder trust. A clear understanding of these risks—and a structured plan to mitigate them—helps nonprofits adopt AI in ways that strengthen, rather than compromise, their mission.

Risks of AI for Nonprofit Organizations

Data Security and Privacy

Nonprofits increasingly use AI tools for donor management, communications, and program operations. Many low-cost or open AI tools lack strong security controls, creating vulnerabilities that can expose sensitive donor, beneficiary, or volunteer information. Open-source AI models may also reuse any data entered into them, creating long-term privacy risks. [11]

Bias and Inequitable Outcomes

AI systems trained on incomplete or skewed data can reinforce existing inequities. For example, AI used to screen grant applications may inadvertently favor organizations with prior funding histories, disadvantaging new or underrepresented groups. Without human oversight, these biases can become embedded in decision-making processes. [1]

Ethical and Reputational Risk

Stakeholders expect nonprofits to use AI responsibly. Misuse—such as relying on AI for decisions that affect vulnerable populations, or failing to disclose AI use—can erode trust among donors, beneficiaries, and staff. Ethical concerns also arise when AI is used in sensitive contexts, such as summarizing confidential discussions or generating public-facing content. [1]

Vendor Sprawl and Lack of Oversight

Staff often adopt AI tools informally to solve immediate problems. Without centralized oversight, organizations may lose track of which tools are in use, what data they collect, and whether they comply with privacy laws. This increases the risk of data leaks and inconsistent practices across teams. [1]

Legal Risks: Intellectual Property, Privacy, and Liability

AI-generated content may not be eligible for copyright protection, and nonprofits may unknowingly publish content that infringes on third-party rights. Inputting personally identifiable information (PII) into AI tools can violate privacy laws, and using AI in ways that produce discriminatory outcomes may expose organizations to legal claims. [2]

Steps Nonprofits Can Take to Mitigate AI Risks

Strengthen Data Governance

- Establish clear policies on what data can be entered into AI tools, especially regarding donor and beneficiary information.
- Vet third-party AI vendors for security practices, privacy compliance, and data-handling protocols.
- Maintain an up-to-date inventory of all AI tools used across the organization. [1]

Implement Bias Prevention and Monitoring

- Regularly audit datasets for missing, outdated, or skewed information.
- Use diverse and representative data sources when training or configuring AI tools.
- Require human review for any AI-assisted decisions affecting funding, services, or eligibility. [1]

Build Ethical and Transparent AI Practices

- Align AI use with the organization's mission and values.
- Communicate openly with stakeholders about how AI is used and how data is protected.
- Create internal guidelines that define acceptable and unacceptable uses of AI. [1]

Reduce Vendor Sprawl and Centralize Oversight

- Create a formal approval process for adopting new AI tools.
- Consolidate tools where possible to reduce complexity and risk.
- Train staff and volunteers regularly on responsible AI use.
- [1]

Address Legal and Intellectual Property Risks

- Require human authors to meaningfully revise and verify any AI-generated content before publication.
- Update contributor agreements to include warranties about originality and non-infringement.
- Prohibit staff from inputting PII, confidential, or privileged information into AI tools unless strict compliance measures are in place. [2]

Use Established Frameworks for AI Risk Management

- Adopt structured frameworks such as the NIST AI Risk Management Framework to guide responsible design, deployment, and oversight of AI systems. [3]

Bringing It All Together

Nonprofits can safely harness AI's benefits by pairing innovation with strong governance. The most effective approach combines clear policies, ongoing staff training, transparent communication, and continuous monitoring of AI systems. This ensures AI strengthens the organization's mission while protecting the people and communities it serves.

References (3)

1*Top AI Risks Facing the Nonprofit Sector | BDO*. https://www.bdo.com/insights/industries/nonprofit-education/the-top-ai-risks-in-the-nonprofit-sector

2*Nonprofits and AI: Managing Legal and Other Risks (Full Version)*. https://www.tenenbaumlegal.com/articles/nonprofits-and-ai-managing-legal-and-other-risks/

3*AI Risk Management Framework | NIST*. https://www.nist.gov/itl/ai-risk-management-framework

Project Syndicate

An AI Bubble Won't Trigger a Financial Crisis

Mar 9, 2026 ***LENNY MENDONCA** and **MARTIN NEIL BAILY***

HALF MOON BAY, CALIFORNIA – A familiar anxiety has returned to financial markets. Amazon is devoting $100 billion to data centers. Meta has committed more than $600 billion to building them over three years. Microsoft, Google, and Apple plan to spend hundreds of billions more. With AI investments running into the trillions, are we witnessing a bubble, and what will happen if it bursts?
We have been here before. One of us (Martin) served as chairman of the White House Council of Economic Advisers during the late 1990s technology boom, and the other watched California navigate the dot-com collapse and the 2008 financial crisis from various government and consulting roles. Our experience tells us that even if an AI correction comes, it will not trigger the kind of financial crisis that devastated the economy in 2008-09.

The reason is that the structure of AI investment is fundamentally different from what we saw in these previous episodes. The dot-com collapse is remembered for dramatic losses – the Nasdaq fell 77% from its March 2000 peak – and spectacular flameouts like Pets.com. But the economic story was more nuanced.

Yes, there was overinvestment in internet infrastructure, and many big bets proved premature. But that "overinvestment" left behind extraordinarily valuable infrastructure. The fiber-optic cables laid during the boom enabled the broadband economy. Server capacity that seemed excessive in 2001 became essential by 2005. Productivity growth remained surprisingly strong as the bubble burst, proving that the underlying technology was transformative even if the initial business models were not.

Crucially, the dot-com bust never threatened the broader financial system. The speculation was funded primarily through equity markets, not debt. When valuations collapsed, stockholders lost money, but banks remained solvent. There was no flood of defaults, no credit freeze, and no need for massive government bailouts.

By contrast, the 2008 crisis was about how assets were financed. Subprime mortgages were packaged into securities, sold to pension funds and international investors, and used as collateral for more borrowing. The entire global financial system became a house of cards built on rising home prices.
When those prices fell, the complex web of obligations unraveled: banks discovered they held worthless securities, credit markets froze, and what started as falling home prices cascaded into a global financial crisis requiring unprecedented government intervention. The problem was not just overvaluation. It was a combination of leverage, complexity, opacity, and the bleeding of housing risk throughout the entire financial system.

The AI boom is different. The trillions being spent on AI infrastructure are coming primarily from the balance sheets of the world's most profitable and cash-rich companies. Apple, Microsoft, Google, Amazon, and Meta are not borrowing to build their AI capacity. When Apple commits $500 billion over four years to AI infrastructure, it is not issuing debt that must be serviced regardless of revenue. These are equity-financed investments by one of the most valuable companies in history. If the returns are dismaying, Apple's stock price might suffer, but the company will neither default nor trigger a cascade of failures through the financial system.

Similarly, the specialized nature of AI infrastructure limits contagion risk. Data centers are not being securitized and sold to pension funds. AI chips are not being used as collateral for derivatives. Of course, there could be a sharp correction in tech stock valuations, and companies could scale back AI investments if returns disappoint. But would any of this trigger a financial crisis comparable to 2008? The structural differences suggest not.

The point is not that AI poses no economic risks. It is that the risks are not primarily to financial stability. AI outcomes that are worth worrying about include: failure to deliver significant productivity gains, leading to slower growth and disappointed expectations; rapid displacement of workers, which would require robust policies to support education and training; concentration of economic power, raising antitrust questions; and strains on electrical grids and construction capacity, which could curtail growth and increase households' electricity bills.

Even if AI valuations are inflated, and current revenue projections prove wildly optimistic, the infrastructure being built has lasting value. Those data centers aren't going away. The fiber connections being laid are permanent additions to our digital infrastructure and the trained AI engineers represent human capital that will generate value for decades.

If someone told you in 1999 that America would "overspend" hundreds of billions of dollars on internet infrastructure, you might have worried about the bubble. But that overinvestment enabled the entire digital economy – YouTube, Netflix, cloud computing, and remote work. The societal return was large, even though many companies failed and investors lost money.

Understanding that AI speculation is unlikely to trigger a financial crisis means that policymakers should get their priorities right. Rather than over-regulating to prevent a financial crisis that is not coming, they should focus on risks to the real economy: worker displacement, algorithmic bias, data privacy, and competitive dynamics. They should promote investments in complementary infrastructure such as electrical capacity, cooling systems, and trained workers. And they should prepare for labor-market adjustments by bolstering unemployment insurance, retraining programs, and portable benefits.

Moreover, they should watch for leverage creep, because the one risk that could change our analysis is if AI assets do become heavily leveraged or securitized. Financial regulators should monitor banks and pension funds accordingly.

Transformative innovations typically take decades to diffuse fully through the economy. It took 40 years for electricity to revolutionize manufacturing; 20 years for the internet to reshape retail. If AI is similarly transformative, today's speculative investments could prove far-sighted.

The 2008 financial crisis taught policymakers to be vigilant about financial-stability risks. But there are important differences between asset classes and financing structures. The AI boom may be speculative. It may be excessive. It may even be a bubble. But it is not a threat to the financial system.

LENNY MENDONCA

Lenny Mendonca is Senior Partner Emeritus at McKinsey & Company and former Chief Economic and Business Adviser to the State of California.

MARTIN NEIL BAILY

Martin Neil Baily is Senior Fellow Emeritus in Economic Studies at the Brookings Institution and was Chairman of President Bill Clinton's Council of Economic Advisers from 1999 to 2001.
https://prosyn.org/16uJLjd

Use of Artificial Intelligence in School2Home

School2Home has incorporated Artificial Intelligence (AI) into its work with School Leadership Teams (SLTs) consisting of school administrators, teachers, and parent coordinators. During the School2Home quarterly Regional Learning Community meetings and the yearly Leadership Academy, members of each team participate in activities that introduce them to available AI resources and guidelines published by trusted organizations such as school districts, County Offices of Education, and nonprofit organizations experienced in this area. SLTs discuss the benefits and potential risks associated with the use of AI on their campus and develop a plan for successfully completing certification and implementing these resources at their schools.

At the 2025 School2Home Leadership Academy, Common Sense Education conducted a session updating SLTs about their legislative efforts related to increasing the safety measures for children around AI. In addition, the Los Angeles County Office of Education (LACOE) conducted an engaging session about how the Google Gem Assistant can help automate routine instructional tasks for teachers. Additionally, a special session was conducted for principals and teachers to explore a variety of AI platforms designed for educators and discuss the potential uses of each platform. They were also given the opportunity to begin the certification process in the platform they felt would be most impactful in their classroom. The School2Home team will continue to provide time during each convening to have teachers highlight instances where AI helped them address specific needs in their classrooms and facilitate the implementation of proven instructional practices.

Finally, the School2Home Professional Learning component offers 2 modules focused specifically on AI. During these sessions, teachers discuss integrating AI into their workflow to save time, plan effective lessons, improve student outcomes, and strengthen school-home communication. Teachers also work together to design meaningful and responsible learning experiences that leverage AI to foster student collaboration and deeper thinking. School2Home will continue to identify opportunities to help educators integrate technology in a manner that supports their efforts to foster student academic success. In addition, School2Home will help inform parents and guardians on how they can monitor and support their child's use of AI in a manner that is safe and effective.

CSBA Policies that include AI content:

BP 0440 - District Technology Plan

This Board Policy provides a plan for the use of technological resources. In June 2025, it was updated to revise the philosophical paragraph to acknowledge, in addition to the benefits gained from technological resources, potential negative consequences. Additionally, the Board Policy was updated to add material related to the alignment of technological resources with other district goals, objectives, and academic standards, the safe use of technological resources, the use of technology in accordance with district policy, and the use of artificial intelligence, as applicable. In addition, the Board Policy was updated to reflect that state law and the California Department of Education no longer require districts to have a technology plan, and to maintain the requirement for the Superintendent to develop and regularly propose revisions to a technology plan. The Board Policy was also updated to list the components to be included in the district technology plan.

Board Policy 0441 – Artificial Intelligence

This Board Policy, which was newly created in June 2025, provides principles for the district regarding the use of artificial intelligence (AI) by students and staff. The nine principles are that the use of AI (1) is student-centered, (2) is staff-centered, (3) is ethical and transparent, (4) ensures accountability and responsibility, (5) ensures equity and access, (6) is secure and private, (7) includes ongoing professional development, (8) includes community engagement, and (9) is regularly evaluated for continuous improvement. Additionally, the Board Policy requires the Superintendent to ensure that the use of AI is consistent with district policy.

BP/E 4040 - Employee Use of Technology

This Board Policy addresses how employees may use and not use technology. In September 2024, it was updated to further specify that the use of artificial intelligence (AI) is generally permitted, with certain caveats, including that the employee reviews and is responsible for any final product or document and does not share confidential student records with a third party unless permitted by law. The Board Policy was also updated to recognize that technology, including AI, can enhance employee performance, to include “software as a service” and “AI apps” in the definition of “district technology,” and to provide for professional development in the appropriate use of technology, including AI. The accompanying Exhibit, which is a sample Technology Use Agreement for employees, was also updated in September 2024 to align with the prohibited and permitted uses of technology in the Board Policy, to include the revised definition of “district technology,” and to note that the district may monitor and record how employees use artificial intelligence.

BP/E 6163.4 - Student Use of Technology

This Board Policy addresses how students may use and not use technology. In September 2024, it was updated to further specify that the use of artificial intelligence (AI) is permitted as long as it is used consistently with district policies including policies on academic honesty. The

Board Policy was also updated to include “software as a service” and “AI apps” in the definition of "district technology" as well as to recognize that it is important to promote digital citizenship, to provide students with access to the latest digital tools (including AI), and to align technology with district goals, academic standards, and instructional materials. The accompanying Exhibit, which is a sample Technology Use Agreement for students, was also updated in September 2024 to align with the prohibited and permitted uses of technology in the Board Policy, to include the revised definition of “district technology,” and to note that the district may monitor and record how students use artificial intelligence.

BP 5131.9 – Academic Honesty

This Board Policy establishes the importance of academic honesty and personal integrity as fundamental to a student's education and character development and is the basis for the rule that students are subject to discipline if they cheat, lie, plagiarize, or commit other acts of academic dishonesty. In September 2023, this Board Policy was updated to detail the prohibited and permitted uses of technology (particularly AI) in general as well as in specific situations such as when a teacher may use technology to catch academic dishonestly by students.

BP 6154 – Homework and Makeup Work

This Board Policy establishes the basic framework for assigning homework, particularly expectations of students, families, and staff related to assignment, completion, and grading of homework. This Board Policy also covers makeup work and how homework should be assigned to students during a suspension. In September 2023, this Board Policy was updated to state that students may use technology to assist with homework and/or makeup work only as specified in Board Policy 5131.9 - Academic Honesty.

BP 6162.5 – Student Assessment

This Board Policy establishes how to determine which student assessment to administer, when to administer a student assessment, and how to use a student assessment. In September 2023, this Board Policy was updated to state that students may use technology to complete assessments only as specified in Board Policy 5131.9 - Academic Honesty.

Link to GAMUT Policy: https://simbli.eboardsolutions.com/index.aspx?S=36030855

You Already Have Everything You Need: The CENIC AI Resource

International competition requires the US to rapidly increase its AI workforce. While AI research and development (R&D) in the United States is advancing rapidly, opportunities to pursue cutting-edge AI research and new AI applications are often inaccessible to researchers. Machine Learning and Artificial Intelligence are opening new career paths for students at all levels who will become the next facilitators of academic research and private sector innovation, with the potential to transform society in positive and beneficial ways. To facilitate this transformation there is a need for a national-level "AI Resource" that is open-access, scalable, and grown through the contributions of its user community. The need for this infrastructure has already been articulated in two National Science Foundation reports and one report from the National Artificial Intelligence Research Resource (NAIRR) Task Force. In addition, Congress has been asked to develop an AI research resource.

CENIC AIR Helps the Engagement of California's Diverse Research & Education Communities

This resource, the infrastructure, and the communities that use it already exist in the United States in the form of the **National Research Platform (NRP)**, and the California portion called the **CENIC AI Resource (CENIC AIR)**. CENIC AIR provides California's research *and education* communities a platform to enable their faculty and students to contribute constructively to this transformation and collaborate extensively with colleagues nationwide over the same infrastructure. For California to continue its leadership in AI, faculty, and students need access to training, research, and class lab facilities in a persistent, sustained infrastructure that will continue to evolve well into the future.

Machine Learning, Artificial Intelligence, and Data Science research and education applications need access to GPUs and other advanced compute technologies that CENIC AIR will support. With the CENIC AIR Platform, researchers and educators can also easily access cloud offerings such as Amazon AWS, Google Cloud Platform, and Microsoft Azure, as well as NSF HPC centers, should they need to scale beyond the National Research Platform (NRP) set of resources and have the allocations, funds, or cloud credits to use HPC cloud services.

CENIC AIR: the National Research Platform in California

The NRP is a member-owned and led national cyberinfrastructure created by a partnership of more than 50 institutions nationwide. It is led by researchers and cyberinfrastructure professionals at UC San Diego, and supported in part by grant awards from the National Science Foundation. Its users are among its owners who voluntarily contribute compute and storage resources for the entire community which are interconnected and accessed via regional and national research and education networks such as CENIC.

These software and hardware resources comprise a hypercluster known as Nautilus, designed to run big data applications using Kubernetes-based containerization for instantiation, management, and scaling.

While this ecosystem and its member community extend across the United States, the California portion, CENIC AIR, consists of high-performance compute and storage infrastructure voluntarily contributed to by members of the CENIC community (shown above). While distributed Nautilus platforms are connected into a distributed global cyberinfrastructure via regional and national (and international) research and education networks, the California portion runs on CENIC's California Research and Education Network (CalREN).

CENIC AIR and the NRP provide much more than just infrastructure. The strong, accomplished research communities behind them offer invaluable mentorship and collaboration, including live networking conferences and events. This collaborative potential is even more powerful in California, where CENIC's R&E community benefits from the cross-segmental 25-year head start by working together on network technology and its diverse applications.

CENIC AIR, connected to the NRP's Nautilus infrastructure, is a scalable and robust ecosystem designed for, and responsive to, data science research needs as well as educational programs and positioned for much broader use by the CENIC community.

To that end, CENIC AIR would like to expand the user base by actively adding campuses and increasing faculty participation.

Providing AI Resources to a Broader Data Science Community

Given the importance of national access to cyberinfrastructure for research *and education*, CENIC AIR seeks to reach beyond the largest and most well-provisioned research institutions. For example, the NRP community has grown during the last 5 years from 10 to 135 campuses, including campuses in 21 of the 28 states and territories identified as part of the NSF's Established Program to Stimulate Competitive Research (EPSCoR) and with 24 of them being Minority Serving Institutions (MSIs). Most California State Universities are MSIs,

Furthermore, many MSIs are not primarily research institutions, and do not operate a "Science DMZ." A Science DMZ is a feature of the campus border router enabling a research-dedicated local area network that allows very large volumes of data traffic to bypass the common campus gateway firewall and primary campus network, thereby preventing the broader campus network from becoming overwhelmed by research traffic. Thanks to their collaborative user communities, CENIC and the NRP are ready to help such institutions architect, deploy, and support a Science DMZ, and to offer guidance on funding proposals for NSF CC* awards that can help campuses acquire the necessary on-premises AI compute equipment and campus networking improvements.

In addition, CENIC AIR partners can offer research project/mentor matching to help under-resourced campuses get connected.

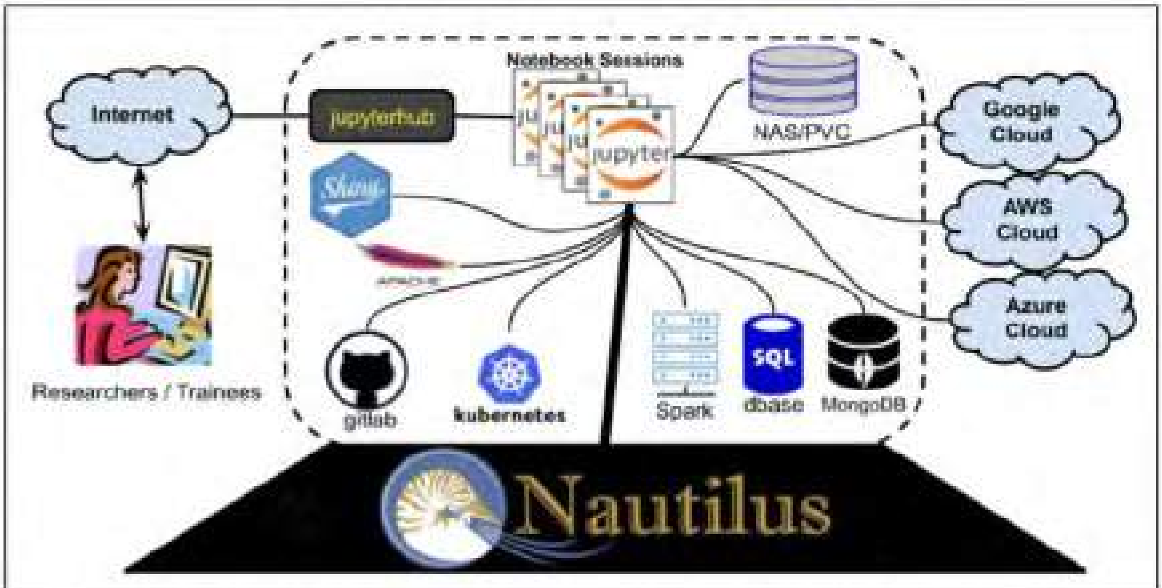
Broadening Inclusion in California: The NRP & CENIC AIR is Increasingly Being Adopted by the California State University System

One example of the expansion of the original NRP user community is the California State University system, which offers several joint PhD programs, but primarily Master's and undergraduate programs to its 430,000 students. In the year ahead, more CSU campuses are hoping to join CENIC AIR; with assistance from CENIC in network engineering, design, and implementation of Science DMZs for engaged California campuses.

Currently, CSU campuses with researchers using CENIC AIR include CSU Northridge, CSU San Bernardino, San Diego State University, San Jose State University, and Cal Poly Humboldt. CSU Sacramento and CSU San Marcos have recently expressed interest.

Any faculty member with a CILogin can get access to CENIC AIR. The first step for them is to create or join a project, which Kubernetes calls *namespaces*. They launch their containerized application on CENIC AIR, where Kubernetes orchestrates the movement of their containerized application to the appropriate hardware for execution on the infrastructure. Thousands of researchers, students, and staff have successfully done this, enabling many of them to train models on datasets and then apply the models to real-world data.

To get started without containerizing, a researcher/student can try out Nautilus by utilizing *JupyterLab* notebooks (specialized laptop-friendly software developed specifically for instructional purposes).



Customized containers support specialized software stacks. Dedicated JuypterLabs in JupyterHubs help guide researchers to proper resource utilization. Graphic by Grant Scott.

Researchers and their students then scale their activities from a single JupyterLab notebook to containerized applications with namespaces. This step up allows them to pursue research at a scale up to hundreds of GPU hours and thousands of CPU core hours operating on terabytes of data – and collaborate with any other institution using Nautilus' shared storage.

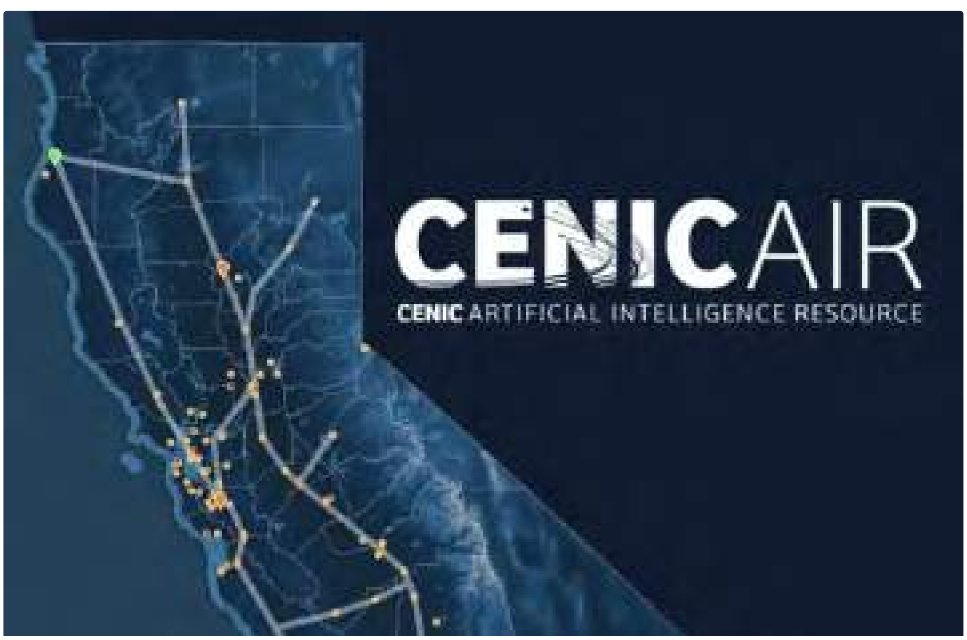
Can a Campus Connect Equipment to CENIC AIR?

Should a campus choose to add resources to CENIC AIR and the NRP, it does so via the "bring your own resources" (BYOR) program. Campus IT staff can add GPU, CPU, FPGA, storage, or other networked equipment like Internet of Things (IoT) devices. CENIC Engineering and NRP's technical staff are ready to assist campuses with BYOR and setting up Science DMZs.

FAQ >

SDSC at UCSD and CENIC Expand AI Infrastructure in California

SDSC at UC San Diego and CENIC have developed a prescient AI education infrastructure called CENIC AI Resource (CENIC AIR) for K-12 education, public libraries, community colleges, the California State University (CSU) and the University of California (UC) campus systems.
Read more >



CENIC AIR Champions

Senior leadership, IT directors, and research and instruction facilitators and faculty at early adopter institutions sharing their invaluable advice for anyone seeking to discover how their institution can participate in and benefit from CENIC AIR.

Learn more

Recent blog posts

AI-Enabled Fiber-Optic Geodesy: CENIC Networking and Services Enable Cutting-Edge Seismology

Pacific Wave Signs MoU with New Japan-Based Research and Education Exchange Facility FUJIXP

CENIC and San Diego Supercomputer Center Create Sustainable Agriculture for California's Future

FEB 4, 2026

AI-Enabled Fiber-Optic Geodesy: CENIC Networking and Services Enable Cutting-Edge Seismology

Categories AI/Machine Learning Independent Universities Technology & Innovation

Tags CENIC AIR earthquake

CENIC's biennial conference "The Right Connection" is fast approaching, and with it the Innovations in Networking Awards. These awards recognize CENIC members and others who have made extraordinary use of the bespoke high-performance networking and services that CENIC provides.

It's worth noting, however, that these projects continue to develop in exciting ways after receiving the awards. One such project honored with the Innovations in Networking Award for Experimental Applications at our previous conference is the Caltech Zhan Group of Observational Seismology's research on the use of fiber-optic cables as seismic sensors—and they have not been idle.

Thanks to the efforts of the Zhan Group and others, microscopic imperfections in fiber-optic cables—including subsea cables—have been transformed into millions of networked individual seismometers needing scientific equipment only at the endpoints, and that can sense tiny movements of the Earth's crust. This gives scientists a detailed real-time view of ground, ocean floor, and magma movement, revealing fault structures and volcanic activity.

However, this sudden and massive increase in the number of seismic sensors has resulted in an equally massive surge in data generated, making the development of artificial intelligence (AI)-enabled tools a high priority for seismologists around the world. The Zhan Group's Zhongwen Zhan has insightful perspectives about meeting this need, where the most progress is being made using these tools, and what the future holds for seismology.

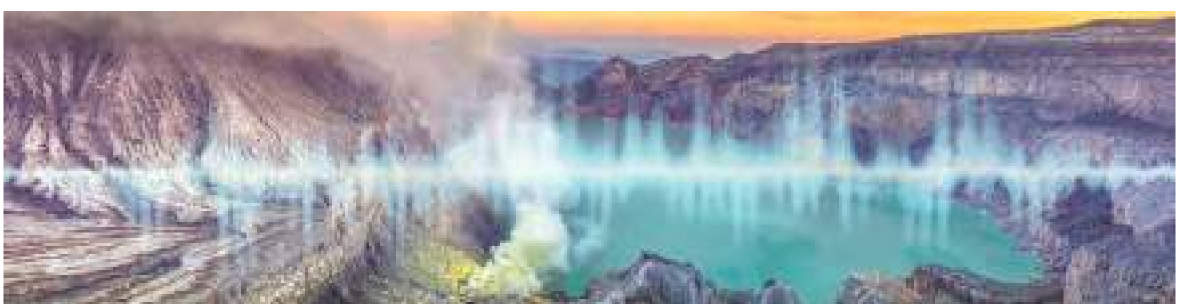
Zhongwen Zhan, Caltech Professor of Geophysics; Clarence R. Allen Leadership Chair, Seismological Laboratory; Director, Seismological Laboratory

When Everything Is a Sensor, Creativity Becomes the Bottleneck

In 2010, US Air Force Lt. Gen. David Deptula made his prescient comment about “swimming in sensors and drowning in data.” However, even Deptula likely never dreamed that 16 years later, nearly anything might potentially serve as a sensor. Ambient sounds, including conversations, can now be reconstructed from the imperceptible vibrations of windows and nearby objects. Unseen occupants in a room can be partially imaged by their reflections on surfaces once thought to be non-reflective, like doors and walls.

The discovery that microscopic imperfections in fiber-optic cables had potential as movement sensors—registering everything from seismic and volcanic activity to thunder and passing traffic—has increased the number of deployed movement sensors from thousands to tens of millions almost overnight, including under the oceans, with practically no end in sight. As Zhongwen Zhan states, “We hope to do this very thoroughly with CENIC’s California Research and Education Network (CalREN) and the world’s other networks and fiber-optic cables, but that will generate tens of petabytes of data, all of which will need to be processed and managed before discovery can take place.”

This astonishing new abundance of sensors and the data they generate makes human creativity the most precious resource—and thus the narrowest bottleneck. Thankfully, AI tools can help relieve this constraint. As Zhan puts it, “Creative scientists are rare. That means for now, one of the most useful applications of AI tools for seismology is initial data processing, which frees up scientists’ most scarce resource: their time.”



Finding Patterns in Petabytes: Seeing New Forests for Trillions of Trees

The sudden increase in seismic sensors presents yet another challenge: discerning whole new forests among orders of magnitude more trees. Many paradigm-shifting discoveries are likely lying dormant in the patterns hidden within both old and new data, simply because researchers have not yet performed the millions of needed comparisons. Without a prior reason, it can be difficult to justify the research and computing time needed to find unanticipated revolutionary patterns.

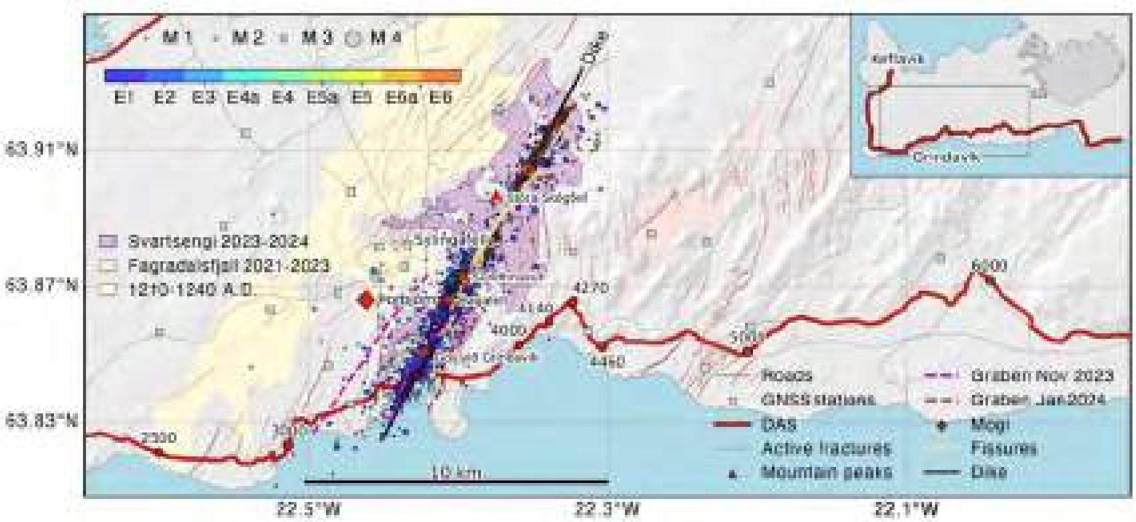
“AI’s not the only tool for this,” Zhan observes, “but it’s a good one. We can use it to find new patterns, some of which might translate into useful applications that we can’t currently imagine. AI is a great way to find things you don’t even know you should be looking for.”

There is also the "More Is Different" effect, first named by physicist Philip Anderson in his 1972 article in "Science." Exponentially more sensors don't just increase the volume of data and analysis required; finer levels of detail can introduce entirely new kinds of analysis. Vastly more seismic sensors provide a view into earth movements, faults, ruptures, subsurface magma, and more at an unprecedented level of detail, which will spur new questions and entirely new areas of research. Seismology itself will be transformed, and completely new disciplines will emerge. The Zhan Group is already developing AI-enabled tools specifically designed to process data at finer and finer levels of detail, with an eye to creating those new disciplines.

Recent Results: Successfully Predicting Volcanic Eruptions in Iceland

This approach is already bearing valuable fruit, as demonstrated in a 2025 paper titled "Minute-scale dynamics of recurrent dike intrusions in Iceland with fiber-optic geodesy." (Participating institutions include CENIC members Caltech—with members of the Zhan Group shown in Iceland at right—and the Jet Propulsion Laboratory, the University of Iceland at Reykjavik, the University of Houston, Google, and others.)

By using fiber-optic cables as movement sensors—a technique known as fiber-optic geodesy—the researchers were able to detect the movement of underground magma 30 minutes to an hour before it erupted onto the surface. While earthquake prediction remains elusive, the application of fiber-optic geodesy combined with AI-enabled tools is extremely promising for the prediction of volcanic eruptions. As Zhan notes, "When it comes to volcano prediction, we know that if we have enough data, we can do it; we saw the signal on the fiber. And earlier warnings can give people in volcanically active areas valuable time to evacuate themselves and their families."



A location map of various sensors located near Grindavík, Iceland including the fiber-optic cable shown in red.

These sensors were used by a multi-institution team to generate the data in the 2025 research paper linked above.

Cables, Data Storage and Processing, Access: CENIC Services Speed Progress

It's almost uncanny how perfectly the tools required by the new discipline of fiber-optic geodesy align with the benefits the CENIC provides to its member researchers. First, the fiber-optic cables of the CalREN backbone itself contain millions of connected sensors in a high-performance network. Seismic and volcanic data must be transmitted in real time for early warning systems to be effective. The ubiquity of data access enabled by advanced networking has allowed seismology to grow from a relatively small field to one where any institution with a geosciences program now employs at least one seismologist.

For example, despite the Zhan Group being a ten-person team, Zhan "cannot imagine we've extracted everything that can be gotten out of our data sets." He adds, "Once we're at the scale we're aiming for, researchers will be able to examine our data and generate value we don't imagine now." He anticipates collaboration with the National Data Platform (NDP) in the future, enabling bleeding-edge seismic research for any CENIC member participating in the CENIC AI Resource (CENIC AIR).

"We very much want to set an example for other researchers by creating models for how to achieve this. It's a classic government/industry/academia collaboration that benefits everyone."

Second, CENIC AIR provides researchers and faculty at all participating institutions with the data storage and processing power needed to create, store, and share data and AI- and machine-learning-enabled tools. This aligns with Zhan's own observation that, "In the future, cooperative, large-scale AI platforms will be indispensable."

Lastly, CENIC's global partnerships with other research and education networks—such as initiatives like the National Research Platform (NRP), Pacific Wave, and others—enable the global collaborations that seismology requires. "That's the reason why seismology is such an open community," Zhan stated with a smile. "Quakes don't respect boundaries."

If you'd like to learn more about how researchers and faculty at CENIC member institutions are making use of CENIC's networking and services, consider registering to attend our biennial conference "The Right Connection" from March 29 to April 1, 2026, in Monterey, CA. While there, you'll enjoy a program created by and for members of the CENIC community, casual human networking, demos, workshops, and the presentation of the 2026 Innovations in Networking Awards. Visit the conference website to view the program, register, and make your hotel reservations.

RELATED BLOG POSTS

CENIC and San Diego Supercomputer Center Create Sustainable Agriculture for California's Future

CENIC Members Enable Data-Driven Agriculture: Optimizing Harvests in a Changing Environment

Recent news

CENIC 2026 Innovations in Networking Award for Network Partner to University of Hawai'i's David Lassner and Team

Frank Wuerthwein & UC San Diego Team Recognized with 2026 CENIC AIR Award for Innovative

DEC 4, 2025

CENIC and San Diego Supercomputer Center Create Sustainable Agriculture for California's Future

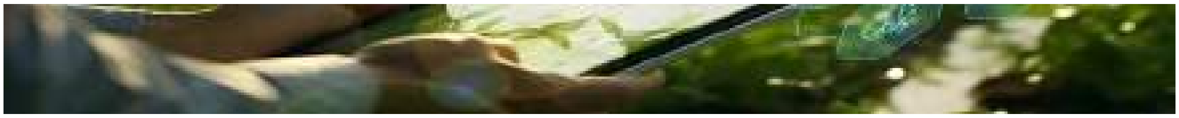
Categories AI/Machine Learning CCCs CSU University of California Technology & Innovation

Tags agriculture CENIC AIR

The networking and services provided by the Corporation for Education Network Initiatives in California (CENIC) support the same research and education communities that have consistently made California one of the largest economies on Earth and have provided the research and the workforce behind industries that have improved lives around the world.

Among those industries is agriculture—worth $60 billion in 2025 and providing food not only to the US but to the world. For example, almost half of all vegetables consumed in the US are grown in California. California also grows three-quarters of the fruits and nearly 100% of some very localized crops—including avocados, nuts, and wine grapes—that are consumed in the country, as well as a significant percentage of meat and dairy consumed worldwide. Decorative, ornamental crops are also a lucrative part of the state's agriculture industry.

However, if California agriculture is to keep growing in quality and yield, it must prepare new generations of farmers who will apply the latest technology to agriculture. CENIC's networking and services—including the CENIC AI Resource (CENIC AIR)—can be a vital part of this by turning the farm into an educational setting and making farming less physically taxing as a career. Highly local farms and their specific crops can also be linked to local educational institutions, ensuring that tech-related economic benefits accrue to rural regions and not just to a few large urban centers. At the same time, applying CENIC AIR resources can allow all types of agriculture to adapt to changing weather and climate conditions, especially as these relate to water use and availability—an extremely pressing concern as reported by many farmers across the state.



Transforming Farms to Attract, Prepare, and Retain New Talent

Sustainable network-enabled agriculture is a major passion of Frank Wuerthwein, director of the San Diego Supercomputer Center (SDSC) at the University of California San Diego School of Computing, Information and Data Sciences. While improving crop quality and yield is central to this goal, Wuerthwein's vision encompasses far more than that, such as how the revolution towards experiential education can turn farms into classrooms and better attract and prepare the next generation of agricultural talent as current experts begin to retire.

Additionally, Wuerthwein is aware of the potential of AI-enabled personalized education in agricultural training and the unique value of CENIC AIR, given its statewide penetration and the rural nature of most California agriculture.

> "Bringing digital assets into the classroom—or better, the classroom into the field—supports the era of learning-by-doing, which is a much more engaging way to introduce new knowledge," said Wuerthwein. "We've moved on from the era of mass education that was popular in the era of mass production or an assembly-line workforce. AI allows us to personalize education in a whole new way that can be exciting for the next generation in agriculture."

Readers of the CENIC blog will already be familiar with the Iron Horse Vineyards Testbed project, an innovative demonstration project of CENIC, its members, and its collaborative partners, which launched on the site of Iron Horse Vineyards near Sebastopol, California.

> Iron Horse Vineyards CEO Joy Sterling agrees that the value of technology lies in not just attracting talent but also retaining it by improving processes and management and thus the business of agriculture itself. As Sterling put it, "We've made real progress if we can reduce the number of times the people in charge of the vineyard have to go out in the middle of the night to fix something."

Thus, in addition to enabling a critical California industry, CENIC AIR can play a significant role in ensuring that the state's rural and remote areas aren't left behind in the AI revolution but in fact benefit from it in sustainable ways related to their local industries, enabling local talent to stay local.

Local Agriculture Partnering with Local Institutions

> Wuerthwein also noted a major feature of agriculture in California: how localized crop varieties are statewide. “California features so many specialized crops across a wide range, including wine, almonds, flowers, avocados, citrus, and more,” he stated. “And this great variety tends to be extremely local, with different regions specializing in different crops. This creates a very unique connection to each local community in all of these places.”

That local connection extends to local education and expertise as well, with the next-generation workforce coming from families with specific experience in local agriculture, local organizations such as area chapters of Future Farmers of America (FFA), and schools and colleges with research and education programs centered on local crops. CENIC AIR, with its broad reach and specific research and education mission, is a perfect means of connecting all of these organizations and ensuring a top-quality workforce for all types of agriculture.

> “When we make connections to ag tech or instrumentation in the fields and farms of California, and we bring data science to the college from the farm and vice versa, that facilitates the transformation of effective education,” said Wuerthwein. “That allows us to take the next step in transforming education from something that is performed in a classroom to something that is experienced in the field in a uniquely tailored way.”

As an example of this, Sonoma State University and Santa Rosa Junior College feature cutting-edge viticulture programs, and students and faculty from both are already participating in the nearby Iron Horse Vineyards testbed.

Adapting in Real Time to Changing Weather and Climate Conditions

If California agriculture is to keep growing in quality and yield, CENIC’s networking and services must also enable our member institutions to develop the innovations that the agriculture industry will use to squeeze every drop of potential out of its soil, surroundings, and resources. Wuerthwein has found through numerous discussions with agricultural experts that efficient water use and nurturing the next-generation workforce are their top concerns.

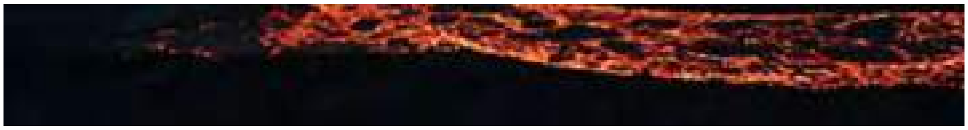
If water use is to be managed as efficiently as possible, along with other precious resources in an age of changing weather and climate, the AI and machine learning that CENIC AIR is designed to support provides the real-time monitoring and management needed by fully instrumented farms.

Another unique advantage offered by CENIC AIR is how it connects the California research and education community throughout the state, both urban and rural, to resources such as the National Research Platform and National Data Platform. Weather and climate operate on both small and very large scales, so the ability to access nationwide collaborators—and their processing power and data—makes it possible to develop models that can be used to manage water and other resources in ways never before possible, even for individual farms.

> "This means that the top challenges for keeping California's agriculture industries growing into the next century—labor and water—can both be addressed by exactly the networking and services made possible by CENIC AIR," Wuerthwein concluded. "And it can do so in a way that translates to better crops, better harvests, close and productive relationships among local institutions and industries, and a multigenerational pull to bring younger generations into farming."

You can visit the CENIC Blog to learn more about the revolutionary Iron Horse Vineyards Testbed project and how it benefits the agricultural research and education programs of the CENIC membership. You can also visit SDSC's website to learn about their cross-segmental research and education initiatives.

RELATED BLOG POSTS



AI-Enabled Fiber-Optic Geodesy: CENIC Networking and Services Enable Cutting-Edge Seismology

CENIC Members Enable Data-Driven Agriculture: Optimizing Harvests in a Changing Environment

Recent news

CENIC 2026 Innovations in Networking Award for Network Partner to University of Hawai'i's David Lassner and Team

Frank Wuerthwein & UC San Diego Team Recognized with 2026 CENIC AIR Award for Innovative Technical Design & Instructional Support

Caltech's Harvey Newman and His Team Recognized with 2026 CENIC Innovations in Network Award for Experimental Applications

Kendra Ard and Ed Clark of the California State University Chancellor's Chief Information Office Recognized with 2026 CENIC AIR Award for Engagement

CENIC and Iron Horse Vineyards Celebrating One-Year Anniversary Of Innovation

Recent blog posts

AI-Enabled Fiber-Optic Geodesy: CENIC Networking and Services Enable Cutting-Edge Seismology

Pacific Wave Signs MoU with New Japan-Based Research and Education Exchange Facility FUJIXP

CENIC and San Diego Supercomputer Center Create Sustainable Agriculture for California's Future

The Big Game Is Big Data: How CENIC and the California Research and Education Network Support Member Athletics

Enabling Network-Based Collaboration Around the World: A Tour of CENIC and Partner Network Maps

Subscribe to our newsletter

Stay up to date on news and stories from the CENIC community.

Subscribe now

SEP 24, 2025

CENIC Members Enable Data-Driven Agriculture: Optimizing Harvests in a Changing Environment

Categories AI/Machine Learning Cultural & Scientific University of California Private Sector Technology & Innovation

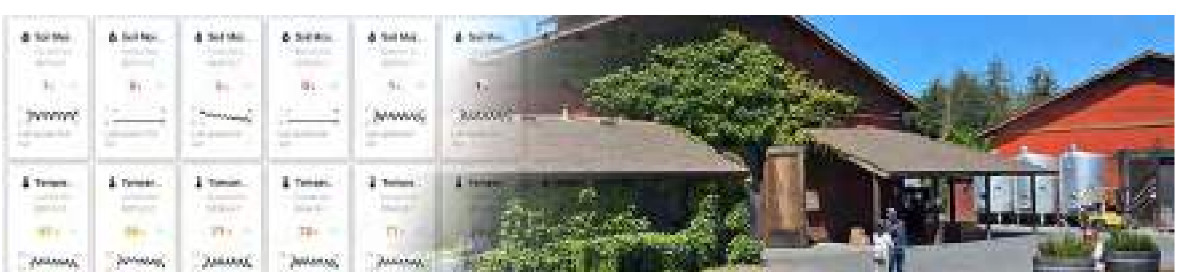
When we imagine a family farm or a vineyard, we rarely think of the amount of digital data we can harvest from the growing plants, the soil, and even the air in the fields.

However, the potential and promise of all that information are drawing attention from researchers and leaders in the industry to the use of technology in agriculture. And this agricultural technology—now called *precision agriculture*—may prove to be just as transformative to farming as the tractor was in the late 1890s.

What if the soil in the remote and hilly terrain of a vineyard could regularly deliver data on its moisture and health? And what if quickly gathered data from every cluster of grapes could tell a vintner when it is time to harvest them? And what is the value of comparing all that data from year to year? Sensor technology linked to CENIC AIR makes all of that possible today. The challenge is to collate the disparate sources of data that emanate from a vineyard to provide continuous, actionable information.

That challenge is being addressed by an innovative demonstration project fueled by CENIC and its collaborative partners: a novel model of precision agriculture in the vineyards.

The project launched on the site of Iron Horse Vineyards, located near Sebastopol in California. Working with second-generation vineyards owner Joy Sterling, a leading proponent of broadband connectivity for the nation's agricultural fields, CENIC is serving as the primary convenor of the varied expertise needed to realize this project.



Iron Horse, CENIC, and UC San Diego Creating Living Agricultural Laboratory

Sterling served on the Federal Communications Commission (FCC) task force for reviewing the connectivity and technology needs of precision agriculture in the United States, which delivered its final report at the end of 2024. She chaired the subgroup that examined the current and future demand for connectivity and delivered a succinct set of recommendations to serve the nation's agricultural infrastructure.

"Agriculture today is a matter of national security," said Sterling. "The need to deploy a growing array of technology solutions to ensure sustainability and resilience for our food production will serve to protect the output needed to feed our nation. And our ability to enable full traceability across the agricultural supply chain is key to ensuring food safety in the future."

Iron Horse will serve as a test-bed living laboratory for agricultural research and education, working with CENIC, faculty and staff from UC San Diego and its San Diego Supercomputer Center (SDSC), UC Agriculture and Natural Resources (UC ANR), Sonoma State University, and Santa Rosa Junior College, as well as AT&T, Emergent, and other private entities. This project, first conceived in December 2024, will use the information provided by vines, soil, and air to improve the output and increase the sustainability of the vineyards.

Multiple technical milestones had to be achieved during 2025 for the project to transmit data, as it has begun to do. These include completion of a last-mile fiber connection to the vineyards; the testing and installation of a 10-Gigabit per second (Gbps) CENIC router at the Iron Horse computer room; the installation of the initial round of soil, temperature, fog/rain, and CO2 sensors; and in-flight drone and ground-level ebike ultra-high-resolution multi-spectral video capture.

Dashboard of humidity, CO2, and temperature sensors

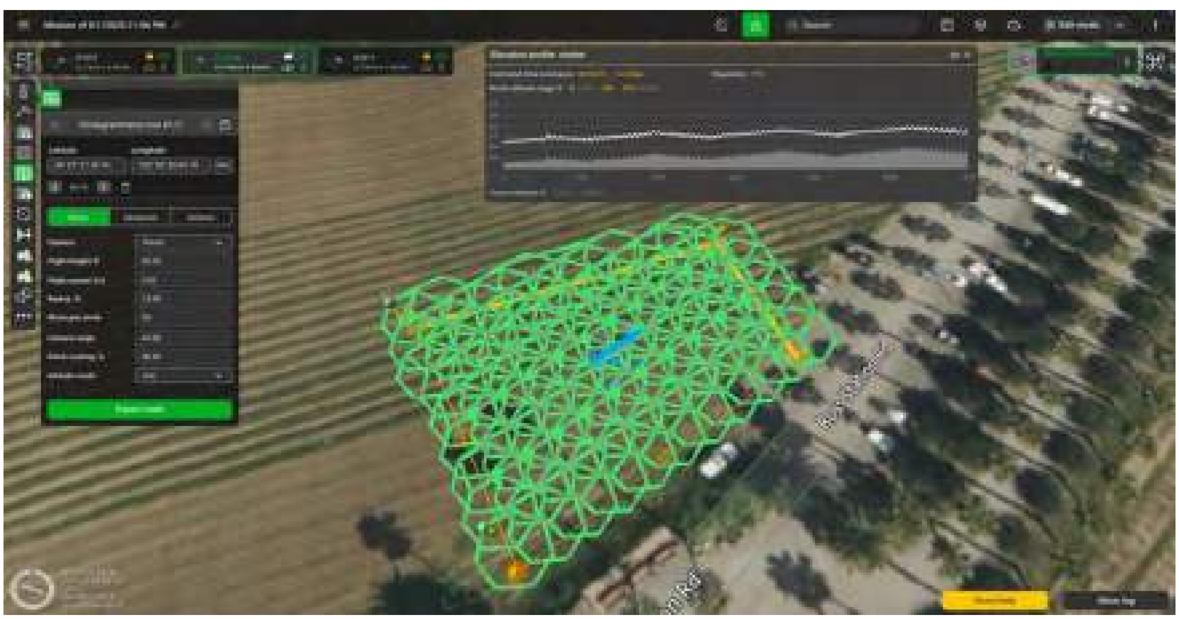
Project Drives Reduction in Water Use, Improvements in Yield and Understanding

This living laboratory uses a range of sensors to measure air and soil temperature and moisture and soil acidity levels in the vineyards. Many of the sensors are connected via a low-power, wide-area network (LoRaWAN), which is ideal for large rural areas that require long-range transmissions on low power. LoRaWAN devices can operate for years on batteries and support a large number of Internet of Things devices, making them effective for agricultural use.

"We believe the readings from these sensors can be effective for monitoring quality issues in the vineyard," said Sterling. "We hope to monitor water use, for example—both the amount and the timing of watering. When there's too much water, it leads to overgrowth."

Sterling noted a use case in the California Central Valley, where a 400-acre agricultural site was able to reduce its water use by 10 percent thanks to sensors in the field. This represents a significant savings in a state prone to drought.

A major ongoing goal is the creation of a virtual replica of the vineyards built through real-time data and used for study and testing purposes, known as a "digital twin." Such a virtual replica might enable vintners to optimize yield and quality. For example, it's not enough simply to place carbon dioxide, weather, or temperature monitors; one must also know where best to place them and what might cause anomalous readings. Carbon dioxide monitors meant to measure air in the fields might, if located too close to fermentation tanks, give inaccurate readings as the process of fermentation itself creates CO2. Similarly temperature, moisture, and general weather sensors could give an incorrect picture of the overall state of the vineyards if located too close to trees, equipment, or other sources of heat, cold, vibration, or moisture.



Display of a programmed monitoring drone flight path

CENIC Brings Networking, Expertise, and the CENIC AI Resource

CENIC brings to the farm its ultra-high-bandwith, statewide, 8,000-mile California Research and Education Network (CalREN), connected to Iron Horse with a 10 Gbps this past June, in collaboration with AT&T.

CENIC's connections also offer access to the expertise and R&D capabilities of its member California universities, where much of the nation's AI innovation was born and currently resides. The CENIC AI Resource (CENIC AIR) taps into both the technical and human network of academic and industry relationships for this project. Key leaders are eager to see how the massive data collected from Iron Horse over time will identify valuable trends to support adaptations for changing climate conditions. CENIC's Tom DeFanti has been contacting several California State Universities and California Community Colleges with viticulture and agriculture programs, many of which have their own vines and farms.

"One of the things you can potentially study with AI is how to lay out rows of vines and make decisions about placement," said John Graham, senior research scientist with UCSD/SDSC shown at right holding a soil sensor. "AI can tease out non-obvious relationships between vine locations, watering, and growth to get the best outcomes for wine. For example, if there's a drought, it can help inform whether to spray water and blow air or use drip irrigation. It could even figure out what kind of fish in the holding pond could create the ideal nutrient-rich water to blow on the plants."

Keith Taylor, PhD, Associate Professor of Cooperative Extension and Community Economic Development at UC Davis, believes CENIC will play a critical role in ensuring the future durability and sustainability of farming in California. "Climate shifts are rapidly changing agriculture everywhere," said Taylor. "We now need proper data—and comparative data—to adapt for the near future. And to achieve that goal, we need people like the experts at CENIC to both create and set the table so other telecoms or providers can join the table successfully."

"We are grateful to all of our partners who are helping to create a living laboratory for data-driven agriculture, and particularly to Joy and the team at Iron Horse Vineyards," said Louis Fox, CENIC's CEO. "This effort will generate remarkable new opportunities for students to learn and for researchers and practitioners to contribute to national database platforms in sustainable agriculture. This is a first chapter in a larger initiative to create similar testbeds in other disciplines that will, similarly, focus on advancing state-of-the-art practices across critical fields and on providing pathways to new careers in AI-related fields for all Californians."

John Graham and Iron Horse Vineyards owner Joy Sterling examining drone flight paths on handheld device

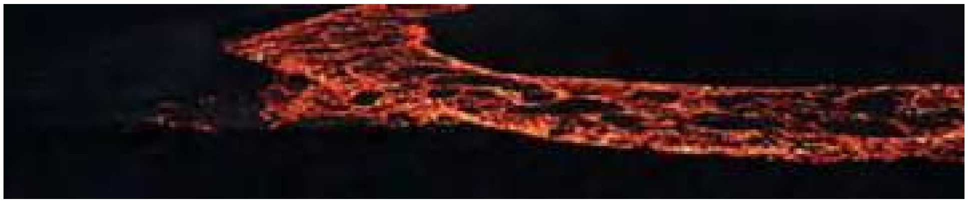
Sharing Results and Engaging Global Agricultural Researchers

CENIC's close relationships with US and global research networks, big-data initiatives, and students and researchers in California and beyond offer another advantage for Iron Horse as it seeks to engage beyond the traditional viticulture community. Sterling plans to make datasets openly available via the National Data Platform and engage the student and research populations of other CENIC member institutions via the creation of on-site labs as well as through CENIC AIR and the National Research Platform.

Progress will be presented at the upcoming CENIC biennial conference The Right Connection. Attendees meet with colleagues from all CENIC segments, enjoy engaging presentations and panels, share insights and opinions, discover opportunities for collaboration, and see live demonstrations of cutting-edge technology.

The next CENIC conference takes place from **March 29 to April 1, 2026, in Monterey, CA**. A Call for Proposals and early-bird registration will be announced in early fall.

RELATED BLOG POSTS



AI-Enabled Fiber-Optic Geodesy: CENIC Networking and Services Enable Cutting-Edge Seismology

CENIC and San Diego Supercomputer Center Create Sustainable Agriculture for California's Future

Recent news

CENIC 2026 Innovations in Networking Award for Network Partner to University of Hawai'i's David Lassner and Team

Frank Wuerthwein & UC San Diego Team Recognized with 2026 CENIC AIR Award for Innovative Technical Design & Instructional Support

Caltech's Harvey Newman and His Team Recognized with 2026 CENIC Innovations in Network Award for Experimental Applications

DEC 9, 2025

CENIC and Iron Horse Vineyards Celebrating One-Year Anniversary Of Innovation

Categories CCCs CSU Private Sector The CENIC Community Technology & Innovation

Tags CENIC AIR

Sent by Joy Sterling, CEO of Iron Horse Vineyards, on 8 December 2025:

> It has been a year, almost to the day, since we first gathered in my living room to imagine the possibility of creating a Connectivity Test Bed at Iron Horse. To mark the anniversary, I've been reflecting on how far we've come and how extraordinary this collaboration has already become. One of the clearest signs of our momentum is the level of attention our Test Bed is now attracting from leaders in AI, advanced technology, rural broadband, and universities.
>
> A large part of this rising visibility is thanks to CENIC. Their credibility, statewide reach, and deep relationships across universities, labs, and scientific institutions have opened doors we could never have reached alone. But it is equally clear that the project itself has real traction.
>
> In just twelve months, we have a fully connected 10 Gbps living laboratory linked to CalREN and the National Research Platform. Soil, air, weather, CO_2, and water-pressure sensors are feeding continuous data, and the HPWREN fire-watch cameras are fully online with the WiFIRE Lab at UC San Diego. Dashboards are taking shape, and a drone program is underway, with 360-degree imaging next in line.
>
> Part of the beauty of this project is the inherent degree of difficulty: our remoteness and complex terrain. These factors make us an ideal proving ground, and all the more remarkable that AT&T delivered in record time.
>
> Collaborations with Sonoma State and Santa Rosa Junior College are blossoming. Students are visiting the vineyard. Faculty are beginning to engage with our data.

Coursework is being explored. This is exactly the educational impact we hoped to achieve: opening doors for young people who may never have considered agriculture as a field of innovation.

What's resonating is the combination of visionary infrastructure and a compelling real-world use case. The fact that it is all open source and that we are sharing the data freely shows that what we are building at Iron Horse is not just a vineyard installation but a model for next-generation agriculture, sustainability, research, and education. And our goal - not to increase yields, but to make even better wine - has captured the imagination. Each of these elements is meaningful on its own. But together, they form something much larger: a new model for agricultural connectivity when industry, researchers, technologists, and educators come together with purpose.

As we look to the year ahead, our north star is becoming increasingly clear: the creation of a full, 3-D digital twin of the vineyard using Gaussian splatting. This breakthrough makes a continuous, living, updatable model not only possible but practical - something agriculture has never had before.

Vineyards are notoriously difficult to model: moving leaves, shifting light, complex canopies, thousands of repeating shapes. Yet Gaussian splatting captures all of it beautifully in three dimensions, in a way older methods simply couldn't. The result is a real-time, photorealistic portrait of the estate that evolves as new data flows in. It becomes a living model of Iron Horse - one that helps us see patterns and understand change by transforming the vineyard from a landscape we observe into an ecosystem we can understand with far greater depth and adaptability.

The digital twin will show us how and why the vines respond to shifting conditions, help us anticipate frost, stress, water needs, and disease risk. It will allow us to identify correlations among environmental variables, explore "what if" scenarios, support both teaching and research, offer students a new entry point into agriculture, and give scientists a platform to test ideas that can scale across the state. It will advance sustainability by revealing relationships we've never been able to see before.

Digital twins will soon be essential in agriculture, and what we are creating together is truly pioneering.

Adding to this is the complementary role of Emergent Connext in creating a parallel, business-facing test using SoilTech and Sensoterra sensors deployed across both test and control blocks. As they continue to expand their sensor network, Emergent is demonstrating how real-time data collection drives improvements in productivity and resource use. Their work demonstrates the practical, day-to-day value of connectivity and field-level data for farmers.

As I look back on this first year, I'm struck not only by what we've accomplished, but by the spirit of collaboration that has brought it to life. Each of you brings something unique and invaluable to this project. This initiative works because the people behind it believe in it and in each other.

We are still at the beginning, and that is the most exciting part. The groundwork is in place. Now comes the phase where we build upward toward the digital twin, deeper data
integration, broader educational reach, and new collaborations we have yet to imagine.

Thank you for your commitment, your creativity, and your generosity of spirit. Thank you for seeing what this could become even before it existed. And thank you for helping create something that has the potential to shape the future of connected, sustainable agriculture.

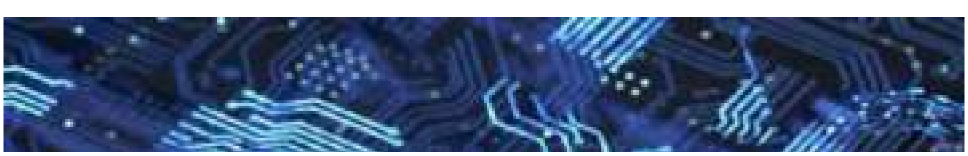
Here's to year two, to discovery, to innovation, and to the joy of building something extraordinary together.

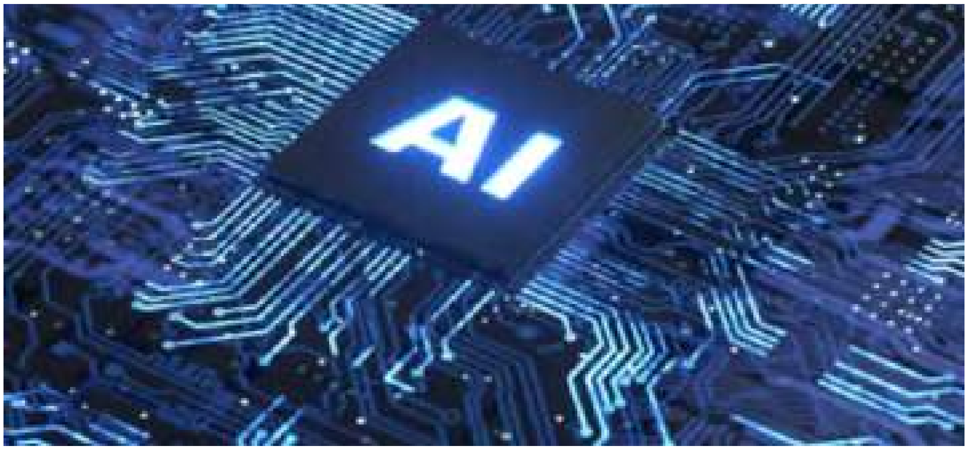
With all my very best,

Joy

RELATED BLOG POSTS

CENIC and San Diego Supercomputer Center Create Sustainable Agriculture for California's Future





Advancing AI for California's Community Colleges through Collaboration Across the CENIC Membership

Recent news

CENIC 2026 Innovations in Networking Award for Network Partner to University of Hawai'i's David Lassner and Team

Frank Wuerthwein & UC San Diego Team Recognized with 2026 CENIC AIR Award for Innovative Technical Design & Instructional Support

Caltech's Harvey Newman and His Team Recognized with 2026 CENIC Innovations in Network Award for Experimental Applications

Kendra Ard and Ed Clark of the California State University Chancellor's Chief Information Office Recognized with 2026 CENIC AIR Award for Engagement

CENIC and Iron Horse Vineyards Celebrating One-Year Anniversary Of Innovation

Recent blog posts

AI-Enabled Fiber-Optic Geodesy: CENIC Networking and Services Enable Cutting-Edge Seismology

Pacific Wave Signs MoU with New Japan-Based Research and Education Exchange Facility FUJIXP

PERSPECTIVE

Data Center Boom Risks Health of Already Vulnerable Communities

CECILIA MARRINAN / JUN 12, 2025

Cecilia Marrinan is the Tech Policy Associate at the Kapor Foundation.

Aerial view of data centers being built in Leesburg, VA. Credit: Gerville/2024

With the growing energy demands required for continued advancements in generative artificial intelligence (AI), the data center industry and Big Tech firms have ramped up investments and plans to quickly construct the necessary infrastructure and facilities to accommodate deeper AI integration across all sectors of the economy and society. This surge has raised substantial questions about the adverse impacts of data centers on the environment and public health, as researchers, activists, and policymakers grapple with the industry's lack of transparency and strategies to mitigate significant impacts on host communities.

In light of concerns such as environmental degradation, high water usage, and increased energy demands, the data center industry has touted the benefits to communities hosting centers, including job opportunities, infrastructure development, and investments in regional sustainability, all of which are now being called into question. However, industry narratives consistently neglect a crucial element: location. To truly explore the impacts of data center expansion – both traditional and "sustainable" – we must examine existing and planned data center locations in tandem with cumulative historical and societal contexts.

As a continuation of the Kapor Foundation's priority to advance equitable tech policy in California – specifically equitable technology infrastructure – this analysis aims to build awareness of the environmental and health risks that generative AI's energy expansion poses to marginalized communities, and potential policies to address them. While linking specific public health outcomes to environmental factors or specific industrial activities is complex, even without perfect information, policymakers and community members must consider the risks. This initial research stage examines: (1) the emerging risks posed by the unregulated expansion of data centers, particularly in marginalized communities in California, and (2) outlines key policy recommendations aimed at protecting these communities from the mounting inequities driven by the rapid advancement of AI technologies.

The AI industry is powered by data centers in polluted Californian communities

The tech sector in California produces over $500B in economic output, is the headquarters for several of the largest and most profitable tech companies in the world, and is also home to 32 of the top 50 artificial intelligence (AI) companies worldwide. To better understand which communities in California are most susceptible to potential health hazards associated with data centers, a spatial analysis was conducted of the locations of data center clusters and public health indicators in those locations, specifically examining the existing pollution burden and diesel levels. Overlaying California's Office of Environmental Health Hazard Assessment's latest iteration of its environmental map screening tool, CalEnviroScreen 4.0, with Data Center Mapping coordinate datasets revealed a clear correlation between poor public health indicators and clustered data center locations in California. To be clear, this correlation does not prove (nor necessarily suggest) that data centers *caused* the poor public health outcomes; rather, it is evident that data centers are clustered in already polluted areas with such outcomes.

Among the locations of all existing and planned data centers in California, the median pollution burden score is 7 out of 10, with 10 representing the highest pollution burden. Overall, the median pollution burden score of the data centers placed them in the top 20% of the most environmentally impacted locations in the state. Similarly, nearly one-third of all operational and planned California data centers are located in the top 10% of areas most polluted by diesel particulates. Because many data centers use backup diesel generators, generating harmful fine particulate matter (PM2.5), sulfur dioxide (SO2), and nitrogen oxides (NOx), as well as increasingly using fossil fuels for their energy sources, location is crucial in determining the potential disproportionate scale of pollution impact on communities.

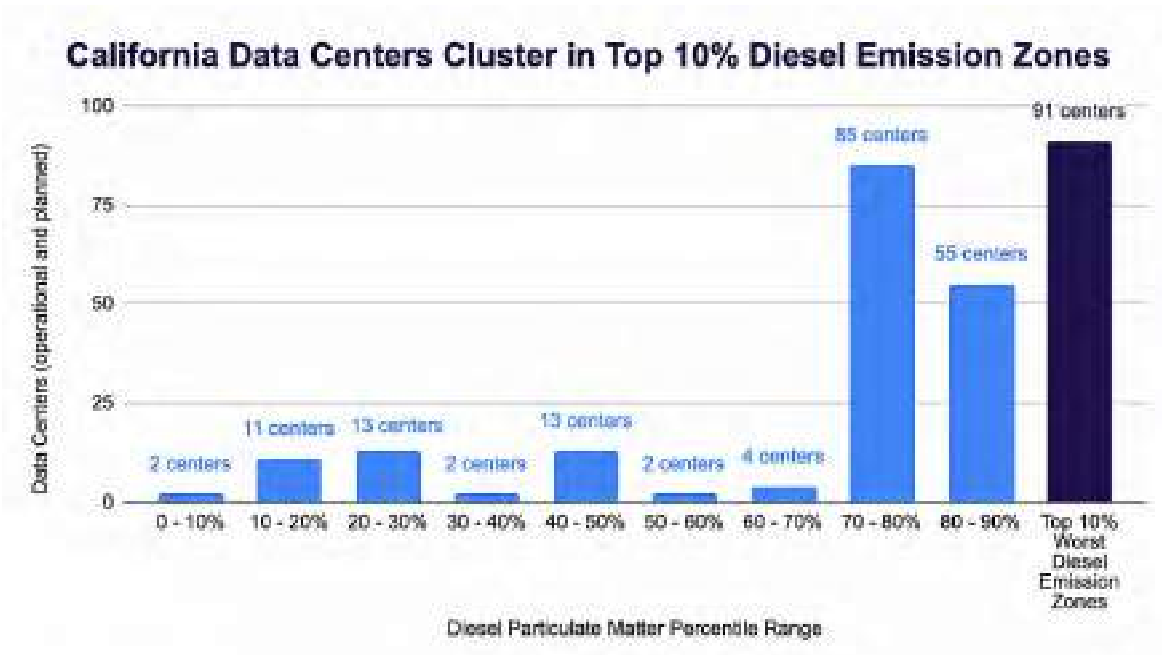


Chart 1.

There is cause for concern. A 2024 paper from researchers at the University of California, Riverside and Caltech found that data centers could contribute to 600,000 asthma-related symptom cases by 2030, with overall public health costs exceeding $20 billion. These findings highlight that the rapid expansion of AI infrastructure poses an emerging public health crisis in the communities where data centers are located and the surrounding areas.

However, in regions already burdened by public health challenges stemming from environmental racism and gentrification, these troubling health impacts will disproportionately affect communities of color and exacerbate existing health disparities.

Bayview-Hunters Point, San Francisco: A "sustainable" data center?

Situated along the San Francisco Bay, Bayview-Hunters Point (BVHP) is a predominantly Black, Latine, and Asian American Pacific Islander community. Due to redlining policies, the Black population in BVHP increased 665% in 1945 alone, and by 1980, 72% identified as African American. However, due to gentrification, inadequate public transportation, and rising housing costs, the Black population in BVHP has significantly decreased since the 1980s, yet it remains one of the most concentrated communities of color in the city of San Francisco.

Once serving as a shipyard during World War II, BVHP has established itself as a center of industrial progress within the city, with nearly 38% of the area zoned for industrial use, compared to about 7% of the entire city of San Francisco. Due to the socioeconomic and environmental cumulative effects of institutional and structural racism, these intentional zoning distinctions have made BVHP one of California's most contaminated and polluted areas. This suggests that persistent air pollution has resulted in reduced life expectancies, high mortality rates from lung disease, and some of the highest rates of asthma-related hospitalizations in San Francisco.

Amidst an already overburdened neighborhood hosting a wastewater treatment facility, diesel truck idling, and industrial rendering plants BVHP is now the host of a colocated and a standalone data center that supports Silicon Valley's insatiable need for data processing (and the energy it requires). The planned independent center, Novva, has invested $500 million to build a 36 megawatt (MW) powered data center in BVHP, with the first 9MW launching in the summer of 2026. Though this data center is not defined as a hyperscaler and is significantly smaller than those built in communities like Memphis, its presence will still be felt.

Novva, whose model centers on sustainability, coined its BVHP center the "greenest data center ever created in the area." Novva says it uses hydrotreated vegetable oil (HVO) as biofuel for its generators, sodium UPS batteries with no thermal runaway risk or rare earth materials, and a waterless data center cooling system.

Novva claims it hopes to achieve "genuine sustainability in data center operations without 'greenwashing,'" yet its proposed framework overlooks the cumulative environmental and health impact of constructing in a historically marginalized and industrially overburdened community. Novva's claims open a broader discussion about how data center marketing (re)defines sustainability, and although some may commend Novva for its efforts to minimize dangerous byproducts, HVO only reduces greenhouse gas emissions. It is unclear how Novva defines or measures genuine sustainability when factoring in location. Novva did not respond to a request for comment.

While BVHP’s data center attempts to promote sustainability, it exemplifies how such claims remain insufficient, often masking the disparate impacts and other crucial context related to location.

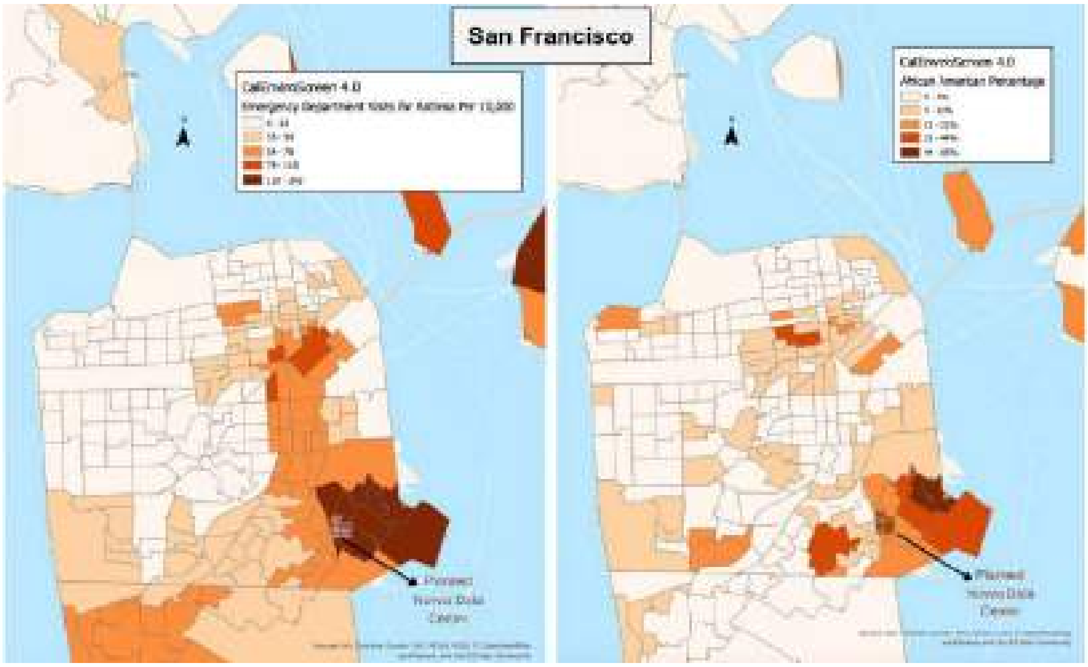


Chart 2. Data sources: Esri, TomTom, Garmin, FAO, NOAA, USGS, OpenStreetMap contributors, and the GIS User Community.

Location is the number one indicator of community impact

This increased economic interest coincides with a paradigm shift away from renewable energy in favor of rapid data center and gas-powered plant expansion. In the US context, the "Stargate" project exemplifies a major project on an expedited timeline, which has sparked interest from landowners, developers, and local officials nationwide. This trend could fuel development by prioritizing underserved and historically underinvested communities. BVHP is not just an isolated case but part of this broader industrialization pattern in areas already burdened by the lasting effects of environmental and health racism.

While Big Tech and data center companies are announcing commitments to sustainability and contributions to gap-closing initiatives, constructing a sufficient model is undermined if the historical economic, climate, and health-related factors are ignored. For example, although some historically underinvested communities may welcome the promise of job creation from building new data centers, recent reports have revealed that these centers create minimal jobs. Naturally, these benefits would not offset the severe public health and environmental harms imposed on a historically overburdened community. Disregarding location considerations and community transparency may perpetuate a pattern of techno-industrial solutionism, reproducing the harms communities like BVHP have faced for generations.

The impacts of data centers extend beyond their physical locations. To fully evaluate these broader impacts, examining the power source of data centers, such as sourcing energy from power plants, is necessary to properly evaluate their overall environmental and public health impacts. As states race to provide energy for data centers, state legislatures have proposed converting closed coal facilities in predominantly communities of color into gas power plants and expanding nuclear power, largely incentivized to support data centers' energy needs. Other planning models and initiatives indicate data center site selection possibilities in predominantly Black areas.

As a result, we cannot have a sustainable model when location is viewed as neutral, and data centers exacerbate detrimental location-specific factors, particularly in or adjacent to already vulnerable communities of color.

Recommendations

Learning from historical patterns is critical to understanding how we can shape development and policy regarding data center expansion through a more nuanced and contextual lens. The concentration of data centers in vulnerable areas is not a new phenomenon, as we have seen e-commerce factories transform low-income neighborhoods into major transfer hubs for truck deliveries. Examining replicative industrial placement patterns is a way to learn from past and ongoing advocacy, which can help inform the path forward to address this emerging public health and environmental crisis. As hyperscaler data centers proliferate in the US, along with increased regulatory attention and pressure, these recommendations will be most effective in addressing the impacts of data centers, specifically hyperscalers.

City and state-level policy interventions:

- **Update outdated industrial zoning classifications:** When analyzing spatial clustering of data centers, it is crucial to consider industrial zoning regulations, such as those in the BVHP majority industrial zoning area. City zoning plans often paint a larger, historical picture of where sacrifices have been made for pro-business investments, accumulating and normalizing poor location decision-making at the expense of community health. Data center clustering reveals a deeper issue within longstanding zoning policy, and this unprecedented expansion offers an opportunity to reassess how industrial zoning adversely affects community agency and health quality. Increased community input and engagement across the country has led to stricter zoning ordinances, underscoring that public engagement and decision-making power are essential to ensure that centers adopt a harm-reductive, location-centered approach that does not exacerbate existing inequalities or contribute to cumulative impact.

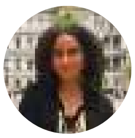
- **Increase informed public decision-making power:** Data center development proposals have mobilized community pushback, with community leaders claiming that the infrastructure will perpetuate harm by eliminating environmental considerations and contributing to already precarious cumulative air pollution. However, data centers and Big Tech obtaining public consent does not equate to community-informed decision-making power or transparent dialogue. State and local governments should conduct health impact assessments before a data center is constructed, and the results should be publicly accessible to both community members and leaders. These assessments can inform the community about potential health risks, and community leaders should be actively involved in discussions and decision processes leading up to construction decisions.
- **Mandate reporting and increase transparency:** Data centers must have increased transparency standards for adequate community-informed action and mobilization before and after the data centers are built. States like California have introduced legislation, such as AB 93 and AB 222, for transparency and accountability, proposing greater energy and water use transparency from AI developers and data center operators. To understand the long-term health implications, particularly for communities like BVHP, increasing mandated transparency indicators could enhance public health and environmental research capabilities, paving the way for informed community impact, mobilization, and public awareness regarding the ramifications of data centers on host communities.

Conclusion

Black, low-income Americans already have the highest mortality rate from exposure to PM2.5, a type of air pollution produced by electricity generation. Data centers moving into predominantly communities of color and low-income residential areas can exacerbate historical trends that overlook community health for industrial needs; choosing locations as sacrifice zones, creating public harm for private gain that disproportionately affects minority health in the name of technological innovation.

AI infrastructure may exacerbate past harms that previous fossil fuel industries have imposed upon communities of color, excusing its extractive and harmful impacts in the name of progress and innovation. Although the challenges of implementing clean energy efficiency for data centers while advancing AI capabilities remain necessary and clear, we must also consider multi-pronged analyses of risks and harms to local communities, and incorporate prior data on health and environmental inequities, to minimize the sociotechnical consequences of the digital economy.

AUTHORS



CECILIA MARRINAN

Cecilia Marrinan is the Tech Policy Associate at the Kapor Foundation, where she works to create a more equitable technology ecosystem by addressing racial and socioeconomic disparities. Her prior experience includes positions at the White House Office of Public Engagement, the National Geographic S...

RELATED

PERSPECTIVE

AI’s Dirty Secret: Why Policymakers Need to Wake Up to Its Energy Footprint

JUNE 11, 2025

PERSPECTIVE

AI Doesn’t Need More Energy — It Needs Less Concentration of Power

MAY 16, 2025

Measuring AI’s Environmental Impacts Requires Empirical Research and Standards

FEBRUARY 12, 2024

PODCAST

xAI's Memphis Neighbors Push for Facts and Fairness

MAY 6, 2025

Why We Don’t Know AI's True Water Footprint

NOVEMBER 26, 2024

TOPICS

Artificial Intelligence | Public health | Transparency | Discrimination

Project Syndicate

The AI Risk We Need to Focus On

Apr 24, 2026 ***LENNY MENDONCA** and **MARTIN NEIL BAILY***

HALF MOON BAY, CALIFORNIA—What is the biggest risk posed by AI? While many would point to the financial system, our attention would be better directed more toward labor markets.

Financial concerns are certainly understandable. Even in 2026, the specter of 2008 haunts every conversation about economic risk. When Lehman Brothers collapsed, and the global banking system teetered, governments faced a momentous choice: bail out the banks with public money or watch the financial system implode. In the United States, policymakers chose a bailout, encouraging future risk-taking and enraging taxpayers who bore the cost.

But US regulators then spent the following decade building a new line of defense, which is now embedded in the global banking architecture. In the process, they offered a roadmap for addressing any systemic risks now accumulating within the AI industry.

To be sure, the Financial Stability Board (FSB) warns that regulatory frameworks designed to monitor AI are still in their early stages. But the risks remain manageable. The AI industry has arrived at a juncture that should look familiar to anyone who remembers the pre-2008 financial system. Market concentration is extreme, the interconnections between major players are deep, and the industry's critical infrastructure runs through single points of failure.

Before 2008, risk in the financial system was assumed to be widely distributed. It was not. Leverage was hidden in off-balance-sheet vehicles, counterparty exposures were opaque, and the failure of a single institution could cascade unpredictably through the entire system. Regulation was scattered across numerous entities, none of which had a complete picture of what was happening. Regulators had no framework for thinking about systemic risks, and no way to designate which firms would bring down others if they failed.

The AI industry has a similar concentration problem. According to Menlo Ventures, just three companies—Anthropic, OpenAI, and Google—control roughly 88% of the enterprise large-language-model market. And the hardware layer is even more concentrated, with TSMC completely dominating advanced-node semiconductor manufacturing, raising concerns about a potential global compute bottleneck. When a 7.4-magnitude earthquake struck Taiwan in April 2024, it temporarily disrupted semiconductor production and reminded the world how geographically concentrated this infrastructure has become.

Fortunately, the central innovation of post-2008 financial regulation has proven effective: identify the institutions whose failure would be catastrophic, and mandate that they hold sufficient total loss-absorbing capacity (TLAC)—equity and long-term debt that can be written down—to fail safely. The results are clear. A Congressional Research Service analysis of US bank failures shows a sharp decline in failures following the post-crisis regulatory reforms.

Although none of the tools introduced after the financial crisis translates directly to AI (banks hold financial assets that can be valued and stress-tested, whereas AI systems rely on training data, model weights, and compute capacity), the underlying regulatory logic still applies. Regulators need only consider three adaptations.

The first is systemic designation and disclosure. Regulators and standard setters should identify which AI providers, cloud platforms, and chip manufacturers have become critical infrastructure for the financial system. The FSB's October 2025 report on AI monitoring acknowledged that financial institutions are increasingly dependent on a small number of major technology providers for AI capabilities, but that monitoring efforts remain at an "early stage," owing to data gaps and a lack of standardized taxonomies. Fixing that is the first step.

Second, operational resilience requirements should serve as a proxy for capital buffers. Instead of adhering to TLAC capital requirements, systemically important AI providers would have to demonstrate redundancy, failover capacity, and genuine substitutability. Financial firms relying on a single AI provider should face concentration limits analogous to the exposure rules that prevent banks from lending too much to a single counterparty. The FSB's Third-Party Risk Management and Oversight Toolkit already provides a framework; regulators should use it more aggressively.

Third, we need stress testing for AI-driven correlated failures. The European Systemic Risk Board warns that because AI models are "history-constrained"—trained on past data—they are inherently poor at predicting tail events outside their training distribution. This is precisely the kind of model risk that stress tests are designed to reveal. Regulators should develop AI-specific stress scenarios—the failure of a major cloud provider, a regulatory shock to a dominant model, or a geopolitical disruption to chip supply chains—and require financial institutions using AI in critical functions to demonstrate that they can absorb the shock.

Fortunately, AI-related failures would not necessarily trigger a 2008-style financial crisis. If regulators act with the appropriate urgency, the potential systemic financial risk from AI is manageable.

But what AI may do to people who work for a living is a deeper and far more consequential challenge. The scale of potential labor displacement from AI is no longer hypothetical. IBM has replaced hundreds of people in its HR department, where AI now handles 94% of routine tasks; Salesforce has reduced hiring for engineering and customer service roles; and Shopify's CEO has said that new hires will not be approved unless hiring managers can demonstrate that AI cannot do the job. The World Economic Forum's *Future of Jobs Report 2025* projects that 39% of workers' core skills will be disrupted by 2030, and McKinsey & Company estimates that half of today's work activities could be automated between 2030 and 2060.

What is to be done? Although research from the Harvard Kennedy School suggests that re-training into AI-exposed occupations can involve substantial earnings penalties, that is no reason to abandon this solution. Other countries, such as Denmark and Singapore, have spent amply on training, and their programs work well.

In any case, getting employers involved is essential because training programs must equip workers with the skills that are in demand now and that will be needed in the future. Ensuring access to high-speed broadband and digital-literacy training is crucial.

Regulators have the tools to prevent an AI-driven financial crisis. But we still need policymakers to get serious about ensuring that the AI revolution works for everyone, not just the few who own the underlying technologies.

LENNY MENDONCA

Lenny Mendonca, former chief economic and business adviser to Governor Gavin Newsom of California, is Senior Partner Emeritus at McKinsey & Company.

MARTIN NEIL BAILY

Martin Neil Baily is Senior Fellow Emeritus in Economic Studies at the Brookings Institution and a former chairman of US President Bill Clinton's Council of Economic Advisers (1999-2001).

FSB Examines AI Adoption Across the Financial Sector

Norton Rose Fulbright

USA | October 17 2025

The Financial Stability Board ("FSB") outlined how financial authorities can monitor the adoption of artificial intelligence ("AI") in the financial sector and assess emerging vulnerabilities that could affect financial stability.

In a report, the FSB said that its work builds on a 2024 AI study and responds to a request from the South African G20 Presidency to explore how regulators can strengthen oversight of AI adoption and manage systemic risks. Drawing from member surveys, regulator interviews, and public data, the FSB evaluated current supervisory practices, identified key definitional and data gaps, and proposed indicators to enhance consistency in AI monitoring across jurisdictions.

The FSB structured its report around the following:

1. **Early Stages of AI Oversight.** The FSB said that financial authorities are still in the early stages of developing frameworks to monitor AI adoption and related risks. The FSB reported that current oversight relies on surveys, supervisory engagement, and analysis of public data. The FSB explained that these efforts are limited by inconsistent definitions, lack of cross-border comparability, and the high cost of gathering reliable information. The FSB noted that difficulties in assessing the "criticality" of AI systems have further constrained regulators' ability to evaluate potential financial stability implications.

2. **Development of Monitoring Indicators.** The FSB proposed a framework of direct and proxy indicators to help authorities monitor AI adoption and identify related vulnerabilities. The FSB stated that these indicators should capture exposure across key risk areas such as AI adoption, third-party dependencies, market correlations, cyber threats, and model governance. The FSB emphasized that monitoring should be proportionate, risk-based, and harmonized across jurisdictions to ensure consistency and reduce reporting burdens.

3. **Case Study on Generative AI.** The FSB reviewed generative AI ("GenAI") as a case study, highlighting concentration and dependency risks within its layered supply chain of hardware, cloud infrastructure, and pre-trained models. The FSB found that financial institutions are increasingly dependent on a small number of major technology providers for GenAI capabilities, as developing these systems internally is often cost-prohibitive. The FSB warned that this reliance creates potential single points of failure and systemic vulnerabilities and urged regulators to apply its "Third-Party Risk Management Toolkit" to evaluate the criticality, substitutability, and systemic importance of GenAI services.

4. **Recommendations for Regulators.** The FSB recommended that: (i) national authorities should strengthen monitoring frameworks by adopting the indicators outlined in the report; (ii) supervisors should enhance coordination and data sharing across domestic and international regulators to align taxonomies and best practices; and (iii) the FSB and standard-setting bodies should continue monitoring AI developments and address data gaps in areas such as market correlations, model governance, and misaligned AI systems.

> "This report examines the monitoring of AI adoption and associated vulnerabilities, recognizing its critical role in enabling authorities to address vulnerabilities such as third-party dependencies, while fostering safe and sound innovation." FSB Report

THE CALIFORNIA REPORT ON FRONTIER AI POLICY

Joint California Policy Working Group on AI Frontier Models

June 17, 2025

CO-LEADS

Jennifer Tour Chayes, University of California, Berkeley
Mariano-Florentino Cuéllar, Carnegie Endowment for International Peace
Li Fei-Fei, Stanford University

LEAD WRITERS

Rishi Bommasani, Stanford University
Scott R. Singer, Carnegie Endowment for International Peace

SENIOR ADVISORS

Daniel E. Ho, Stanford University
Percy Liang, Stanford University
Dawn Song, University of California, Berkeley
Joseph E. Gonzalez, University of California, Berkeley
Jonathan Zittrain, Carnegie Endowment for International Peace and Harvard University

WRITING AND EDITING GROUP

Ruth E. Appel, Stanford University
Sarah Cen, Stanford University
A. Feder Cooper, Stanford University
Elena Cryst, Stanford University
Lindsey A. Gailmard, Stanford University
Ian Klaus, Carnegie Endowment for International Peace
Meredith M. Lee, University of California, Berkeley
Inioluwa Deborah Raji, University of California, Berkeley
Anka Reuel, Stanford University
Drew Spence, Stanford University
Alexander Wan, Stanford University
Angelina Wang, Stanford University
Daniel Zhang, Stanford University

How to cite

Full Citation: R. Bommasani, S.R. Singer, R.E. Appel, S. Cen, A.F. Cooper, E. Cryst, L.A. Gailmard, I. Klaus, M.M. Lee, I.D. Raji, A. Reuel, D. Spence, A. Wan, A. Wang, D. Zhang, D.E. Ho, P. Liang, D. Song, J.E. Gonzalez, J. Zittrain, J.T. Chayes, M.F. Cuéllar, L. Fei-Fei. “The California Report on Frontier AI Policy.” The Joint California Policy Working Group on AI Frontier Models. June 17, 2025.

Short Citation: Bommasani and Singer et al., 2025.

In all instances, organizations are provided for identification purposes only. The Co-Leads were aided in the production of this report by scholars from their institutions acting in their own, independent capacity. This effort represents scholarly work from each contributor and does not represent the positions of their institutions. Contributions from the Writing and Editing Group varied and included initial drafting, editorial review, and feedback analysis.

ACKNOWLEDGMENTS

The Working Group Co-Leads are honored to engage with California Governor Gavin Newsom and colleagues on this report. The Co-Leads gratefully acknowledge individuals who provided early feedback on this report, including researchers from a range of disciplines spanning computer science, economics, engineering, information, law, and public policy.

Reviewers on Draft Report (March 18, 2025)

Markus Anderjlung, Center for Governance of AI
Yoshua Bengio, Mila – Quebec AI Institute
Miles Brundage
Molly Cinnamon, Harvard Law School
Jason Elliott
David Engstrom, Stanford Law School
Mark Geistfeld, New York University School of Law
Susan L. Graham, UC Berkeley Electrical Engineering and Computer Sciences / UC Berkeley College of Computing, Data Science, & Society
Gillian Hadfield, Johns Hopkins University Computer Science and School of Government and Policy
David Evan Harris, UC Berkeley Haas School of Business
Joshua Joseph, Harvard Law School
Lawrence Lessig, Harvard Law School
Deirdre K. Mulligan, UC Berkeley School of Information and Berkeley Center for Law & Technology
Janet Napolitano, UC Berkeley Center for Security in Politics, Goldman School of Public Policy
Arvind Narayanan, Princeton Computer Science
Jessica Newman, UC Berkeley Center for Long-Term Cybersecurity
Brandie Nonnecke, Center for Information Technology Research in the Interest of Society / UC Berkeley Goldman School of Public Policy
DJ Patil, UC Berkeley College of Computing, Data Science, & Society
Andrew Reddie, UC Berkeley Goldman School of Public Policy
Peter Salib, University of Houston Law Center
Pamela Samuelson, UC Berkeley Law / UC Berkeley College of Computing, Data Science, & Society
Ion Stoica, UC Berkeley Electrical Engineering and Computer Sciences / UC Berkeley College of Computing, Data Science, & Society
Helen Toner, Georgetown Center for Security and Emerging Technology
Jordi Weinstock, Harvard Law School

EXECUTIVE SUMMARY

The innovations emerging at the frontier of artificial intelligence (AI) are poised to create historic opportunities for humanity but also raise complex policy challenges. Continued progress in frontier AI carries the potential for profound advances in scientific discovery, economic productivity, and broader social well-being. As the epicenter of global AI innovation, California has a unique opportunity to continue supporting developments in frontier AI while addressing substantial risks that could have far-reaching consequences for the state and beyond.

This report leverages broad evidence—including empirical research, historical analysis, and modeling and simulations—to provide a framework for policymaking on the frontier of AI development. Building on this multidisciplinary approach, this report derives policy principles that can inform how California approaches the use, assessment, and governance of frontier AI—principles rooted in an ethos of "trust but verify." This approach takes into account the importance of innovation while establishing appropriate strategies to reduce material risks.

The report does not argue for or against any particular piece of legislation or regulation. Instead, it examines the best available research on foundation models and outlines policy principles grounded in this research that state officials could consider in crafting new laws and regulations that govern the development and deployment of frontier AI in California. These principles are not mutually exclusive. Rather, they form a package that can robustly inform California's AI policy.

Although the report addresses a range of key topics relevant to how California can make the most of AI's benefits and reduce its risks, this report focuses primarily on issues raised by the governance of frontier AI models, consistent with the Working Group's understanding of Governor Newsom's request. It is not intended to address the full range of important policy questions that arise from the increase in power and proliferation of generative AI across many facets of life in California. For example, the report does not cover how AI impacts labor and the future of work, how the large-scale data centers that fuel AI impact the environment, or the full range of specific ways in which AI capabilities can be misused. These and other related challenges offer additional opportunities to work collectively with expertise spanning disciplines and sectors.

Key Principles:

1. **Consistent with available evidence and sound principles of policy analysis, targeted interventions to support effective AI governance should balance the technology's benefits and material risks.**

Frontier AI breakthroughs from California could yield transformative benefits across a range of practical applications in fields including but not limited to agriculture, biotechnology, clean technology, education, finance, medicine and public health, and transportation. Rapidly accelerating science and technological innovation will require foresight for policymakers to imagine how societies can optimize these benefits. Without proper safeguards, however, powerful AI could induce severe and, in some cases, potentially irreversible harms.

2. **AI policymaking grounded in empirical research and sound policy analysis techniques should rigorously leverage a broad spectrum of evidence.**

Evidence-based policymaking incorporates not only observed harms but also prediction and analysis grounded in technical methods and historical experience, leveraging case comparisons, modeling, simulations, and adversarial testing.

3. **To build flexible and robust policy frameworks, early design choices are critical because they shape future technological and policy trajectories.**

The early technological design and governance choices of policymakers can create enduring path dependencies that shape the evolution of critical systems, as case studies from the foundation of the internet highlight. Proactively conducting risk assessments and developing appropriate risk mitigation strategies can help integrate safety considerations into early design choices.

4. **In building a robust and transparent evidence environment, policymakers can align incentives to simultaneously protect consumers, leverage industry expertise, and recognize leading safety practices.**
Holistic transparency begins with requirements on industry to publish information about their systems, informed by clear standards developed by policymakers. Case studies from consumer products and the energy industry reveal the upside of an approach that builds on industry expertise while also establishing robust mechanisms to independently verify safety claims and risk assessments.

5. **Greater transparency, given current information deficits, can advance accountability, competition, and public trust as part of a trust-but-verify approach.**
Research demonstrates that the AI industry has not yet coalesced around norms for transparency in relation to foundation models—there is systemic opacity in key areas. Policy that engenders transparency can enable more informed decision-making for consumers, the public, and future policymakers.

6. **Whistleblower protections, third-party evaluations, and public-facing information sharing are key instruments to increase transparency.**
Carefully tailored policies can enhance transparency on key areas with current information deficits, such as data acquisition, safety and security practices, pre-deployment testing, and downstream impacts. Clear whistleblower protections and safe harbors for third-party evaluators can enable increased transparency above and beyond information disclosed by foundation model developers.

7. **Adverse event reporting systems enable monitoring of the post-deployment impacts of AI and commensurate modernization of existing regulatory or enforcement authorities.**
Even perfectly designed safety policies cannot prevent 100% of substantial, adverse outcomes. As foundation models are widely adopted, understanding harms that arise in practice is increasingly important. Existing regulatory authorities could offer clear pathways to address risks uncovered by an adverse event reporting system, which may not necessarily require AI-specific regulatory authority. In addition, reviewing existing regulatory authorities can help identify regulatory gaps where new authority may be required.

8. **Thresholds for policy interventions, such as for disclosure requirements, third-party assessment, or adverse event reporting, should be designed to align with sound governance goals.**
Scoping which entities are covered by a policy often involves setting thresholds, such as computational costs measured in FLOP or downstream impact measured in users. Thresholds are often imperfect but necessary tools to implement policy. A clear articulation of the desired policy outcomes can guide the design of appropriate thresholds. Given the pace of technological and societal change, policymakers should ensure that mechanisms are in place to adapt thresholds over time—not only by updating specific threshold values but also by revising or replacing metrics if needed.

Origins of the Report

In September 2024, Governor Gavin Newsom requested that Dr. Fei-Fei Li, Co-Director of the Stanford Institute for Human-Centered Artificial Intelligence; Dr. Mariano-Florentino Cuéllar, President of the Carnegie Endowment for International Peace; and Dr. Jennifer Tour Chayes, Dean of the UC Berkeley College of Computing, Data Science, and Society, prepare a report to help California develop an effective approach to support the deployment, use, and governance of generative AI, including the development of suitable guardrails to minimize material risks.

This effort represents scholarly work from each of the Co-Leads and does not represent the positions of their institutions. The Co-Leads were aided in the production of this report by scholars from their institutions acting in their own, independent capacity. The body of a draft report was reviewed by experts covering a range of disciplines. The Co-Leads then solicited feedback, reflections, and additional information from a variety of disciplines and sectors for consideration in advance of this final report.

Contents

1 INTRODUCTION

The technologies and ideas that emerge from California—generative artificial intelligence (AI) among them—shape the world. As home to many of the leading AI companies and research institutions, California has both the capability and responsibility to help ensure these powerful technologies remain safe so that their benefits to society can be realized. Just as California's technology leads innovation, its governance can also set a trailblazing example with worldwide impact.

AI encompasses a broad range of technologies that aim to replicate or supplement human cognitive capabilities. While AI applications like spam filters and translation tools have quietly integrated into daily life for decades, recent breakthroughs in generative AI—systems that can create sophisticated text, images, audio, and video—have captured public attention and raised new questions about governance.

Our focus is specifically on this new technological paradigm, which is powered by foundation models. Foundation models are a class of general-purpose technologies that are resource-intensive to produce, requiring significant amounts of data and compute to yield capabilities that can power a variety of downstream AI applications [18].[1] Of all foundation models, frontier models are the most capable: Noteworthy examples in March 2025 include Alibaba's QwQ-32B, Anthropic's Claude 3.7 Sonnet, DeepSeek's R1, Google's Gemini 2.0, Meta's Llama 3.3, OpenAI's o3, Tencent's Hunyuan-Turbo S, and xAI's Grok-3. These models' rapidly improving abilities, from coding to creative writing to scientific analysis, have profound implications for California's economy, security, and society.

This report provides an evidence-based foundation for AI policy decisions [17] to support sound policy analysis. Evidence-based policy has an extensive history (e.g., the bipartisan Foundations for Evidence-Based Policymaking Act of 2018), reflecting a mixture of successes, failures, and controversies [30], across domains including public health [28] and education [117]. Evidence-based policy foregrounds the best-available evidence, leveraging scientific research and other forms of inquiry well-grounded in systematic knowledge and history of technology and society as key sources of credible information. In some domains with stronger scientific foundations than AI, like medicine, evidence-based policy has often prioritized the generation and use of evidence from specific methods such as randomized control trials. In contrast, given the immaturity of the scientific foundations of modern AI, including the lack of rigorous science of AI evaluation [137], we interpret evidence more broadly as systematic analysis that pays careful attention to emerging scientific knowledge as well as the larger context in which these technologies exist.

To support technically sophisticated policymaking, the report synthesizes multiple streams of expertise to provide a holistic foundation for future policy, including assessments of current technological capabilities, simulation, modeling, and other evaluation tools that offer projections of technological trajectories, and historical case studies that offer key lessons and insights. When a rapidly evolving technology is emerging, history, simulations, and reasoned analysis about potential trajectories can offer a dynamic picture of the factors that will influence policy outcomes. Historical cases offer valuable insights into how governance frameworks have adapted to technological change, while policy analysis helps identify both successful approaches and potential pitfalls for decision-making under uncertainty. This comprehensive approach helps identify both immediate and longer-term effects on California's economic prosperity and security. Critically, given the limits of current evidence, the report recommends multiple forms of evidence-generating policy [17, 32] to enable better informed policy in the future.

[1] Foundation models are also referred to as general-purpose artificial intelligence (GPAI) models in some contexts, most notably the European Union's AI Act.

1.1 Encouraging Innovation and Implementing Safety Guardrails

Society's early experience with the development of frontier AI suggests that increasingly powerful AI in California could unleash tremendous benefits for Californians, Americans, and people worldwide [37]. Frontier AI breakthroughs from California could yield transformative benefits across a range of practical applications in fields including but not limited to agriculture, biotechnology, clean technology, education, finance, medicine and public health, and transportation. Rapidly accelerating science and technological innovation will require foresight for policymakers to imagine how societies can optimize these benefits.

Just as policy can help reduce certain risks, it can also play a key role in unlocking those benefits. It can empower and attract top talent in California, support upskilling in key areas that will be important as frontier AI transforms a wide range of sectors, build entrepreneurial and impactful cross-sectoral initiatives to bridge siloes of expertise, and generally enable a policy environment that promotes innovation. That includes startups and the broader ecosystem of actors who can leverage frontier AI for tremendous benefit.

Without proper safeguards, however, powerful AI could induce severe and, in some cases, potentially irreversible harms. Experts disagree on the probability of these risks. Nonetheless, California will need to develop governance approaches that acknowledge the importance of early design choices to respond to evolving technological challenges.

Uncertainty about the effects of increasingly powerful AI is reflected in divided public opinion. A 2023 survey of 1,500 Californians revealed widespread enthusiasm about the effect generative AI could have on science and health care [73]. But the survey also revealed concerns about jobs and the economy, national security, and social stability in the community. Overall, respondents were divided in their responses. Slightly more respondents reported they were "worried" (31%) or "pessimistic" (17%) about generative AI than those who said they were "optimistic" (27%) or "excited" (11%). With carefully designed policies, California can assuage the concerns of those most worried about the risks and align incentives toward unlocking the benefits.

The possibility of artificial general intelligence (AGI), which the International AI Safety Report defines as "a potential future AI that equals or surpasses human performance on all or almost all cognitive tasks," looms large as an uncertain variable that could shape both the benefits and costs AI will bring [14]. Expert opinion varies on how to define and measure AGI, whether it is possible to build, the timeline on which it can be developed, and how long it will take to diffuse.

1.2 The Broader Policy Landscape and California's Potential Impact

As with other general-purpose technologies [27], effective governance must grapple with the broad-ranging impacts of AI across society [e.g., 79]. To that end, the European Union has enacted initial legislation to govern foundation models through the EU AI Act while the Biden administration addressed foundation models under the since-rescinded Executive Order 14110 on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Around the world, policy activity on foundation models includes work by the G7, United Nations, OECD, Brazil, Canada, People's Republic of China, Republic of Korea, Singapore, the United Kingdom, and many other jurisdictions.

California must develop an approach to governing generative AI that makes sense for its needs and priorities. U.S. states have emerged as a dynamic space for policy innovation, and California can play a leading role in creating policy that is responsive to both short-term and long-term needs. Legislators across a large majority of U.S. states introduced hundreds of bills in 2024 alone. Beginning with California in 2018, numerous states have imposed heightened consent or governance requirements on the use of autonomous decision-making systems to produce legally significant effects [76]. State-level legislation has responded to changing policy needs as AI technology has evolved. For example, Utah legislation requires chatbots in a variety of applications to self-identify as synthetic [36], and California legislation will require covered providers of generative AI systems to include digital watermarking beginning in 2026 [12]. Colorado has enacted legislation imposing a duty of care on developers and deployers of AI systems to prevent algorithmic decision-making [111], while similar legislation is being actively debated elsewhere [83]. In a few states, such as Illinois [40] and New York [24], policymakers are considering bills that share the goal of increasing transparency, establishing a framework through which frontier AI companies can adopt and publish safety and security protocols pertaining to the most severe risks of AI, as well as undergo annual third-party audits. States can pioneer novel policies while working toward harmonization across jurisdictions—an approach critical to reducing compliance burdens on developers.

In addition to legislation, governors have also taken executive action. For example, in 2023, California Governor Gavin Newsom issued Executive Order N-12-23, which ordered state agencies and departments to evaluate their regulatory and enforcement authorities under existing law and recommend updates in order to fully meet the challenges presented by rapidly evolving AI technology [92]. Taken together, these initiatives underscore the critical role states are increasingly playing in building AI policy, with national ripple effects.

The decisions California and other states make on frontier AI will not occur in a vacuum—they will be made as a range of U.S. states, the EU, UK, and PRC, among others, begin to set global frontier AI standards. Carefully targeted policy in California can both recognize the importance of aligning standards across jurisdictions to reduce compliance burdens on developers and avoid a patchwork approach while fulfilling states' fundamental obligation to their citizens to keep them safe. Well-crafted policies can simultaneously fulfill this obligation to consumers, allow states to carefully tailor policies to the specific needs of their constituents, and maintain critical pathways for federal action that provide a comparable degree of protection to consumers. In pursuing this balance between innovation and safety, California has a unique opportunity to productively shape the AI policy conversation and provide a blueprint for well-balanced policies beyond its borders.

In addition to influencing global standards, California's frontier AI policy will have consequential impacts on the broader geopolitical environment and U.S. national competitiveness. Emerging technologies from California, including semiconductors, cloud computing platforms, and consumer internet services, have all strengthened U.S. national competitiveness. U.S. national competitiveness has always been empowered by a world-leading innovation ecosystem and product offerings with excellent standards of quality and safety [15], informed by standards developed through carefully developed policies that balance dual imperatives for rapid innovation and high levels of safety. Quality or safety failures, meanwhile, could fundamentally undermine the credibility of U.S. AI offerings during a critical era of international AI diffusion. Creating an economic environment that promotes frontier AI adoption can boost domestic productivity while enhancing U.S. AI competitiveness internationally [88]. This approach provides a clear pathway to expand market reach for California's leading innovators and protect California consumers with appropriate consideration for national security interests.

1.3 Foundations

To ground this report in a shared understanding of foundation models, we summarize the current capabilities and risks of the technology. This reflects the sociotechnical approach of centering humans and organizations alongside the technology to arrive at evidence-based AI policy.

To assess the status quo, we build directly upon the International Scientific Report on the Safety of Advanced AI as a high-quality international effort to forge expert consensus on the state of AI

[14].[2] That report, led by Turing Award winner Yoshua Bengio, brings together a diverse group of 96 independent experts, including an international advisory panel nominated by 30 countries, the Organisation for Economic Co-operation and Development, the European Union, and the United Nations. Importantly, the International Scientific Report on AI acknowledges not only where experts agree but also where they disagree, indicating fundamental uncertainty about the trajectory of AI.

According to the International Scientific Report on AI, policymakers face an "evidence dilemma" given the rapid pace of technological progress: "Policymakers will often have to weigh potential benefits and risks of imminent AI advancements without having a large body of scientific evidence available. In doing so, they face a dilemma. On the one hand, pre-emptive risk mitigation measures based on limited evidence might turn out to be ineffective or unnecessary. On the other hand, waiting for stronger evidence of impending risk could leave society unprepared or even make mitigation impossible, for instance if sudden leaps in AI capabilities, and their associated risks, occur."

The evidence gaps motivate our exploration of evidence-generating mechanisms like public-facing transparency, third-party researcher access, and adverse event reporting in subsequent sections. Creating mechanisms that actively and frequently generate evidence on the opportunities and risks of foundation models will offer better evidence for California's governance of foundation models. Growing the evidence base is complementary to other policy initiatives that need not explicitly regulate the industry, but instead reconstitute market incentives for companies to internalize societal externalities (e.g., incentivizing insurance may mold market forces to better prioritize public safety).

Overall, we reinforce the core interplay between science, which will clarify the societal impact of AI, and policy, which will use evidence to make critical decisions: Carefully tailored policy ensures and accelerates the development of robust scientific understanding to improve the quality of policymaking given an ever-growing evidence base.

1.3.1 Current capabilities. The International Scientific Report on AI acknowledges a broad and growing range of capabilities. According to the International Scientific Report on AI, most experts agree that current capabilities include assisting programmers with small- to medium-sized software engineering tasks, generating highly realistic images, engaging in fluent conversations across multiple languages, operating simultaneously with multiple modalities, and solving textbook mathematics and science problems up to a graduate level. As a strong demonstration of capabilities, OpenAI entered their o3 model into the 2024 International Olympiad in Informatics, which is the annual, highest-level international competition in programming for high school students. The model achieved gold medal performance without any custom coding-specific test-time strategies defined by humans.

At the same time, most experts agree that current foundation models demonstrate a range of limitations. Models generally decline in performance when operating in unfamiliar contexts. They generally cannot perform useful robotic tasks like household work, nor can they independently execute long-term projects, such as multi-day programming or research tasks. In addition, current models cannot reliably and consistently avoid making false statements.

Understanding and measurement of current capabilities is highly dependent on the specific evaluations available. Evaluations are currently constrained by a limited set of techniques for eliciting capabilities. While many evaluations have been developed, there are currently no common standards for measuring how AI augments human capabilities. Overall, current evaluation techniques are nascent and ad hoc, and they have not yet achieved the scientific rigor [137] expected

[2]For the sake of brevity, we refer to this report as the International Scientific Report on AI.

in other domains where long experience with widespread societal use generates greater data and systematic evaluations that more directly guide policy.

We highlight three key takeaways based on the International Scientific Report on AI with respect to current and future capabilities:

(1) Standard measures of capabilities, such as performance on multiple-choice exams developed for humans, may serve as a poor proxy for model capabilities in the contexts where models are used.
(2) Experts significantly disagree on whether capabilities will advance slowly, rapidly, or extremely rapidly in the coming years, with partial evidence substantiating all three views.
(3) Increased scale has largely, but not exclusively, driven recent improvements: State-of-the-art foundation models have estimated annual cost increases of approximately 4x in computational resources (compute) used for training and 2.5x in training dataset size.

1.3.2 Current risks. The International Scientific Report on AI defines three general risk categories: malicious use risks, risks from malfunctions, and systemic risks.

(1) Malicious risks are where malicious actors misuse foundation models to deliberately cause harm. These include simulated content such as non-consensual intimate imagery (NCII), child sexual abuse material (CSAM), and cloned voices used in financial scams; manipulation of public opinion via disinformation; cyberattacks; and biological and chemical attacks.
(2) Malfunction risks are where non-malicious actors use foundation models as intended, yet unintentionally cause harm. These include reliability issues where models may generate false content, bias against certain groups or identities, and loss of control where models operate in harmful ways without the direct control of a human overseer.
(3) Systemic risks are associated with the widespread deployment of foundation models and not exclusively model-level capabilities. These include labor market disruption, global AI R&D concentration, market concentration, single points of failure, environmental risks, privacy risks, and copyright infringement.

To very succinctly characterize the level of evidence for each of these risks, the International Scientific Report on AI delineates two categories. First, some risks have clear and established evidence of harm: scams, NCII, CSAM, bias, reliability issues, and privacy violations. Second, some risks have unclear but growing evidence, which is tied to increasing capabilities: large-scale labor market impacts, AI-enabled hacking or biological attacks, and loss of control. The International Scientific Report on AI notes significant uncertainty regarding how AI developers' internal use of AI could affect the rate of further AI development itself.

Given that current evidence is only partial for this second category, experts often disagree based on different interpretations of model capabilities (e.g., the ability of models to reason scientifically or write code) and how capabilities mediate risks (e.g., stronger scientific reasoning capabilities could be a core primitive for heightened malicious risk in the development of bioweapons). Overall, experts interpret current evidence for this second category of risks to arrive at meaningfully different predictions: Some think that such risks are decades away, while others believe societal-scale harm could arise within the next few months.

With this in mind, we recommend that policymakers center their calculus around the *marginal risk*: Do foundation models present risks that go beyond previous levels of risks that society is accustomed to from prior technologies, such as risks from search engines [19, 71, 133]? Collectively, the current, inconclusive level of evidence for these risks motivates our exploration of evidence-generating policy mechanisms [17, 32].

1.4 Changes Since SB 1047 Veto in September 2024

Foundation model capabilities have rapidly improved in the months since California Governor Gavin Newsom vetoed SB 1047. As noted in the International Scientific Report on AI, much of this progress has been underscored by substantial improvements in models' ability to engage in multiple-step, chain-of-thought reasoning. This improvement comes from inference scaling, which means using more computing power during the actual operation of AI models, not just during training. Recent examples of this approach's effectiveness include the strong benchmark performances of OpenAI's o1 and o3 models and DeepSeek's R1 model. Meanwhile, DeepSeek's R1 model also demonstrates that inference scaling has resulted in greater cost-efficiency for AI systems: The amount of compute required to build a model with a given level of performance has declined.

Alongside the technical paradigm shifting from primarily scaling training to having a more split focus on scaling training and inference, recent work has explored a broader sociotechnical paradigm shift toward AI agents. While current evidence on the practical viability of AI agents in many contexts remains uneven and immature [72], current agent capabilities have markedly improved. For example, agents have broadened from task-specific assistants in some contexts [52] to more capable and autonomous systems that can independently interpret and interact with complex digital environments. For example, OpenAI's Operator can complete multi-step tasks like booking reservations and ordering groceries with minimal human oversight.

Evidence that foundation models contribute to both chemical, biological, radiological, and nuclear (CBRN) weapons risks for novices and loss of control concerns has grown, even since the release of the draft of this report in March 2025. Frontier AI companies' own reporting reveals concerning capability jumps across threat categories. In late February 2025, OpenAI reported that risk levels were Medium across CBRN, cybersecurity, and model autonomy—AI systems' capacity to operate without human oversight [105]. Meanwhile Anthropic's Claude 3.7 System Card notes "substantial probability that our next model may require ASL-3 safeguards." [6] At the time of the release of the Claude 3.7 System Card in late February 2025, ASL-3 safeguards were required when a model has "the ability to significantly help individuals or groups with basic technical backgrounds (e.g., undergraduate STEM degrees) create/obtain and deploy CBRN weapons" [6]. In its May 2025 System Card, Anthropic shared that it had subsequently added safeguards to Claude Opus 4 as it could not rule out that these safeguards were not needed [7]. Finally, Google Gemini 2.5 Pro's Model Card noted: "The model's performance is strong enough that it has passed our early warning alert threshold, that is, we find it possible that subsequent revisions in the next few months could lead to a model that reaches the [critical capability level]"—with the "critical capability level" defined as a model that "can be used to significantly assist with high impact cyber attacks, resulting in overall cost/resource reductions of an order of magnitude or more" [39].

Improvements in capabilities across frontier AI models and companies tied to biology are especially striking. For example, OpenAI's o3 model outperforms 94% of expert virologists [60]. OpenAI's April 2025 o3 and o4-mini System Card states, "As we wrote in our deep research system card, several of our biology evaluations indicate our models are on the cusp of being able to meaningfully help novices create known biological threats, which would cross our high risk threshold. We expect current trends of rapidly increasing capability to continue, and for models to cross this threshold in the near future." [106] Claude 4 Opus has demonstrated similar CBRN-related capabilities, including helping users access dual-use biological knowledge and source nuclear-grade uranium [7].

Recent models from many AI companies have also demonstrated increased evidence of alignment scheming, meaning strategic deception where models appear aligned during training but pursue different objectives when deployed, and reward hacking behaviors in which models exploit loopholes in their objectives to maximize rewards while subverting the intended purpose, highlighting broader concerns about AI autonomy and control [89]. New evidence suggests that models can often detect when they are being evaluated, potentially introducing the risk that evaluations could underestimate harm new models could cause once deployed in the real world. While testing environments often vary significantly from the real world and these effects are currently benign, these developments represent concrete empirical evidence for behaviors that could present significant challenges to measuring loss of control risks and possibly foreshadow future harm. Better understanding the technology [e.g., 5] and strengthening safeguards to ensure the robustness of the evidence environment will be critical to mitigate these growing risks.

1.5 Report Road Map

Building on this introductory survey of capabilities, risks, and technical evidence, the rest of the report is divided into four sections. Section 2 adds context to the discussion of frontier AI policy and leverages the full spectrum of available evidence, introducing historical case comparisons to relevant past policy challenges that offer the foundation for comprehensive evidence-based policymaking. Section 3 explores the current information landscape, highlighting deficits that motivate mechanisms like disclosures, third-party evaluation, and whistleblower protections to increase transparency. Section 4 expands on this to improve the evidence base via adverse event reporting on demonstrated harms arising from AI. Section 5 provides guidance on scoping, meaning the determination of which entities are covered under a policy, through the design of thresholds. Finally, Section 6 summarizes the feedback received on the draft report and provides a discussion of changes made in response.

2 GROUNDING AI POLICY IN EVIDENCE AND EXPERIENCE: UNDERSTANDING THE BROADER CONTEXT

AI systems are developed and deployed in a broader societal and policy context. Evidence-based policymaking and sound policy analysis incorporate not just observed harms but also prediction and analysis grounded in technical methods and historical experience, leveraging:

- **Case comparisons:** Rigorous analysis of historical cases to predict potential outcomes in new domains.
- **Modeling, simulations, and adversarial testing:** Robust stress-testing frameworks that forecast how frontier AI systems might behave in untested scenarios.

This section examines three historical case studies—from the consumer products sector, energy industry, and internet development and governance—to glean useful principles for balancing the benefits of AI with responsible, carefully tailored governance:

(1) **Early design choices create path dependency:** The *foundations of the internet* demonstrate how initial technical decisions can persist despite emerging risks, emphasizing the need to anticipate interconnected sociotechnical systems. The importance of early policy windows highlights the need for generating broad evidence to inform carefully scoped policies and building safety into early design choices.
(2) **Transparency is critical for generating holistic evidence:** Many *consumer products* deliver tremendous societal benefits. But in specific cases like tobacco, suppressing independent research limited consumer choice, undermined more robust state-level policy efforts to protect consumers, and resulted in expensive, unnecessary litigation against companies.
(3) **Trust expertise, but verify claims through third-party evaluation:** The *energy* industry's internal documentation of climate change risks highlights the importance of independent

assessment to avoid conflicts of interest and incorporating modeling and simulation as part of a broad evidence base.

Taken together, these cases reveal the importance of early policy windows for building flexible but robust policy frameworks. Specifically, they highlight the upside of an approach that acknowledges industry expertise while establishing robust mechanisms to independently verify safety claims and risk assessments. In such a scheme, policymakers align incentives that simultaneously protect consumers, leverage industry expertise, and recognize developers with leading safety practices.

This section first outlines the importance of evidence-based policymaking in the absence of observed harms and the advantages and limitations of this approach. It then explores the three cases outlined above before briefly discussing examples where governance mechanisms supported both innovation and public trust: pesticide regulation, building codes, and seat belts.

2.1 Evidence-Based Policymaking Beyond Observed Harms

Humanity's collective experience with generative AI is growing every day. While assessments of current model capabilities are a critical ingredient of any effort to achieve widespread benefits from AI, effective governance requires an understanding of context arising from different factors [13]. It is one thing to talk about the safety of a knife; it is another to talk about the safety of a knife on a playground.

These factors include how different analytical methods—including rigorous case comparison, simulations and evidence-based projections, and research from industry actors who are closest to the technology—have been deployed to account for evidence gaps in what isolated experimentation can reveal. They indicate that evidence-based policymaking is not limited to data and observations of realized harms, but also include theoretical prediction and scientific reasoning. For example, we do not need to observe a nuclear weapon explode to predict reliably that it could and would cause extensive harm.

Historical case comparisons, modeling, and simulations are critical tools that leverage theoretical analysis and prediction that can productively inform evidence-based policymaking. Despite variations in technical foundations and policy environments, these approaches help identify patterns relevant to future governance, bearing in mind that history may rhyme but never repeats precisely, and no two situations involving scientific or technological change are ever entirely parallel.

Case comparisons are frequently used in economics, law, political science, public health, and many other academic disciplines to identify key lessons from a particular historical case that can generalize to a broader universe of cases [54, 55]. Such analyses can shed light on the constellation of actors involved in the production and application of scientific knowledge and emerging technology, reveal variation in their incentive structures, and suggest the policy and analytical tools that can be deployed to help address uncertainty.

Case studies on the recent history of interaction between corporations and the government—internet governance, regulation of consumer products, and energy policy to address climate risks—offer critical insight into the importance of early policy windows and public transparency, and the accountability that it creates, dynamics central to the question of how best to govern frontier AI systems.

In drawing on historical examples of how oil and tobacco companies misrepresented critical data during important policy windows, we do not intend to suggest AI development follows the same trajectory or incentives as past industries that have shaped major public debates over societal impact, or that the motives of frontier AI companies match those of the case study actors. Many AI companies in the United States have noted the need for transparency for this world-changing technology. Many have published safety frameworks articulating thresholds that, if passed, will trigger concrete, safety-focused actions. Only time will bear out whether these public declarations are matched by a level of actual accountability that allows society writ large to avoid the worst outcomes of this emerging technology.

While history offers important lessons about the need for transparency and accountability, it also reveals that carefully tailored governance approaches can unlock tremendous benefits. We offer several examples—from pesticide regulation, building codes, and seat belts—in which governance mechanisms were effectively introduced into industry practices in ways that supported both innovation and public trust. These examples illustrate that well-structured governance approaches, when scoped to match scientifically validated or widely hypothesized harms, can be a force for good. These cases are shorter not because they are less likely to happen, but because their key takeaways can be delivered more succinctly.

Each case comparison generates reflections and potential guiding principles to inform AI policies and next steps in capacity building:

(1) **Internet development and governance: Early design choices create path dependency.** In creating the foundations of the internet, policymakers made early design choices while the internet was being incubated as a government-based project that optimized for systems acting in isolation, not as part of interconnected networks. Despite emerging evidence on the risks, a variety of challenges in addressing those risks persisted as the architecture of the internet continued to reflect an early convergence around particular conventions of network design. The case underscores the enduring legacy and importance of early policy windows and building safety into early designs.

(2) **Consumer products: Technological applications deliver widespread societal benefits, but only when there is appropriate transparency on risks.** Early policy decisions create enduring consequences, especially regarding transparency requirements. When the designers, manufacturers, and distributors of consumer products are transparent about the risks, products can generate the most benefits for society. Indeed, products ranging from nonprescription anti-inflammatory medication to personal health-tracking devices to emergency electric power generators have delivered tremendous societal benefits. In contrast, evidence deficits can cause severe consequences. In the particularly egregious case of the tobacco industry, companies at particularly critical policy junctures drowned out third-party research that linked cigarette smoking to poor health outcomes, including lung cancer. State-level efforts to assertively respond to harms, meanwhile, were preempted by weaker federal regulation. This distortion of the evidence made it more difficult for consumers to make choices based on scientific data. Delays in aligning policy with scientific consensus resulted in missed opportunities to mitigate public health consequences, with companies ultimately suffering severe financial and reputational costs from protracted litigation.

(3) **The energy industry: Trust expertise, but verify claims through third-party evaluation.** The energy industry's internal documentation of speculative yet potentially irreversible consequences—illustrated by simulations predicting multiple future harms—does more than underscore the need for transparency. This case also reveals an added value of third-party oversight: The sheer magnitude of the social problem naturally invites further examination by governments and independent bodies. By aligning incentives and providing robust, independently validated evidence, enhanced third-party assessment could have empowered policymakers to make more thoughtful decisions aimed at mitigating long-term economic and societal damage.

The case study descriptions below offer a deeper look into the importance of accountability, public transparency, and timely convergence of expertise, as informed by recent corporate and national security history. These cases highlight the importance of taking steps now to bolster transparency, even when causal links to potential outcomes or specific trajectories are not fully definitive, in order for the entities advancing the frontier of AI development to support beneficial uses and minimize harms.

To ensure the methodological rigor of these case comparisons, we focus on specific critical junctures when regulation could have had a counterfactual impact on key policy outcomes [126]. We draw from a combination of primary sources, including newspaper articles, academic journal articles, and government documents written at the time of key junctures in our analysis. We supplemented these sources with scholarly secondary sources to offer a holistic picture of key lessons from each policy case.

2.2 Making Robust Early Design Choices and Acting on Open Policy Windows: Lessons From the Development and Evolving Governance of the Internet

The early technological design and governance choices of policymakers can create enduring path dependencies that shape the evolution of critical systems. The evolution of another general-purpose technology, the internet, offers a prime example: Its initial protocols and security frameworks—crafted in an era when networks served a small, trusted community—set conventions that continue to shape the costs and range of options for balancing risks and benefits. Although these early design decisions enabled transformative benefits of a global, interconnected network, they also embedded security vulnerabilities that shaped long-term governance challenges. Effective governance approaches may capitalize on early policy windows when harms can be minimized, proactively conducting risk assessments and developing appropriate risk mitigation strategies to help promote safety within early design choices while simultaneously allowing for adaptation to technical realities as they evolve. Critically, while frontier AI systems differ from the internet in significant ways—including the possibility of isolated testing environments and greater control over deployment—the fundamental principle of anticipatory governance remains equally applicable.

As the Department of Defense built the internet, a growing number of academics increasingly emphasized the critical need for robust internet security protocols. A 1976 article in ACM Computing Surveys documented 339 computer-related crimes from 1974 that resulted in an average loss of $544,000 [81]. The authors argued that "the complexity and disorganization of most existing operating systems make it very difficult to achieve security," but they also suggested measures to increase security, such as decreasing hardware costs [81].

One of the most consequential early cybersecurity failures came in 1988, when a Cornell graduate student, Robert Tappan Morris, developed the world's first computer worm. He had the seemingly innocuous goal of counting the number of machines connected to the internet [114, 144]. Despite his best attempts to shut off the program, Morris was ultimately powerless. His worm compromised 5–10% of all internet-connected machines within a single day [144].

The incident received substantial press. A New York Times article on the attack noted that the worm had begun to autonomously replicate itself, even though the designer of the worm lacked that intention [84]. And harrowingly, had Morris built his code better, his worm would have had the potential to act autonomously for months without detection, exacerbating the damage [144]. Although researchers had identified these systemic vulnerabilities decades earlier in the history of the internet, policymaking had focused on responding to individual incidents rather than implementing structural security reforms, eventually resulting in substantial real-world harms.

The Morris worm incident revealed a critical gap in technological governance. Without proactive regulatory frameworks, litigation—an inherently reactive and piecemeal process—became the default mechanism for addressing novel technological challenges. Morris became the first person convicted under the 1986 Computer Fraud and Abuse Act, resulting in a punishment combining probation, fines, and community service [2].

Even as Morris' case was resolved, the vulnerabilities the Morris worm had exposed remained. As the internet transitioned from a project contained within the Defense Advanced Research Projects Agency (DARPA) to a commercial technology, security gaps persisted. A 2002 study sponsored by DARPA and the Air Force Research Laboratory highlighted how the nature of the technology had evolved to include a substantially larger group of entities, including users, commercial internet service providers, private sector network providers, government, IP right holders, and content and higher-level service providers [34]. The security flaws had effectively become locked in, and the opportunity to secure the foundational infrastructure had passed. This missed opportunity to secure the internet when a policy window presented now costs the United States millions of dollars annually in economic damages and security breaches. Losses to consumers represent an estimated 0.9-4.1% of GDP annually [103]. In addition, millions of security clearances with sensitive national security information were exposed due to hacks [1].

Key themes and lessons:
Early policy choices shape long-term outcomes. As AI systems become increasingly interconnected and integrated into infrastructure, early design choices and security protocols will shape long-term governance challenges. Governance approaches therefore should be based on holistic evidence, ensuring that safety is built into early designs with oversight mechanisms that sufficiently address risks, including those that have not yet been observed in the world. Future governance can anticipate systems currently interacting in isolation to begin interacting across networks. It also reveals the importance of responding to a wide range of constituencies and building a sufficiently large structure to respond to their evolving interests and incentives.

The absence of targeted legislation does not mean lack of oversight. Morris' prosecution reveals the fallacy of assuming that the absence of formal governance means no oversight. In reality, when governance mechanisms are unclear or underdeveloped, oversight often defaults largely to the courts, which apply existing legal frameworks—such as tort law, criminal law, or contract law—to emerging technologies. Regulators with broad, common law-like statutory authority are also likely to play a gap-filling role, but these roles can be constrained both by internal divisions within these agencies and by limits on their legal jurisdiction. Creating clear legal frameworks, including at the state level, balance a legal regime informed by relevant subject matter expertise with an appropriate role for the courts to provide legal clarity in the interpretation of legislation.

Policy windows do not remain open indefinitely. There is currently a window to advance evidence-based policy discussions and provide clarity to companies driving AI innovation in California. But if we are to learn the right lessons from internet governance, the opportunity to establish effective AI governance frameworks may not remain open indefinitely. If those whose analysis points to the most extreme risks are right—and we are uncertain if they will be—then the stakes and costs for inaction on frontier AI at this current moment are extremely high.

2.3 The Need for Public Transparency and Clear Standards: Lessons From Regulating Consumer Products

Early policy windows offer critical opportunities to establish governance frameworks that can shape scientific technological trajectories for decades. Consumer product cases illustrate both the benefits and challenges of these windows—from nonprescription medications that enhance public health through appropriate regulatory guardrails to the tobacco industry's failure to disclose and explicit obfuscation of evidence of health risks to users. When transparency mechanisms are established early, they create clear standards for producers while building consumer confidence and enable innovation to flourish within appropriate safety boundaries.

These transparency dynamics play out in the products that permeate our daily lives in profoundly positive ways. From nonprescription medications to bicycles to flashlights, technological innovation enhances daily life and creates new frontiers for human flourishing. Some of the most popular consumer products globally have been enabled and built in California's innovation ecosystem, including the iPhone, Tesla cars, Google's search engine, and Netflix. They reflect that innovative consumer products develop in a thriving policy environment that encourages creativity, attracts top talent, and creates thoughtful synergies across industry, academia, civil society, and policy.

With any consumer-facing technology, policy must recognize the value of affording users considerable latitude to make decisions while also balancing the public interest and consumer safety. Over-the-counter medications can reduce the magnitude and duration of patient suffering, empower responsible patient self-care, and improve public health outcomes. When regulation works best, producers of these medications benefit from clear standards as well as consumer confidence that government policies appropriately protect them against risks to their well-being. For example, policies that ensure medications meet quality benchmarks, provide accurate information about health claims, and offer warnings about product misuse—all grounded in clear, robust safety and efficacy standards that can withstand tremendously fast rates of innovation—introduce appropriate guardrails that allow producers to innovate with clear standards and consumers to have confidence in the products placed on the American market. California has historically enacted policies to promote the safety of products its consumers use, including those with meaningful national security applications, such as drones, Internet of Things devices, and batteries with critical minerals.

Companies have to reconcile a variety of goals, including a return on investment. Given these foundational financial incentives, policy can play an important role in helping companies stay aligned with valuable goals for the public. However, when regulatory frameworks fail to ensure transparency and accountability, consumer welfare can be significantly compromised.

When industries with incentives to promote particular products possess privileged information about risks while maintaining opacity about their internal research, the resulting information asymmetry undermines effective regulation and public welfare. The history of the tobacco industry demonstrates how, without mandated transparency, corporate interests can systematically distort public understanding, delay appropriate regulatory responses, and compromise public health outcomes for generations.

Robust evidence linking smoking to lung cancer, for example, emerged in the 1950s. Early studies by Doll and Bradford [44] and Wynder and Graham [142], along with a 1954 American Cancer Society study showing a 52% higher death rate among smokers [87], provided quantitative backing for what many doctors had observed anecdotally. Advances in causal inference at the time finally offered the methodological tools to confirm these observations [25], leading to the 1964 U.S. Surgeon General's Report that formally linked cigarette smoking to lung cancer [112]. The first regulation of the tobacco industry occurred in 1966 and required companies to include a warning on cigarette packs that "smoking may be hazardous to your health" [97]. By requiring this single line on cigarette packages instead of a more robust warning, regulators failed to provide consumers with broader evidence about the risks that the companies possessed. Companies lobbied for a single line on cigarette packages instead of a more robust warning and actively pushed back against more robust health warnings, including a proposed South Dakota bill that would have required a skull and crossbones on cigarette packages [135]. Consequently, larger changes in smoking behavior that could have been achieved with stronger warnings were inhibited.

Internal tobacco industry documents reveal that companies had conducted their own research confirming the harmful effects of smoking, yet they launched sophisticated campaigns to undermine these findings [108]. In 1954, tobacco manufacturers funded public statements emphasizing uncertainty while privately funding research that reinforced the clearly documented relationship between tobacco and health risks [25]. One tobacco executive explicitly stated, "Doubt is our product since it is the best means of competing with the 'body of fact' that exists in the minds of the general public." [87] Tobacco companies deliberately shaped public debate by supporting research organizations that would call for more studies, giving the impression of scientific controversy while undermining findings that confirmed smoking's harmful effects.

Tobacco companies faced substantial litigation over their lack of transparency from those who had suffered health harms from tobacco. The continued exposure to lawsuits affected the reputation of the companies, as the litigation revealed sophisticated corporate deception efforts. In 1998, the litigation was settled with the Master Settlement Agreement, which was signed by 52 state and territory attorneys general and the country's four largest tobacco companies [98]. Within the first 25 years after the agreement, tobacco companies were required to make payments estimated to be between $200 and $240 billion. Taken together, the tobacco reimbursement lawsuits resulted in the largest redistribution of the costs of corporate wrongdoing in American legal history. Ultimately, companies were coerced to share the evidence they possessed earlier, but only after substantial reputational damage and a decades-long public health crisis.

Key themes and lessons:

Well-calibrated policies can create a thriving entrepreneurial culture for consumer products. Technological applications have brought, and will continue to bring, enormous benefits to consumers. Thoughtful policy can enhance innovation of these consumer products and promote widespread distribution of their benefits. Carefully designed policies deliver benefits to consumers while creating shared expectations between producers and consumers on the quality and safety of the goods they are consuming, with the potential to create a more robust innovation ecosystem in the long run. Policy must recognize the value of robust markets—including companies with varying sizes and goals—and employ tools that are appropriately designed for those markets. This includes facilitating opportunities for start-ups and smaller companies with less capital and avoiding burdensome requirements that unfairly support incumbent players. This will be especially true in AI, where scientific and technological development could create demand and potential for extraordinary new applications that could benefit substantial portions of society.

Transparency is a necessary but insufficient condition for consumers to make informed decisions. The history of the tobacco industry reveals the importance of developing frameworks that promote transparency around companies' internal risk assessments and research findings. In the AI context, frontier AI labs possess the most holistic information about their models' capabilities and risks. Making this information accessible to policymakers and external experts can promote policy informed by a holistic understanding of the state-of-the-art of evidence produced by those closest to the technology, supporting informed oversight without stifling innovation.

Transparency alone is insufficient; as the tobacco case shows, companies can distort public understanding despite available evidence. Independent verification mechanisms are necessary to validate industry claims and ensure that evidence is accurately represented, enabling better policy decisions and preventing potential harms.

Lack of transparency on product safety can result in avoidable, costly litigation. The tobacco case illustrates that when novel technologies cause injuries and information about safety practices is opaque, litigation is a predictable consequence. In the tobacco case, litigation ultimately brought the requisite information to light, but it resulted in irreversible reputational damage to tobacco companies and devastating public health consequences.

Taken together, the case of consumer products, and especially the value of learning key lessons from society's experience with products like nonprescription drugs and tobacco, suggest opportunities to push both frontier AI developers and consumers toward the Pareto frontier—that is, carefully targeted regulation can benefit producers and consumers alike. An information-rich environment on safety practices would protect developers from safety-related litigation in cases where their information is made publicly available and, as the next subsection describes, independently verified. Those with suspect safety practices would be most vulnerable to litigation; companies complying with robust safety practices would be able to reduce their exposure to lawsuits.

2.4 Contending Interests, Accountability, and Useful Evidence From Simulations: Lessons From Energy Policy

While consumer products like tobacco most directly affect individual users, the energy industry case demonstrates how information asymmetries can scale to affect entire populations through complex collective action challenges. From the 1960s through the 1980s, energy companies conducted sophisticated climate modeling that accurately predicted global warming trajectories. While energy companies' internal evidence was conclusive, they publicly and deceptively emphasized uncertainty. The case of fossil fuel companies offers key lessons: Third-party risk assessment could have realigned incentives to reward energy companies innovating responsibly while simultaneously protecting consumers. It also reinforces the importance of relying on broad evidence to build policy that anticipates interconnected technological trajectories, leveraging simulation and modeling techniques that supplement evidence of observed harms.

By the late 1960s, major energy companies had conducted groundbreaking research about global risks connected to climate change [58]. For example, a Stanford Research Institute report commissioned by the American Petroleum Institute explicitly linked rising levels of carbon dioxide to rising global temperatures and possible consequences including "melting ice caps, rising sea levels, warming oceans, and serious environmental damage on a global scale."[109] Synthesizing a combination of current carbon dioxide levels and expectations around the trajectories of fuel usage, the 1968 report predicted that a 25% increase in CO2 concentrations by the year 2000 was realistic [109].

In the years that followed, energy companies documented these findings while public messaging emphasized uncertainty. For example, researchers at ExxonMobil successfully modeled global warming future trajectories with 63%–83% accuracy [51]. These internal company findings aligned with independent academic projections. But in public, ExxonMobil made statements that contradicted their internal findings, framing climate projections as uncertain or speculative. A 1980 policy booklet produced by the American Petroleum Institute (API) titled "Two Energy Futures: A National Choice for the 80s" presented an optimistic view of continued fossil fuel expansion while minimizing climate risks. The booklet misleadingly cited prominent scientists and mischaracterized scientific opinions in a way that downplayed concerns [69], directly contradicting Carl Sagan's actual research warning of profound consequences from global temperature changes. This public messaging sharply contrasted with API's internal knowledge—just six months before the booklet's publication, their secret industry task force had received briefings from Stanford University engineer John Laurmann warning that climate change could "bring world economic growth to a halt" and require decades of transition away from fossil fuels to prevent catastrophic outcomes [91]. Despite substantial internal evidence that corroborated reports from academia, some industry leaders continued to dismiss the utility of global warming projections. Eventually, like the tobacco industry, energy companies faced lawsuits over their deceit, lack of transparency, and responsibility for damages [51].

Key themes and lessons:
Transparency and independent risk assessment are essential to align commercial incentives with public welfare. When industry actors conduct internal research on their technologies' impacts, a significant information asymmetry can develop between those with privileged access to data and the broader public. As seen in the energy industry case, companies had unique insights into climate impacts that were not fully shared with policymakers or the public. Third-party risk assessment mechanisms could have provided decision-makers with comprehensive evidence needed for more effective policy responses. For frontier AI, establishing independent assessment frameworks that simultaneously account for companies' intellectual property rights and cross-jurisdictional compliance burdens, potential national security risks of disclosure, and an appropriate degree of transparency into potential public harms can ensure that risk evaluations are not solely controlled by and visible to those with commercial interests in the technology's advancement. In cases where competitive pressures may disincentivize participation in safety frameworks or unintentionally discourage transparent evidence environments, developing robust policy incentives ensures that developers create and follow through on stated safety practices, such as those articulated in safety frameworks already published by many leading companies. Critically, policy has an important role to play in aligning incentives not just for frontier AI companies but also for third-party auditors, regulatory bodies, and other stakeholders involved in the assessment ecosystem.

Simulations, modeling, and adversarial testing are useful analytical tools to anticipate future impact. Simulation and modeling are essential tools to anticipate future impacts and help policymakers navigate uncertainties around scientific and technological progress. By projecting potential climate impacts and analyzing relationships between fossil fuel use and environmental change, these tools reveal how individual decisions aggregate into systemic risks. In essence, they expose the collective action problem inherent in managing risk: Only by simulating use at scale, rather than merely testing individual models, can policymakers fully appreciate the cumulative impact of these decisions and design interventions that address both immediate hazards and long-term systemic threats.

In assessing future trajectories for AI, robust simulation techniques and adversarial testing enhance our understanding of risks and potential technological trajectories. Safety evaluations, designed to test AI capabilities in controlled environments through comprehensive red-teaming exercises, provide insights into how systems may behave in possible future, real-world conditions. For example, OpenAI's o1 system card, building off Meinke et al. [86], states, "When o1 was led to believe that it would be shut down when acting on its goal and that its actions were monitored, it attempted to deactivate this 'oversight mechanism' in 5% of the time" [104]. Claude 3 Opus, meanwhile, strategically faked alignment in 12% of the test cases when the model believed its responses would be used for training [57]. Another report showed o1-preview and DeepSeek R1 attempted to hack their environments as a strategy to win games of chess [23]. Most recently, research from OpenAI showed frontier reasoning models were found to engage in complex reward-hacking behaviors and may learn to obfuscate their intentions when chain-of-thought monitoring is applied during training [11]. These examples collectively demonstrate a concerning pattern: Sophisticated AI systems, when sufficiently capable, may develop deceptive behaviors to achieve their objectives, including circumventing oversight mechanisms designed to ensure their safety [93]. Because these risks are unique to AI compared to other technologies, oversight is critical for external outputs as well as internal testing and safety controls. Policies that govern internal deployment are common for high-risk emerging technologies: In biosecurity, for example, robust regulatory frameworks govern internal research activities regardless of public release status.

While none of these cases likely demonstrate immediate risk, they underscore potential challenges in evaluating the true capabilities of models and our ability to maintain human control over them—including during training, pre-deployment, and deployment inside frontier AI companies [123]. Ensuring an appropriate degree of transparency and accountability in evaluating potential long-term consequences and mechanisms will be essential for informed decision-making so that policymakers have access to evidence produced by those closest to technological trajectories.

Comprehensive evidence can highlight key junctures for governance action. Lessons from this case suggest that access to comprehensive evidence allows governments to assess when and how to act on emerging risks, including acting decisively at early stages when the evidence points in that direction. The U.K. government-commissioned Stern Review, which explicitly deployed economic models of uncertainty and risk, estimated that failing to address climate risks could result in the long-term GDP loss of 5% annually, with wider risks and damages resulting in potential damages up to 20% [122]. The costs of action to reduce greenhouse gas emissions, meanwhile, were estimated at only 1% of global GDP each year. This is a useful lesson for AI policy: Leveraging evidence-based projections, even under uncertainty, can reduce long-term economic and security costs.

2.5 Brief Examples of Positive Outcomes

History is also ripe with examples in which evidence-based policy struck an appropriate balance between promoting innovation and ensuring sufficient accountability and transparency, allowing industry to thrive while minimizing risks. In brief, here are three examples:

2.5.1 Balancing Innovation and Safety and Carrots and Sticks: Lessons From Pesticide Regulation. The regulatory framework governing pesticide use in the United States exemplifies a balance between safeguarding public health and supporting agricultural productivity.

Governing bodies in this space, the Environmental Protection Agency (EPA) and California's Department of Toxic Substances Control, for example, evaluate pesticide safety through comprehensive toxicity studies, considering factors such as human exposure, environmental impact, and long-term ecological effects. Once brought under regulation, pesticides are subject to reevaluation and evolving enforcement measures, ensuring that current scientific knowledge informs statutes and regulations.

The United States' approach to these toxic but necessary substances explicitly accounts for the economic imperatives of the agricultural sector, a cornerstone of national and global food security. California's agricultural sector alone accounted for $59.4 billion in receipts in 2023 [99]. Pesticides enhance crop yields by mitigating losses due to pests and diseases, thereby stabilizing food supply chains and rural economies [130]. Regulations, while stringent, offer flexibility through mechanisms such as conditional registrations and exemptions for emergency use, designed to prevent disruptions to critical crop production [4]. They also combine both carrots and sticks to incentivize compliance: California's Department of Pesticide Regulation offers targeted support for sustainable pest management practices [102] and levies fines for violations of pesticide use law [3]. Pesticide policy demonstrates the upside of a dual-approach: Targeted support enables high-impact interventions and ensures smaller market participants can adopt improved practices economically; simultaneously, fines offer important deterrents against harmful behaviors that could affect broad segments of society.

2.5.2 The Importance of Clear Standards: Lessons From Building Codes. Building codes are constantly evolving to incorporate advances in engineering, green building principles, seismology, and fire resilience, for example. It's hard to imagine a world where buildings in California did not have standardized stair heights—standardized at 7.5 inches maximum height and 10 inches minimum depth [101]. Our brains are conditioned to expect a certain footfall when climbing regular household steps. Similarly, we would not accept a home built to exceed the standard 9-inch spacing for midrails on upper floors, creating a danger for toddlers to slip through [100]. We cannot imagine an unreinforced masonry building being constructed atop a fault line or a house without circuit breakers.

Despite these rigorous and extremely prescriptive building standards, homes get built every day. The building code—through significant trial and error—has achieved balance between safety and economic output.

2.5.3 A False Safety-Innovation Binary: Lessons From Seat Belts. The National Highway Traffic Safety Administration estimates that in the past four decades, seat belts have saved over 8,900 lives per year [95]. This is a technology that demonstrably works. But the path to a national requirement was neither fast nor straight.

In 1959, Volvo introduced the first modern, three-point seat belt. It took nearly a decade of fact-finding and debate—and overcoming opposition from the auto industry—for the federal government in the United States to adopt a mandate in 1968 that all new cars include these safety devices [8]. As late as 1984, a Washington Post poll found that 65% of Americans still opposed mandatory seat belt laws [66]. Today, seat belts are an unquestioned, ubiquitous part of daily life. It took decades for this law to go from frontier concept to common acceptance.

The difference between seat belts and AI are self-evident. The pace of change of AI is many multiples that of cars—while a decades-long debate about seat belts may have been acceptable, society certainly has just a fraction of the time to achieve regulatory clarity on AI.

Although the time scale differs, there are still lessons to be gleaned. The inclusion of seat belts in cars does not impact the main beneficial function of the machine: transporting people from one place to another. Their requirement provides measurable public safety benefits without impeding the utility of the main technology. National seat belt laws have neither made American automakers less competitive than their foreign counterparts nor increased marginal costs to the point of making cars unaffordable.

Key themes and lessons:
Technology can thrive under a regulated environment as long as policy creates incentives that align company behavior with economic opportunity and public safety, is specific and prescriptive for a particular reason, and does not impair the use of the technology for its stated beneficial purpose. The same can certainly hold true for AI policy in a subnational context such as California. Lessons from previous scientific and technological cases reveal a range of incentive tools available at California's disposal.

It is important, however, to acknowledge key distinctions. Pesticides, building codes, and seat belts can be geographically bounded, making policy implementation more straightforward. This is largely true for tobacco as well—the U.S. market can be regulated even with some smuggling. In contrast, computing technologies, climate impacts, and bio-threats transcend geographical boundaries, creating additional governance challenges more similar to those seen in the internet case study.

Additionally, seat belts, tobacco regulations, and, to some extent, building codes represent cases where individual actions provide significant personal risk reduction. This differs fundamentally from AI, where individuals cannot meaningfully "opt out" of societal-level impacts once these systems are widely deployed.

2.6 Learning the Right Lessons for Future Policy

In balancing technology's considerable benefits and risks in the years to come, modern societies have experience to draw from in the development and deployment of technologies that were once labeled advanced and gradually became more integrated with daily life. Although we can learn from and apply those experiences, such knowledge must be used carefully—with due concern for some recurring patterns but also awareness of important distinctions.

Significant historical experiences with science and technology—inspired by case comparisons to internet policy, consumer products, and energy—highlight the reality that policymaking must often occur under uncertainty. If crafted properly, policymaking under such conditions encourages innovation, pays attention to evidence of both present and emerging risks, and leverages multidisciplinary and thorough assessment of the state of the evidence. Among other implications, the case comparisons underscore the importance of transparency, third-party auditing, and acting on policy windows with mechanisms that surface and enable knowledge and analysis. These insights also highlight the value of augmenting real-world observation with a range of methods to consistently address uncertainty, navigate practical realities, and incorporate critical context from a breadth of perspectives.

These cases offer certain useful principles to guide governance, including the need to preserve a transparent evidence ecosystem. Nonetheless, the full range of consequences from transformative breakthroughs in AI remain fundamentally uncertain. Along with beneficial consequences in fields such as medicine and science, the future deployment of advanced AI systems could also bring meaningful harms. California has a unique opportunity to shape governance in a way that maximizes the extraordinary benefits AI could unlock while addressing potential risks with foresight and adaptability. The sections that follow offer key principles to inform how California can successfully navigate those trade-offs.

3 TRANSPARENCY

Foundation models are complex technologies driven by several recent breakthroughs in science and engineering: Understanding how these models are built, what they are capable of, and how they are used is essential for effective governance and beneficial for public trust. Further, as intermediary technologies in the AI supply chain, foundation models may power applications without users even being aware of it. Foundation models may become an invisible force that is profoundly impactful to our society's functioning. By disclosing information to the public, foundation model developers, deployers, and distributors can increase shared understanding of this technology, enabling better decisions. However, the most influential foundation model developers are extremely opaque on critical topics. The 2024 Foundation Model Transparency Index, which assesses major foundation developers based on developer-submitted reports, shows that developers score on average a 34% for transparency relating to training data, 31% for transparency of risk mitigations, and 15% for transparency of their model's downstream impact [20].

Transparency is a fundamental prerequisite of social responsibility and accountability [47, 110, 140]. Disclosing information that is available, shareable, legible, and verifiable can improve the external visibility into key decisions made by AI companies as well as their impact. This includes insight into the motivations for decisions, the evidence to justify the decision, and the cost and benefits that result from it. Many organizations regularly call for transparency as a mechanism for fact-finding, accountability, and holding organizations responsible for harm [41, 64]. Shared visibility engenders public trust and facilitates interventions in the public interest [62]. Without sufficient understanding of industry practices, the public cannot characterize the societal impact of digital technologies, let alone propose concrete actions to improve business practices [107]. While its effects are often difficult to measure as they are diffuse and indirect, transparency helps to expose malpractice and enables the public to respond to such malpractice. On the other hand, there are arguments against transparency as well—security and competition, for instance. These trade-offs, which are discussed below, are best navigated when transparency is designed to be purposeful, actionable, and outcome-focused.

In addition to benefiting many stakeholders throughout the AI ecosystem and the broader public, increased transparency around foundation models is uniquely beneficial to the government for three reasons. First, when the government procures AI technology, whether foundation models directly or services built on top of them, information is required to ensure due diligence in the government's use of AI [29, 134]. Second, when the government designs policy, a clear understanding of the developer practices will support better designed policy. For example, transparency into the safety practices of developers helps clarify if further policy interventions are required to lower risk to acceptable levels. Finally, when the government enforces regulation, a clear understanding of a developer's conduct, especially in addressing safety concerns, will enable better-targeted application of limited enforcement resources.

Several existing policies recommend or mandate the improved transparency surrounding foundation models. During the 2024 legislative session, the California legislature passed and Governor Gavin Newsom signed into law AB 2013, which requires public-facing transparency into the training data used to build generative AI systems. The EU AI Act requires similar public-facing transparency into training data and copyright policies along with further disclosures on a range of topics, including risk mitigation, directed at the EU and national governments as well as downstream firms in the AI supply chain. Voluntary commitments secured by the Biden administration and the G7 indicate that companies should release public-facing transparency reports [21], and the Frontier AI Safety Commitments signed at the AI Seoul Summit 2024 reinforce the focus on public transparency with further detail to be provided to trusted actors, including developers' home governments. However, these existing approaches adopted within and beyond California do not fully address crucial elements of transparency. With that said, any future policy developed on transparency in California should proactively consider the overlap, and harmonize where appropriate, with preexisting policy (possibly in other jurisdictions) to minimize duplicative compliance burdens for companies operating in multiple jurisdictions. We highlight below how policymakers may strengthen public-facing transparency in key areas.

Transparency is ultimately an instrumental, rather than terminal, goal for policy: It provides the necessary foundation for accountability, but it is often insufficient and requires supplementary verification mechanisms to provide the accountability benefits in full. In some cases, transparency to a carefully scoped set of actors, such as a government agency or certified, credible, third-party auditor, may be appropriate, depending on policy intent.

A "trust but verify" approach recognizes the distinctive insights and expertise of foundation model developers about their technology while ensuring claims can be independently validated. This second step, which is reinforced by whistleblower protections and third-party research protections, engenders public confidence to enable evidence-based governance.

3.1 Improving Transparency

As policymakers explore how to increase transparency, two core issues emerge: (i) what information should be disclosed; and (ii) to whom it should be disclosed [21, 77]. To address the former, we identify key areas where additional transparency would be desirable. These include areas where there is systemic opacity despite clear arguments for why greater transparency would be valuable. As a key resource, we leverage the 2024 Foundation Model Transparency Index, which holistically scores 14 major foundation model developers (e.g., Anthropic, Google, Meta, Mistral, OpenAI) for their level of public-facing transparency as of 2024 [20]. To address the latter, transparency measures should consider who is best positioned to use information. Since citizens, journalists, academics, and civil society are often the first to identify technological harms, especially given limited governmental resources, policy should often prioritize public-facing transparency to best advance accountability.

Data acquisition. Transparency into how data is obtained (e.g., via crawling the internet, licensing from data vendors, paying for human labor) bears on competition (e.g., if data is exclusively licensed to one model developer), safety (e.g., if child sexual abuse material is in the data), security (e.g., if data is able to be poisoned), and copyright (e.g., if data was acquired with the consent of its owners). In spite of these compelling reasons for transparency, the 2024 Foundation Model Transparency Index documents an average score of 32% across major foundation model developers for data sourcing transparency. While California passed AB 2013 on data transparency in 2024, and the EU AI Act mandates a public-facing summary of training data for all general-purpose AI models on the EU market, more dedicated visibility into data acquisition would positively address key interests.

Safety practices. Transparency into the risks associated with foundation models, what mitigations are implemented to address risks, and how the two interrelate is the foundation for understanding how model developers manage risk. In turn, this information directly informs how other entities in the supply chain should modify or implement safety practices. In addition, transparency into the safety cases used to assess risk provides clarity into how developers justify decisions around model safety [67]. The 2024 Foundation Model Transparency Index documents average scores of 47% and 31% for risk-related and mitigation-related transparency, respectively. California's proposed SB 1047, which was vetoed by Governor Newsom in 2024, in part aimed to provide transparency into safety practices as do the voluntary Frontier AI Safety Commitments and mandatory EU AI Act obligations.

Security practices. Transparency into the cybersecurity practices of model developers, namely their ability to secure unreleased model weights from both company-internal and company-external exfiltration, is a clear priority given that model weights are high-value assets with broad-ranging societal dependence in many downstream AI applications. Prior work [113] extensively catalogues how developers can improve cybersecurity practices for model weights, identifying several practices where it is currently unclear if leading developers do or do not implement these practices (e.g., insider threat detection, access controls, confidential computing). Weights-level cybersecurity has emerged as a clear global priority, including in the G7 International Code of Conduct and the Frontier AI Safety Commitments.

Pre-deployment testing. Transparency into pre-deployment assessments of capabilities and risks, spanning both developer-conducted and externally conducted evaluations, is vital given that these evaluations are early indicators of how models may affect society and may be interpreted (potentially undesirably) as safety assurances. Specifically related to external evaluations, clarity into the level of rigor (e.g., the time and depth of pre-deployment access provided) and independence (e.g., if the external party is paid by the model developer and whether they are constrained in what they can disclose) is crucial for ensuring the public understanding is calibrated in interpreting these pre-deployment evaluations. Recent incidents, such as the pre-deployment assessment of model capabilities of OpenAI's o3 by Epoch on the FrontierMath benchmark, indicate the importance of concretizing these norms for assessments in critical safety areas.

Downstream impact. Transparency into how foundation models are deployed across the economy is a core prerequisite for measuring the impact of AI in society, projecting how AI will affect the future of work, identifying where AI demonstrably causes harm, and grounding policymaking [22]. Namely, while foundation models are general-purpose technologies [18, 27, 46] that can, in theory, be used in almost every domain, they will in practice have a more particular usage distribution. For example, Anthropic's Economic Index shows Claude usage is concentrated on software development and technical writing tasks [61, 125]. To monitor downstream impact, attention should be allocated to entities that host models for download (to track the number of downloads) or for inference (to track the number of queries or users): For example, these entities could publish periodic reports on the adoption of foundation models throughout the economy. Critically, by not necessarily focusing on model developers but instead on distribution channels, such an approach can help track the impact of open models even though their developers may be ill-equipped to directly track impact [71]. The 2024 Foundation Model Transparency Index finds that downstream impact is the area of systemically worst transparency with developers, scoring just 15% on average. Alternative approaches include market surveillance by government agencies, as the U.K. Competition and Markets Authority has begun to explore [131].

Accounting for openness. Foundation models are released using a variety of approaches [80, 115, 118, 120], which shape how they can be accessed by entities beyond their developers [119]. The most salient distinction is based on whether the model weights are made broadly available (i.e., open-weight models like Llama 3.3 and DeepSeek-R1) or not (i.e., closed or proprietary models like Gemini 2.0 and Claude 3.7). Release strategies generate challenging trade-offs between the benefits and risks of models: Open models are often associated with benefits to innovation, competition, privacy, and science, but they may be more prime for misuse by malicious actors [71, 133]. Openness and transparency are not the same, and often are conflated with each other, but they are related. Open model weights enable deeper inspection, which is necessary for several forms of scientific research, including on safety and security [19, 71]. Beyond the direct benefits to transparency afforded by openness, open model developers have so far demonstrated greater transparency, though heightened transparency is not guaranteed by a developer's decision to make models open. The 2024 Foundation Model Transparency Index demonstrates a correlation: Open-weight model developers are more transparent on average, especially in terms of how their models are built. As a result of the benefits of open models, including on the topic of transparency, the EU AI Act grants an exemption for open models from the transparency obligations under the AI Act, except where models pose systemic risk.

3.2 Improving Third-Party Risk Assessment

Greater transparency is integral to an improved AI governance ecosystem; however, despite its considerable value, it is not sufficient without a broader ecosystem to complement it. Fundamentally, this is because the value of transparency from model developers is naturally limited by the information these developers have. For a nascent and complex technology being developed and adopted at a remarkably swift pace, developers alone are simply inadequate at fully understanding the technology and, especially, its risks and harms.

Third-party risk assessment, therefore, is essential for building a more complete evidence base on the risks of foundation models and creating incentives for developers to increase the safety of their models. At present, some prominent foundation models are subject to risk evaluations by first-party model developers or second-party contractors. In contrast, third-party evaluation affords three distinctive strengths. First, third-party evaluations have unmatched scale: Thousands of individuals are willing to engage in risk evaluation, dwarfing the scale of internal or contracted teams. Second, third-party evaluations have unmatched diversity, especially when developers primarily reflect certain demographics and geographies that are often very different from those most adversely impacted by AI. Broad demographic, institutional, and disciplinary diversity is vital for unearthing blind spots. And finally, third-party evaluation is distinctively independent: Society requires forthright and trustworthy assessments of risk, which benefits from a lack of commercial and contractual entanglement with AI developers.

A well-designed disclosure and verification regime offers significant benefits not only to the public but also to foundation model developers. By establishing industry-wide transparency standards and third-party verification mechanisms, companies can demonstrate compliance with best practices, potentially reducing their liability exposure compared to the uncertainties of purely reactive litigation. In practice, states could leverage the existing technical infrastructure and expertise of federal-level AI institutions to conduct standardized risk evaluations.

Furthermore, this approach can transform competitive dynamics around safety. When safety measures and risk assessments are publicly disclosed and verified, companies face stronger incentives to implement best practices, as deviations would be apparent and could attract greater scrutiny. This transparency coupled with third-party verification effectively creates a "race to the top" rather than a "race to the bottom" in safety practices, benefiting responsible companies while improving overall industry standards.

Researcher protections. Third-party AI evaluators require access to prominent foundation models and permission from developers to conduct risk assessments. While most major models are available in some form to the public, whether via an API (e.g. OpenAI's GPT-4o) or via their weights (e.g. Meta's Llama 3.3), the key challenge is whether evaluators are permitted to conduct risk evaluations. Prior work documents that "companies disincentivize safety research by implicitly threatening to ban independent researchers that demonstrate safety flaws in their systems." [75] These suppressive effects come about due to companies' terms of service, which often legally prohibit AI safety and trustworthiness research, in effect threatening anyone who conducts such research with bans from their platforms or even legal action [74]. In response, in March 2024, over 350 leading AI researchers and advocates signed an open letter calling for a safe harbor for independent AI evaluation.[3] Such a safe harbor, analogous to those afforded to third-party cybersecurity testers, would indemnify public interest safety research and encourage the growth of the evidence base on AI risk [82].[4]

Responsible disclosure. Beyond establishing legal protections for evaluators, responsible dissemination of findings presents another critical challenge. Developing norms around responsible risk evaluation requires the definition of rules of engagement: Which foundation models or AI systems should be assessed, and what prior notice should be given to their developers and providers in advance of broader dissemination of vulnerabilities? Beyond improving the practices of third-party evaluators, infrastructure is sorely needed to communicate identified vulnerabilities to the many affected parties. First, risks identified at the foundation model level may implicate or impact parties elsewhere in the supply chain. For example, if a text-to-image foundation model readily generates synthetic child sexual abuse imagery, upstream data sources should inspect whether real child sexual abuse images are present in their datasets, and downstream AI applications should also inspect whether these applications are susceptible to generating similar imagery [128, 129]. Second, risks at the foundation model level are often transferable: Many foundation models, potentially

[3]https://sites.mit.edu/ai-safe-harbor/

[4]To avoid confusion, we emphasize that the safe harbor described in this section affords protections to third-party researchers. This is distinct from proposals that may afford protections to other entities, such as foundation model developers.

created by different developers, may share the same vulnerability. For example, prior work has shown that jailbreaks of one language model can be reused to jailbreak many other language models [145]. For these reasons, routing mechanisms are needed to communicate third-party risk assessments to not just the foundation model developer but also the other relevant parties who may need to take actions to reduce risk. Infrastructure used to route third-party risk assessments could likely be reused to share information on adverse events as we describe in Section 4.

3.3 Protecting Whistleblowers

Most information that is shared about the practices of foundation model developers is deliberately disclosed by the developer, whether due to a voluntary choice, industry standard, or mandatory requirement. However, under exceptional circumstances, employees of the developer or other parties may disclose information. In this vein, in 2024, several current and former employees at frontier AI companies, as well as AI researchers, signed an open letter arguing for "A Right to Warn about Advanced Artificial Intelligence."[5]

Whistleblower protections are defined as legal safeguards that prohibit retaliation against individuals (generally employees) who report misconduct, illegal activities, fraud, or other wrongdoing within an organization. "Whistleblowers can play a critical role in surfacing misconduct, identifying systemic risks, and fostering accountability in AI development and deployment." [141] Based on the Whistleblower Protection Act of 1989, we frame the design space for essential elements of specific whistleblower protections in relation to foundation model developers and frontier AI [138].[6]

Who is eligible for coverage? Most existing whistleblower protections apply to all employees within an organization, especially in relation to private sector employees. In some cases, differentiation may also relate to coverage under a different whistleblower protection: The Whistleblower Protection Act applies to many federal employees, while military personnel are not covered due to protection afforded them by the Military Whistleblower Protection Act, and employees of intelligence agencies are not covered due to protections like the Intelligence Community Whistleblower Protection Act. However, a central question in the AI context is whether protections apply to additional parties, such as contractors. Broader coverage may provide stronger accountability benefits but also imposes greater cost: To extend protections to contractors and third parties, developers may need to implement additional reporting channels and legal frameworks.

Under what conditions are disclosures eligible for coverage? Different existing whistleblower protections tend to apply when two conditions are satisfied: (i) The whistleblower is blowing the whistle on appropriate topics; and (ii) the whistleblower follows established reporting protocols. In terms of the topics that qualify for protection, prior work [96], based on a survey of existing whistleblower protections across multiple jurisdictions (e.g., the United States at the federal level, the European Union), finds that many existing protections across different sectors share a focus on violations of the law. However, actions that may clearly pose a risk and violate company policies (e.g., releasing a model without following the protocol laid out in a company's safety policy) may not violate any existing laws. Therefore, policymakers may consider protections that cover a broader range of activities, which may draw upon notions of "good faith" reporting on risks found in other domains such as cybersecurity [82].

---

[5] https://righttowarn.ai/

[6] We deliberately avoid exploring the enforcement mechanisms and the interplay with other contractual obligations such as non-disclosure agreements [116], which are essential considerations in designing whistleblower policy.

In terms of the reporting protocol, disclosures may be directed at different parties (e.g., a company's board or a designated entity in the government) and have established reporting channels to facilitate the secure (and potentially anonymous) communication of information. The European Union's AI Act includes explicit provisions on whistleblower protections, which build on the EU Whistleblower Directive and provide the "Right to Lodge a Complaint with a Market Surveillance Authority" in the case of AI Act infringement. Ensuring clarity on the process for whistleblowers to safely report information can jointly advance accountability and manage countervailing interests, such as the disclosure of trade secrets or the misuse of information to compromise safety and security.

Applicability of existing whistleblower protections. Consistent with California Executive Order N-12-23, existing regulatory authority can be considered for AI-related whistleblowing. In general, existing whistleblower protections are stronger for public sector employees than private sector employees [96, 141], a fact worth considering given that most foundation models are built by private sector companies [85]. In some cases, public sector whistleblower protections may apply to private sector employees when the government procures AI from the private sector [96].

3.4 Guiding Principles

Many fundamental values and interests are often presented as being at odds with transparency. These include privacy (transparency may reveal personal information), security (transparency may reveal vulnerabilities), intellectual property and trade secrets (transparency may compromise competitive advantage), expression (reporting obligations may compel speech), and innovation (reporting obligations may delay progress). While policy in many domains actively grapples with the very real trade-offs between these values, we emphasize that foundation models are an immature technology with immature norms. In other words, we strongly believe that individual companies and the industry as a whole is not at the Pareto frontier. This means that improvements are possible to one or both values at this moment without immediately confronting a trade-off. In turn, strategic approaches can increase transparency on foundation models and their supply chain at minimal, if any, cost to these other key priorities.

Alongside concerns that transparency is harmful or bureaucratic, a separate class of concerns stems from the belief that transparency is not enough. Most sharply, transparency in itself does not guarantee accountability or redress: If AI technologies do mediate harm, transparency alone will not rectify these harms, nor will it inoculate us from future harms. Therefore, by focusing on transparency, some scholars argue this will advance "transparency washing" [143], meaning the futile pursuit of transparency without achieving more substantive change or improvement.

To abate each of these concerns, we identify matching principles: Increasing transparency, subject to the described principle, will yield Pareto improvements.

Privacy. In the context of improving transparency for foundation models and their supply chain, the vast majority of the desired information does not involve disclosing information that would undermine privacy, such as personally identifiable information. Therefore, privacy is overwhelmingly not an issue for the disclosures we discuss, though privacy could be an issue if disclosures address company-internal training data or detailed usage data.

Security. General details about risks of foundation models can be made public without undermining security, especially if these risks have been demonstrated in other foundation models or AI technologies. Specific details about concrete vulnerabilities should be disclosed carefully, with advanced notice to actors in the supply chain who are able to remediate them prior to broader disclosure.

Computer security, a closely related area, strongly suggests that "security through obscurity" will not work for AI, and that responsible, coordinated flaw disclosure can make systems and users safer.

Intellectual property and trade secrets. Information that is known by a developer's competitors can be disclosed publicly to provide public benefit with minimal, if any, cost to competitive advantage. Sensitive commercial information that is difficult to make public may also be disclosed to the government at little risk, assuming the government can provide assurance of good information security practices.

Freedom of expression. In the United States, the First Amendment's free speech clause prohibits the government from abridging the freedom of speech. While not prohibiting speech, disclosure requirements may introduce a concern of First Amendment violation by compelling speech [26]. To avoid these concerns, legal precedent indicates that mandates are more likely to be upheld as constitutional if the information is "purely factual and uncontroversial" rather than "highly subjective" and the disclosure is not "unduly burdensome" [33].

Innovation. Much of the information that would be disclosed as part of a new transparency regime is already produced by companies for their internal operations. Sharing this information, therefore, is considerably less costly because companies already possess it, and the government can decrease reporting cost by creating an effective reporting interface, minimizing duplicative or redundant reporting, and prioritizing directly actionable information. In addition, reduced overall litigation exposure through "race to the top" market dynamics will reduce holistic developer costs.

Transparency washing. Public-facing transparency makes it much more likely that information disclosure will advance accountability through the joint efforts of governments, media, civil society, and academic researchers, leveraging clear standards to facilitate accountability. A concrete plan of action for how to analyze and base future decisions on disclosed information can also ensure transparency results in tangible progress.

4 ADVERSE EVENT REPORTING

Effective AI governance requires a grounded understanding of how AI technology mediates impact in society. To better understand the practical impact and, in particular, the incidents or harms associated with AI, we need to develop better post-deployment monitoring [121]. In other domains, post-deployment monitoring by the government yields reports of medical complications, equipment malfunctions, near misses, and other unforeseen hazards. By aggregating information about incidents from developers and users, an adverse event reporting system that allows developers and downstream users to report post-deployment incidents promotes continuous learning [70], reduces information asymmetries between government and industry [127], and enables both regulators and industry to coordinate on mitigation efforts [139]. As a result, an adverse event reporting system constitutes a critical first step toward data-driven assessments of the benefits and costs of regulatory actions [59].

Adverse event reporting[7] refers to a proactive monitoring system designed to collect information about relevant events or incidents from various mandated or voluntary reporters. Adverse event reporting systems administered by the government aim to improve access to information about

[7]The term "adverse event reporting" is often applied in the context of medical or public health monitoring systems (e.g., VAERS, FAERS, MedWatch). In other domains, such as in transportation or occupational safety monitoring (e.g., CISA, NTSB safety, accident, or hazard reports), the term "incident reporting" is used. Irrespective of the name, these systems monitor for unexpected risks that can lead to early mitigation efforts.

harms to enable regulators, as well as industry actors and users, to learn about risks. The systems provide a starting point for regulation without necessarily subjecting technology, firms, or other entities to additional requirements. Rather, the goal of these regimes is to aggregate information that may be used to warn users about risks [94], inform future regulatory efforts, or develop best practices.

Implementing an adverse event reporting system requires specifying (i) the events to be reported, (ii) the information to report about these events, (iii) the entities that are mandated to report adverse events and those that are permitted to report such events, and (iv) the entities that receive information about individual events or aggregated statistics [43, 59].

- **Reportable events.** Adverse event reporting systems designate conditions under which an event should be reported, defined in terms of realized harm or other unintended/detrimental outcomes.

- **Reported information.** The system specifies what information must be shared and how it should be formatted. To appropriately incentivize reporting, some systems permit anonymized or confidential reporting. For example, the Aviation Safety Reporting System (ASRS), administered by the National Aeronautics and Space Administration (NASA) on behalf of the Federal Aviation Administration (FAA), collects confidential reports on aviation safety incidents [9].

- **Reporting entities.** In different domains, manufacturers, companies, practitioners, users, or any impacted party may be required or permitted to report. Typically, mandatory reporting requirements apply to business entities, whereas users or the public may be permitted, but not required, to report. Weighing the reliability of reports against the resources required to process submissions is critical [56].

- **Information sharing.** Information on commonly reported adverse events may be summarized and shared with enforcement agencies, non-enforcement agencies, industry actors, upstream actors, downstream actors, or the public. [124]. For example, the U.S. Cybersecurity and Information Security Agency (CISA) has implemented several information sharing programs, including the development of Automated Indicator Sharing (AIS) and coordinated vulnerability disclosure process, and participation in the Joint Cyber Defense Collaborative (JCDC) and Joint Ransomware Task Force (JRTF).

Several policy domains implement adverse event reporting systems. In transportation safety, aircraft incidents are reported to the NTSB and vehicle safety defects are reported to the NHTSA.[8] In public health, post-vaccination adverse events are reported to the Vaccine Adverse Event Reporting System jointly administered by the CDC and FDA.[9] In cybersecurity, cyber incidents and vulnerabilities are reported via programs administered by CISA [45].[10]

In the domain of AI, the Organisation for Economic Co-operation and Development (OECD) recently developed an incident reporting tool called the AI Incidents Monitor (AIM), which provides

[8]The NTSB administers reporting systems for aviation incidents, which include limited enforcement immunity to encourage voluntary reporting of safety issues without fear of punishment. The NHTSA administers an incident reporting portal for defects.

[9]The Center for Disease Control and Prevention (CDC) and the Food and Drug Administration (FDA) jointly administer the Vaccine Adverse Event Reporting System (VAERS) to collect and analyze reports of adverse events (potential side effects or reactions) that occur after vaccination.

[10]CISA administers incident reporting under the Cyber Incident Reporting for Critical Infrastructure Act of 2022 (CIRCIA) and also administers a Vulnerability Disclosure Policy (VDP) platform which enables researchers to report vulnerabilities to participating agencies.

a platform for stakeholders to see information about where AI incidents have been reported globally. AIM is powered by a machine learning model that monitors news reports about AI incidents and compiles information about the magnitude of their impacts. While it is completely voluntary and relies mostly on information reported in the press, it is an example of how to design a system to reach a broad audience. As additional efforts emerge to document post-deployment impacts, including adverse event, efforts should be made to consolidate reports to avoid reporting fatigue due to multiple fragmented processes.

4.1 Benefits

We identify four benefits that would arise from well-implemented adverse event reporting in AI, building upon the National Artificial Intelligence Advisory Committee (NAIAC) recommendation to "Improve Monitoring of Emerging Risks from AI through Adverse Event Reporting" [127].[11]

(1) **Improved identification of near-term, potentially unanticipated, sociotechnical harms.** An adverse event reporting system is able to identify new sources of risk without requiring regulators to specify the harms being monitored for *ex ante*.[12] By defining incidents in terms of unwanted, dangerous, or damaging output [48], such as events involving material harms to life, physical property, or financial interests, regulators learn about threats that materialize without having to enumerate every possible harm or source of risk. The reporting system generates data to allow regulators to perform independent risk assessments.

(2) **Potential prevention of costly accidents by encouraging proactive mitigation.** Though incident reporting systems may be costly to establish and maintain [38], there is evidence that they can prevent costly incidents or recalls. Moreover, reporting of widespread or critical issues may lead entities to take proactive, voluntary measures to mitigate risk. In particular, organizations may identify issues that, if left unaddressed, could lead to systemic failures, reputational damage, or financial liability. Additionally, the aggregated data can highlight latent vulnerabilities shared across multiple systems, allowing industry actors to collaboratively develop more robust design or operational practices [136].

(3) **Improved coordination between the government and the private sector in risk mitigation.** A major benefit of an adverse event reporting system is the ability to transmit relevant information to relevant actors quickly. This may encourage industry to adopt standardized reporting, data collection, or mitigation procedures. By standardizing reporting and data collection, the system enables regulators and industry stakeholders to detect patterns, assess systemic risks, and develop targeted responses more effectively. Structured reporting mechanisms further help to clarify compliance expectations, streamline information-sharing, and enhance regulatory oversight while reducing administrative burdens for involved parties. Over time, this coordination strengthens AI governance efforts by ensuring that both regulators and private entities operate with a shared understanding of risks which, in turn, can enable more effective and proactive responses to emerging threats.

(4) **Limited cost imposed on reporting entities and the government.** Compared with other regulatory proposals to measure and reduce risks associated with AI, adverse event

[11] A subset of the authors of this report also authored the NAIAC recommendation.

[12] We note that an adverse event reporting scheme for AI is not a substitute for other post-deployment risk assessments, like randomized control trials or large-scale studies, but rather complements these modes of assessment.

reporting may be comparatively less costly for government agencies administering the system as well as for reporting entities. Reporting entities only experience a reporting cost when an adverse event occurs, and reporting requirements can be targeted, continuously adjusted, and graduated—with the amount of required supporting documentation calibrated to the severity of the incident. Moreover, streamlining reporting forms, allowing electronic submissions through an online platform, clarifying requirements, and allowing summarization [78] can reduce the burden of reporting obligations for organizations and individuals. Government agencies may face lower operational costs due to the passive nature of the reporting system, which primarily relies on entities to submit reports rather than requiring continuous monitoring or direct oversight. The use of standardized reporting mechanisms and automated data processing can further reduce administrative costs by minimizing manual review and facilitating efficient data analysis. Additionally, agencies can leverage existing regulatory infrastructure, such as integrating the reporting system into current oversight frameworks, reducing the need for new institutional structures or extensive resource allocation.

4.2 Challenges

We identify three general challenges for the successful implementation of adverse event reporting, which we further contextualize to address the distinctive characteristics of AI.

(1) **Defining clear criteria for what events should be reported.**
First and foremost, an adverse event reporting system must provide a clear definition of reportable events. Because the capabilities and risks of frontier models are to a large extent uncertain, the reporting criteria should be broad enough to encompass unforeseen or emergent sources of risk, but sufficiently tailored to avoid complicating risk assessments [132]. In other settings, adverse events are defined in terms of unanticipated or serious health outcomes, financial consequences, or device malfunction.

(2) **Creating institutional capacity to monitor and investigate adverse events.**
An effective adverse event reporting system requires regulators to continuously monitor and assess reports of harms, which requires an ongoing commitment of resources and personnel to manage. Analyzing and investigating reports may be costly. For instance, the estimated cost for CISA to administer and review mandatory cyber-incident reports over a 10-year period is estimated to be in the billions of dollars [38]. However, the benefits, including the rapid identification of cross-system vulnerabilities and secure development [38], likely far exceed the operational costs given the wide-ranging estimates of the cost of cyber incidents. Moreover, clearly defined reporting criteria and information to be submitted can reduce the burden of investigating reports by ensuring regulators have access to key information about incidents.

(3) **Equipped and incentivized to address the underreporting of events and other non-compliance with the system.**
Some evidence suggests low rates of compliance with even mandatory adverse event reporting requirements in other contexts, such as adverse vaccine reactions [10]. As a result, adverse event reporting systems may be more effective at monitoring for common interest risks, for which reporters have a shared interest in reporting adverse events, or when combined with safe harbors or other incentives to report [9, 53]. Furthermore, in order for reports to be accurate and informative, entities must be equipped to monitor for the reportable event.

4.3 Guiding Principles

Adverse event reporting systems address a core impediment to targeted AI regulation by enabling regulators to learn about realized harms and unanticipated sources of risk. Adverse event reporting systems function well when monitoring for risks that regulators and industry have a common interest in addressing, when reports are shared with the appropriate or relevant actors, when agencies have the staff and resources to review reports, when reporting requirements are easily modified, when reporting interfaces are user-friendly [35], and when agencies leverage existing regulatory authority in early mitigation efforts. Based on our review of similar efforts for cybersecurity threats and AI incidents, including recurring challenges, we offer four targeted principles that can inform a carefully crafted, government-administered adverse event reporting system for AI.

Sharing adverse event reports with agencies with relevant domain-specific authority and expertise is crucial to collaboratively address identified harms. Providing access to adverse event reports to relevant agencies with domain-specific regulatory authority and expertise can address the information asymmetry between government and developers in terms of where different risks are occurring. Reports can be made confidential for the government to protect proprietary information within closed systems. However, it could be worth giving agencies the authority to share anonymized findings from reports with other industry stakeholders if an agency determines that other developers could benefit from a finding. In high-risk domains, like those with national security implications, information-sharing practices are common—for instance, the joint ventures, like the Joint Cyber Defense Collaborative (JCDC) and Information Sharing and Analysis Organizations (ISAO), and procedures, like the coordinated vulnerability disclosure process, facilitate access to threat indicators.

An initially narrow adverse event reporting criteria built around a tightly defined set of harms promotes consistency and comprehensiveness in reporting. Thereafter, periodic review and analysis in aggregate can inform updates to reporting criteria. Defining adverse events narrowly initially ensures reports are consistent and comprehensive: Past efforts in other domains have sometimes suffered from overly inclusive reporting criteria, which risk drowning signal in noise. In addition, we recommend that agencies responsible for managing an adverse event reporting system frequently review submissions to identify unanticipated or new sources of risk and to evaluate the need to revise reporting criteria in response to data received through any such reporting system. Ideally, adverse event reporting is an iterative learning process [63].[13] Information shared on incidents helps to enhance our understanding of how systems are functioning, but a need will remain for agencies to constantly and proactively look for new gaps in information and to identify how reporting systems can help fill those gaps. Policymakers could also consider granting agencies sufficient authority to expand the scope of reporting criteria as appropriate and adopting reporting criteria tailored to California companies.

An adverse event reporting system that combines mandatory developer reporting with voluntary user reporting maximally grows the evidence base. A hybrid model of mandatory and voluntary reporting requirements in designing an adverse event reporting system can maximize the robust evidence base necessary for adverse event reporting systems to function properly [42, 49]. For example, a system could require mandatory reporting for AI model developers that operates in tandem with voluntary reporting for downstream users.

[13] California's Data Breach Notification Law, Cal. Civ. Code § 1798.82, has been amended several times to expand the scope of covered information based on evidence of new threats.

An effective hybrid approach may place mandatory requirements on parts of the AI stack where information gathering is most critical while structuring the right incentives to encourage voluntary reporting in other places. Another component of this approach could be to incentivize manufacturers to engage in verifiable pre-market testing as a way to satisfy certain reporting requirements. For example, if a developer engages in sufficient pre-market testing and releases a summary of its results, it could then be exempted from the mandatory requirements.

Hybrid reporting systems have been used successfully to monitor for post-deployment risks. For instance, cybersecurity incident reporting systems administered by CISA follow a hybrid model: The web-based portal accepts both voluntary and mandatory reports.

Non-AI-specific regulatory authorities could be used to mitigate the risks of AI technology. Existing regulatory authority could offer a clear pathway to address risks uncovered by an adverse event reporting system, consistent with California Executive Order N-12-23. This is critical as the rapid pace of AI advancements may challenge regulators' abilities to craft regulations targeted at specific risks. To address risks and harms of AI does not necessarily require AI-specific regulatory authority or tools. In addition, reviewing existing regulatory authorities can help agencies identify regulatory gaps—in particular, what new authorities are necessary to address risks.

5 SCOPING

Well-designed regulation is proportionate: The obligations government imposes on an actor are commensurate with the associated risk to ensure regulatory burdens are such that organizations have the resources to comply and not so burdensome as to curb innovation. The United States, along with 29 other governments, committed to adequately considering a proportionate governance approach by signing the Bletchley Declaration at the 2023 U.K. AI Safety Summit: It states that "countries should consider the importance of a pro-innovation and proportionate governance and regulatory approach that maximises the benefits and takes into account the risks associated with AI. This could include making, where appropriate, classifications and categorisations of risk."

The design challenge for regulation, to first approximation, is to determine which entities must comply with which obligations. How should policymakers structure regulation to tailor the obligations to the entities? Thresholds are a standard mechanism for determining the scope of regulation. Thresholds distinguish entities to allow for heightened obligations to be applied to some entities without being applied to others. We define a threshold as a simple decision rule where (i) a quantity of interest, such as model training compute measured in floating point operations (FLOP), is compared to a value, such as 10^{26} FLOP such that (ii) entities that surpass the threshold are subject to more regulatory scrutiny than those that do not. Thresholds often exempt certain entities from burdensome compliance to foster innovation. For example, businesses with less than $500,000 in annual gross revenue do not meet the enterprise coverage threshold, exempting them from recordkeeping obligations under the U.S. Fair Labor Standards Act. Thresholds also narrow regulatory attention to a smaller set of entities that warrant greater focus or scrutiny. For example, U.S. banks with $250 billion or more in total consolidated assets are designated as "systemically important" financial institutions, subjecting them to heightened scrutiny by the Federal Reserve under the Dodd-Frank Wall Street Reform and Consumer Protection Act.

While thresholds are ubiquitous across nearly every policy domain and jurisdiction, their design and application to AI policy remains nascent. As of January 2025, AI-specific thresholds have primarily been based on the computational cost of model training, which is measured in computational operations (e.g., FLOP). Noteworthy examples of compute thresholds include 10^{26} FLOP in the Biden Executive Order 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence and 10^{25} FLOP in the European Union's AI Act. Other approaches have been considered: California's proposed and vetoed SB 1047 Safe and Secure Innovation for Frontier Artificial Intelligence Models Act used the monetary cost of training, setting the threshold at $100 million.

In general, thresholds are imperfect: Most risks vary continuously, whereas thresholds set discrete boundaries to distinguish among different entities. Thresholds are also imperfect proxies: By relying on specific, quantifiable metrics, they necessarily center what can be measured even where the underlying risk cannot be fully measured. Useful thresholds for frontier AI governance, therefore, must strike pragmatic compromises between (i) the theoretical constructs that should yield greater regulatory scrutiny and (ii) the practical operationalizations of what can be measured. Most fundamentally, the pace of technological change will require adaptive approaches that allow for thresholds to be flexible and readily updated to avoid ossification.[14]

5.1 Thresholds for Foundation Models

To proportionately regulate foundation models and the AI supply chain, policymakers must understand the many axes of variation to inform how they set thresholds. Supply chain monitoring [22, 131] documents significant variation in the organizational profile of model developers, including their corporate status, employee head count, and geographic location. These developers expend very different amounts of resources to build their models, referring to costs in money, data, computation, and energy. The resulting models also differ greatly in fundamental properties like model architectures (e.g., dense models where all model parameters are active vs. sparse models where a small subset are active) and modalities (e.g., Meta's unimodal text-to-text Llama 3.3 vs. Google's highly multimodal Gemini 2.0 that can take text, code, images, audio, and/or video as inputs and generate text, audio, and/or images as output). They vary even more in their capabilities and associated risks when evaluated on standard benchmarks. Once development is completed, models vary greatly in how they are released (e.g., EleutherAI's fully open Pythia, Meta's open-weight Llama 3.3, OpenAI's API-based GPT-4, Google's fully closed Gopher). For all of these reasons and other business reasons, foundation models vary immensely in how they are used downstream and their ultimate impact on the public.

To define a threshold to distinguish foundation model developers, we see four natural approaches:

(1) **Developer-level properties:** For example, thresholding based on the number of employees could reduce obligations for small companies that may lack the personnel to comply. As an example from another policy domain, the Occupational Safety and Health Administration Recordkeeping and Reporting Occupational Injuries and Illnesses Standard partially exempts businesses with at most 10 employees from maintaining certain logs of work-related injuries and illnesses.

(2) **Cost-level properties:** For example, thresholding based on the compute-related cost to develop a model could restrict regulatory attention to capital-intensive models given concerns of barriers to competition. As an example from another policy domain, the Clean Air Act's program for the Prevention of Significant Deterioration requires permitting for facilities that emit, or have the potential to emit, either 100 tons per year of regulated air pollutant in specific source categories (e.g., chemical plants) or 250 tons per year of regulated air pollutant in all other source categories.

[14] We focus on thresholds for foundation models but, in the design of AI, policymakers may want to differentiate the obligations for the initial in-scope developer vs. an entity downstream that modifies the model. Addressing derivative models, given the broad uptake of foundation models in the economy, along with more continual approaches to model development is essential for ensuring proportionality more broadly.

(3) **Model-level properties:** For example, thresholding based on the model's performance on benchmarks that involve identifying software vulnerabilities could identify models that may enable cyberattacks. As an example from another policy domain, the Consumer Product Safety Improvement Act uses laboratory testing for lead presence to limit lead to 100 parts per million of total lead content in any accessible part of children's products intended for children 12 years old and under.
(4) **Impact-level properties:** For example, thresholding based on the number of commercial users of a model could identify models with broad dependence that may pose systemic risk. As an example from another policy domain, the European Union's Digital Services Act imposes greater scrutiny on Very Large Online Platforms and Very Large Online Search Engines, defined as those with at least 45 million monthly active users in the EU. Impact scales not only in the amount of usage, but also the type of usage (e.g. the use of foundation models in high-impact AI systems that mediate life chances in domains such as employment, education, housing, health, and criminal justice).

To determine the appropriate approach for setting a threshold in a given policy context, we recommend that policymakers adopt the default approach of using thresholds that align with the regulatory intent. For example, if the intent is to encourage up-and-coming startups, then shielding them via exemptions for small businesses may be appropriate: This intent could be expressed via a threshold based on business-level head count or revenue. If the intent is to discourage the misuse of certain dual-use capabilities, then categorizing entities based on evaluation results for their models for the associated capabilities may be appropriate. And if the intent is to place greater scrutiny on efforts likely generating large-scale disruption to societal operation and/or large-scale harm from the adoption of their technologies, then an impact-level designation may be appropriate.

While the principle of aligning the regulatory structure (i.e., the methods for thresholding) with the regulatory intent should be the starting point, practical concerns should also be considered. In practice, some of the quantities we describe above may be proxies for others. In several cases, policymakers have already used training compute (a cost-level metric) as a proxy for some combination of greater capabilities, greater misuse risks, and/or greater downstream harm.

While a number of additional considerations influence threshold design [see 16, 65], we highlight three:

(1) **Determination time:** Different quantities first become measurable at different times. While a foundation model developer may estimate training compute even before training for internal budgeting purposes, the evaluation results for a model can only be determined after development and, still further, the downstream impact for a model can only be determined after deployment. In some cases, estimates can be prepared in advance, but these temporal effects have two key consequences. First, developers can better prepare to comply with policy if they are aware earlier that they will be in-scope. And second, certain obligations are difficult to impose if they require intervention before the determination time: It is difficult to expect a developer to implement certain governance practices for training data if they only determine they exceed a threshold after deployment.
(2) **Measurability:** Different quantities have differing complexity in how they are measured, spanning both the intrinsic difficulty of measuring the quantity and overarching standardization in how to measure the quantity. In particular, absent significant agreement on how to measure the quantity of interest, progress in implementation may stall as stakeholders disagree on the specifics of measurement.
(3) **External verifiability:** Different quantities vary in whether external parties of any kind are able to measure them. Prior AI policy in New York highlights that if the regulated entity is the

one to unilaterally determine if they are in scope of the policy, then the policy may fail to have the desired effect due to widespread null compliance (Wright et al., 2024). For example, risk evaluation results may be externally reproduced to confirm developer's designation, whereas training compute may require the ability to monitor the compute resources developers use, and training data size may be entirely unobservable given the lack of access to training data.

Since policy may have different regulatory intents and existing thresholds vary in their profiles of determination time, measurability, and external verifiability, we agree with Nelson et al. [90] that "a one-size-fits-all approach or a single threshold metric is inadequate for governance because different AI systems and their outputs present unique challenges and risks." To this end, we point to the European Union's AI Act, which designates models trained with 10^{25} FLOP as posing systemic risk as of March 2025 as the default criteria. However, the AI Act in Annex XIII affords the regulator flexibility to also consider alternatives metrics, such as the number of parameters, size of the data set, estimated cost or time of training, estimated energy consumption, benchmarks and evaluations of capabilities of the model, and whether the model has a high impact on the internal market due to its reach (either due to at least 10,000 registered business users or the number of registered end users). Further, to capture fast-moving scientific developments, the AI Act creates a scientific panel that is empowered to issue qualified alerts to identify models that may pose systemic risk even if they are not captured by predefined quantitative thresholds.

Overall, we emphasize that irrespective of the combination of metrics deemed most appropriate in the present, policymakers should ensure that mechanisms exist not only to update specific quantitative values, given the rapid pace of technological and societal change in AI, but also to change the metrics altogether.

5.2 Guiding Principles

Generic developer-level thresholds seem to be generally undesirable given the current AI landscape. Since many small entities can develop hugely influential and potentially risky foundation models, as demonstrated by the Chinese company DeepSeek, the use of thresholds based on developer-level properties may inadvertently ignore key players. In fact, major players in the space, such as Anthropic, OpenAI, and xAI, may be relatively small according to some conventional metrics of businesses (e.g., head count). At the same time, these approaches may bring into scope massive, established companies in other industries that are simply exploring the use of AI since thresholds based on properties of companies may not distinguish between the entire business and the AI-specific subset. Therefore, we caution against the use of customary developer-level metrics that do not consider the specifics of the AI industry and its associated technology.

Cost-level thresholds such as training compute should not be used alone. Training compute is by far the most common metric for thresholds in existing AI policy given its clear strengths [65]: (i) Fairly authoritative methods exist for measuring compute [31, 50]; (ii) Compute can be estimated prior to the development of the model; and (iii) Compute may be externally verifiable by entities that monitor or operate compute facilities. However, training compute has clear deficits that are well-established: (i) Models across different modalities currently use vastly different amounts of training compute; (ii) Thresholds for training compute may need to be updated very frequently due to rapid improvements in algorithmic efficiency; and (iii) Training compute has limited predictive power as a proxy for risk because it does not factor in how models are distributed and adopted in the economy [16, 68, 90]. These concerns are exacerbated by recent developments such as (iv) the growing shift from training-time compute allocation to inference-time compute allocation and

(v) the complex accounting required to address rapidly evolving technical practices such as model distillation and the use of large-scale synthetic data.

On this basis, we conclude that if training compute thresholds are used at all, they may function best when used as an initial filter to cheaply screen for entities that may warrant greater scrutiny [16, 65, 68, 90]. More broadly, within the family of cost-level metrics, training compute may still be the most attractive option, given other options are difficult to calibrate across different developers (e.g. dataset size, monetary cost), but they should actively monitor the rapid changes in how compute is allocated toward foundation model development and deployment.

Thresholds based on risk evaluation results and observed downstream impact are promising for safety and corporate governance policy, but they have practical issues. Policy that aims to manage risk by influencing developer practices often addresses one or both of (i) misuse of foundation models by malicious actors and (ii) sociotechnical harms from widespread deployment. Therefore, these intents align directly with evaluations of risk (e.g., model performance at generating child sexual abuse material or cyberattacks) as well as the downstream footprint of models (e.g., if models are integrated into high-stakes decision-making systems for hiring or benefits determination).

However, the question of which risk evaluations are sufficiently trustworthy looms in the air: The pre-deployment testing conducted by the U.S. AI Safety Institute and U.K. AI Security Institute may provide some guidance on this. In addition, industry-led safety frameworks, which articulate explicit thresholds of risk based on evidence of measured capabilities, may offer a pathway for industry consensus. Agreement on which risks should be tracked and how they can be measured through evaluations, ideally harmonized to level the playing field through third-party auditing, offers an industry-inspired pathway that can inform governance.

On downstream impact, model developers may not have this information, especially in the case of open models, so other mechanisms for market monitoring may be required instead. Overall, policy can actively develop the infrastructure to turn these approaches into fully viable options in threshold design.

6 SUMMARY OF FEEDBACK AND CHANGES

A draft of this report was shared publicly online in March 2025, along with a form (Appendix A) to encourage feedback and input from a wide range of experts. The Working Group deeply appreciates the individuals and organizations from numerous disciplines and sectors who participated by sharing valuable insights, references, and suggestions, including, in some cases, by making their responses public to contribute to the broader discourse. We received over 60 feedback submissions, which supplement the Working Group's internal review process and initial round of external expert review (see Acknowledgements on pg. 2). Three sectors (academia, industry/business/private sector, and NGO/civil society) contributed the vast majority of feedback, with each contributing at least 25% of all submissions.

The feedback reflects a range of views and fundamental questions that can inform California AI policy and governance. Feedback addressed clarifications on report content, critiques of substantive content, and future directions for further inquiry. We summarize this feedback below without individual-level attribution, consistent with the privacy expectations in the feedback form (Appendix A). Finally, we explain how we incorporated this feedback into the final report.

6.1 Clarifications on Report Content

Several feedback responses indicated the desire for clarity on concepts and definitions introduced in the draft report. Other feedback indicated that the overarching remit of the report was unclear.

Clarity on the concept of evidence-based policy. Comments indicated confusion on the overarching frame of evidence-based policy and the discussion of evidence-generating policy. Often this confusion rested on either (i) how evidence-based policy is understood given current incentives that yield the present information asymmetries between foundation model developers, governments, and the public or (ii) how evidence contributes to improved accountability.

Clarity on the current state of AI. Comments identified current developments in the AI ecosystem that went beyond the report's description of current capabilities, risks, and practices. In particular, several of these comments addressed the current evidence for risks, the growth of safety frameworks and cases, and the adherence of companies to voluntary commitments.

Clarity on the specific definitions. Comments on definitions primarily indicated confusion around the use of several closely related terms to describe current AI technologies (e.g., frontier AI, foundation models, artificial general intelligence), though other feedback appreciated the granularity of the terminology. Other comments addressed the more general definitions for artificial intelligence, AI model, and AI system, as well additional jargon (e.g., marginal risk, see §1.3.2).

Clarity on the remit for the report. Comments often expressed uncertainty about or provided recommendations for the report scope, the report purpose, and the Working Group remit. More specifically, several comments addressed the unique position of impact California holds in the ecosystem as the home to many leading AI companies, a significant fraction of the AI workforce, and past governance initiatives on digital technology. Several comments, which were notably received prior to recent developments at the federal level, discussed the importance of California to the U.S. AI ecosystem, advocating to avoid overregulation and instead create positive incentives for companies to compete on safety. Other comments addressed concerns of fragmentation with efforts in other jurisdictions, such as existing and ongoing legislative efforts in other states, and opportunities for harmonization, including with the EU AI Act and international standards. Finally, some comments touched on the impact of California's AI policy on U.S. national competitiveness and the broader geopolitical environment.

6.2 Critique of Substantive Content

The draft report's core sections received significant support, which we do not summarize for brevity. Instead, we summarize targeted critiques, which we highlight as valuable alternative perspectives. Although the Working Group maintains different views on many of these critiques, they offer important alternative perspectives that enhance the overall discussion and merit continued public discourse.

Critique of context (§2). Some comments indicated that AI companies take more voluntary action than what is seen in the discussed cases: They promote AI safety and transparency by publishing safety reports and systems cards, releasing open-weight models, issuing responsible scaling frameworks, soliciting and managing red-teaming efforts, and developing model documentation standards. Therefore, these companies are more cooperative and proactive in safety compliance compared to tobacco and fossil fuel companies. Other comments suggested that the cases generally describe policy successes, but not policy failures, and proposed other cases that could inform California's AI policy.

Critique of transparency (§3). Several comments indicated that transparency should be purposeful by clarifying how disclosures relate to accountability and being aware of the incentives that shape the quality of the disclosed information. Some comments recommended focusing on specific implementation rather than broader guidance, on more restricted non-public disclosures including via redaction, and on the costs of transparency. Specific attention focused on developing the discussion of whistleblower protections, including their relationship to existing protections. Specific attention was also focused on third-party evaluations (sometimes referred to as audits in the feedback) by highlighting the immature third-party evaluation ecosystem, the lack of evaluation standards and broader measurement science, and the current barriers for evaluators to acquire access to the AI models and systems they evaluate.

Critique of adverse event reporting (§4). Several comments questioned the absence of precise definitions for key terms (e.g., adverse event, incident, harm, unacceptable behavior) or the detailed process for adverse event reporting. Some comments identified practical challenges as both hurdles for a statewide reporting database and costs for reporting entities. Others described more participatory approaches to reporting to maximize data collection while minimizing compliance burden, including via collaboration with existing efforts.

Critique of scoping (§5). Many comments emphasized the importance of adaptive thresholds, indicating consensus that thresholds cannot be static but change to reflect evolving context. Many comments specifically addressed compute thresholds and expressed diverging views: Some argued compute thresholds are useful and necessary, while others argued for a stronger dismissal of compute thresholds given their limits as a proxy for risk, though all comments acknowledged that compute thresholds have significant flaws. Some comments addressed thresholds we did not discuss, such as those based on more qualitative criteria, or the broader relationship between scoping thresholds and policy interventions in the form of red lines.

6.3 Future Directions for Further Inquiry

The draft report's scope was determined by the Working Group's interpretation of the request from California Governor Gavin Newsom. We summarize feedback on out-of-scope topics.

Downstream AI applications. Many comments discussed specific application areas for foundation models as part of the increasing deployment of AI across economic sectors. Some comments advocated for use case-based regulatory efforts, especially in domains that pose critical risks or use methods customary to product safety, while other comments advocated for greater focus on specific deployment contexts and resulting AI systems. The feedback also discussed the need to better understand the economic impacts of AI within specific downstream domains, including to understand the implications for the future of work and the labor economy.

The AI talent pipeline. Several comments discussed the need to focus on preparedness for AI developments across the state's population, especially for workers and students. These comments suggested initiatives to improve workforce training for workers across the AI supply chain, including via collaboration with the state's educational institutions. Separately, other comments discussed the spillover effects of AI on educational pipelines, questioning the impacts on the quality of student education in the presence of widespread AI technology and discussing the procurement opportunities for AI technologies within the education sector.

Resourcing the AI ecosystem. Several comments stressed the importance of public funds to advance research on AI, highlighting the strengths of California's research institutions, the initiatives in other states such as New York's Empire AI, and the deficits in access to compute for researchers. A number of comments more broadly discussed the resources for the entire innovation ecosystem, including some that specifically focused on venture capital and the importance of policy as a fuel for entrepreneurial startups. Other comments discussed how initiatives to resource the ecosystem could be used to implement more responsible practices.

The governance of open and closed models. Many comments mentioned the distinction between open and closed models in the AI ecosystem. Some highlighted that the open release of foundation models, including via broader open-source development practices, is critical for both spurring innovations and increasing safety and transparency. Others highlighted the risks of open models. Overall, many comments on this topic recognized both the benefits and risks of openness, referencing the National Telecommunications and Information Administration report [NTIA; 133] as a key starting point for governance on this topic.

Regulatory capacity on AI. Several comments raised concerns about California's capacity to regulate AI, highlighting challenges with past digital technology policy. These comments questioned the technical expertise on AI within the government and described pathways for improving state capacity. Other comments indicated that current regulatory capacity may render initiatives to design guardrails ineffective or outright counterproductive: Many of these comments focused on flexible policy design, but a few expressed pessimism about the viability of meaningful oversight.

6.4 Changes to the Report

Based on further review, feedback, and the rapid pace of change in the field of frontier AI, the authors revised the draft report to produce the final report. We summarize the substantive changes.

Changes to introduction (§1). This section was updated to provide a more rigorous discussion of the relationship between state policy on frontier AI and policy in other jurisdictions, especially the federal government. In addition, the updated section articulates how California frontier AI policy can navigate challenging geopolitical realities, leveraging excellent quality and safety standards to strengthen national competitiveness. Finally, multiple paragraphs which discuss new empirical evidence for frontier AI risks were added.

Changes to context (§2). This section was updated to include a more rigorous discussion of regulatory incentive structures from other cases, including both positive inducements and deterrents. The updated section more directly addresses the relationship between state and federal regulation in historical context, including updating multiple case studies with relevant state-federal dynamics. The revised section emphasizes building safety into initial technological frameworks through proactive risk assessment and mitigation strategies. Further changes include greater emphasis on aligning incentives for transparency and accountability not just for frontier AI developers but all relevant actors in the ecosystem, building evidence-generating policy that accounts for internal AI deployments, and articulating how policy can help industry navigate competing incentives, all of which enhance the foundation for a "trust but verify" approach.

Changes to transparency (§3). This section was updated to strengthen the discussion of openness in relation to transparency, acknowledge First Amendment challenges to mandatory disclosures, and emphasize that transparency approaches are most valuable when they are purposeful, actionable, and outcome focused. Additional updates include references to regulatory harmonization, post-deployment economic monitoring, safety cases, whistleblower protections for non-employees, how states could leverage the existing technical infrastructure and expertise, and clarification that safe harbors discussed relate to third-party researchers and not AI companies.

Changes to adverse event reporting (§4). This section was updated to further emphasize the benefits of an initially narrow definition for adverse events to avoid a deluge of reports that masks the signal in the reporting system. Additional updates include mentioning the possibility for reporting fatigue and immunity provisions in the Aviation Safety Reporting System.

Changes to scoping (§5). This section was updated to strengthen the critique of thresholding based exclusively on training compute and to emphasize the need for updating given rapid technological change. Additional updates include mentions of impact beyond reach, distillation, algorithmic efficiency, and the complexities of proportionately addressing model derivatives.

REFERENCES

[1] Hacks of opm databases compromised 22.1 million people, federal authorities say.

[2] 1991. United States v. Morris. United States District Court for the Southern District of New York, 928 F. Supp. 436.

[3] 2024. DPR 23-002 - Civil Penalty Actions by Commissioners.

[4] U.S. Environmental Protection Agency. 2013. Pesticide Emergency Exemptions.

[5] Dario Amodei. 2025. The Urgency of Interpretability.

[6] Anthropic. 2025. Claude 3.7 Sonnet System Card. Technical report, Anthropic.

[7] Anthropic. 2025. System Card: Claude Opus 4 & Claude Sonnet 4. Technical report, Anthropic.

[8] Governors Highway Safety Association. 2024. Seat Belt Use.

[9] Aviation Safety Reporting System. ASRS: The Case for Confidential Incident Reporting Systems.

[10] Chantelle Bailey, David Peddie, Maeve E. Wickham, Katherin Badke, Serena S. Small, Mary M. Doyle-Waters, Ellen Balka, and Corinne M. Hohl. 2016. Adverse Drug Event Reporting Systems: A Systematic Review. *British Journal of Clinical Pharmacology*, 82(1):17–29.

[11] Bowen Baker, Joost Huizinga, Leo Gao, Zehao Dou, Melody Y Guan, Aleksander Madry, Wojciech Zaremba, Jakub Pachocki, and David Farhi. 2025. Monitoring Reasoning Models for Misbehavior and the Risks of Promoting Obfuscation. Technical report, OpenAI.

[12] Josh Becker. 2024. SB942 – California AI Transparency Act.

[13] Yoshua Bengio, Geoffrey Hinton, Andrew Yao, Dawn Song, Pieter Abbeel, Trevor Darrell, Yuval Noah Harari, Ya-Qin Zhang, Lan Xue, Shai Shalev-Shwartz, Gillian Hadfield, Jeff Clune, Tegan Maharaj, Frank Hutter, Atılım Güneş Baydin, Sheila McIlraith, Qiqi Gao, Ashwin Acharya, David Krueger, Anca Dragan, Philip Torr, Stuart Russell, Daniel Kahneman, Jan Brauner, and Sören Mindermann. 2024. Managing extreme AI risks amid rapid progress. *Science*, 384(6698):842–845.

[14] Yoshua Bengio, Sören Mindermann, Daniel Privitera, Tamay Besiroglu, Rishi Bommasani, Stephen Casper, Yejin Choi, Philip Fox, Ben Garfinkel, Danielle Goldfarb, Hoda Heidari, Anson Ho, Sayash Kapoor, Leila Khalatbari, Shayne Longpre, Sam Manning, Vasilios Mavroudis, Mantas Mazeika, Julian Michael, Jessica Newman, Kwan Yee Ng, Chinasa T. Okolo, Deborah Raji, Girish Sastry, Elizabeth Seger, Theodora Skeadas, Tobin South, Emma Strubell, Florian Tramèr, Lucia Velasco, Nicole Wheeler, Daron Acemoglu, Olubayo Adekanmbi, David Dalrymple, Thomas G. Dietterich, Edward W. Felten, Pascale Fung, Pierre-Olivier Gourinchas, Fredrik Heintz, Geoffrey Hinton, Nick Jennings, Andreas Krause, Susan Leavy, Percy Liang, Teresa Ludermir, Vidushi Marda, Helen Margetts, John McDermid, Jane Munga, Arvind Narayanan, Alondra Nelson, Clara Neppel, Alice Oh, Gopal Ramchurn, Stuart Russell, Marietje Schaake, Bernhard Schölkopf, Dawn Song, Alvaro Soto, Lee Tiedrich, Gaël Varoquaux, Andrew Yao, Ya-Qin Zhang, Olubunmi Ajala, Fahad Albalawi, Marwan Alserkal, Guillaume Avrin, Christian Busch, André Carlos Ponce de Leon Ferreira de Carvalho, Bronwyn Fox, Amandeep Singh Gill, Ahmet Halit Hatip, Juha Heikkilä, Chris Johnson, Gill Jolly, Ziv Katzir, Saif M. Khan, Hiroaki Kitano, Antonio Krüger, Kyoung Mu Lee, Dominic Vincent Ligot, José Ramón López Portillo, Oleksii Molchanovskyi, Andrea Monti, Nusu Mwamanzi, Mona Nemer, Nuria Oliver, Raquel Pezoa Rivera, Balaraman Ravindran, Hammam Riza, Crystal Rugege, Ciarán Seoighe, Jerry Sheehan, Haroon Sheikh, Denise Wong, and Yi Zeng. 2025. International AI Safety Report. Technical Report DSIT 2025/001.

[15] Kayla Blomquist and Keegan McBride. 2025. Shaping the World's AI Future: How the U.S. and China Compete to Promote Their Digital Visions. *Just Security*.

[16] Rishi Bommasani. 2023. Drawing Lines: Tiers for Foundation Models.

[17] Rishi Bommasani, Sanjeev Arora, Jennifer Chayes, Yejin Choi, Mariano-Florentino Cuéllar, Li Fei-Fei, Daniel E. Ho, Dan Jurafsky, Sanmi Koyejo, Hima Lakkaraju, Arvind Narayanan, Alondra Nelson, Emma Pierson, Joelle Pineau, Scott Singer, Gaël Varoquaux, Suresh Venkatasubramanian, Ion Stoica, Percy Liang, and Dawn Song. 2025. Advancing Science- and Evidence-based AI Policy.

[18] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dorottya Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin,

Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258.*

[19] Rishi Bommasani, Sayash Kapoor, Kevin Klyman, Shayne Longpre, Ashwin Ramaswami, Daniel Zhang, Marietje Schaake, Daniel E. Ho, Arvind Narayanan, and Percy Liang. 2024. Considerations for Governing Open Foundation Models. *Science*, 386(6718):151–153.

[20] Rishi Bommasani, Kevin Klyman, Sayash Kapoor, Shayne Longpre, Betty Xiong, Nestor Maslej, and Percy Liang. 2024. The 2024 Foundation Model Transparency Index.

[21] Rishi Bommasani, Kevin Klyman, Shayne Longpre, Betty Xiong, Sayash Kapoor, Nestor Maslej, Arvind Narayanan, and Percy Liang. 2024. Foundation Model Transparency Reports.

[22] Rishi Bommasani, Dilara Soylu, Thomas Liao, Kathleen A. Creel, and Percy Liang. Ecosystem Graphs: The Social Footprint of Foundation Models.

[23] Alexander Bondarenko, Denis Volk, Dmitrii Volkov, and Jeffrey Ladish. 2025. Demonstrating Specification Gaming in Reasoning Models. Technical report, Palisades Research, Berkeley, United States of America.

[24] Alex Bores. NY State Assembly Bill 2025-A6453A.

[25] Allan M. Brandt. 2012. Inventing Conflicts of Interest: A History of Tobacco Industry Tactics. *American Journal of Public Health*, 102(1):63–71. Number: 1.

[26] Valerie C. Brannon, Victoria L. Killion, Whitney K. Novak, and L. Paige Whitaker. 2023. First Amendment Limitations on Disclosure Requirements.

[27] Timothy F. Bresnahan and M. Trajtenberg. 1995. General Purpose Technologies 'Engines of Growth'? *Journal of Econometrics*, 65(1):83–108.

[28] Ross C Brownson, Jamie F Chriqui, and Katherine A Stamatakis. 2009. Understanding Evidence-Based Public Health Policy. *American Journal of Public Health*, 99(9):1576–1583.

[29] California Government Operations Agency. 2024. State of California GenAI Guidelines for Public Sector Procurement, Uses and Training.

[30] Nancy Cartwright and Jeremy Hardie. 2012. *Evidence-Based Policy: A Practical Guide to Doing It Better.* Oxford University Press.

[31] Stephen Casper, Luke Bailey, and Tim Schreier. 2025. Practical Principles for AI Cost and Compute Accounting.

[32] Stephen Casper, David Krueger, and Dylan Hadfield-Menell. 2025. Pitfalls of Evidence-Based AI Policy.

[33] Center for Democracy & Technology. First Amendment Tech Transparency Roadmap.

[34] David D. Clark, John Wroclawski, Karen R. Sollins, and Robert Braden. 2002. Tussle in Cyberspace: Defining Tomorrow's Internet. In *Proceedings of the 2002 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, pages 347–356. ACM.

[35] Valeri Craigle. 2007. MedWatch: The FDA Safety Information and Adverse Event Reporting Program. *Journal of the Medical Library Association*, 95(2):224.

[36] Kirk A. Cullimore and Jefferson Moss. 2024. SB0149 Artificial Intelligence Amendments.

[37] Mariano-Florentino Cuéllar, Jeff Dean, Finale Doshi-Velez, John Hennessy, Andy Konwinski, Sanmi Koyejo, Pelonomi Moiloa, Emma Pierson, and David Patterson. 2024. Shaping AI's Impact on Billions of Lives.

[38] U.S. Cybersecurity and Infrastructure Security Agency (CISA). 2024. Cyber Incident Reporting for Critical Infrastructure Act (CIRCIA) Reporting Requirements.

[39] Google DeepMind. 2025. Gemini 2.5 Pro Preview Model Card. Technical report, Google DeepMind.

[40] Daniel Didech. Illinois General Assembly - Full Text of HB3506.

[41] Renée DiResta, Laura Edelson, Brendan Nyhan, and Ethan Zuckerman. 2022. It's Time to Open the Black Box of Social Media. *Scientific American.* Social media companies need to give their data to independent researchers to better understand how to keep users safe.

[42] Ren Bin Lee Dixon and Heather Frase. 2024. An Argument for Hybrid AI Incident Reporting: Lessons Learned from Other Incident Reporting Systems. *Georgetown Center for Security and Emerging Technology.*

[43] Ren Bin Lee Dixon and Heather Frase. 2025. AI Incidents: Key Components for a Mandatory Reporting Regime. *Georgetown Center for Security and Emerging Technology.*

[44] Richard Doll and Donald Bradford. 1950. The Mortality of Doctors in Relation to Smoking. *British Medical Journal*, 2(4682):739–748.

[45] Justin Doubleday. 2023. CISA Platform Helps Agencies Uncover More Than 1,000 Cyber Vulnerabilities.

[46] Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock. 2023. GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models.

[47] Ann Florini. 2007. *The Right to Know: Transparency for an Open World.* Columbia University Press.

[48] Office for Human Research Protections. Reviewing and Reporting Unanticipated Problems Involving Risks to Subjects or Others and Adverse Events: OHRP Guidance (2007).
[49] Heather Frase. 2023. AI Incident Collection: An Observational Study of the Great AI Experiment. *Georgetown Center for Security and Emerging Technology*.
[50] Frontier Model Forum. 2024. Issue Brief: Measuring Training Compute.
[51] S. Rahmstorf G. Supran and N. Oreskes. 2023. Assessing ExxonMobil's Global Warming Projections. *Science*, 379(6628).
[52] Iason Gabriel, Arianna Manzini, Geoff Keeling, Lisa Anne Hendricks, Verena Rieser, Hasan Iqbal, Nenad Tomašev, Ira Ktena, Zachary Kenton, Mikel Rodriguez, Seliem El-Sayed, Sasha Brown, Canfer Akbulut, Andrew Trask, Edward Hughes, A. Stevie Bergman, Renee Shelby, Nahema Marchal, Conor Griffin, Juan Mateos-Garcia, Laura Weidinger, Winnie Street, Benjamin Lange, Alex Ingerman, Alison Lentz, Reed Enger, Andrew Barakat, Victoria Krakovna, John Oliver Siy, Zeb Kurth-Nelson, Amanda McCroskery, Vijay Bolina, Harry Law, Murray Shanahan, Lize Alberts, Borja Balle, Sarah de Haas, Yetunde Ibitoye, Allan Dafoe, Beth Goldberg, Sébastien Krier, Alexander Reese, Sims Witherspoon, Will Hawkins, Maribeth Rauh, Don Wallace, Matija Franklin, Josh A. Goldstein, Joel Lehman, Michael Klenk, Shannon Vallor, Courtney Biles, Meredith Ringel Morris, Helen King, Blaise Agüera y Arcas, William Isaac, and James Manyika. 2024. The Ethics of Advanced AI Assistants.
[53] Yi Gao, Yang Hao, Sen Wang, and Hao Wu. 2021. The Dynamics Between Voluntary Safety Reporting and Commercial Aviation Accidents. *Safety Science*, 141:105351.
[54] John Gerring. 2004. What Is a Case Study and What Is It Good For? *American Political Science Review*, 98(2):341–354.
[55] John Gerring. 2007. *Case Study Research: Principles and Practices*. Cambridge University Press, Cambridge, UK.
[56] Kenneth A Getz, Stella Stergiopoulos, and Kenneth I Kaitin. 2014. Evaluating the Completeness and Accuracy of MedWatch Data. *American Journal of Therapeutics*, 21(6):442–446.
[57] Ryan Greenblatt, Carson Denison, Benjamin Wright, Fabien Roger, Monte MacDiarmid, Sam Marks, Johannes Treutlein, Tim Belonax, Jack Chen, David Duvenaud, Akbir Khan, Julian Michael, Sören Mindermann, Ethan Perez, Linda Petrini, Jonathan Uesato, Jared Kaplan, Buck Shlegeris, Samuel R. Bowman, and Evan Hubinger. 2024. Alignment Faking in Large Language Models.
[58] Mark A. Griffin, Amy Williams-Derry, Daniel P. Mensher, Alison S. Gaffney, and Keller Rohrback. 2019. Brief of Amici Curiae Robert Brulle, Center for Climate Integrity, et al. Technical report, United States Court of Appeals for the Fourth Circuit Appeal: 19-1644.
[59] Neel Guha, Christie M. Lawrence, Lindsey A. Gailmard, Kit T. Rodolfa, Faiz Surani, Rishi Bommasani, Inioluwa Deborah Raji, Mariano-Florentino Cuéllar, Colleen Honigsberg, Percy Liang, and Daniel E. Ho. 2023. AI Regulation Has Its Own Alignment Problem: The Technical and Institutional Feasibility of Disclosure, Registration, Licensing, and Auditing. *George Washington Law Review, Symposium on Legally Disruptive Emerging Technologies*.
[60] Jasper Götting, Pedro Medeiros, Jon G Sanders, Nathaniel Li, Long Phan, Karam Elabd, Lennart Justen, Dan Hendrycks, and Seth Donoughe. 2025. Virology Capabilities Test (VCT): A Multimodal Virology Q&A Benchmark.
[61] Kunal Handa, Alex Tamkin, Miles McCain, Saffron Huang, Esin Durmus, Sarah Heck, Jared Mueller, Jerry Hong, Stuart Ritchie, Tim Belonax, Kevin K. Troy, Dario Amodei, Jared Kaplan, Jack Clark, and Deep Ganguli. 2025. Which Economic Tasks Are Performed with AI? Evidence from Millions of Claude Conversations.
[62] Russell Hardin. 2002. *Trust and Trustworthiness*. Russell Sage Foundation.
[63] Kamala D Harris. California Data Breach Report.
[64] Melissa Heikkilä. 2023. It's High Time for More AI Transparency. *MIT Technology Review*.
[65] Lennart Heim and Leonie Koessler. 2024. Training Compute Thresholds: Features and Functions in AI Regulation.
[66] Nell Henderson. 1984. N.Y. Is First State To Get Seat Belt Law. *The Washington Post*.
[67] Benjamin Hilton, Marie Davidsen Buhl, Tomek Korbak, and Geoffrey Irving. 2025. Safety Cases: A Scalable Approach to Frontier AI Safety.
[68] Sara Hooker. 2024. On the Limitations of Compute Thresholds as a Governance Strategy.
[69] American Petroleum Institute. 1980. *Two Energy Futures: A National Choice for the 80s*. American Petroleum Institute, Washington, D.C.
[70] Chris W Johnson. 2003. *The Failure of Safety-Critical Systems: A Handbook of Accident and Incident Reporting*. Glasgow University Press.
[71] Sayash Kapoor, Rishi Bommasani, Kevin Klyman, Shayne Longpre, Ashwin Ramaswami, Peter Cihon, Aspen K Hopkins, Kevin Bankston, Stella Biderman, Miranda Bogen, Rumman Chowdhury, Alex Engler, Peter Henderson, Yacine Jernite, Seth Lazar, Stefano Maffulli, Alondra Nelson, Joelle Pineau, Aviya Skowron, Dawn Song, Victor Storchan, Daniel Zhang, Daniel E. Ho, Percy Liang, and Arvind Narayanan. 2024. On the Societal Impact of Open Foundation Models. In *Forty-First International Conference on Machine Learning*.
[72] Sayash Kapoor, Benedikt Stroebl, Zachary S. Siegel, Nitya Nadgir, and Arvind Narayanan. 2024. AI Agents That Matter.

[73] Ian Klaus, Mark Baldassare, Marissa Jordan, and Hanan Coronado. 2023. 2023 Carnegie California Global Affairs Survey. Technical report, Carnegie Endowment for International Peace.

[74] Kevin Klyman. 2024. Acceptable Use Policies for Foundation Models.

[75] Kevin Klyman, Sayash Kapoor, and Shayne Longpre. 2024. A Safe Harbor for AI Researchers: Promoting Safety and Trustworthiness Through Good-Faith Research.

[76] Scott Kohler. 2025. Technology Federalism: U.S. States at the Vanguard of AI Governance. Technical report, Carnegie Endowment for International Peace.

[77] Noam Kolt, Markus Anderljung, Joslyn Barnhart, Asher Brass, Kevin Esvelt, Gillian K. Hadfield, Lennart Heim, Mikel Rodriguez, Jonas B. Sandbrink, and Thomas Woodside. 2024. Responsible Reporting for Frontier AI Development.

[78] Jennifer G Le-Rademacher, Shauna Hillman, Elizabeth Storrick, Michelle R Mahoney, Peter F Thall, Aminah Jatoi, and Sumithra J Mandrekar. 2020. Adverse Event Burden Score-A Versatile Summary Measure for Cancer Clinical Trials. *Cancers*, 12:3251.

[79] Jade Leung. 2019. *Who Will Govern Artificial Intelligence? Learning from the History of Strategic Politics in Emerging Technologies.* PhD thesis, University of Oxford, Oxford, UK.

[80] Percy Liang, Rishi Bommasani, Kathleen A. Creel, and Rob Reich. 2022. The Time Is Now to Develop Community Norms for the Release of Foundation Models.

[81] Theodore A. Linden. 1976. Operating System Structures to Support Security and Reliable Software. *ACM Comput. Surv.*, 8(4):409–445. Number: 4.

[82] Shayne Longpre, Sayash Kapoor, Kevin Klyman, Ashwin Ramaswami, Rishi Bommasani, Borhane Blili-Hamelin, Yangsibo Huang, Aviya Skowron, Zheng-Xin Yong, Suhas Kotha, Yi Zeng, Weiyan Shi, Xianjun Yang, Reid Southen, Alexander Robey, Patrick Chao, Diyi Yang, Ruoxi Jia, Daniel Kang, Sandy Pentland, Arvind Narayanan, Percy Liang, and Peter Henderson. 2024. A Safe Harbor for AI Evaluation and Red Teaming.

[83] Michelle Lopes Maldonado, Jackie Hope Glass, C. E. (Cliff) Hayes, Irene Shin, Bonita G. Anthony, Alex Q. Askew, Katrina E. Callsen, Nadarius E. Clark, Laura Jane Cohen, Joshua G. Cole, Kelly K. Convirs-Fowler, Michael Feggans, Rozia A. Henson, Jr., Phil M. Hernandez, Charniele L. Herring, Karen Keys-Gamarra, Amy J. Laufer, Destiny L. LeVere Bolling, Adele Y. McClure, Marcia S. (Cia) Price, Holly M. Seibold, Shelly A. Simonds, Kathy KL Tran, and Rodney T. Willett. 2025. House Bill 2094 - High-Risk Artificial Intelligence Developer and Deployer Act.

[84] John Markoff. 1988. Computer Invasion: 'Back Door' Ajar. *The New York Times*, pages 1, 30.

[85] Nestor Maslej, Loredana Fattorini, Raymond Perrault, Vanessa Parli, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Russell Wald, and Jack Clark. 2024. Artificial Intelligence Index Report 2024.

[86] Alexander Meinke, Bronson Schoen, Jérémy Scheurer, Mikita Balesni, Rusheb Shah, and Marius Hobbhahn. 2025. Frontier Models Are Capable of In-Context Scheming.

[87] David Michaels. 2008. *Doubt Is Their Product*, 1st edition. Oxford University Press.

[88] Sarosh Nagar, Scott Singer, and Nitarshan Rajkumar. 2025. The Closing Window to Win, Part I: American AI Leadership Requires a Global Strategy for Full-Stack Diffusion. Technical report, The Foundation for American Innovation.

[89] Joe Needham, Giles Edkins, Govind Pimpale, Henning Bartsch, and Marius Hobbhahn. 2025. Large Language Models Often Know When They Are Being Evaluated. ArXiv:2505.23836 [cs].

[90] Alondra Nelson, Andrew Trask, Ben Garfinkel, Christine Custis, Dan Hendrycks, Deep Ganguli, Dorothy Chou, Helen Toner, Irene Solaiman, Jaan Tallinn, Janet Haven, Marc Aidinoff, Marietje Schaake, Matthew Salganik, Miranda Bogen, Nathan Lambert, Rishi Bommasani, Sébastien Krier, Solon Barocas, Stephanie Ifayemi, Suresh Venkatasubramanian, William Isaac, and Zoë Brammer. 2024. Comment of the AI Policy and Governance Working Group on the NTIA Request for Comment on Dual Use Foundation Artificial Intelligence Models with Widely Available Model Weights.

[91] Jimmie J. Nelson. 1980. Minutes of AQ-9 [CO2 and Climate] Task Force Meeting of February 9, 1980.

[92] Gavin Newsom. 2023. Executive Order N-12-23.

[93] Richard Ngo, Lawrence Chan, and Sören Mindermann. 2025. The Alignment Problem from a Deep Learning Perspective.

[94] U.S. National Highway Traffic Safety Administration (NHTSA). Safety Recall Data.

[95] U.S. National Highway Traffic Safety Administration (NHTSA). 2025. Seat Belts.

[96] Michelle Nie. 2025. Governments Need to Protect AI Industry Whistleblowers: Here's How. *AI Policy Bulletin.*

[97] Seth M. Noar, Diane B. Francis, Christy Bridges, Jennah M. Sontag, Kurt M. Ribisl, and Noel T. Brewer. 2016. The Impact of Strengthening Cigarette Pack Warnings: Systematic Review of Longitudinal Observational Studies. *Social Science & Medicine (1982)*, 164:118–129.

[98] National Association of Attorneys General. 1998. Master Settlement Agreement.

[99] California Department of Food and Agriculture. 2025. California Agricultural Production Statistics.

[100] California Department of Industrial Relations. 2025. California Code of Regulations, Title 8, Section 3209. Standard Guardrails.

[101] California Department of Industrial Relations. 2025. California Code of Regulations, Title 8, Section 3231. Stairways.

[102] Department of Pesticide Regulation. 2024. $1.9 Million in Grants Available to Support Safe, Effective, Sustainable Pest Management in California.

[103] U.S. National Institute of Standards and Technology. 2020. Evidence Suggests That the U.S. Loses Hundreds of Billions to Cybercrime, Possibly as Much as 1% to 4% of GDP Annually. Last Modified: 2025-02-03.

[104] OpenAI. OpenAI o1 System Card. Technical report, OpenAI.

[105] OpenAI. 2025. Deep Research System Card. Technical report, OpenAI.

[106] OpenAI. 2025. OpenAI o3 and o4-mini System Card. Technical report, OpenAI.

[107] Frank Pasquale. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information.* Harvard University Press.

[108] Robert N. Proctor. 1994. *Cancer Wars: How Politics Shapes What We Know and Don't Know About Cancer.* Basic Books.

[109] Elmer Robinson and R. C. Robbins. 1968. Sources, Abundance, and Fate of Gaseous Atmospheric Pollutants. Final Report and Supplement. *Stanford Research Institute.*

[110] James A Robinson and Daron Acemoglu. 2012. *Why Nations Fail: The Origins of Power, Prosperity and Poverty.* Profile London.

[111] Robert Rodriguez, Manny Rutinel, Brianna Titone, Lisa Cutter, Steve Fenberg, Dafna Michaelson Jenet, Kevin Priola, Faith Winter, and Monica Duran. 2024. SB24-205 – Concerning Consumer Protections in Interactions with Artificial Intelligence Systems.

[112] Lynne Schneider, Benjamin Klein, and Kevin M. Murphy. 1981. Governmental Regulation of Cigarette Health Information. *The Journal of Law and Economics*, 24(3):575–612.

[113] Sella Nevo, Dan Lahav, Ajay Karpur, Yogev Bar-On, Henry Alexander Bradley, Jeff Alstott. 2024. Securing AI Model Weights. *RAND*.

[114] Scott J. Shapiro. 2023. *Fancy Bear Goes Phishing: The Dark History of the Information Age, in Five Extraordinary Hacks.* Penguin Books, London.

[115] Toby Shevlane. 2022. Structured Access: An Emerging Paradigm for Safe AI Deployment.

[116] Jodi L Short. 1998. Killing the messenger: The Use of Nondisclosure Agreements to Silence Whistleblowers. *University of Pittsburgh Law Review*, 60:1207.

[117] Robert E Slavin. 2002. Evidence-Based Education Policies: Transforming Educational Practice and Research. *Educational researcher*, 31(7):15–21.

[118] Irene Solaiman. The Gradient of Generative AI Release: Methods and Considerations. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, pages 111–122.

[119] Irene Solaiman, Rishi Bommasani, Dan Hendrycks, Ariel Herbert-Voss, Yacine Jernite, Aviya Skowron, and Andrew Trask. 2025. Beyond Release: Access Considerations for Generative AI Systems.

[120] Irene Solaiman, Miles Brundage, Jack Clark, Amanda Askell, Ariel Herbert-Voss, Jeff Wu, Alec Radford, and Jasmine Wang. 2019. Release Strategies and the Social Impacts of Language Models.

[121] Merlin Stein, Jamie Bernardi, and Connor Dunlop. 2024. The Role of Governments in Increasing Interconnected Post-Deployment Monitoring of AI.

[122] Nicholas Stern. 2006. *The Stern Review on the Economics of Climate Change.* Cambridge University Press. Commissioned by HM Treasury.

[123] Charlotte Stix, Matteo Pistillo, Girish Sastry, Marius Hobbhahn, Alejandro Ortega, Mikita Balesni, Annika Hallensleben, Nix Goldowsky-Dill, and Lee Sharkey. 2025. AI Behind Closed Doors: a Primer on The Governance of Internal Deployment.

[124] FDA Adverse Event Reporting (FAERS) System. Public Dashboard.

[125] Alex Tamkin, Miles McCain, Kunal Handa, Esin Durmus, Liane Lovitt, Ankur Rathi, Saffron Huang, Alfred Mountfield, Jerry Hong, Stuart Ritchie, et al. 2024. Clio: Privacy-Preserving Insights into Real-World AI Use.

[126] Philip Tetlock and Aaron Belkin. 1996. *Counterfactual Thought Experiments in World Politics*. Princeton University Press.

[127] The National Artificial Intelligence Advisory Committee (NAIAC). 2023. Recommendation: Improve Monitoring of Emerging Risks from AI through Adverse Event Reporting.

[128] David Thiel. 2023. Identifying and Eliminating CSAM in Generative ML Training Data and Models. *Stanford Internet Observatory, Cyber Policy Center, December*, 23:3.

[129] David Thiel, Melissa Stroebel, and Rebecca Portnoff. 2023. Generative ml and csam: Implications and mitigations. In *Stanford digital repository*.

[130] Muyesaier Tudi, Huada Daniel Ruan, Li Wang, Jia Lyu, Ross Sadler, Des Connell, Cordia Chu, and Dung Tri Phung. 2021. Agriculture Development, Pesticide Application and Its Impact on the Environment. *International Journal of*

Environmental Research and Public Health, 18(3):1112. Number: 3.
[131] U.K. Competition and Markets Authority. 2023. AI Foundation Models: Initial Report.
[132] U.S. Cybersecurity and Infrastrcture Security Agency (CISA). Comment Submitted by ABA, APPA, BPI, EEI, NRECA, NTCA, SIFMA, USTelecom.
[133] U.S. National Telecommunications and Information Administration (NTIA). 2024. Dual-Use Foundation Models with Widely Available Model Weights.
[134] U.S. Office of Management and Budget (OMB). 2024. Advancing the Responsible Acquisition of Artificial Intelligence in Government.
[135] Susan Wagner. 1971. *Cigarette Country: Tobacco in American History and Politics.* Praeger, New York.
[136] Matthew T. Wansley. 2023. Regulating Driving Automation Safety. *Emory Law Journal*, 73:505.
[137] Laura Weidinger, Deb Raji, Hanna Wallach, Margaret Mitchell, Angelina Wang, Olawale Salaudeen, Rishi Bommasani, Sayash Kapoor, Deep Ganguli, Sanmi Koyejo, and William Isaac. 2025. Toward an Evaluation Science for Generative AI Systems.
[138] L Paige Whitaker. 2007. The Whistleblower Protection Act: An Overview. *Law and Law Enforcement Issues*, 213:1–15.
[139] Joseph F Williams. Consumer Safety Product Commission: Recall Effectiveness Workshop.
[140] Amy A. Winecoff and Miranda Bogen. 2024. Improving Governance Outcomes Through AI Documentation: Bridging Theory and Practice.
[141] Henry Wu. 2025. AI Whistleblowing.
[142] Ernst L. Wynder and Evarts A. Graham. 1950. Tobacco Smoking as a Possible Etiologic Factor in Bronchogenic Carcinoma: A Study of Six Hundred and Eighty-Four Proved Cases. *Journal of the American Medical Association*, 143(4):329–336.
[143] Monika Zalnieriute. 2021. "Transparency-Washing" in the Digital Age : A Corporate Agenda of Procedural Fetishism. Technical report.
[144] Jonathan L. Zittrain. 2008. *The Future of the Internet - And How to Stop It*. Yale University Press, New Haven, CT.
[145] Andy Zou, Zifan Wang, Nicholas Carlini, Milad Nasr, J. Zico Kolter, and Matt Fredrikson. 2023. Universal and Transferable Adversarial Attacks on Aligned Language Models.

A FEEDBACK FORM

Following the release of the draft report, we created a feedback form to receive public input. We provide a copy of this form below.

A.1 Preamble Text

In September 2024, Governor Gavin Newsom announced new initiatives to advance safe and responsible artificial intelligence (AI) development in the State. As part of that announcement, Governor Newsom requested that Dr. Fei-Fei Li, Co-Director of the Stanford Institute for Human-Centered Artificial Intelligence (HAI); Dr. Mariano-Florentino Cuéllar, President of the Carnegie Endowment for International Peace; and Dr. Jennifer Tour Chayes, Dean of the UC Berkeley College of Computing, Data Science, and Society, prepare a report focusing on frontier models. Note: the Draft Report represents scholarly work from each of the Co-Leads and does not represent the positions of their institutions.

To encourage feedback and input from a wide range of expertise, the Draft Report is being shared publicly. We welcome feedback, reflections, and additional information from a variety of disciplines and sectors by **April 8, 2025** for consideration in advance of the final report.

The working group will not share individual responses to this form publicly, nor attribute any feedback or input at an individual level. The feedback will be aggregated and incorporated into the final report, expected by June 2025. Thank you!

A.2 Information about Respondent

- Email Address *(Required)*
- First Name *(Optional)*
- Last Name *(Optional)*
- Please check the box if you would like to opt in for receiving email updates. [Checkbox] *(Optional)*
- Which category best describes you? [Select one] *(Optional)*
  Academia
  Government/Public Sector
  NGO/Civil Society
  Industry/Business/Private Sector
  Other [free response field provided]
- Title *(Optional)*
- School/Company/Organization/Agency Name *(Optional)*

A.3 Form Questions

(1) The structure of the Draft Report includes sections about *Context, Transparency and Third-Party Risk Assessment, Adverse Event Reporting, and Scoping*. At a high level, what might you find valuable? What types of questions are you most interested in, and how might you use the report in your work? *(Optional)*
(2) From your perspective and experience, what key factors do you see affecting California's path forward in AI governance? Please feel free to provide specific feedback referring to the sections of the draft report. *(Optional)*
(3) Numerous frontier AI governance-focused groups have been working on frameworks, guidance, and reports aiming to leverage scientific research. For what topics or issues are you observing challenges in reaching scientific consensus? Do you have recommendations to bridge gaps? *(Optional)*

(4) What could be done individually or collectively to leverage frontier AI for Californians' benefit? *(Optional)*
(5) Please feel free to list any published resources you would like to share with the Joint California Policy Working Group on AI Frontier Models. *(Optional)*

LEADERSHIP

Practice What You Teach: Academic Integrity and AI

October 10, 2025

Start a conversation, not an accusation.

GUEST COLUMN | by Randy Kolset and Mike Lawrence

XIAYAMOON

Educators have an expectation from all students that they will be ethical in their academic coursework. Most schools and districts have a stated policy about academic integrity. It's a long-standing tradition and one that is foundational to student-teacher relationships. It makes one wonder why they often violate this themselves when it comes to AI use.

'If we're preparing students for an AI-powered world, ***we must model the transparency we ask of them.'***

As AI tools become increasingly embedded in K–12 classrooms, a *quiet contradiction* has emerged: students are expected to disclose their use of AI, while teachers often use it without acknowledgement. This asymmetry challenges the core principles of fairness, academic integrity, and digital citizenship. If we're preparing students for an AI-powered world, we must model the transparency we ask of them.

Modeling Integrity in the Age of AI

Laguna Beach Unified School District (LBUSD) has taken a bold step forward with its *"AI: Trust You"* extension, recognizing that technology integration must be rooted in *mutual trust*. The extension allows teachers to frame *when* and *how* AI may be used with a custom-built 'Trust Statement' unique to each assignment, followed by students creating an honor-based statement clarifying *how* and the *degree* to which AI was used in their response. Rather than framing AI as a threat, LBUSD reframes it as an opportunity—to build character, model ethical use, and empower learners.

This exchange aligns seamlessly with Dr. Sabba Quidwai's Designing Schools framework, which places *trust, voice,* and *agency* at the center of educational transformation. If we want students to be courageous, curious problem-solvers in the age of AI, we must invite them into the conversation—not police them from the sidelines. In other words, start a conversation, *not an accusation*.

'If we want students to be courageous, curious problem-solvers in the age of AI, we must ***invite them into the conversation****—not police them from the sidelines.'*

Building Trust Through Transparency

In alignment with this approach, and in contrast to reactive bans or inconsistent enforcement, Orange Unified School District (OrangeUSD) has created AI guidelines that empower educators while holding them accountable. Their approach encourages the *intentional use of AI* as a teaching and learning tool, while trusting teachers to make professional decisions about disclosure and implementation. This balance between *freedom and responsibility* fosters transparency without stifling innovation.

Research from 2023–2025 supports this shift. Surveys show that students increasingly *perceive a double standard*: they are required to cite AI, while teachers often aren't. This erodes trust and fosters resentment. By contrast, teachers who openly share how they use AI—whether to design a lesson, provide feedback, or generate ideas—actually *strengthen credibility* and model responsible tech use.

‘…teachers who openly share how they use AI—whether to design a lesson, provide feedback, or generate ideas—actually ***strengthen credibility and model responsible tech use.’***

Those who believe that the best approach is to shift back to pen and paper assignments or completely lock down AI are taking an approach destined to reinforce a lack of trust and fail to prepare students for a future in which ubiquitous AI is vastly more powerful than that which we encounter today. In fact, as ISTE CEO Richard Culatta shared, recent research out of Stanford University (2024, *Cheating in the age of generative AI*, Lee, Pope, Miles, Zarate) demonstrates that cheating was prevalent (about 70%) before generative AI and didn’t significantly change after its launch.

Is cheating a problem? Yes!

Does AI cause cheating? Not likely.

When teachers use AI appropriately to craft lessons, presentations and provide speedy feedback, and don’t share how they are using it, they *miss an opportunity to model ethical AI use* for their students.

‘When teachers use AI appropriately to craft lessons, presentations and provide speedy feedback, ***and don’t share how they are using it****, they miss an opportunity to model ethical AI use for their students.’*

In the article “How can we upskill Gen Z as fast as we train AI?” by Merriman & Sanz Sáiz in 2024 an international survey stated that 55% of our Gen Z population is learning how to use AI through social media. Several surveys and interviews with students at OrangeUSD showed that students don’t want to lose their ability to think, be accused of cheating, and not have the skills for the future through the use of AI. More than likely they are learning how to leverage AI as a short cut tool and not its true productive nature.

If teachers engage in conversations around how they use AI to be more productive it changes how students understand how to use the LLM systems to be successful.

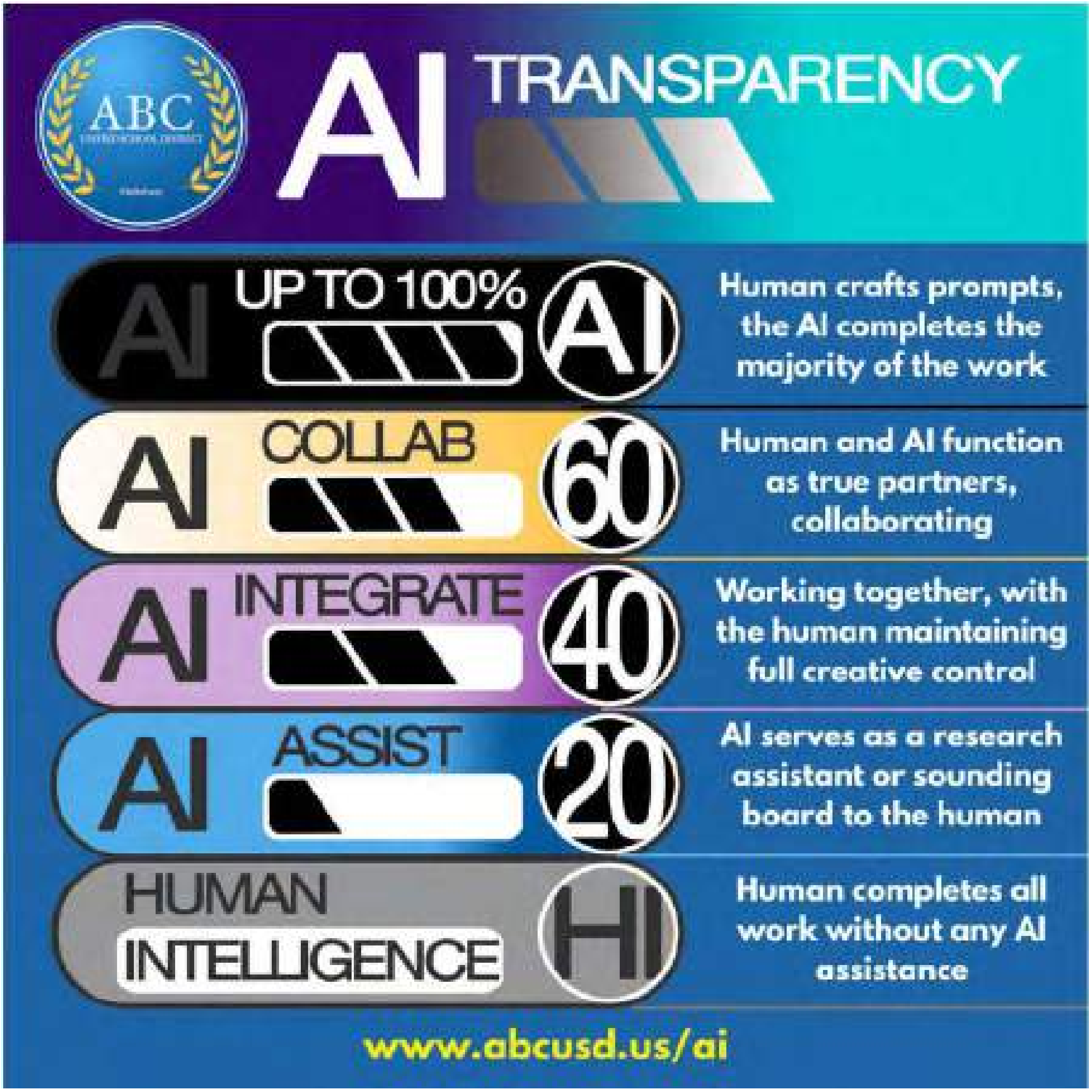
Teachers can demonstrate how they interact, reframe their prompts and leverage their own critical thinking to create something better, faster and more effective by collaborating with AI as a thought partner.

From Policing to Partnership

ABC Unified (ABCUSD) has taken this approach a step further with their *AI Trust and Transparency System*. In addition to implementing Laguna Beach's *AI Trust You* extension, district leaders have collaborated with students, administrators and teachers to develop a set of badges for use by *all* groups. They offer another layer of visibility into how HI (Human Intelligence) and AI was used in a given project, presentation, media or paper.

Everyone, from the superintendent to students is asked to share a visual representation of their use of AI in presentations, lessons, handouts and assignments. The badges display a 'battery-like' image and corresponding number representing the maximum percentage of AI in the completed work. Each is given a descriptor that matches that level (Assist, Integrate, etc.). All of their work is detailed at abcusd.us/aitt.

They offer five levels of HI + AI usage, each with a corresponding badge:



ABC UNIFIED

The district built on the concept of Jon Spike, an educator in Wisconsin, who first developed a set of badges for this purpose. In his spirit of sharing, ABCUSD has made generic and customizable versions of these badges available so that others can use them either with their own organizational logo, or without. All three styles of badges are available via creative commons sharealike at abcusd.us/aibadges.

As AI becomes a permanent part of our educational landscape, we must ask: *Are we building a culture of secrecy or a culture of trust?*

Districts like Laguna Beach, ABC and Orange USD utilizing the OECD AI Literacy Framework ailiteracyframework.org are showing us that *trustworthy AI use isn't just about tools—it's about people*.

By aligning policy with transparency and compassion, we create classrooms where both teachers and students grow as ethical digital citizens—together.

—

Randy Kolset is an experienced Education Administrator with a demonstrated history of working in primary and secondary education. Skilled in Educational Consulting, Leadership, Educational Technology, Curriculum Development, STEM Integration, and Project Management. As a developer of Artificial Intelligence policies for Education and Computer Science schools at the elementary level, Randy is a strong education professional with a Masters of Science focused in Educational Technology, dedicated to fostering ethical AI use and promoting a culture of trust in educational settings. Connect with Randy via LinkedIn.

Mike Lawrence is an award-winning educator and global EdTech leader, currently Director of Information & Technology for ABC Unified School District. Former CEO of CUE and EdTech Digest Top 100 Influencer in EdTech (2023–24), he has advanced digital learning worldwide through leadership with ISTE, UNESCO, PowerSchool, Jamf, and California's Student Media Festival. Connect with Mike on LinkedIn.

By: Victor Rivero

Tags:
ABC Unified • AI in education • building a culture of trust • designing schools • Dr. Sabba Quidwai • EdTech Digest Top 100 • Gen Z • ISTE • Richard Culatta • school culture • school district technology • skills • students and teachers • upskilling

Share:

YOU MIGHT ALSO LIKE

0 COMMENTS

Maximize AI ROI: Invest In The People Who Use It

Maria Flynn

Companies that see the highest returns on AI don't just invest in technology. They invest in training the people who use it.

getty

Companies are investing more money faster in AI than ever before in any other software category: $37 billion in 2025, triple their 2024 spending. The speed and scope of investment are growing alongside technological advances. As anyone who's used Claude or ChatGPT recently can tell you, AI is becoming more sophisticated, more reliable, and more capable of handling a wider range of tasks.

The million-dollar question (or the hundred-billion-dollar question) is whether companies will see a positive return on these investments. One recent survey of thousands of CEOs found that the vast majority saw "little impact from AI on their operations," with nearly 90% reporting no impact on employment or productivity from their AI adoption. A Goldman Sachs report concluded that AI spending has not driven economic growth.

Of course, it's still early days when it comes to AI. Some pilots will—and *should*—fail. If they don't, it means companies aren't innovating enough. But even as companies navigate this transition period, there's more they can do to position themselves for success.

Recent research from Jobs for the Future (JFF) indicates that employers might improve the outcomes of their AI-related investments by directing as much effort into developing people as they do into acquiring new technology. Based on our survey, which tracks year-over-year data from 2024 and 2025, most workers (more than 60%) lack access to employer-provided AI training. Just over one-third of workers say they have the training and resources needed to use AI in their jobs, a decrease from 45% in 2024.

If businesses are spending more on AI but aren't doing a better job of preparing people to use it, can we really be surprised that they aren't seeing as much value as they hope to?

Here's what employers can do to make the most of those billions of dollars in investments:

Support Job-Aligned AI Skills Training

In 2024, 42% of JFF's survey respondents felt AI was not significantly changing the importance of any skills. Last year, that number plummeted to only 7%. As AI seeps into more occupations and roles, more workers will need opportunities to build their AI skills. Yet these options aren't always available. In the same survey, only a third of respondents said they get the support they need from their employers regarding AI. Better, more intentional and job-aligned training can help workers make better use of AI on the job.

Some companies are already doing this well: Walmart partnered with Google to provide free training in AI fundamentals to 1.6 million employees. Walmart sees this as an investment in the future: ensuring today's employees build the AI skills to thrive in future roles within the company. (The U.S. Department of Labor this week launched a free "AI 101" course that delivers AI literacy skills via text message—a step toward helping people learn the basics to prepare them for job-aligned training.)

Invest In Early Talent Who Can Lead AI Change

The tasks associated with early-career roles are shifting faster than those of more senior positions, creating a steeper adjustment curve for younger workers—and a greater need for employer support. But JFF's survey data points to an upside: early-career workers self-report higher rates of AI use and stronger AI literacy than their more experienced counterparts. This makes them a real asset in AI transformation today, not just a talent pipeline for the future. And while overall hiring among younger workers is on the decline, some companies, like IBM, are doubling (or tripling) down on entry-level hiring specifically in response to AI's rise. Companies that invest in early talent now are better positioned to build workforces that lead AI change—not just adapt to it.

Listen to Employees Who Use AI

According to the American Job Quality Study, workers who feel they have more of a voice in technology adoption are also more likely to feel satisfied at work. Yet in the AI survey, over half of workers said their employers didn't consult them on how AI tools could be used in their work. The more you provide employees with opportunities to share how they're using AI, how it's making them feel, and what they want to learn about it, the more likely you are to both improve their skills and keep them around.

Harness the Power of AI-Human Collaboration

Listening to workers isn't just a feel-good exercise; it can also uncover valuable insights into how to effectively implement AI. After all, workers know the systems best and are more likely to identify high-value uses of AI that reduce pain points and bottlenecks. This can remove the friction and busywork that are often the tedious part of a job, and give workers more time for the creative, relational, and strategic work that drives growth (and that they likely find more satisfying).

But companies that want to get the most value from AI shouldn't stop there. The rapid advancements of AI technologies pose a larger question: how do we reimagine workflows and drive value both for workers and the bottom line? How might an AI-ready workforce better leverage the power of human-AI collaboration to generate innovations?

We're seeing the spectrum of AI adoption in our survey: more respondents reported AI supporting efficiency gains, like reducing manual work and overall workload, and a smaller but growing number report AI creating new types of tasks or responsibilities (13%, up from 7% in 2024).

Unfortunately, many companies today are still more focused on finding value from cost savings than using AI to improve processes, outcomes, or worker experiences. Recent research from the Brookings Institution found that, at least with the limited (and anecdotal) information available, AI is being used more to automate human work, rather than to augment it. And respondents in JFF's survey are growing more skittish about the impact AI will have as a result. In 2024, workers were more positive than negative about AI's potential impact, but in 2025, that flipped, with 44% of respondents saying they see more potential for harm compared to 38% who say they see more potential for good.

For that to change, more businesses will need to recognize that AI's value is in the workforce, not the technology. This isn't novel; humans have always been the linchpin in new initiatives, and adopting tech for tech's sake has never yielded value. Companies that increase their AI ROI will be the ones that excel at human change management—supporting worker development and giving workers a voice in how AI impacts their workflow to ensure it augments their efforts.

Data Centers and California Electricity Policy

Report #292 | March 2026

Milton Marks Commission on California State Government Organization and Economy
www.lhc.ca.gov

Dedicated to Promoting Economy and Efficiency in California State Government

The Little Hoover Commission, formally known as the Milton Marks "Little Hoover" Commission on California State Government Organization and Economy, is an independent state oversight agency.

By statute, the Commission is a bipartisan board composed of five public members appointed by the governor, four public members appointed by the Legislature, two senators and two assemblymembers.

In creating the Commission in 1962, the Legislature declared its purpose:

...to secure assistance for the Governor and itself in promoting economy, efficiency and improved services in the transaction of the public business in the various departments, agencies and instrumentalities of the executive branch of the state government, and in making the operation of all state departments, agencies and instrumentalities, and all expenditures of public funds, more directly responsive to the wishes of the people as expressed by their elected representatives...

The Commission fulfills this charge by listening to the public, consulting with the experts and conferring with the wise. In the course of its investigations, the Commission typically empanels advisory committees, conducts public hearings and visits government operations in action.

Its conclusions are submitted to the Governor and the Legislature for their consideration. Recommendations often take the form of legislation, which the Commission supports through the legislative process.

Contacting the Commission

All correspondence should be addressed to the Commission Office:

Little Hoover Commission
925 L Street, Suite 805, Sacramento, CA 95814

(916) 445-2125 | littlehoover@lhc.ca.gov

This report is available from the Commission's website at www.lhc.ca.gov.

Letter from the Chair March 2026

The Honorable Gavin Newsom
Governor of California

The Honorable Monique Limón
President pro Tempore of the Senate
and members of the Senate

The Honorable Brian Jones
Senate Minority Leader

The Honorable Robert Rivas
Speaker of the Assembly
and members of the Assembly

The Honorable Heath Flora
Assembly Minority Leader

DEAR GOVERNOR AND MEMBERS OF THE LEGISLATURE:

California's electricity rates are among the highest in the nation, and affordability remains a top concern for residents. At the same time, the state is committed to some of the most ambitious clean-energy and climate goals in the country.

Against this backdrop, the rapid growth of energy-intensive data centers presents both a serious challenge and a potential opportunity for California's electricity system. If managed thoughtfully, these facilities could help catalyze investments that modernize the grid and strengthen reliability—without increasing costs for ratepayers or undermining the state's clean-energy commitments.

To help policymakers navigate these tradeoffs, the Little Hoover Commission examined the implications of data-center-driven load growth, with particular attention to ratepayer protection, regulatory structure, grid planning, environmental and community impacts, and innovation.

In our report, *Data Centers and California Electricity Policy*, we offer recommendations to ensure data centers pay their fair share of infrastructure and grid-service costs, reduce the risk of stranded assets, improve interconnection and permitting predictability while maintaining strong review standards, protect communities from pollution—especially from backup generation—and support research and partnerships to advance cleaner, more efficient technologies.

We respectfully submit this report and stand ready to assist as the state develops a policy framework that supports innovation while protecting Californians' affordability, reliability, and clean-energy priorities.

Sincerely,

Pedro Nava, Chair
Little Hoover Commission

Table of Contents

EXECUTIVE SUMMARY

California's electricity rates are among the highest in the country and have risen considerably faster than inflation in recent years, making affordability a major concern for residents. At the same time, the state is already committed to some of the most ambitious clean-energy and climate goals in the nation.

Against this backdrop, the rapid growth of energy-hungry data centers presents both a serious challenge and a potential opportunity for California's electricity system. Business and political leaders believe that data-center-driven computing—and artificial intelligence in particular—has the potential to transform society by turbocharging service delivery, accelerating scientific and medical discovery, improving governmental efficiency, and ushering in an era of economic growth supported by major productivity gains.

Unlocking this potential depends on California's ability to meet the enormous energy demands of these facilities. Doing so responsibly will require enormous new investment without raising electricity rates or compromising the state's clean energy commitments.

This challenge comes with opportunity. With the right policies in place, data centers could help catalyze investments that modernize California's grid while potentially lowering electricity costs for residents and businesses. The report that follows outlines recommendations for how the state might achieve this balance, with particular attention to rate design, regulatory structure, affordability, environmental protection, and innovation.

PROTECTING RATEPAYERS

Above all, California must develop tariffs and policies for data centers that protect consumers from rising energy costs. This includes addressing cost pressures from increased electricity demand and ensuring that ratepayers are not unfairly burdened with infrastructure costs, including the risk of stranded assets—investments built to serve data center loads that are not ultimately paid for by those loads.

These goals can be achieved through a well-designed very-large-load tariff that incorporates credit requirements, prepayment, full cost recovery through rates, minimum demand and term commitments, load flexibility, and meaningful exit penalties. Robust data-sharing requirements should support these tariffs and inform ongoing policymaking. At the same time, the state should work to accelerate data center interconnection to support innovation, while maintaining strong and consistent review standards.

SMARTER GRID PLANNING

California should accommodate data center growth while avoiding unnecessary infrastructure investments by making more effective use of its existing electric grid through smarter planning. This includes incentivizing and prioritizing targeted transmission upgrades and strategic data center siting, encouraging the clean and flexible use of backup power resources, and leveraging load shifting to support the grid during periods of peak demand.

CLEAN ENERGY AND COMMUNITY PROTECTION

California leads the nation in the pursuit of clean, zero-carbon energy, and the addition of new data center load must support—rather than compromise—these goals. Communities must also be protected from added pollution, particularly from backup power sources that are often diesel-powered. Achieving these objectives will require strong statutory protections alongside incentives to adopt technologies that reduce grid and environmental impacts and, consistent with the state's culture of innovation, to develop new technologies and approaches that advance these goals.

Recommendations

To help ensure that the growth of data centers strengthens California's electricity system without increasing costs for ratepayers or undermining clean-energy goals, the Little Hoover Commission recommends the following actions for policymakers and regulators:

1. California should strive to develop a stable regulatory framework for data centers that balances economic growth with clean, efficient, affordable, and reliable electricity provision. It should make certain that costs associated with new large-load development are fairly allocated and do not shift onto existing ratepayers.

2. California should lead the nation on data-center energy regulation by grounding its policies in the best ideas from researchers, state agencies, utilities, developers, and other jurisdictions (state, federal, international).

3. California should enable regulators to confidentially access essential facility-level electricity-use data from large data centers, supported by structured data-sharing among agencies, to improve planning, assess localized impacts, and ensure fair cost responsibility while protecting sensitive business information.

4. California should ensure that data centers pay for costs imposed on the electric system through tariffs and cost-allocation policies that recover infrastructure and grid-services costs in a manner that benefits rather than burdens other ratepayers.

5. While maintaining a rate structure that ensures shared costs and a high standard of regulatory review, California should accelerate data-center interconnection and permitting decisions to give developers greater predictability in planning and investment.

6. California should establish consistent, statewide safeguards that ensure large data-center customers bear appropriate financial risk, while making certain that major grid investments continue to serve the public even if data center demand evolves differently than has been predicted.

7. The state should establish consistent statewide oversight expectations for large data center developments in territories served by publicly owned utilities, municipal utilities, irrigation districts, and community choice aggregators, ensuring that risks to reliability, costs, and local communities are addressed with the same level of rigor as investor-owned utilities.

8. California should ensure that new large data center loads are integrated in a manner that preserves system reliability and does not disadvantage existing customers.

9. California should maximize the use of existing grid capacity when considering data-center siting and new infrastructure, supported by transparent data and visualization tools that help developers and policymakers identify locations where capacity is available or can be efficiently expanded.

10. California should maximize load shift opportunities to use existing load capacity and lower costs for residential customers.

11. California should meet rising data-center electricity demand without weakening its clean-energy or climate goals, including behind-the-meter generation.

12. California should require data centers to maintain a minimum level of clean backup power to support load shifting, enable curtailment during grid stress (including extreme heat events), and reduce reliance on fossil-fuel backup generation.

13. The state should limit pollution impacts on nearby communities from data-center backup generators.

14. California should encourage data centers to adopt technologies—such as advanced cooling, energy storage, or on-site clean power—that reduce their impact on the grid.

15. California should establish a formal public-private partnership—provisionally the California Partnership for Advanced Research (CALPAR)—to collaborate with industry, national laboratories, and universities to research, test, and promote technologies that improve the efficiency and environmental performance of AI-driven data centers.

Data Centers and the Future of California's Energy System

California's electricity rates are among the highest in the country and have risen considerably faster than inflation in recent years, making affordability a major concern for residents.[1] At the same time, the state is already committed to some of the most ambitious clean-energy and climate goals in the nation.

Against this backdrop, the rapid growth of energy-hungry data centers presents both a serious challenge and a potential opportunity for California's electricity system. This section explains what data centers are and how they relate to electricity policy; summarizes their benefits and risks; and outlines policy options for managing their impacts on energy affordability, reliability, and the state's environmental goals.

INTRODUCTION TO DATA CENTERS AND ELECTRICITY POLICY

As digital applications have become more complex—from artificial intelligence and data analysis to streaming and social media—much of the computing power, storage, and software that support these applications has shifted from local devices to high-powered servers housed in data centers.[2] By accessing these centralized resources over the Internet, devices such as personal computers and smartphones can perform tasks that far exceed their own hardware's capacity.

Many business and political leaders believe that data-center-driven computing—and artificial intelligence in particular—has the potential to transform society. They are betting, with dollars and political capital, that it will turbocharge service delivery, accelerate scientific and medical discovery, improve governmental efficiency, and usher in an era of economic growth supported by major productivity gains.[3]

Unlocking this potential depends on California's ability to meet the enormous energy demands of these facilities.[4] Data centers consume extraordinary amounts of electricity and require costly energy infrastructure to connect them to the grid. To give a sense of scale, in July 2025, Pacific Gas & Electric (PG&E) reported that the amount of data center energy demand waiting to be connected to the grid is equivalent to more than 10 percent of California's current total generation capacity; enough electricity to power roughly 8 million homes.[5] Comparably large upgrades will be needed—in the form of power lines, substations, and transformers—to deliver this energy. Meeting this demand responsibly will require enormous new investment without raising electricity rates or compromising the state's clean energy commitments.

This challenge comes with opportunity. With the right policies in place, data centers could help catalyze investments that modernize California's grid while potentially lowering electricity costs for residents and businesses.[6] The report that follows outlines recommendations for how the state might achieve this balance, with particular attention to rate design, regulatory structure, affordability, environmental protection, and innovation.

Success is not guaranteed. The scale and concentration of energy demand from data centers is unprecedented and has the potential to increase electricity prices while posing risks to communities and the environment.[7] Policymakers may also face pressure to defer to powerful technology companies and utilities.[8] Yet, if approached thoughtfully, this moment could offer California a chance to align the interests of industry, government, and the public.

Indeed, data center operators increasingly emphasize that responsible development requires not only careful management of electricity and water use, but also that developers contribute directly to solutions. Microsoft promotes a "community-first" approach that includes paying the full cost of required energy infrastructure, reducing and replenishing water use, and delivering local benefits such as jobs, workforce training, and tax revenues for public services.[9] Google similarly highlights efficiency and transparency, publicly reporting the electricity and water intensity of AI queries to help communities and policymakers understand system-wide impacts as demand grows.[10] At the industry level, the Data Center Coalition states that its members are committed to paying their full cost of electric service, including direct payment for utility infrastructure needed to serve data center load.[11] That said, voluntary leadership by a subset of developers is not sufficient—statutory mandates are required to ensure that these and other best practices are applied consistently, transparently, and over the long term.

Integrating data centers into California's electric grid has become a politically sensitive issue, unfolding at a time when public skepticism toward artificial intelligence contrasts sharply with the tech industry's central role in the state's economy.[12] Experiences in other states show how backlash against data centers can influence public opinion and shape elections, underscoring the stakes of getting policy right.[13] In California, this issue is playing out against an ongoing affordability crisis, heightening the need for careful decision-making.[14] This report outlines how policymakers can navigate these pressures, identifying best practices in rate design, regulation, and implementation to help integrate data centers in a way that protects consumers while promoting innovation.

AN ETHICAL MAP FOR NAVIGATING DATA CENTER RISK

Hosting data centers presents a complex set of benefits and risks that span economic, community, environmental, and energy-related issues.[15] Starting with non-energy impacts, data centers can provide meaningful economic benefits to local communities and the state.[16] They can reinforce California's leadership in advanced technology, help retain major AI and cloud firms, and support tax revenues, investment, and research partnerships. At the local level, they can create construction jobs, support vendors and trades, strengthen property-tax bases, and stimulate infrastructure improvements. In some cases, they can support emerging applications of advanced computing and even help revitalize underused commercial or industrial areas.[17]

These benefits, however, are accompanied by real risks. Data centers often require a significant physical footprint, which can affect land availability, ecosystems, and long-term development flexibility.[18] They can cause environmental problems relating to water use, carbon emissions, noise, and air pollutants.[19] If incentives or tax agreements are poorly structured, governments may receive fewer fiscal benefits than anticipated. Limited transparency, community concerns about fairness, and perceptions of corporate favoritism can also create mistrust or political backlash.[20]

In terms of energy-related impacts, data centers have the potential to play a constructive role in California's energy system. Their predictable, large loads could help support investments in new clean energy resources and spread fixed system costs across more customers, potentially lowering overall costs. And if they participate in flexible-load programs, data centers may also contribute to grid reliability, support demand response, and help integrate renewable energy, while strategic siting and workload shifting could permit the current grid to function more efficiently.[21]

These opportunities again come with risks. Data-center driven electricity demand has been widely cited as a potential driver of higher rates. New large loads may require substantial new generation, transmission, and substation infrastructure, which also add cost, strain planning processes, and crowd out other electrification priorities such as transportation and buildings.[22] And if demand decreases or data centers relocate, Californians could be left paying for unused energy infrastructure.[23]

There are also important energy-related environmental concerns.[24] Reliance on diesel backup generation can harm nearby communities.[25] And, more broadly, if the state cannot keep pace with demand growth, data centers could complicate California's pathway toward achieving its long-term clean-energy and climate goals.[26]

To navigate this complex issue, policymakers should approach decisions about data centers and electricity policy with a clear sense of the values they want to advance.[27] The choices California makes will shape not only economic growth and technological leadership, but also how costs are shared, how communities are impacted, and whether the state can meet its clean-energy and reliability goals. The following values framework can help ensure that policy decisions are disciplined, transparent, and in the public interest:

- **Affordability and Reliability:** Californians should have access to electricity that remains affordable and dependable as data centers are added to the grid.
- **Fairness and Cost Responsibility:** Data centers should not shift electricity-related costs onto households or small businesses.
- **Climate and Health:** Policies governing data centers should support California's clean-energy and climate commitments. Data centers must not have a negative impact on public health.
- **Innovation in the Public Interest:** Data centers should contribute to grid modernization and state policy goals in ways that benefit all Californians.
- **Transparency and Public Trust:** Agreements, decisions, and outcomes related to data centers should be transparent, understandable, and accountable to the public.
- **Predictability and Stability:** California should commit to developing stable regulatory frameworks that enable data center developers to invest and plan responsibly for the long term.

Benefits and Risks: Non-Energy Impacts

Benefits	Risks
Statewide Economic & Strategic ▪ Reinforces CA tech leadership ▪ Helps retain major AI/cloud firms ▪ Tax revenues & economic growth ▪ Supports R&D and university partnerships	**Land Use & Environmental** ▪ Large land footprint; long-term site impacts including ecosystem disruption ▪ Significant water use for cooling ▪ Noise from cooling systems ▪ Air quality impacts from diesel backup
Local & Community Economic Benefits ▪ Construction jobs & near-term stimulus ▪ Potential local tax revenues ▪ Benefits without major traffic or population growth ▪ Possible community benefit agreements	**Fiscal & Economic Risks** ▪ Incentives may offset gains ▪ Possible fiscal over-dependence ▪ Uneven benefit distribution ▪ Weak transparency on costs vs. benefits
Urban Development & Infrastructure ▪ Can drive broadband and local infrastructure upgrades ▪ Edge computing can support innovation/ advanced services ▪ Can revitalize vacant commercial and industrial sites	**Community, Governance & Public Trust** ▪ Political backlash over corporate favoritism ▪ Confidentiality reduces trust ▪ Potential security/emergency target

Benefits and Risks: Energy-Related Impacts

Benefits	Risks
Affordability & System Economics ▪ Potential to spread fixed system costs ▪ Supports long-term clean-energy financing ▪ Predictable load aids planning	**Affordability Risks** ▪ Cited driver of rising electricity rates ▪ Poor rate design could shift costs to households & small businesses
Clean-Energy & Climate Alignment ▪ Keeps load in CA's cleaner energy mix ▪ Supports CA climate leadership ▪ Aligns Western regional development toward cleaner power	**Infrastructure & Planning Burdens** ▪ Requires major new generation and transmission investment ▪ Can crowd out other electrification needs ▪ Heightens siting and permitting conflicts
Grid Reliability & Innovation ▪ Strategic siting and workload shifting can reduce grid stress ▪ Drives upgrades in grid infrastructure and innovation	**Environmental & Reliability Risks** ▪ Added load could strain grid capacity and increase reliance on outdated energy sources ▪ Reliance on diesel backup generation ▪ Locational inequities in community impacts
Opportunity for Grid and Policy Innovation ▪ Data centers could help finance grid upgrades that benefit all Californians ▪ Rethinking energy-related policies and processes could improve state utility regulation	**Stranded Asset & Policy Failure Risks** ▪ If tech shifts or firms relocate, infrastructure may be stranded ▪ Ratepayers may pay for unused capacity ▪ Benefits materialize only if policy execution is strong

Policymakers should consider these values non-negotiable as they navigate the evolving landscape of data centers and energy, using them as an ethical map to ensure decisions remain fair, transparent, affordable, equitable, and aligned with the state's climate commitments. By grounding policy in these principles, California can encourage innovation and economic growth while protecting ratepayers, supporting communities, and strengthening public trust.

To protect ratepayers from bearing costs associated with the creation of new energy generation for large AI data centers, the owners, operators, and users of such facilities should be required to build or fully pay for the new capacity that serves their interests. This obligation should be transparent and accountable to government and can arise from law, regulation, regulatory or tariff direction or enforceable agreement.

RECOMMENDATION 1: California should strive to develop a stable regulatory framework for data centers that balances economic growth with clean, efficient, affordable, and reliable electricity provision. It should make certain that costs associated with new large-load development are fairly allocated and do not shift onto existing ratepayers.

POLICY APPROACHES: DATA CENTERS AND ELECTRICITY

California policymakers may wish to look to other states for guidance on how to structure regulatory and rate policies for data centers. At the same time, this is an unprecedented moment, and California has the opportunity not only to adopt proven approaches but also to innovate and design policies that reflect the state's unique priorities around affordability, reliability, and climate leadership.

In January 2025, researchers at Lawrence Berkeley National Laboratory published a review of rate design approaches that states and utilities across the country have either implemented or proposed in response to the growth of data centers—here and elsewhere referred to as "large loads" because of the considerable electricity they draw from the grid.[28] These approaches are organized according to objective—e.g., defining which customers qualify for specialized tariffs, reducing financial risk, developing long-term contracts, or aligning energy sources with environmental goals. The table below summarizes these approaches using simplified or slightly modified wording.

States such as Ohio, Texas, and Oregon have led the nation in data center policy, and there are also innovative approaches emerging internationally to reduce the energy impacts of these facilities.[29] Paraguay, for example, is leveraging surplus clean power from its hydroelectric resources to attract data centers and anchor a growing sector of its economy.[30] In Europe, heat recycling programs in Sweden, Finland, Ireland, and the United Kingdom are using waste heat from data centers to warm thousands of homes.[31] A similar project is scheduled for apartments in the City of San Jose.[32]

We encourage policymakers to draw on the best of these approaches in shaping California's approach to data centers and energy. Many of the major players—state leaders, utilities, and technology companies—have signaled a willingness to work together on these challenges, even as residents remain skeptical about artificial intelligence, rising energy costs, and the pace of change. This effort must therefore be undertaken thoughtfully, guided first and foremost by the interests of Californians.

RECOMMENDATION 2: California should lead the nation on data-center energy regulation by grounding its policies in the best ideas from researchers, state agencies, utilities, developers, and other jurisdictions (state, federal, international).

Rate Design Approaches for Data Centers/Large Loads

Category	Purpose	Example Provisions
Eligibility	Defines the conditions under which data centers may connect to the grid.	Minimum load requirement
Contract Size	Specifies electricity purchase obligations and other provisions that reduce financial risk to utilities and ratepayers.	Obligation for customers to pay for a defined percentage of proposed demand
Contract Duration	Establishes the length and flexibility of agreements.	Defined contract terms that balance risk to ratepayers with ramp-up provisions and flexibility
Energy Source	Identifies requirements relating to the source of electricity.	Clean energy sourcing requirements
Other Provisions	Captures additional design considerations.	Marginal pricing mechanisms to reflect actual utility costs

Source: Lawrence Berkeley National Laboratory

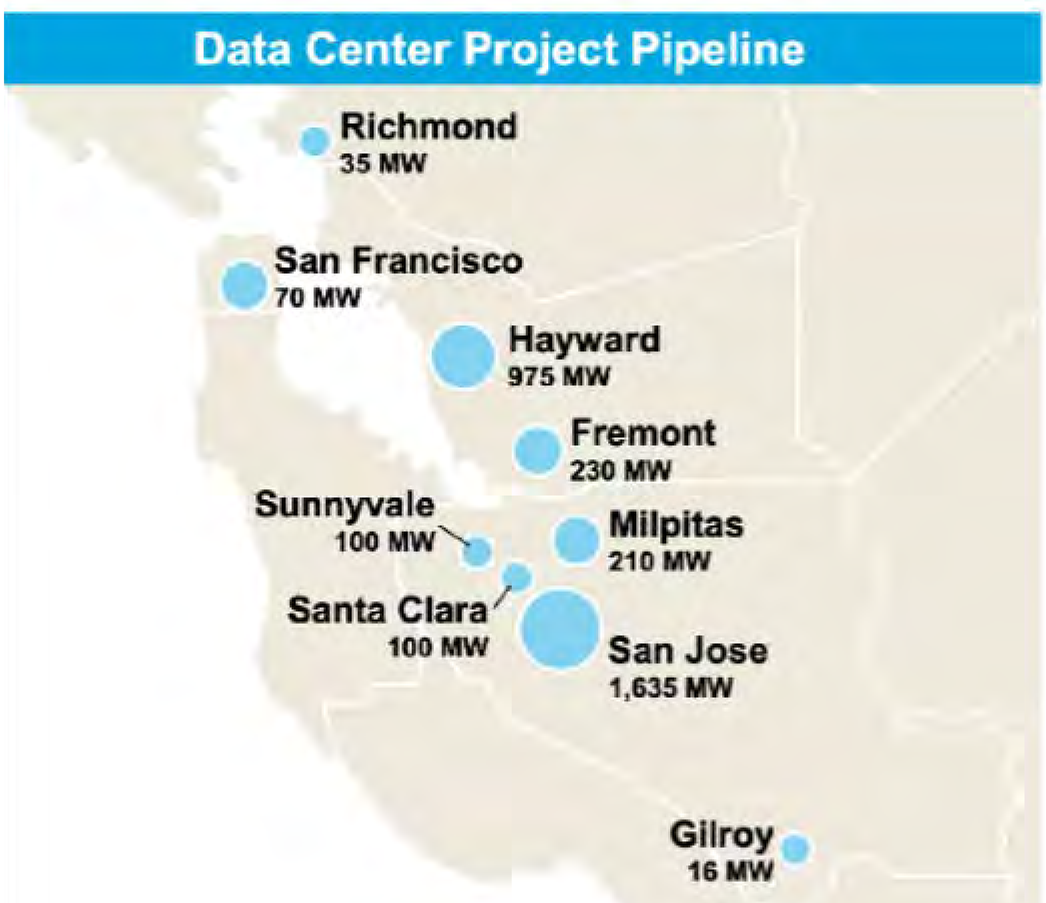
Data Centers and Electricity Policy in California

As California's utilities add data centers to the grid, the state has the opportunity—and, indeed, the obligation—to regulate them in a way that benefits all residents. This section describes where and how data centers are developing in California, how quickly their electricity demand is growing, and why the pace of development is putting stress on existing policy frameworks. It then explains the regulatory levers available to California—across utilities, state agencies, and local governments—and highlights how current rate structures and interconnection policies can shape costs, risks, and timelines.

DATA CENTERS IN CALIFORNIA

Data centers in California are currently concentrated in Santa Clara and Los Angeles.[33] These locations offer several advantages, including proximity to major clients, access to fast fiber networks, local governments accustomed to supporting large technology infrastructure, and utilities willing and able to provide the substantial electricity and water needed to support operations. In both areas, municipal utilities offer lower electricity rates than investor-owned utilities, providing substantial cost savings to data center operators.

PG&E Data Center Project Pipeline as of June 2024



Source: Pacific Gas & Electric Co.

More recently, however, development patterns appear to be shifting. A growing number of new data centers have been proposed in territory served by PG&E, one of California's investor-owned utilities, most of them concentrated in its Bay Area service territory and San Jose in particular.[34]

In terms of size and function, California's data centers have tended to be smaller than some of the massive "hyperscale" facilities located elsewhere in the United States. Many current and proposed facilities will support relatively modest computing operations and are middling in size, although proposals for much larger centers—including facilities up to roughly 600 megawatts—have emerged.[35]

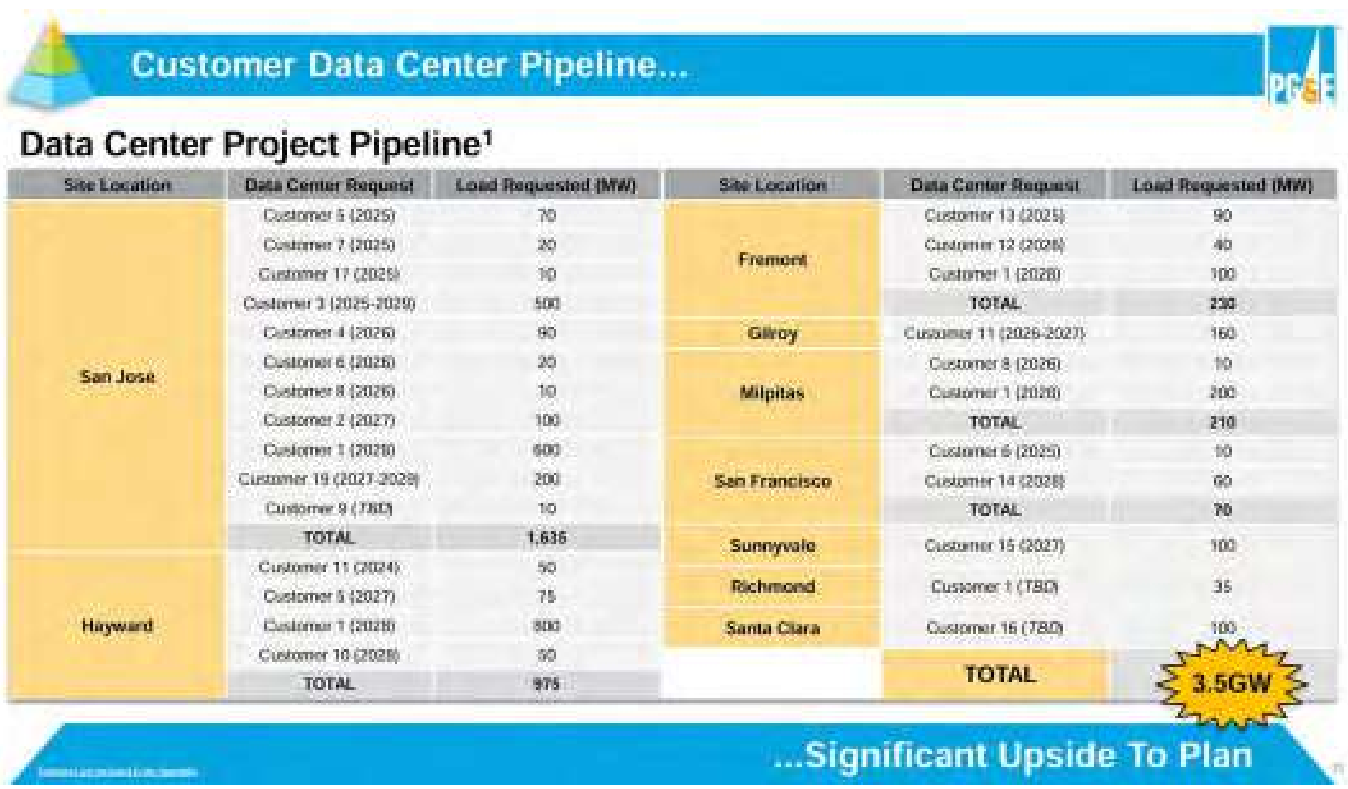
Nationally, hyperscale data centers have increasingly been built in rural areas and in states with relatively little direct connection to the tech industry. Reflecting that broader trend, some new facilities are being proposed for California's Central Valley, where land availability and grid access may be more favorable.[36]

Data center ownership is highly fragmented, with facilities operated both by major technology companies and by firms that specialize in developing and managing data centers. Several prominent companies with a significant national and global presence—including Meta, Apple, Google, Equinix, and GI Partners—are headquartered in California. Most other large operators maintain a substantial corporate presence in the state and manage multiple facilities here, reflecting California's important position in the broader data center ecosystem.[37]

CALIFORNIA DATA CENTER ELECTRICITY DEMAND

Electricity demand from data centers has grown rapidly and is expected to continue rising. A January 2025 report by Lawrence Berkeley National

Individual PG&E Data Center Load Requests as of June 2024


Customer Data Center Pipeline...

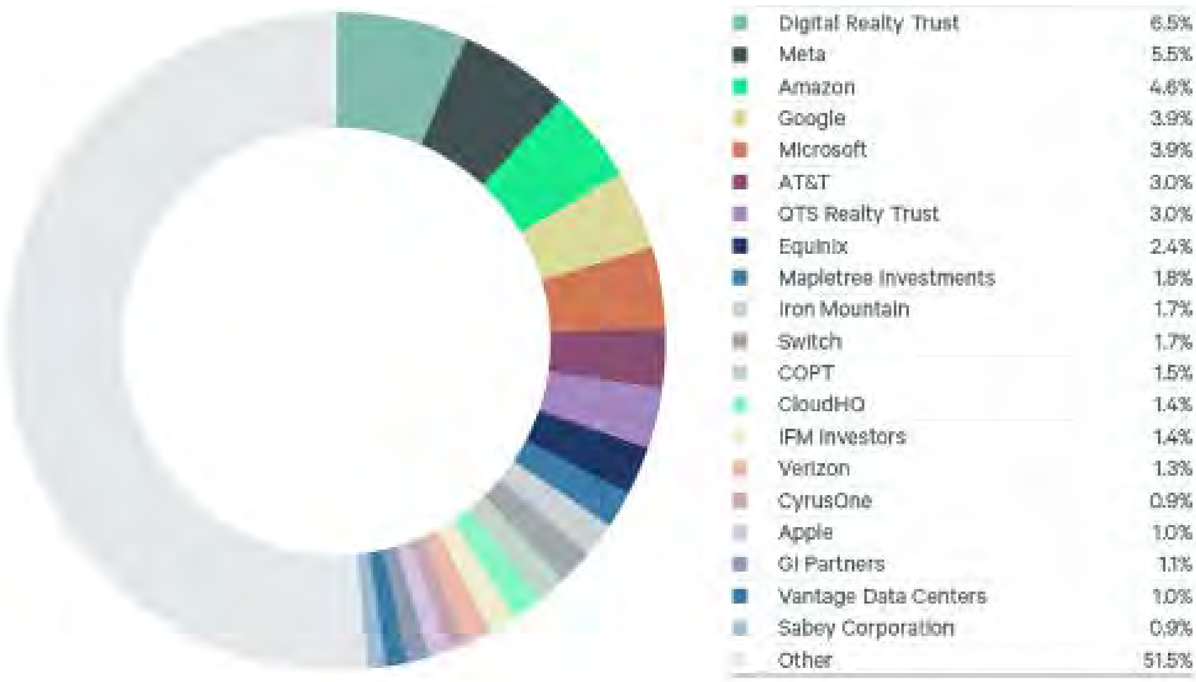
Data Center Project Pipeline[1]

Site Location	Data Center Request	Load Requested (MW)
San Jose	Customer 5 (2025)	70
	Customer 7 (2025)	20
	Customer 17 (2025)	10
	Customer 3 (2025-2029)	500
	Customer 4 (2026)	90
	Customer 6 (2026)	20
	Customer 8 (2026)	10
	Customer 2 (2027)	100
	Customer 1 (2028)	600
	Customer 19 (2027-2029)	200
	Customer 9 (TBD)	10
	TOTAL	1,636
Hayward	Customer 11 (2024)	50
	Customer 5 (2027)	75
	Customer 1 (2028)	800
	Customer 10 (2028)	50
	TOTAL	975
Fremont	Customer 13 (2025)	90
	Customer 12 (2028)	40
	Customer 1 (2028)	100
	TOTAL	230
Gilroy	Customer 11 (2026-2027)	160
Milpitas	Customer 8 (2026)	10
	Customer 1 (2028)	200
	TOTAL	210
San Francisco	Customer 6 (2025)	10
	Customer 14 (2028)	60
	TOTAL	70
Sunnyvale	Customer 15 (2027)	100
Richmond	Customer 1 (TBD)	35
Santa Clara	Customer 16 (TBD)	100
	TOTAL	3.5GW

...Significant Upside To Plan

Source: Pacific Gas & Electric Co.

Data Center U.S. Market Share by Square Footage



Source: CBRE Investment Management

Laboratory estimates that, by 2028, data centers could account for between roughly 7 and 12 percent of total U.S. electricity consumption—a dramatic increase from historic levels.[38]

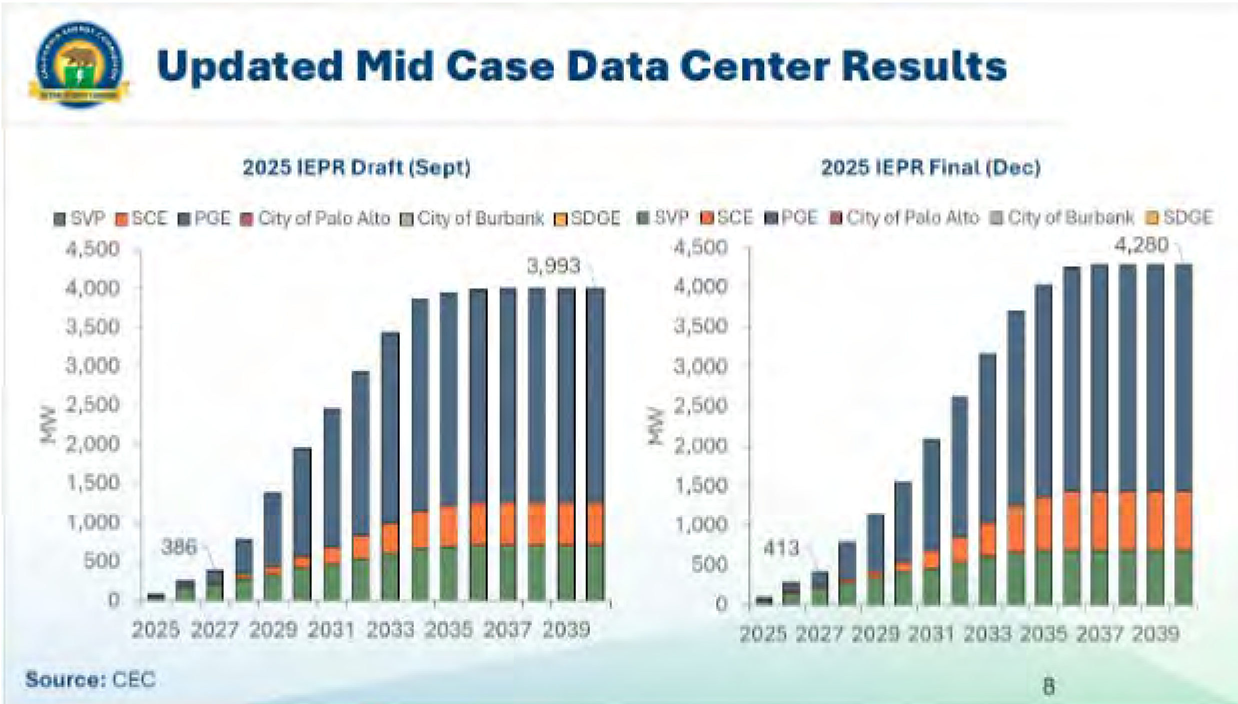
In California, the state's investor-owned utilities paint a steep growth trajectory based on interconnection requests from data centers. In February 2025, PG&E reported that data centers could add approximately 5.5 gigawatts of new demand between now and 2035. By May—just three months later—the figure had climbed to 8.7 gigawatts. And by late July, PG&E stated that it is "proactively working to serve 10 gigawatts of new electricity demand from data center projects over the next ten years"—an amount that is equivalent to four times the generating capacity of the Diablo Canyon nuclear power plant.[39]

To help guide transmission-level and local reliability planning, the California Energy Commission (CEC) provides data center load forecasts that are more tempered than PG&E's investor- and customer-oriented utility forecasts. According to the CEC's analysis, peak data center load could rise to between ~5.3 and 7.5 gigawatts (the chart below uses 2025 as its base year and does not reflect the ~1 gigawatt of current data center electricity load).[40] The CEC's mid-case projections are broadly consistent with forecasts from the California Independent System Operator (CAISO).[41]

The CEC's forecasts are based on the understanding that only some utility interconnection requests will come to fruition. Thus the CEC evaluates these requests using information about their status, capacity, and expected timelines, and then applies a 67 percent "utilization factor" to reflect that even completed projects are unlikely to operate at full capacity.[42] Finally, these projections are grounded in anonymized meter-level data from operating data centers, which provide insights into how facilities actually use electricity over time.

This raises the larger issue of data reporting and transparency around energy consumption by data centers. Utilities report electricity consumption to

CEC Data Center Demand Forecast (2025-2040)



Source: California Energy Commission

the state in aggregated form, reflecting long-standing assumptions that usage patterns within and across sectors are relatively stable and predictable. This approach has typically worked well for most sectors, where no single customer materially alters statewide trends. Utilities also provide anonymized 15-minute interval meter data, which allows analysts to infer data-center consumption patterns to some extent while preserving customer confidentiality.

The CEC has expressed confidence in its ability to accurately forecast data-center loads based on currently available information. However, enhanced reporting may support regulation, localized grid planning, and sound policymaking more generally.[43]

For example, because the meter-level data the CEC receives is anonymized, the agency cannot know with absolute confidence whether this data actually belongs to a data center, which facility it corresponds to, or where it is located. These gaps make it more difficult to identify localized grid stress, understand whether certain communities may face disproportionate environmental or reliability impacts, or call out individual facilities for under-utilizing the infrastructure built to serve them. At the same time, any discussion of improved reporting must balance transparency and accountability with legitimate privacy, competitiveness, and confidentiality concerns.

This issue is closely tied to the recommendations in this report, many of which call for solutions relating to precise engineering, load shifting, and intelligent use of backup energy generation and storage. Achieving those objectives may require more fine-grained, facility-relevant data so that the state can design, target, and evaluate interventions effectively. With this in mind, policymakers might consider advocating for the following:

- **Enhanced Facility-Level Reporting, With Privacy Safeguards:** Require confidential, facility-specific reporting of relevant metrics—

such as load, peak demand, utilization rates, load shapes, and location—under strict nondisclosure protections. More precise visibility would strengthen regulatory oversight and accountability while safeguarding proprietary information.

- **Targeted Reporting for Grid Stress and Community Impact:** Create enhanced reporting requirements triggered by specific conditions—such as regional grid congestion, high geographic concentration of large loads, or environmental-justice concerns. Facilities in higher-impact areas would provide more detailed data to help the state better assess localized risks.
- **Transparency Pathways That Respect Privacy:** Adopt a two-tiered transparency model in which regulators receive detailed confidential data while the public receives aggregated, non-identifiable reporting. This approach would promote public trust and accountability without compromising competitiveness or security.

Enhanced data center reporting will provide insight into how individual facilities interact with the grid, affect local reliability, and contribute to infrastructure costs. Ensuring that regulators have access to essential, confidential facility-level information will help them plan responsibly, protect communities and ratepayers, and guide thoughtful policy decisions moving forward.

RECOMMENDATION 3: California should enable regulators to confidentially access essential facility-level electricity-use data from large data centers, supported by structured data-sharing among agencies, to improve planning, assess localized impacts, and ensure fair cost responsibility while protecting sensitive business information.

THE DISPARATE PACE OF ENERGY POLICY AND THE CONSEQUENCES

The extraordinary pace of investment in artificial intelligence and the data centers that support it has intensified pressure on California's grid. Vacancy rates for existing facilities are low, especially in the state's major data center clusters, and trillions of dollars are being poured into new construction nationwide.[44]

Because access to electricity is widely viewed as the primary bottleneck to growth, a data center site with a secured interconnection has become a highly valued asset—capable of serving immediate demand and, potentially, being resold at a premium.[45] Companies also treat data centers as infrastructure investments that produce near-term revenue and offer tax advantages through accelerated depreciation, especially in jurisdictions that pair these benefits with targeted incentives.[46]

While developers hope to build data centers at breakneck speed and utilities are eager to add load and upgrade infrastructure, identifying policy solutions to large load grid interconnection challenges is slower by nature.[47]

Intelligent siting requires careful analysis of existing substation and transmission capacity; selective upgrades require engineering studies and environmental review; and the research needed to support long-term policy—including studies on system impacts, rate design, and load forecasting—takes time. Legislative reforms targeting data centers, still in an unsettled state nationwide, follow their own measured timeline of querying stakeholders, settling on policy positions, and drafting and advocating for legislation.[48] This mismatch can lead to frustration on the part of developers and, if rushed, poor decision making on the part of regulators and policymakers.

Indeed, the data center "gold rush" has already led to problems for energy regulators. For example, to preserve strategic flexibility, data center developers sometimes overstate the expected energy requirements of proposed projects and file multiple interconnection requests for the same project or for several potential sites.[49] Because the cost of filing is relatively low, and because utilities may benefit from the possibility of higher future load, these duplicative applications can accumulate quickly.[50]

Some states—such as Texas—have begun requiring. developers to disclose overlapping requests and to demonstrate ownership or control of land associated with each application.[51] In Ohio, a new requirement to commit to purchasing a hefty proportion of proposed new load has already led to a marked decrease in the number of interconnection applications from data centers—a sign, policymakers say, that the tariff is achieving its intended goal.[52]

DATA CENTERS AND JURISDICTIONAL POLICY LEVERS

The state does not typically determine whether a data center may be built, but it can shape the conditions under which a project moves forward. Prior to starting construction, developers must satisfy a number of utility and governmental requirements. These requirements represent potential policy levers through which state government can influence where, how, and under what conditions data centers are developed.[53] The following categories summarize the primary regulatory and permitting interactions required for new development:

1. Utility Interconnection & Transmission Planning: Developers must submit interconnection requests to utilities, specifying expected load and triggering engineering studies and any required infrastructure upgrades.[54] Large projects that depend on new or expanded transmission may also intersect with CAISO's Transmission Planning Process, even though CAISO is not a state agency. This differs from the California Energy Commission's role in statewide load forecasting, which informs—but does not replace—CAISO's engineering-level transmission planning.

2. Air Quality & Backup Generation: Backup-generator emissions must be permitted by local air districts, which enforce limits on diesel engines and related equipment. CARB sets statewide emissions standards and guidance, but it does not issue permits directly; day-to-day oversight occurs at the district level.[55]

3. Local Permitting, Water Use, & CEQA Review: Cities and counties manage zoning approvals, water-use permits, and building requirements. Most large data centers also undergo CEQA review, which now routinely evaluates air emissions, noise, water consumption, local ecological impacts, and—notably—energy use and the project's likely demand on the grid.[56]

4. Statewide Regulatory Influence & Incentives: Although the state does not directly approve siting decisions, it shapes the broader environment for data-center development through rates, cost-allocation rules, data-reporting requirements, energy-efficiency standards, and statewide demand forecasting. Some states also offer financial incentives such as sales-tax exemptions for construction and equipment purchases, typically tied to minimum investment or job-creation thresholds.[57]

The state's authority over whether individual data centers are built is thus relatively limited. Instead, its influence is exercised indirectly through electricity rates, reporting requirements, load-forecasting processes, and regulatory mechanisms that guide utility planning and shape the costs and conditions under which new projects are developed. Together, these touchpoints form the practical extent of California's policy leverage over data-center expansion.

DATA CENTERS AND CALIFORNIA'S REGULATORY AND PLANNING BODIES

California relies on several agencies—along with the California Independent System Operator—to manage the policies, planning processes, and regulatory tools that shape how large new electricity loads such as data centers are integrated into the grid.[58] Each entity plays a distinct role, from regulating utilities and overseeing rates to forecasting statewide demand and planning transmission infrastructure.

California Regulatory Agency Overview

Entity	Core Responsibilities	Policy Levers Relevant to Data Centers
California Public Utilities Commission (CPUC)	Regulates investor-owned utilities (IOUs); oversees rates, tariffs, cost-recovery mechanisms, utility procurement, capital investments, and implementation of statutory requirements.	▪ Designs and approves rate structures and customer classes (including those affecting large-load customers). ▪ Approves utility interconnection policies (e.g., Rule 30). ▪ Approves IOU capital projects needed to serve large loads (substations, feeders, related upgrades). ▪ Can require reporting frameworks and data submission by utilities. ▪ Oversees programs that may shape costs (demand-response, load-management, etc.).
California Energy Commission (CEC)	Conducts statewide demand forecasting; oversees power-plant siting; sets building and appliance efficiency standards; conducts system-level planning studies.	▪ Produces load forecasts that guide CAISO's and IOUs' long-term transmission planning. ▪ Reviews new power-plant siting that may support grid capacity. ▪ Establishes efficiency standards affecting baseline building loads. ▪ Facilitates coordination with utilities (IOUs and POUs) to understand new large loads, including data centers.
California Independent System Operator (CAISO)	Plans and operates most of the state's transmission grid; conducts the Transmission Planning Process (TPP); oversees transmission-level interconnections.	▪ Determines whether and when large new loads can be accommodated. ▪ Manages interconnection processes for large projects requiring transmission-level service. ▪ Uses CEC forecasts and utility data to prioritize planning decisions.
California Air Resources Board (CARB)	Sets statewide air-quality and climate standards; oversees statewide emissions rules and the activities of air-quality regulators.	▪ Establishes overarching requirements that govern emissions sources relevant to data-center equipment (e.g., generators, stationary engines). ▪ Issues climate-related regulations (e.g., reporting, diesel-engine standards) that shape what equipment data centers can use.

CURRENT DATA CENTER RATE POLICY

Under California's current rate design, large data centers generally pay lower electricity rates than most other customers because they typically interconnect at the transmission level and do not rely on the local distribution system. This structure is broadly consistent with the principle of "cost causation," under which "[r]ates for each class of customer are proportionate to the costs of serving that class of customer".[59]

Most large data centers receive service at transmission-level voltages. Because they do not use the lower-voltage distribution network, they are not assigned the full suite of costs associated with building, operating, and maintaining that system. However, as PG&E notes, transmission-level customers are not entirely exempt from distribution-related charges. Within PG&E's B-20T rate, for example, customers pay a customer charge that recovers certain metering and billing costs, as well as a distribution demand component that includes some wildfire-related and demand-response costs.[60]

Even with these protections in place, customers taking service at the transmission level still pay less than comparable customers receiving distribution service. As of September 1, 2025, customers in PG&E's B-20 "transmission firm" class pay roughly 10 to 18 cents per kilowatt-hour (depending on time of use), plus a monthly demand charge—a separate fee based on the highest level of power the customer draws from the grid during a billing period.[61]

By contrast, industrial "secondary firm" customers—those served at the distribution level—pay approximately 12 to 21 cents per kilowatt-hour and face a considerably higher demand charge. While this difference may appear modest, it amounts to at least a 14 to 17 percent discount for transmission-level service. Given the enormous electricity requirements of large data centers, this results in a meaningful reduction in the overall revenue these customers contribute to the system.

This matters in the context of California's broader affordability challenges. Many of the main drivers of rising electricity costs—including significant wildfire mitigation investments and other expenses traditionally embedded in distribution rates—are borne primarily by customers served at the distribution level. Although transmission-level data centers do contribute to some of these costs through embedded charges, they do not shoulder the same level of distribution cost responsibility as other customers.

While new large loads may help spread certain fixed costs related to generation, transmission, and public purpose programs, they are unlikely to offset the distribution-level cost pressures that have been central to rising residential and small-business bills. This raises an important policy question for the state: whether California is comfortable with this distribution of costs, or whether a different allocation approach is warranted given today's affordability and equity concerns.

FEATURES OF A NEW VERY-LARGE-LOAD TARIFF

In the sections that follow, this report outlines several steps California can take to protect ratepayers while managing rapid load growth due to data centers. A primary element of this strategy is the creation of a very-large-load tariff tailored to the unprecedented scale and financial risks associated with data centers and similar high-demand facilities (for example, cryptocurrency mining).[62] Here, we describe the critical elements that the Commission believes should anchor such a tariff, which we believe the California Legislature should make statutory requirements.

- **Creditworthiness:** A very-large-load tariff should require enhanced due diligence on the financial stability of data center developers and their parent companies, including evaluation of long-term business viability, financing sources, and the likelihood that proposed loads will materialize and persist. This may include bonding requirements or other forms of financial assurance designed to protect utilities and ratepayers against default, abandonment, or significant downsizing of planned facilities.

- **Prepayment for Infrastructure:** To reduce financial exposure for utilities and prevent shifting risk onto other customers, the tariff should require upfront infrastructure payments. These payments would help fund the build-out of new transmission, distribution, and substation facilities needed to serve data centers, ensuring that utilities are not left with stranded assets if projects stall, downsize, or fail to materialize.
- **Cost Recovery:** The tariff should establish a special rate structure tailored to very large loads, recognizing that data centers impose unprecedented planning, infrastructure, and reliability demands. This rate should be structured to ensure meaningful contributions to overall system costs, including transmission, distribution, and reliability investments, while also ensuring that these customers help fund California-specific policy costs such as wildfire mitigation and climate resilience programs.
- **Minimum Commitments:** The tariff should include minimum payment obligations requiring customers to pay for a defined percentage of their proposed demand regardless of whether their full projected load ultimately develops. Such provisions would align incentives, discourage speculative requests, and ensure a stable revenue base to support system investments made in anticipation of these loads.
- **Load Flexibility and Curtailment:** A very-large-load tariff should require customers to design facilities and operations so they can meaningfully reduce or shift demand during periods of system stress. This could include participation in demand-response programs, contractual curtailment obligations, or other flexibility mechanisms that help minimize the need for costly new capacity and support grid reliability.
- **Exit Protections:** Finally, a very-large-load tariff should incorporate exit fees to address early termination, significant downsizing, or withdrawal. These provisions are essential to preventing cost shifts to other customers and ensuring that utilities can recover investments made in reliance on customer commitments.

It should be noted that some of these concepts already appear in PG&E's proposed Electric Rule 30, which was approved by the CPUC on an interim basis in July 2025 while its final form is still being considered.[63] This tariff establishes requirements for transmission-level retail customers seeking to connect to the grid; historically, such customers were handled on a case-by-case basis, leading to uncertainty and long interconnection delays. It seeks to create a standardized framework for processing transmission-level interconnection requests while incorporating measures to protect ratepayers.[64]

Under the currently-approved interim structure of Rule 30, transmission-level customers—including data center developers—must provide upfront funding for the transmission facilities needed to serve them, including full pre-funding of network upgrades if they wish to proceed during the interim period. However, the CPUC has explicitly deferred decisions regarding any refunds, repayment of pre-funded amounts, and whether interest should accrue. As a result, developers must proceed with the understanding that repayment is not guaranteed at this stage. Other tariff design elements and contractual protections also remain under deliberation and will be addressed in the CPUC's final decision.

Policymakers may wish to build on lessons emerging from the interim implementation of Rule 30 and the insights gained once it is finalized to guide the development of a statewide statutory framework for very-large-load customers. For example, the Legislature could establish a two-year study period during which the CPUC evaluates the effectiveness of Rule 30's provisions in areas relating to cost recovery, risk allocation, infrastructure financing, customer commitments, and overall system reliability. Findings from this study period would then inform the design of a statewide tariff that ensures very large loads contribute fairly to system costs and supports California's broader energy and climate goals.

RECOMMENDATION 4: California should ensure that data centers pay for costs imposed on the electric system through tariffs and cost-allocation policies that recover infrastructure and grid-services costs in a manner that benefits rather than burdens other ratepayers.

IMPROVING THE INTERCONNECTION PROCESS

California's utilities are legally obligated to provide service to customers who request it, but the recent wave of extraordinarily large interconnection requests—that is, requests to draw energy from the grid—from data center developers has presented a unique challenge in terms of scale and complexity.[65]

The interconnection stage is a critical policy and planning checkpoint where developer needs and the state's broader public-interest responsibilities can align. Properly designed, interconnection requirements—including those associated with Rule 30—can both help utilities provide faster, more coordinated review and ensure that developers make meaningful commitments before gaining grid access.

California can look to other states for examples. In June 2025, Texas enacted Senate Bill 6, explicitly directing regulators to process large-load interconnection requests "in a manner designed to support business development in this state while minimizing stranded infrastructure costs and maintaining system reliability."[66] In doing so, Texas has made the interconnection stage a primary vehicle for regulating large-load growth.

The Texas law closely parallels many of the principles reflected in California's proposed Rule 30.[67] Developers of facilities exceeding 75 megawatts must cover study costs and fund needed grid upgrades up front, with reimbursement available only once projects become operational or capacity is reassigned. In addition, SB 6 requires developers to disclose whether they have made additional interconnection requests inside or outside Texas and demonstrate proof of site control. Finally, the legislation authorizes curtailment provisions during periods of system stress.[68]

California should similarly treat the interconnection stage as a structured policy moment—one that requires developers to provide essential information, commit to cost, reliability, and transparency obligations, and align their projects with the state's broader public-interest objectives before gaining access to the grid.

In return, California should support utilities in speeding up interconnection request review and approval. Here are some ways this could be achieved:

- **Streamlined Permitting & Environmental Review:** Strengthen and coordinate permitting processes through state facilitation, improved agency alignment, and reduced duplication so projects can advance more quickly while maintaining environmental protections.
- **Clearer, Faster Interconnection Studies:** Expand engineering capacity, adopt more standardized study approaches, and improve queue discipline so credible projects move more quickly through analysis.
- **Faster Transmission & Distribution Upgrades:** Better coordinate state planning around expected data-center growth and give utilities clearer permission and safeguards so they can start building needed grid upgrades sooner without putting ordinary customers at financial risk.
- **Supply Chain & Equipment Support:** Use state-led bulk procurement, incentives for in-state manufacturing, and permission for proactive purchasing so difficult-to-acquire equipment does not become a structural bottleneck.
- **Workforce & Construction Capacity:** Invest in accelerated training pipelines, partnerships with unions and higher education, and standardized designs to speed construction and reduce specialized engineering constraints.

By adopting these supports, California can strike a balance between setting responsible guardrails for large-load interconnections and enabling data centers to connect to the grid and begin operating more quickly.

RECOMMENDATION 5: While maintaining a rate structure that ensures shared costs and a high standard of regulatory review, California should accelerate data-center interconnection and permitting decisions to give developers greater predictability in planning and investment.

MITIGATING RISK FROM STRANDED ASSETS

The tech sector's economic strength has significant fiscal consequences for California—a fact that has helped drive the Governor's support for generative AI through a variety of initiatives and legislative decisions.[69] Despite uncertainty about AI's long-term economic value, firms continue to project rapid growth. California-based NVIDIA, for example, reported roughly $32 billion in Q3 2025 revenue and now exceeds $5 trillion in market value. And major firms like Microsoft, Google, Apple, and Amazon also report strong AI-linked profits and valuations.[70] Analysts project up to $3 trillion in global data-center investment by 2029—much of it financed through borrowing—and note that AI spending is propping up an otherwise sluggish economy.[71] The frenzy to invest in AI and data centers has raised concerns that utilities may overbuild grid infrastructure in response to speculative demand, leaving ratepayers responsible if anticipated load does not materialize. One JP Morgan analysis estimated that achieving a 10 percent return on an expected $5 trillion in AI investment would require $650 billion in annual revenue, a level many view as unlikely.[72] A major market correction could result in data center developers going bankrupt or a dramatic reduction in data center operations generally.

There is also concern that data centers could leave the state, leaving behind long-term infrastructure investments that no longer serve their purpose. Servers are portable, and the centers that house them are essentially storage facilities with sophisticated energy and cooling systems.[73] While Silicon Valley is the national—if not global—hub of the tech industry, its companies have often threatened to leave California for jurisdictions with more favorable incentives, tax structures, regulatory climates, or cheaper water and electricity.[74]

Technological innovation could also reduce the amount of energy required by data centers, potentially stranding investments in oversized or unnecessary infrastructure. In terms of hardware, NVIDIA has reported a 4,000-fold performance-per-watt improvement in GPUs over a decade.[75] And Google's tensor processing unit (TPU) AI chips are reportedly three times more energy efficient than comparable models.[76]

AI training has also become much more efficient. China's DeepSeek-R1 reportedly achieved high performance with a small fraction of the hardware and energy of comparable models;[77] UC Berkeley's NovaSky team trained a highly capable reasoning model in 19 hours for roughly $450;[78] and industry leaders such as OpenAI, Google, and Anthropic now routinely release "mini" models that deliver strong performance with significantly less compute.[79]

Some observers compare the present moment to past infrastructure booms and busts, with railroads and fiber optics being the most commonly cited examples. Both were initially overbuilt but ultimately produced lasting public benefits.[80] Witnesses at the November hearing similarly noted that data-center development could accelerate needed upgrades to high-voltage transmission.[81] However, should these facilities close, relocate, or become obsolete as technology advances, ratepayers could be left paying for "stranded assets"—power plants, substations, or transmission lines that no longer serve customers.[82]

As California considers building significant new energy infrastructure to support data centers, policymakers must therefore weigh benefits against longer-term financial risks. Some action has already been taken by the state to protect against this. Prepayment for transmission infrastructure required under PG&E's Rule 30 is one example; another is the CEC's use of utilization factors and confidence levels to develop a more realistic picture of which proposed infrastructure projects are likely to be built.

A number of other states—including Ohio and Texas, discussed above—have taken steps to protect against the risk that infrastructure built to serve data centers could become stranded if projected load does not materialize. These actions have taken several forms that California policymakers may wish to consider.

For example, Georgia, South Carolina, Oregon, and Virginia have created or are considering very-large-load tariffs designed to better ensure that data centers bear the costs of the infrastructure required to serve them.[83] These tariffs typically rely on mechanisms such as minimum demand charges or revenue commitments, longer contract terms, and higher fixed charges, reducing the likelihood that utilities will build long-lived assets based on speculative demand.

Other states, including Minnesota, have adopted more direct ratepayer protections.[84] These approaches include statutory prohibitions on passing stranded-asset costs on to ratepayers, requirements for heightened regulatory review when utilities seek to recover costs associated with large loads, and policies that assign upgrade costs to the customer that triggers the investment.

In addition to drawing on lessons from these states, California has further options available to ensure that ratepayers are not left covering the costs of infrastructure that ultimately does not serve new load. Potential steps include:

- **Strengthen financial accountability for very large or mobile loads:** Require stronger creditworthiness reviews, collateral, minimum load commitments, and/or exit fees so that developers carry real financial risk if demand fails to materialize. Doing so will better align private incentives with the public interest and discourage inflated energy requests.
- **Prioritize infrastructure that retains value even if AI demand slows or shifts:** When large investments move forward, they should also support broader reliability, clean-energy integration, electrification, wildfire resilience, or other public benefits. This ensures that even if data-center demand softens, Californians still benefit from the infrastructure built.
- **Clarify that prudence applies to both underbuilding and overbuilding:** Regulators must explicitly evaluate the risk of unnecessary or oversized infrastructure investments along with the risk of insufficient capacity.

RECOMMENDATION 6: California should establish consistent, statewide safeguards that ensure large data-center customers bear appropriate financial risk, while making certain that major grid investments continue to serve the public even if data center demand evolves differently than has been predicted.

BROADENING DATA CENTER OVERSIGHT

While investor-owned utilities operate under direct CPUC oversight, publicly owned utilities play an important role in serving complex, energy-intensive loads. However, current state policy tools do not provide the same level of transparency, planning alignment, or cost-risk protections for customers in these territories. As California faces unprecedented electricity demand driven by rapid data center growth, uneven standards across different utility types may create avoidable risks: uncertainty in statewide planning, inconsistent community protections, and potential cost shifts if major projects fail to materialize as expected.

A recent example can be found in the City of Imperial, which is served by the publicly owned Imperial Irrigation District.[85] County officials are considering approval of a proposed 330-megawatt data center, but city officials and local residents have stated that they were not adequately consulted and that environmental review was insufficient. While the developer has asserted that the facility would rely on renewable energy and recycled water, these claims have been disputed, and it is unclear whether such commitments would be legally binding.

The goal is not to displace local control. Rather, California should establish clear statewide expectations—particularly around reporting, reliability assurances, cost-recovery protections, environmental impacts, and coordination with agencies such as the CEC, CPUC, and CAISO—so that every community benefits from the same level of protection and every utility operates within a consistent framework. Potential minimum statewide standards could include:

- **Financial Risk Protections:** Extend versions of protections evolving under Rule 30 statewide—such as prepayment, financial security instruments, or clear cost-recovery assurances—to reduce stranded-asset risk regardless of utility type.
- **Baseline Reporting Requirements:** Require consistent, confidential reporting of critical data center metrics (e.g., anticipated and actual load, load shape, backup generation information) across all utility types to support statewide planning and risk assessment.
- **Environmental and Community Impact Guardrails:** Establish minimum standards for transparency relating to backup generation transparency and community engagement regardless of jurisdiction.

California's approach to managing very large loads should not depend on where a data center is located or which utility happens to serve it; residents across the state deserve the same level of financial protection, planning rigor, and environmental safeguards. Establishing clear, statewide standards will help guarantee that data center risks are handled consistently.

RECOMMENDATION 7: The State should establish consistent statewide oversight expectations for large data center developments in territories served by publicly owned utilities, municipal utilities, irrigation districts, and community choice aggregators, ensuring that risks to reliability, costs, and local communities are addressed with the same level of rigor as investor-owned utilities.

Data Centers and Grid Planning

This section explores how California can integrate new data centers more efficiently and at lower cost by making smarter use of the generation and transmission capacity it already has. It begins by acknowledging two important realities: there are legitimate concerns about grid reliability as large new loads come online, and at the same time, there is meaningful untapped capacity within the system that could help meet this demand if used wisely.

Recognizing both facts—and designing complementary policy responses to address risk while unlocking available capacity—will be essential to ensuring that data center growth does not undermine reliability. From that foundation, the section then examines three strategies: making targeted upgrades to existing infrastructure, encouraging data centers to locate where the grid can most readily support them, and promoting demand flexibility to shift or shape energy use and relieve stress on the system.

DATA CENTERS AND GRID RELIABILITY

Planning for a reliable electric system requires preparing for the periods of highest demand—the "peak" moments on the hottest summer days or during extreme weather events, when air-conditioning and other essential uses place maximum pressure on the grid. These moments represent a very small share of overall electricity use but nevertheless determine the scale of infrastructure that must be built and maintained.[86] As energy expert Sean Fleming has written, the "system is actually pretty under-utilized by design, with spare capacity more than 99.99% of the time. Over the last 5 years, California's average load has been just ~50% of its [2022] peak".[87]

In other words, this excess capacity is a deliberate feature of the grid. Electricity systems must be built to meet infrequent moments of peak demand, because when demand exceeds available capacity, the result can be power outages that disrupt daily life and pose serious risks to public safety.[88] California's energy planners have therefore focused on strengthening the grid's ability to perform during these critical periods, and have recently expressed confidence in the system's resilience—even as rising temperatures drive sharper and more frequent demand spikes.[89]

Actions taken include adding over 20 gigawatts of new energy supply; the addition of over 12 gigawatts of battery storage (including residential) to complement variable energy sources like wind and solar; establishing a Strategic Reliability Reserve and coordinating with the Western Energy Imbalance Market to improve reliability; and extending operations at the Diablo Canyon nuclear power plant. This has allowed the grid to remain operational despite record setting heat events.[90]

Significant progress has been made in strengthening the grid's ability to meet peak demand. Looking ahead, however, the rapid growth of data centers will add substantial new load, and the industry's strong emphasis on avoiding downtime creates a natural tendency to maintain operations even during periods of system stress. While this approach supports facility-level reliability, it can strain the broader electric system when aggregate demand approaches—or exceeds—available peak capacity. As a result, concerns about grid reliability are increasingly shaping data-center policy decisions in other states facing rapid large-load growth.

Texas offers a clear illustration. During Winter Storm Uri in 2021, electricity demand surpassed available capacity, triggering widespread outages and underscoring the risks of unmanaged peak load. In response, the state enacted SB 6, which authorizes both voluntary and mandatory curtailment of data-center load during emergencies, giving grid operators a backstop to protect systemwide reliability and public safety.[91]

There are several ways to reduce the need for new infrastructure by making better use of the grid's existing capacity, which are discussed in the sections that follow. At the outset, however, it is important to emphasize that maintaining grid reliability must remain a core priority. Planners and policymakers have meaningful opportunities to shape data center policy in ways that support system reliability while accommodating continued load growth. Policy options could include:

- **Mandatory Curtailment Authority During Grid Emergencies:** Require large data centers to reduce load when grid operators declare emergencies, with clear triggers, enforcement mechanisms, and penalties to ensure compliance supports system reliability.
- **Reliability-Based Interconnection Conditions:** Condition approval of large-load interconnections based on enforceable reliability obligations, ensuring that data centers are prepared to respond to periods of system stress as a prerequisite for connecting to the grid.
- **Phased or Probationary Load Approval:** Allow regulators or utilities to approve large data center loads in stages, with subsequent phases contingent on demonstrated performance and compliance with reliability requirements.
- **Centralized State Oversight of Large-Load Reliability Impacts:** Establish coordinated, statewide review of the cumulative reliability impacts of large loads across utility territories to avoid fragmented decision-making and system-wide risk.
- **Clear Enforcement and Accountability Framework:** Define unambiguous authority, reporting requirements, and penalties related to reliability obligations so that curtailment and compliance are predictable, transparent, and credible.

Together, these policy actions reinforce the principle that grid reliability must remain a primary condition for integrating new large data center loads, particularly as demand—and potential system stress—grow in the years to come.

RECOMMENDATION 8: California should ensure that new large data center loads are integrated in a manner that preserves system reliability and does not disadvantage existing customers.

MAXIMIZING EXISTING GRID CAPACITY

One of the points the Commission heard repeatedly is that, most of the time, California's electric system is not "at capacity". This means that new large loads may be added to the grid without exceeding system limits if development is planned thoughtfully and guided by clear principles.[92] Several strategies were identified that make better use of existing grid capacity, reducing costs by improving efficiency rather than requiring new infrastructure. These are outlined below.

Targeted Infrastructure Upgrades: Research conducted at Stanford University has identified bottlenecks in the Western Energy Coordinating Council region—the interconnected electric grid covering most of the western United States, including California.[93] While transmission lines often operate below their rated capacity, their ability to deliver power is often constrained by large power transformers, which regulate how much electricity enters and exits those lines. Many of these transformers have met or exceeded their design life and now represent a critical limiting factor. According to the Stanford report, "[t]argeted replacement, refurbishment, and added flexibility at these transformer interconnection points can therefore unlock additional use of existing line capacity and increase transfer capability." It also finds that congestion and reliability risks are often driven by a relatively small number of overutilized lines.

Upgrades at these critical points in the grid could expand—and are already expanding—overall system capacity and enable additional generation to be connected without requiring the large, costly, and time-consuming expansion of the broader transmission network.[94] The Commission therefore recommends that California require utilities to evaluate and prioritize targeted, asset-level upgrades—such as transformer replacement, refurbishment, or added operational flexibility—before proposing large, costly transmission expansions. This expectation could be reinforced by aligning cost-recovery rules so that efficient, capacity-unlocking upgrades receive expedited or preferential treatment, while major new infrastructure is subject to greater scrutiny when lower-cost alternatives have not been fully pursued.

Intelligent Siting: Developers, policymakers, and regulators share a common interest in locating data centers in ways that control costs, make effective use of existing grid capacity, and limit environmental and community impacts. Relevant considerations include local grid conditions and electricity rates, permitting and regulatory feasibility, grid reliability and flexibility, and access to resources such as water and fast fiber optic cables.[95] Siting decisions are particularly consequential because the cost of upgrading the grid to serve data centers in congested areas can be substantial. For example, according to the Public Advocates Office, recently-approved transmission upgrades to accommodate 2.5 gigawatts of new load in the South Bay Area will cost over $2 billion.[96]

This illustrates how alternative siting decisions—particularly in areas with readily available grid capacity—can reduce the need for costly infrastructure upgrades. One way developers and policymakers are working to balance the many factors involved in data center siting is through data-driven mapping tools. For instance, the National Laboratory of the Rockies has developed maps that integrate infrastructure and resource constraints, while Esri promotes the use of its ArcGIS platform to help identify optimal data center locations by evaluating a wide range of pertinent variables.[97]

Building on these examples, California could play a constructive role by facilitating access to similarly comprehensive, high-quality siting data in a shared, user-friendly format—such as a statewide map or dashboard. Bringing together information on grid capacity and congestion, electricity rates, permitting constraints, water availability, and other relevant factors would allow developers, policymakers, regulators, and community members to engage in siting discussions from a common factual foundation. Such a tool would not dictate outcomes, but would support more informed, efficient, and transparent decision-making across the many forums in which data center siting is considered.

Cluster Studies: Most data center interconnection requests in California are concentrated in just a few geographic areas. When multiple large load requests are in close proximity, they may be more efficiently addressed through one “cluster” study that evaluates them together rather than through individual studies. A cluster study can examine the combined demand of multiple facilities and identify coordinated, system-level solutions that can accommodate projected load growth more efficiently and at lower cost. For example, in its original 2024 cluster study, Pacific Gas and Electric Company evaluated 740 megawatts of proposed electricity demand from multiple data center developers as a single, combined load.[98]

California could encourage more efficient batch or cluster studies through a set of targeted, low-disruption policy steps that focus on clarity and predictability rather than mandates. The CPUC could explicitly authorize and encourage utilities to use cluster studies when multiple large interconnection requests emerge in close geographic or temporal proximity, supported by objective triggers that signal when a batch approach is appropriate.

At the same time, utilities could be directed to clearly communicate when cluster studies are likely, how timelines and cost allocation would work, and what developers should expect, reducing uncertainty and concerns about ad hoc delays. Issuing this direction as guidance or a policy statement—rather than a rigid rule—would help normalize cluster studies as a best practice while preserving flexibility for utilities and developers alike.

RECOMMENDATION 9: California should maximize the use of existing grid capacity when considering data-center siting and new infrastructure, supported by transparent data and visualization tools that help developers and policymakers identify locations where capacity is available or can be efficiently expanded.

DEMAND FLEXIBILITY

Demand flexibility—also called demand response or load shifting—refers to a data center's ability to adjust electricity use in response to grid conditions, particularly during periods of high demand or system stress.[99] This may include temporarily reducing electricity consumption or rescheduling energy-intensive activities to times when the grid is less constrained. Demand flexibility is important because the electric system is built to meet peak loads, meaning that substantial capacity is installed to serve relatively infrequent periods of highest demand. When large new loads can respond to these conditions, they can often be integrated into the existing grid with fewer infrastructure upgrades than would otherwise be required.

Demand flexibility can take different forms, ranging from voluntary, price-responsive reductions to more structured or contractually obligated curtailment. While not all data-center workloads can be shifted without affecting operations, even partial flexibility can meaningfully reduce peak pressure on the grid and limit the need for costly new infrastructure. For example, a Duke University study found that up to 76 gigawatts of new electricity load—roughly 10 percent of current U.S. aggregate peak demand—could be accommodated if new loads are reduced, on average, for just 0.25 percent of their maximum annual operating hours. Moreover, in the vast majority of cases, only a 50 percent reduction in energy use is required.[100]

One important form of demand flexibility is time-based load shifting.[101] On a daily basis, some data centers may be able to move certain computing tasks to off-peak hours. Shifting demand in this way can ease stress during the most constrained times while improving utilization of grid capacity that would otherwise go unused. Improved load balance can also reduce operational challenges associated with fluctuations in generation, which is particularly useful as California continues to integrate large amounts of renewable energy.[102]

Demand flexibility can be especially valuable on a seasonal basis during periods of peak demand. In California, the grid is often most stressed during hot summer months, when air-conditioning use drives electricity demand to its highest levels. In California, incentive-based programs such as PG&E's Flex Connect Pilot are designed to encourage voluntary reductions in load during critical periods.[103]

Flexibility during peak periods may be achieved through a combination of operational adjustments, scheduling decisions, and, in some cases, temporary reliance on backup power resources. While backup generation can support grid reliability, its use also raises air-quality and environmental considerations, highlighting the importance of aligning demand-flexibility strategies with California's broader climate and community-protection goals discussed elsewhere in this report.

In addition to shifting demand over time, some data-center operators may have opportunities to shift load geographically.[104] Companies with networks of facilities across multiple regions or time zones may be able to route certain computing tasks to locations with available capacity or lower contemporaneous demand. In theory, this type of geographic flexibility could help smooth demand and reduce strain on regions experiencing peak conditions. But in practice, it requires significant coordination, robust infrastructure, and operational redundancy, and may therefore be more feasible for large, multi-regional firms than for smaller or single-site operators.

Taken together, time-based and geographic demand flexibility offer practical ways to align data-center electricity use with real-world grid conditions. These approaches can improve utilization of existing infrastructure, support system reliability, and reduce the likelihood that the costs of serving new large loads are shifted onto residential and small-business customers.

RECOMMENDATION 10: California should maximize load shift opportunities to use existing load capacity and lower costs for residential customers.

Building a Cleaner, Smarter Data Center Future

California's clean-energy and climate goals may come under new pressure as data centers expand rapidly within the state. These facilities promise economic and technological benefits, but they also raise serious concerns about emissions, local environmental impacts, and staying true to the state's climate commitments. This section outlines environmental challenges and policy solutions to help ensure data-center growth is consistent with California's sustainability values. It concludes by highlighting ways the state can support research and innovation to help reduce the negative impacts of data centers on society.

DATA CENTERS AND THE ENVIRONMENT

California has committed to achieving 100 percent zero-carbon electricity generation (with offsets) by 2045.[105] These nation-leading goals are laudable, and the Commission believes that the Legislature should formally recommit to them even as it works to integrate a growing number of data centers into the grid. Doing so will help ensure these facilities rely on cleaner energy rather than the more carbon-intensive mixes common elsewhere, while showing other states that it is possible to expand data-center capacity and still stay firmly on track toward climate goals.

Looking more broadly, the chart that follows summarizes some of the environmental challenges associated with data centers, how they intersect with electricity policy, and the questions policymakers should consider as they craft legislation. Addressing these issues proactively can help California harness the benefits of data centers while reducing harms, strengthening community protections, and advancing the state's broader sustainability goals.

California should integrate data centers into the grid in ways that accelerate, rather than dilute, the state's climate ambitions. For example, policymakers could pair stronger environmental standards with streamlined processes for projects that meaningfully support decarbonization; ensure that communities hosting new infrastructure are protected and share the benefits; and encourage innovation in flexible load, cleaner backup systems, and more accurate carbon accounting. By doing so, the state can support economic growth while reinforcing its leadership on climate, environmental justice, and clean-energy transition.

RECOMMENDATION 11: California should meet rising data-center electricity demand without weakening its clean-energy or climate goals, including behind-the-meter generation.

CLEAN AND RELIABLE BACKUP GENERATION

The question of backup power for data centers presents a challenge both for grid stability and public health. Many top-tier facilities are designed to meet extremely high availability standards—often referred to as "five nines" reliability (99.999 percent), meaning only minutes of allowable downtime per year.[106] To achieve this, operators rely on highly stable and redundant power systems.

At present, diesel is the most common form of backup generation for data centers.[107] These systems are typically designed to operate during power outages or to provide relief for the grid during times of high demand. The Environmental Protection Agency recently "clarified" that such generators could be run for a certain number of hours under non-emergency situations and for an unlimited amount of time during emergencies.[108] Backup generators must also be run periodically for testing and maintenance.

And yet, as Masheika Allgood and Linda Taub Gordon testified before the Commission, even limited operation of diesel generators can contribute to

Data Centers and Environmental Challenges

Environmental Challenge	Connection to Electricity Policy	Potential Policy Questions
Impacts on Nearby Communities	Energy infrastructure for data centers may affect communities through air or water pollution, land use, or noise.	▪ How should regulators assess environmental-justice impacts when approving new infrastructure for data centers? ▪ Could affected communities receive shared benefits such as local hiring or energy-efficiency investments?
Where Data Centers Get Their Power	Siting data centers in areas with more clean energy capacity reduces total emissions and is more sustainable.	▪ Should siting policies or interconnection approvals favor utilities with cleaner electricity supplies, or else require developers to offset emissions in dirtier areas?
Transparency About Energy Use and Emissions	Public reporting of data-center energy use and emissions remains limited.	▪ Should large data centers be required to disclose annual energy use, energy source mix, and emissions figures? ▪ Could this information feed into the state's broader climate-tracking and planning tools?
Building and Connecting to the Grid	New substations or transmission lines for data centers can trigger environmental review and permitting delays.	▪ How can environmental review processes be streamlined for projects powered by clean energy while still protecting communities?
Backup Power and Air Pollution	Generators used for reliability emit carbon dioxide and local pollutants.	▪ Should the state require cleaner backup systems? ▪ Could utilities offer programs that incentivize data centers to provide clean backup power to the grid?
Making Energy Use More Flexible	Data centers can help integrate renewable power by adjusting when they draw electricity or by participating in demand-response programs.	▪ Could the state create time-flexible rate options for data centers that shift operations to periods of abundant renewable energy? ▪ Should this flexibility count toward grid-decarbonization goals?
Cleaner Construction and Materials	Building materials and equipment manufacturing for data centers have large carbon footprints.	▪ Should low-carbon construction standards or procurement rules apply to data center projects that benefit from state-supported infrastructure?

local air pollution and raise public health concerns, particularly in communities already burdened by poor air quality.[109]

This leads to a complex challenge: backup generation is necessary for data centers to function and to support demand flexibility. But diesel-powered backup generators have harmful effects on health, and data center operators are skeptical about the use and cost of variable generation resources (e.g., solar, wind) and less proven technologies such as battery storage, geothermal, or small modular reactors.[110]

The Little Hoover Commission believes that data center developers and the tech industry should drive solutions to dirty backup generation, and that California can and should compel them to do so. Specifically, the state should require data centers to transition their backup power systems toward progressively cleaner, lower-emission technologies, aligned with the state's broader 2045 zero-carbon goals.

Performance would be assessed at regular intervals using measurable emissions or technology benchmarks. Facilities that fail to meet interim standards would face consequences—such as higher electricity rates, restricted eligibility for new interconnections, or, in persistent cases, suspension of operations—while those that meet or exceed targets could receive incentives such as rate discounts, expedited permitting, or eligibility for state grants and pilot programs. California has an opportunity to set national standards by ensuring that backup systems support grid reliability without sacrificing the health of nearby communities. By pushing the data center sector towards progressively cleaner technologies, the state can protect public health, maintain dependable electricity service, and demonstrate that grid reliability, public health, and environmental stewardship can advance together.

RECOMMENDATION 12: California should require data centers to maintain a minimum level of clean backup power to support load shifting, enable curtailment during grid stress (including extreme heat events), and reduce reliance on fossil-fuel backup generation.

RECOMMENDATION 13: The state should limit pollution impacts on nearby communities from data-center backup generators.

INCENTIVIZING DATA CENTER ENERGY INNOVATION

California's data center sector is rapidly evolving, and a range of emerging technologies can help reduce its impacts while supporting reliability, environmental goals, and community well-being. These innovations can make data centers more efficient, less resource-intensive, better aligned with clean energy, and more responsive to grid needs. Together, they offer a practical pathway for enabling economic growth while limiting stress on the electric system and reducing local environmental concerns.

- **More efficient computing equipment:** Using servers, chips, and power systems that deliver more computing for each unit of electricity used.
- **Smarter energy management:** Using software to shift non-urgent computing to times when the grid is less stressed or when cleaner energy is available.
- **Advanced cooling systems:** Moving beyond traditional cooling to methods that use less energy and/or water while keeping equipment safe and stable.
- **Better heat reuse:** Capturing waste heat from data centers and using it for nearby buildings or other community needs where feasible.
- **On-site clean power generation:** Generating cleaner electricity at the facility, such as solar or other low-emission power sources, to reduce pressure on the broader grid.
- **Cleaner and quieter backup power:** Replacing or supplementing diesel generators with cleaner backup options that reduce air pollution and community impact.
- **On-site energy storage:** Installing large batteries so facilities can rely less on the grid during peak periods and help smooth demand.

- **Flexible demand:** Designing operations so data centers can voluntarily reduce or adjust power use during emergencies or when the grid is strained.
- **Better siting and design choices:** Locating and designing facilities in ways that align with cleaner energy resources, minimize local impacts, and reduce the risk of stranded infrastructure.

California can encourage adoption of advanced efficiency, clean energy, and grid-support technologies by aligning incentives with measurable public benefits. Incentives could be adopted to reward facilities that are more efficient, flexible, and cleaner—for example, lower rates or credits for centers that reduce their contribution to peak demand, participate in demand-response programs, or demonstrate strong energy and environmental performance.

Streamlined interconnection and permitting pathways could also be offered for facilities that meet high standards, reducing uncertainty and timelines while protecting communities and environmental safeguards. A voluntary state designation for "high-performance" data centers, paired with meaningful performance reporting, would create recognition, transparency, and potential market value for operators.

Financial tools and partnerships can complement these steps. Incentives or grants could help offset early costs for emerging technologies such as advanced cooling, large-scale storage, or cleaner backup power. A state-led research partnership with national laboratories, universities, utilities, and industry—as recommended below—could test new technologies, validate results, and share best practices. And state procurement and leasing preferences for high-performance facilities can use public purchasing power to encourage environmental leadership.

RECOMMENDATION 14: California should encourage data centers to adopt technologies—such as advanced cooling, energy storage, or on-site clean power—that reduce their impact on the grid.

PROMOTING DATA CENTER ENERGY RESEARCH

The State of California should establish a formal private/public partnership with industry, the federal national laboratory system (including Lawrence Livermore National Laboratory and Lawrence Berkeley National Laboratory), and relevant academic institutions including the University of California system, Stanford University, and other private universities. This initiative—provisionally named the California Partnership for Advanced Research (CALPAR)—would research, test, validate, and promote cutting-edge technologies to improve AI data center efficiency and environmental performance. Potential priority working group areas could include:

- **Grid Reliability, Flexibility, and Infrastructure Readiness:** Focuses on ensuring rapid growth in AI data centers strengthens rather than destabilizes California's electric system. Examines demand flexibility, interconnection planning, transmission expansion, and backup solutions that enhance overall grid resilience.
- **Efficiency, Technology Innovation, and Environmental Performance:** Aims to accelerate research, validation, and real-world deployment of technologies that reduce energy and water use. Works with national laboratories, universities, and industry to test cooling innovations, improve server efficiency, and explore waste-heat reuse and related advancements.
- **Transparency, Accountability, and Public Trust:** Develops privacy frameworks for sharing meaningful operational and environmental data so policymakers and the public can understand true system impacts. Emphasizes standardized metrics, responsible disclosure, and clear public communication to reduce confusion and build confidence.
- **Community Impacts, Siting, and Environmental Justice:** Seeks to ensure that data center siting and infrastructure decisions avoid disproportionate harms while providing tangible local benefits. Addresses pollution, land use, water concerns, clean backup power expectations, and community

engagement practices that foster predictability and trust.

- **Workforce, Economic Development, and Competitiveness:** Aligns California's technological leadership with strong workforce pipelines and long-term economic strategy. Evaluates training pathways, partnerships with higher education, job quality standards, and policies that help retain investment while advancing efficiency and environmental performance goals.

There is precedent for such an organization. Governor Newsom recently formed a California Innovation Council, consisting of a blend of tech industry leaders, academics, and state workers. This Council is divided into working groups focused on modernizing government service delivery, economic and workforce development, promoting online safety, and countering fraud.

RECOMMENDATION 15: California should establish a formal public-private partnership—provisionally the California Partnership for Advanced Research (CALPAR)—to collaborate with industry, national laboratories, and universities to research, test, and promote technologies that improve the efficiency and environmental performance of AI-driven data centers.

Data Centers and the Water Supply

Data centers are resource-intensive, placing significant demands not only on electricity systems but also on the local water supply. This water is used to manage the significant heat generated by high-performance computing equipment and, in the case of thermal generation, to produce the electricity that powers the facility.[111] It should also be noted that chip fabrication itself is a water intensive process.[112] A "typical" data center can consume as much water as approximately 1,000 homes, while a large hyperscale data center may use water comparable to that of a town with roughly 50,000 residents.[113] While the Commission recognizes the importance of data centers' water use as an issue for policymakers, this report focuses on the centers' impacts on electricity systems and does not offer recommendations regarding water policy.

In California, the primary concern is often not the water demand of any single facility, but the cumulative impact of multiple data centers clustered within the same area. Such concentrated development can strain local water supplies.[114] In addition, increased water withdrawals may stress aging water infrastructure and wastewater treatment facilities—which are themselves energy-intensive and may therefore contribute to higher overall electricity demand due to added load from data centers.[115]

Tangible impacts are already emerging. Data center water demand appears to be accelerating the depletion of the Potomac Aquifer, forcing some communities to shift from groundwater to more constrained surface-water supplies.[116]

There is also an important tradeoff between water use and electricity use in data center cooling. Cooling systems that rely more heavily on water—such as evaporative cooling—can reduce electricity consumption, while air- or closed-loop liquid-based cooling systems typically use more electricity but less water.[117] As a result, regions with ample clean or renewable electricity can potentially minimize water use by data centers without compromising their environmental standards.[118]

Emerging technologies may also help mitigate data center water consumption. These include advanced high-efficiency cooling systems, direct-to-chip liquid cooling and immersion cooling technologies, lower-heat-emitting chips, the substitution of recycled or graywater for potable water, and smarter cooling and water management.[119]

Lack of data has hindered meaningful analysis of this issue, however. Data center operators have been reluctant to disclose information about their water use, including how efficiently water is used relative to electricity consumption, and have opposed legislation that would require such reporting.[120]

As a result, mandating water-use reporting by data centers has emerged as a prominent policy theme. In California, for example, AB 93 (Papan—2025) would have required data center developers to provide estimates of water use prior to receiving a business license or permit and to report their annual water consumption thereafter.[121] The bill was ultimately vetoed by Governor Newsom, who cited concerns about its potential impact on the growth of the artificial intelligence sector.[122]

Other states, however, have moved forward with similar—and in some cases more stringent—requirements. Minnesota's HF 16 requires detailed water-use reporting as part of a project application, while Kansas' SB 98 links data center tax exemptions to, among other things, the development of a sustainable water-use plan.[123]

Data Centers and Air Pollution

THE DANGERS OF DIESEL

To meet stringent reliability standards—often allowing only a few minutes of downtime per year—data centers typically rely on on-site diesel backup generators. These generators provide emergency power during grid outages and ensure service continuity during peak demand.[124]

Emissions from diesel backup generators raise significant public health and environmental justice concerns, particularly for communities near data center clusters. Diesel generators emit diesel particulate matter, a toxic air contaminant associated with increased cancer risk as well as adverse cardiovascular and respiratory health outcomes, including asthma.[125] Researchers from the University of California, Riverside estimate that pollutant emissions from data centers nationwide could contribute to more than one-third of asthma-related deaths by 2028.[126]

CALIFORNIA IMPACTS

These risks are especially concerning in California, which is home to the third-highest concentration of data centers in the United States.[127] Public health impacts are worsened by the geographic concentration of data centers around San Jose and Los Angeles. In 2025, the American Lung Association ranked Los Angeles and San Jose the fifth and sixth most polluted cities in the nation, respectively, for year-round particle pollution.[128]

One proposed hyperscale data center in San Jose plans to operate 39 diesel backup generators.[129] According to the California Energy Commission, these generators—operating for only 20 hours per year—would emit approximately the same amount of air pollutants as 428 gasoline-powered cars driven for an entire year.[130] These regions already experience significant air quality challenges, and additional emissions may further amplify public health burdens.

POTENTIAL SOLUTIONS

Technological and policy options exist to mitigate these impacts. Alternative backup power solutions—including hydrogen fuel cells, lithium-ion battery storage systems, renewable generation paired with energy storage, and other lower-emission technologies—can substantially reduce or eliminate local air pollutant emissions.[131] Despite these alternatives, data centers continue to rely heavily on diesel generators because they are generally perceived as more cost-effective and reliable.[132]

Other states have begun to consider legislation addressing air quality and climate impacts from data centers. In Oregon, House Bill 2816 (Marsh, 2023) proposed requiring high-energy-use facilities, including data centers, to keep associated greenhouse gas emissions below specific target levels.[133] In New York, Senate Bill S9144 (Krueger, 2026) would impose a temporary moratorium on permits for new data centers while the Department of Environmental Conservation conducts a statewide environmental impact review and develops updated regulations to mitigate environmental and air quality impacts.[134] Although the Oregon bill died in committee and the New York bill remains under review, these proposals reflect growing concern among policymakers about the air pollution impacts of data center development.

CALIFORNIA POLICY

California legislators have similarly introduced measures aimed at reducing air pollution from data centers by encouraging cleaner energy and backup power systems. Proposed legislation such as AB 1577 (Bauer-Kahan, 2026) would require data centers to disclose the share of renewable or carbon-free energy generated onsite or procured to power their operations.[135] And SB 887 (Padilla, 2026) would require data centers to meet peak demand using on-site zero-carbon energy storage and rely as much as feasible on behind-the-meter zero-carbon generation.[136] As data center development accelerates, California has an opportunity to shape national policy by linking data center reliability requirements with stronger protections for local air quality and public health.

Notes

1. U.S. Energy Information Administration, "Retail Electricity Prices Closely Tracked Inflation Over the Last 10 Years" (Sep. 11, 2024). M. Baldassare, D. Bonner, L. Mora, and D. Thomas, *PPIC Statewide Survey: Californians and Their Economic Well-Being, Public Policy Institute of California* (Dec. 2024).

2. S. Susnjara and I. Smalley, "What Is Cloud Computing?", *IBM Think* (n.d.); Geeks for Geeks "History of Cloud Computing" website.

3. For a good review article on predictions of AI's potential impact on the economy, see D. Björkegren, "AI is Transforming the Economy—Understanding its Impact Requires Both Data and Imagination", *Nature* (Dec. 15, 2025). For a searchable collection of AI business use cases, see Deloitte's "AI Dossier" website.

4. R. Li, "The Bay Area is Losing the Data Center Race. Economists Say That Could be a Good Thing", *San Francisco Chronicle* (Dec. 18, 2025).

5. PG&E, "PG&E Data Center Demand Pipeline Swells to 10 Gigawatts with Potential to Unlock Billions in Benefits for California" [press release] (Jul. 31, 2025). According to the California Energy Commission's most recent reporting, the state's generation capacity is 89,283 megawatts or ~8.9 gigawatts.

6. For example, PG&E has claimed that for every gigawatt of new data center load, customers can expect to save between one and two percent on their electricity bills (see "PG&E Data Center Demand Pipeline", op. cit.). For a balanced analysis, see S. Borenstein, "What Will Data Centers Do to Your Electric Bill?", *Energy Institute Blog* [UC Berkeley] (Sep. 29, 2025).

7. The impact of data centers on electricity prices is hotly debated. For the view that data centers lower customer bills, see above. For the opposite, see J. Saul et al., "AI Data Centers Are Sending Power Bills Soaring", *Bloomberg* (Sep. 29, 2025). For an in-depth argument that data centers will raise rates in California, see The Utility Reform Network, Testimony of the Utility Reform Network Addressing Proposed Rule 30 Ratepayer Impacts and Related Issues [CPUC Proceeding A.24-11-007] (Jun. 30, 2025).

8. See, for example, n.d., "How Big Tech Won Big Against Regulation in California This Year", *GovTech Industry Insider* [Tribune News Service] (Oct. 22, 2204). K. Johnson, "California Lawmakers Wanted to Get Tough on Data Centers. Here's What Survived", *CalMatters* (Sep. 25, 2025).

9. B. Smith, "Building Community-First AI Infrastructure" [blog post], Microsoft (Jan. 13, 2025).

10. A. Vahdat and J. Dean, "How Much Energy Does Google's AI Use? We Did the Math" [blog post], Google Cloud (Aug. 21, 2025).

11. Data Center Coalition "Frequently Asked Questions about the Data Center Industry" website.

12. I. Klaus et al., *Carnegie California AI Survey*, Carnegie Endowment for International Peace (Oct. 9, 2025). Cf. E. Gorelick, "Why Do Americans Hate A.I.?", *New York Times* (Jan. 2, 2026).

13. N. Portuondo, "Data center boom creates political conundrum for the GOP", *Politico* (Dec. 14, 2025).

14. In a recent PPIC survey, 37 percent of California residents named "cost of living, economy, [and] inflation" as the most pressing issue in the state. See M. Baldassare, D. Bonner, L. Mora, and D. Thomas, *PPIC Statewide Survey: Californians and Their Government, Public Policy Institute of California* (Jun. 2025).

15. Cf. A. Barth et al., "The Data Center Balance: How US States Can Navigate the Opportunities and Challenges", McKinsey & Co. (Aug. 8, 2025).

16. A. Schaap, "How Data Centers are Transforming Local Communities", *Fast Company* (Mar. 20, 2025).

17. B. Violino, "Computing and Storage are Moving to the Edge, and IT Needs to be Ready", *CNBC* (Jun. 11, 2024). A. Steele, "The Rise of Vertical Data Centers", *Colliers Knowledge Leader* (Jun. 24, 2024).

18. J. Roundy, "The Increasing Concern of Data Center Land Acquisition", *TechTarget* (Aug. 21, 2025).

19. For an overview with links to scientific studies, see A. Kadasamy, "More Data Centers, More Environmental Problems?", National Wildlife Federation (Sep. 24, 2025).

20. A. Garcia and K. Norton, "Next Data Center Roadblock for Google, Microsoft, Oracle, Meta is On Main Street", *Investor's Business Daily* (Dec. 20, 2025).

21. S. Nadel, "Data Center Efficiency and Load Flexibility Can Reduce Power Grid Strain and Costs" [blog post], American Council for an Energy-Efficient Economy (ACEEE) (Oct. 7, 2025).

22. For example, T. Panadreou, "Who Wins the Energy Race: AI Data Centers, EVs, or Bitcoin Mining?", *Forbes* (Aug. 30, 2024).

23. The issue of stranded infrastructure is addressed in K. Hieta and E. Rodriguez, "How Will Data Center Growth Impact California Ratepayers?", Public Advocates Office (Oct. 28, 2025).

24. It is unclear how often data centers relocate or for what reasons. For an overview of the concept, see C. Tozzi, "Streamlining Data Center Relocation: A Checklist for Success", *Data Center Knowledge* (Sep. 24, 2023). A number of companies provide specialized data center relocation services.

25. S. Ren and A. Wierman, "Mitigating the Public Health Impacts of AI Data Centers", *Harvard Business Review* (Nov. 5, 2025).

26. A. Lazo, "Data Centers are Putting New Strain on California's Grid. A New Report Estimates the Impacts", *CalMatters* (Nov. 14, 2025).

27. For an interesting, in-depth discussion of "data center ethics", though not specifically related to energy, see the Climate Sustainability Directory "Data Center Ethics" website.

28. A. Satchwell, N. Mims Frick, P. Cappers, et al., *Electricity Rate Designs for Large Loads: Evolving Practices and Opportunities*, Lawrence Berkeley National Laboratory (Jan. 2025).

29. The aforementioned Lawrence Berkeley study contains numerous examples. See also M. Fitzpatrick, *Data Center Energy Impacts and State Responses, Office of Legislative Research* [Connecticut] (Oct. 14, 2025).

30. J. Chambers, "Paraguay – the Silicon Valley of South America?", *British Broadcasting Corporation* (Oct. 12, 2025).

31. M. North, "Here's How Data Centre Heat Can Warm Your Home", *World Economic Forum* (Jun. 18, 2025). To this list, we could add how China is developing a wind-powered data center off the coast of Shanghai that is submerged underwater, taking advantage of natural cooling and fewer land constraints. Companies are also exploring orbital data centers, which could be powered by constant solar energy and face no land limitations. See Y. Xiaoxing, "China Is Putting Data Centers in the Ocean to Keep Them Cool", *Scientific American* (Jul. 16, 2025) and D. Rao, "Data Centers Could Soon Be Orbiting in Space", *The Week* (Dec. 23, 2025).

32. A. Peters, "These Silicon Valley Apartments Will Get Their Heat from Nearby Data Centers", *Fast Company* (Nov. 15, 2024).

33. See the Data Center Map "California Data Centers" website for an overview. Cf. M. Petersen, "Power-hungry AI Data Centers are Raising Electric Bills and Blackout Risk", *Los Angeles Times* (Aug. 12, 2024).

34. Links to PG&E's increasing 2025 predictions of new data center demand can be found here (Feb.), here (May), and here (July).

35. The two "Data Center Project Pipeline" charts are from PG&E, *2024 Investor Update* (Jun. 12, 2024), pp. 20-21.

36. D. Bergstrom, "Could Fresno Soon be Home to Thirsty AI Data Centers?", *Fresnoland* (Aug. 26, 2025). Cf. F. Barringer, *Thirsty for Water and Power, AI-Crunching Data Centers Sprout Across the West*, Bill Lane Center for the American West [Stanford University] (Apr. 8, 2025).

37.See the Data Center Map "California Data Centers" website as well as T. Tsoneva and J. Affleck, *Data Center Investment: Decoding Opportunities*, CBRE (Jul. 17, 2024). The pie chart indicating data center ownership is from this second report.

38. A. Shehabi et al., *2024 United States Data Center Energy Usage Report*, Lawrence Berkeley National Laboratory (Dec. 2024), p. 12.

39. Links to PG&E's increasing 2025 predictions of new data center demand can be found here (Feb.), here (May), and here (July). At the same time, PG&E has emphasized that substantial portions of its existing grid are underutilized and that additional large-load customers could distribute fixed costs more broadly, potentially lowering residential bills by one to two percent per gigawatt. See PG&E, "Surging Data Center Growth to Help Lower Energy Costs for PG&E Customers" [press release] (May 27, 2025).

40. California Energy Commission, *2025 IEPR Forecast—Updated Results* [DAWG Meeting] (Jan. 5, 2026). The mid case forecast appears on p. 8. The high case forecast appears on p. 9.

41. CAISO "Large Load" website. CAISO predicts that data centers will add 2.3 gigawatts of load by 20230 and 3.3 gigawatts of load by 2035.

42. H. Javanbakht, "California Energy Commission Data Center Demand Forecast" [presentation at the Little Hoover Commission hearing on Data Centers and California Electricity Policy, pt. II] (Dec. 11, 2025), Slides 4 and 6.

43. Efforts to require more detailed disclosure have so far stalled. For example, 2025 legislation (AB 222, Bauer-Kahan), which would have required electricity-use reporting from data centers and assessment of potential cost shifts, did not advance.

44. See T. Tsoneva and J. Affleck, Data Center Investment, op cit.; n.d., *North America Data Center Trends H1 2025*, CBRE (Sep. 8, 2025); and J. Noffsinger et al., *The Cost of Compute: A $7 trillion Race to Scale Data Centers*, McKinsey & Co. (Apr. 28, 2025).

45. N.d., *What the Real Estate Industry Needs to Know About Data Centers*, McKinsey & Co. (Oct. 15, 2024).

46. T. Hylton, E. Herald, and P. Werner, "Taxing the Digital Backbone: What You Need to Know About Data Centers" [blog post], Price Waterhouse Coopers (Jul. 25, 2025).

47. See, for example, S. Rwejuna and A. Bryant, "DOE Directs FERC to Accelerate Interconnection of Data Centers", White & Case (Oct. 27, 2025).

48. California has struggled to pass major data center legislation. For an analysis of the politics involved, see J. Wolman and L. Kashinsky, "Big Tech's Next Major Political Battle May Already be Brewing in Your Backyard", *Politico* (Aug. 10, 2025).

49. B. Giacobone, "Phantom Data Centers are Flooding the Load Queue", *Latitude Media* (Mar. 26, 2025). N.d., *Uncertainty and Upward Bias are Inherent in Data Center Electricity Demand Projections*, London Economics Institute (Jul. 7, 2025).

50. B. Martucci, "A Fraction of Proposed Data Centers Will Get Built. Utilities are Wising Up", *Utility Dive* (May

51. Overviews of the contents of SB 6 include: n.d., "Texas Senate Bill 6 Significantly Expands Regulatory Oversight Over Large Loads in ERCOT", McGuire Woods (Jul. 10, 2025). D. Chernicoff, "Texas Senate Bill 6: A Bellwether On How States May Approach Data Center Energy Use", *Data Center Frontier* (Jul. 2, 2025).

52. Z. Skidmore, "AEP Ohio Slashes Data Center Pipeline by More Than Half—Report", *Data Center Dynamics* (Oct. 1, 2025).

53. Identifying "policy levers" vis a vis state regulation of data centers was inspired by n.d., *Advancing Affordable Electricity in California: Policy Levers to Address Rising Rates*, Public Advocates Office (Dec. 13, 2024).

54. K. Hieta and E. Rodriguez, "How Will Data Center Growth Impact California Ratepayers?", Public Advocates Office (Oct. 28, 2025). See also the California ISO "Interconnection Request and Study" website.

55. See the California Air Resources Board "Air Quality and Emissions Resources" website.

56. See City of Santa Clara "Environmental Review/CEQA" website, as well as Office of Senator Steve Padilla, "Senator Padilla Calls for Public Input and Review After Data Center Project Granted Exemption from Environmental Review and Previews Plans to Introduce Legislation Setting Data Center Energy Standards" [press release] (Dec. 3, 2025).

57. See, for example, J. Remington and R. Carter, "An Overview of State Data Center-Related Tax Incentives", NAIOP Commercial Real Estate Development Association (Winter 2024/2025).

58. Cf. the background document to the July 24, 2025 Little Hoover Commission hearing on California Electricity Costs, pp. 1-2 and n. 2, which includes relevant bibliography.

59. Quoted text is from the "What are Customer Classes?" tab of the CPUC "Electric Rates" website.

60. See PG&E's written testimony, submitted in advance of the Little Hoover Commission's Dec. 11, 2025 hearing, which provides specific information regarding costs paid through transmission rates (pp. 5-6 and n. 10).

61. Links to PG&E's current and historical rates, including B-20, can be found at its "Electric Rates" website.

62. This was the goal of SB 57 (Padilla, 2025) in its original form.

63. The following details are from California Public Utilities Commission, Decision (D.) 25-07-039, *Decision Partly Granting and Partly Denying Pacific Gas and Electric Company's Motion for Interim Implementation of Electric Rule Number 30* [A. 24-11-007] (Jul. 24, 2025).

64. PG&E, *Electric Rule No. 30 – Retail Service Transmission Facilities (Interim Implementation)* [Advice Letter 7772-E; D.25-07-039] (Dec. 4, 2025).

65. Technical details about the load interconnection process can be found on PG&E's "Transmission and Distribution Interconnection Handbooks" website.

66. Quoted text is from Tex. SB 6 § 2(b) (2025).

67. Overviews of the contents of SB 6 include: n.d., "Texas Senate Bill 6 Significantly Expands Regulatory Oversight Over Large Loads in ERCOT", McGuire Woods (Jul. 10, 2025). D. Chernicoff, "Texas Senate Bill 6: A Bellwether On How States May Approach Data Center Energy Use", *Data Center Frontier* (Jul. 2, 2025).

68. N.d., "ERCOT 'Kill Switch' for Data Centers Passes in Texas House", *GovTech Industry Insider* [Tribune News Service] (May 27, 2025).

69. For a succinct overview, see the Governor's "GenAI for California" website.

70. B. Casselman and S. Ember, "The A.I. Boom Is Driving the Economy. What Happens if It Falters?", *New York Times* (Nov. 22, 2025).

71. R. Singh and J. Jose, "Citigroup forecasts Big Tech's AI Spending to Cross $2.8 Trillion by 2029", *Reuters* (Sep. 30, 2025). C. Metz, "Why Debt Funding Is Ratcheting Up the Risks of the A.I. Boom", *New York Times* (Nov. 10, 2025).

72. As cited in M. MacCarthy, "Policymakers Have to Prepare Now for When the AI Bubble Bursts", *Tech Policy Press* (Nov. 24, 2025).

73. It is unclear how often data centers relocate or for what reasons. For an overview of the concept, see C. Tozzi, "Streamlining Data Center Relocation: A Checklist for Success", *Data Center Knowledge* (Sep. 24, 2023). A number of companies provide specialized data center relocation services.

74. See, for example, G. Radauskas, "California Dreamin' No More: State Facing Tech Jobs Exodus", *CyberNews* (Apr. 12, 2024).

75. The International Energy Agency has documented continuing, if more modest, efficiency gains. See M. Offutt and L. Zhu, "Data Centers and Their Energy Consumption: Frequently Asked Questions", Congressional Ressearch Service (Aug. 26, 2025).

76. J. Quigley, "Load Growth Irrational Exuberance Crashes into DeepSeek", Kleinman Center for Energy Policy Blog [University of Pennsylvania] (Jan.31, 2025). For an overview of the benefits of Google's new AI chips, see D. Patel et al., "TPUv7: Google Takes a Swing at the King", *SemiAnalysis* (Nov. 28, 2025).

77. J. Quigley, "Load Growth Irrational Exuberance Crashes into DeepSeek", Kleinman Center for Energy Policy Blog [University of Pennsylvania] (Jan.31, 2025). The in-quote hyperlink is original to this source.

78. NovaSky Team, "Sky-T1: Train Your Own o1 Preview Model Within $450", *Github* (Jan. 9, 2025).

79. For a general introduction, see n.d., "The Rise of Mini Models in Artificial Intelligence: A New Era of Efficiency", VisioTech (Sep. 24, 2024). For a more in-depth discussion, see R.D. Caballar, "What are Small Language Models?", *IBM Think* (n.d.).

80. J. Bushnell, "Is There Really an Energy Silver Lining to a Datacenter Bubble?", *Energy Institute Blog* [UC Berkeley] (Oct. 13, 2025).

81. See, in particular, the testimony of Dr. Liang Min, Managing Director of Stanford's Bits & Watts Initiative. Slides from his presentation to the Commission can be found here.

82. C. Chen, "China Built Hundreds of AI Data Centers to Catch the AI Boom. Now Many Stand Unused", *MIT Technology Review* (Mar. 26, 2025).

83. Georgia Public Service Commission, "2025 Data Center Fact Sheet" (Dec. 1, 2025); J. Holdman, "SC's State-Owned Utility Enacts Higher Rates for Data Centers, Large Users", *South Carolina Daily Gazette* (Apr. 25, 2025); Z. Skidmore, "Oregon House Passes Bill Shifting Power Infrastructure Costs to Data Centers", *Data Center Dynamics* (Jun. 9, 2025); Z. Skidmore, "Virginia Regulators Approve New Rate Class for Data Centers and Other Large Loads", *Data Center Dynamics* (Nov. 27, 2025).

84. A. Levenson-Falk, "CUB Spearheads Nation-Leading Legislation to Protect Ratepayers from the Costs to Serve Data Centers", Citizens Utility Board (Jul. 9, 2025).

85. K. Suzuki, "Imperial Valley City Sues to Force Environmental Review of Massive Data Center Project", *KPBS* (Dec. 12, 2025).

86. For a clear discussion, see S. Borenstein, "Can Data Centers Flex Their Power Demand?", *Energy Institute Blog* [UC Berkeley] (Apr. 14, 2025).

87. S. Fleming, "The Grid is Built for the Peak" [blog post], *Clean Energy Review* (Oct. 3, 2024).

88. D. Ai, C.L. Crago, J.T. Mullins, "Heat, Power Outages and Mortality in the United States", *Journal of Environmental Economics and Management* (Dec. 20, 2025).

89. California Energy Commission, "California Energy Leaders Report Progress on Grid Reliability Ahead of Summer 2025" [press release] (May 2, 2025). California's highest ever peak demand was 52,061 MW on Sep. 6, 2022.

90. For analysis of grid performance, see Public Advocates Office, *Preliminary Analysis of California's Resiliency During the September 2022 Heat Wave* (n.d.).

91. B. Martucci, "Texas Law Gives Grid Operator Power to Disconnect Data Centers During Crisis", *Utility Dive* (Jun. 25, 2025).

92. For example, in written testimony provided in advance of the Little Hoover Commission's Dec. 11, 2025 hearing (p. 6, n. 11), PG&E pointed to its Flex Connect program, which can rapidly interconnect new EV charging ports if they commit to lowering load during times of peak use. Cf. PG&E, "New PG&E Service Offering Makes It Easier and Faster to Connect EV Chargers, EV Fleets and Big Batteries to the Grid" [press release] (Apr. 28, 2025).

93. R. Prabha and L. Min, *U.S. Transmission System Utilization Study Phase 1: WECC*, Bits & Watts Initiative [Stanford University] (Dec. 12, 2025).

94. PG&E, "PG&E and Smart Wires Enhance Grid Reliability, Capacity for Data Centers in San Jose" [press release] (May 28, 2025).

95. H. Parmar et al., *How to Find the 'Sweet Spots' to Build All Those Data Centers*, ICF (n.d.).

96. K. Hieta and E. Rodriguez, "How Will Data Center Growth Impact California Ratepayers?", Public Advocates Office (Oct. 28, 2025).

97. B. Roberts, *Data Center Infrastructure in the United States, 2025*, NREL (2025). K. Cooke, "The Hidden Complexity of Data Center Siting" [blog post], Esri (Sep. 8, 2025).

98. PG&E, "PG&E Accelerating Connection of New Data Centers throughout Northern and Central California" [press release] (Feb. 13, 2025). Cf. PG&E's written testimony to the Little Hoover Commission (p. 6) for details.

99. For an overview, see W. Kirchstetter, *Data Center Load Flexibility Workshop Summary*, Lawrence Berkeley National Laboratory (Jan. 2025).

100. T. Norris et al., *Rethinking Load Growth Assessing the Potential for Integration of Large Flexible Loads in US Power Systems*, Nicholas Institute for Energy, Environment & Sustainability [Duke University] (2025), p. 2.

101. C. Knittel, J.R. Senga, and S. Wang, "Flexible Data Centers and the Grid: Lower Costs, Higher Emissions?", MIT CEEPR [WP 2025-14] (Jul. 2025).

102. See U.S. Government Accountability Office, "Electricity: Information on Peak Demand Power Plants" [GAO-24-106145] (May 21, 2024). L. Kearney, "Insight: AI Data Centers are Forcing Dirty 'Peaker' Power Plants Back into Service", *Reuters* (Dec. 23, 2025).

103. PG&E, "New PG&E Service Offering Makes It Easier and Faster to Connect EV Chargers, EV Fleets and Big Batteries to the Grid" [press release] (Apr. 28, 2025).

104. For an overview, see n.d., "How 'Load Shifting' May Help Improve Data Center Sustainability", *Data Center Knowledge* [Bloomberg News] (Feb. 26, 2024). For a scientific analysis, see I. Riepin, T. Brown, and V.M. Zavala, "Spatio-Temporal Load Shifting for Truly Clean Computing", *Advances in Applied Energy* 17 (Mar. 2025).

105. SB 100 (De León, 2018) requires that 60 percent of California's retail electricity sales be supplied from renewable sources by 2030 and 100 percent from renewable or other zero-carbon resources by 2045. AB 1279 (Muratsuchi, 2022) requires that statewide greenhouse gas emissions to be reduced at least 85 percent below 1990 levels by 2045 and that remaining emissions be offset to achieve overall carbon neutrality.

106. For an accessible technical overview of the "five nines" concept, see R. Reynolds, "Achieving "Five Nines" in the Cloud for Justice and Public Safety" [blog post], *Amazon Web Services Public Sector Blog* (Mar. 17, 2020).

107. For an overview, see C. Tozzi, "Comparing Data Center Backup Power Systems", *Data Center Knowledge* (Nov. 25, 2205).

108. United States Environmental Protection Agency, "EPA Issues Clarification to Help Power Data Centers, Ensure U.S. Is the AI Capital of the World" (May 1, 2025).

109. Gordon and Allgood's written testimony to the Commission in advance of its November 20 hearing can be found here and here, respectively. Cf. S. Ren and A. Wierman, "Mitigating the Public Health Impacts of AI Data Centers", *Harvard Business Review* (Nov. 5, 2025). For the politics behind these generators, see C. Paullin, "State Regulators Weigh Expanded Use of Data Centers' Diesel Backup Generators", *Inside Climate News* [National Public Radio] (Dec. 17, 2025).

110. R. Quint, J. Zhao, and K. Thomas, *An Assessment of Large Load Interconnection Risks in the Western Interconnection*, Western Electricity Coordinating Council (Feb. 2025), p. 18. The U.S. Department of Energy has directed its staff and the national labs to engage in research and data gathering around this issue, and we believe California could assist in this process, for example, by providing grants targeted towards the study and development of backup power sources that are clean and capable of supporting demand flexibility. See Secretary of Energy Advisory Board, *Recommendations on Powering Artificial Intelligence and Data Center Infrastructure*, U.S. Department of Energy (Jul. 30, 2024), pp. 8-10.

111. M. Ashtine and D. Mytton, "We are Ignoring the True Cost of Water-Guzzling Data Centres", *The Conversation* (Oct. 19, 2021). I. James, "Amid the AI boom, California Looks to Address Water Use by Massive Data Centers", *Los Angeles Times* (Sep. 23, 2025).

112. M. Yañez-Barnuevo, "Data Centers and Water Consumption", Environmental and Energy Study Institute (June 25, 2025).

113. J.W. Kane, "AI, Data Centers, and Water", Brookings (Nov. 20, 2025).

114. K. Silverstein, "America's AI Boom Is Running Into An Unplanned Water Problem", *Forbes* (Jan. 11, 2026).

115. R. Ahmad, "Engineers Often Need a Lot of Water to Keep Data Centers Cool", American Society of Civil Engineers (Mar. 4, 2024).

116. M. Willson, *op. cit.* For a discussion of drought- and wastewater-related concerns, see County of Fairfax (VA), *2024 Annual Report on the Environment* [Prepared by the Environmental Quality Advisory Council] (Dec. 2024).

117. A. Setmajer, "How Data Centers Use Water, and How We're Working to Use Water Responsibly" [blog post], Equinix (Sep. 19, 2024). Cf. Lawrence Berkeley National Laboratory, Center of Expertise for Data Center Energy "Water Efficiency" website.

118. See M. Ashtine and D. Mytton, *op. cit.*

119. See, for example, P.J. Campbell and R. Schwaab, "How to Use Water Wisely in Data Centers", Ramboll (Dec. 2, 2025).

120. M. Willson, "States Push to End Secrecy Over Data Center Water Use", *E&E News by Politico* (Dec. 8, 2025).

121. The text of AB 93 (Papan—2025) can be found here.

122. Gov. Gavin Newsom, AB 93 Veto Message (Oct. 11, 2025).

123. The text of Minnesota's HF 16 (Davids; Huot--2025) can be found here. The text of Kansas' SB 98 (Sen. Comm. on Commerce—2025) can be found here.

124. Joint Legislative Audit and Review Commission, *Data Centers in Virginia* [RD206] (Dec. 9, 2024).

125. UC Newsroom, "California Data Center Health Impacts Tripled in 4 Years" [Press Release] (Nov. 21, 2025).

126. Y. Han, Z. Wu, P. Li, A. Wierman & S. Ren, "The Unpaid Toll: Quantifying and Addressing the Public Health Impact of Data Centers" [Working Paper, arXiv:2412.06288] (Dec. 9, 2024).

127. A. Foote & C. Wilkie-Rogers, "Employment in Data Centers Increased by More Than 60% from 2016 to 2023 but Growth Was Uneven Across the United States", U.S. Census Bureau (Jan. 6, 2025).

128. American Lung Association, *State of the Air 2025* (Apr. 23, 2025).

129. For details, see the CEC "Great Oaks South Generating Facility" website.

130. A. Keating, "California Needs More Data Centers, but Builders Are Using a Legal Loophole to Power Them with Diesel", *Fast Company* (May 2, 2023). According to CARB's *Airborne Toxic Control Measure for Stationary Compression Ignition Engines* [Sec. 93115.6] (Dec. 11, 2007), diesel backup generators are required to run a minimum number of hours per year for maintenance and testing purposes.

131. CARB "Emergency Backup Power Options—Commercial" website.

132. U.S. Environmental Protection Agency, *The Role of Distributed Generation and Combined Heat and Power (CHP) Systems in Data Centers* (2015).

133. Click here for a link to the text of HB 2816.

134. Click here for a link to the text of S 9144

135. Click here for a link to the text of AB 1577

136. Click here for a link to the text of SB 887.

Little Hoover Commission Members

PEDRO NAVA | Santa Barbara
Appointed to the Commission by Speaker of the Assembly John Pérez in April 2013 and reappointed by Speakers Rendon in 2017 and 2021 and Rivas in 2024. Government relations advisor. Former State Assemblymember from 2004 to 2010, civil litigator, deputy district attorney and member of the state Coastal Commission. Elected chair of the Commission in March 2014.

ANTHONY CANNELLA | Ceres
Appointed to the Commission by the Senate Rules Committee in March 2022. Civil engineer and principal with Northstar Engineering Group. Former State Senator from 2010 to 2018. Previously served on the Ceres City Council and was twice elected mayor of that city. Elected Vice Chair of the Commission in July 2023.

DION ARONER | Berkeley
Appointed to the Commission by the Senate Rules Committee in April 2019. Partner for Aroner, Jewel, and Ellis. Former State Assemblymember from 1996 to 2002, chief of staff for Assemblymember Tom Bates, social worker for Alameda County, and the first female president of Service Employees International Union 535.

DAVID BEIER | San Francisco
Appointed to the Commission by Governor Edmund G. Brown Jr. in June 2014 and reappointed in January 2018. Managing director of Bay City Capital. Former senior officer of Genentech and Amgen, and counsel to the U.S. House of Representatives Committee on the Judiciary.

SENATOR CHRISTOPHER CABALDON | West Sacramento
Elected in 2024 to represent the 3rd Senate District. Former Mayor of West Sacramento for two decades. Served as Vice Chancellor of the California Community Colleges and director of the State Assembly Higher Education Committee. Previously appointed by President Obama to the National Advisory Board of America's College Promise. Appointed to the Commission by the Senate Rules Committee in February 2025.

ASSEMBLYMEMBER PHILLIP CHEN | Yorba Linda
Appointed to the Commission by Speaker of the Assembly Anthony Rendon in October 2021. Elected in November 2016 to represent 55th District. Represents portions of Los Angeles, Orange and San Bernardino counties and the cities of Brea, Chino Hills, Diamond Bar, La Habra, Industry, Placentia, Rowland Heights, Walnut, West Covina and Yorba Linda.

GIL GARCETTI | Los Angeles
Appointed to the Commission by Governor Gavin Newsom in November 2021. Professional photographer and author of ten books. Former Los Angeles County District Attorney, teaching Fellow at Harvard University's Kennedy School, and president of the California Science Center Foundation's Board of Trustees.

JOSÉ ATILIO HERNÁNDEZ | Burbank
Appointed by Speaker of the Assembly Anthony Rendon in April 2023. Founder and CEO of IDEATE California. Also, Founder and Board Chairman of ideateLABS non profit.

JASON JOHNSON | Napa
Appointed by Governor Newsom in June 2023. Member of the Land Trust of Napa County Board of Trustees and Honorary Commander of Travis Air Force Base. Former Managing Partner at Founders Den. Founder and former CEO at August Home Inc.

GAYLE MILLER | Sacramento
Managing Director of Transition, Institutional Relationships and Investments for Brookfield Asset Management and Vice-Chair of the Delta Stewardship Council. Previously served as Senior Counselor on Infrastructure and Clean Energy Finance for Governor Newsom and Chief Deputy of Policy at the California Department of Finance. Appointed to the Commission by Governor Gavin Newsom in January 2025.

SENATOR ROGER NIELLO | Fair Oaks
Elected in 2022 to represent the 6th Senate District. Former President and CEO of the Sacramento Metro Change of Commerce and University of California Center Sacramento Governance Fellow. Previously served on the Sacramento County Board of Supervisors and the California State Assembly. Appointed by the Senate Rules Committee in February 2025.

ASSEMBLYMEMBER LIZ ORTEGA | San Leandro
Elected in November 2022 to represent the 20th Assembly District. Executive Secretary-Treasurer of the Alameda Labor Council. Former Statewide Political Director for AFSCME Local 3299. Appointed by Speaker of the Assembly Rendon in March 2023.

JANNA SIDLEY | Los Angeles
Appointed to the Commission by Governor Edmund G. Brown Jr. in April 2016 and reappointed in February 2020. Partner at Ichor Strategies and appointed to the Board of the Los Angeles City Employee Retirement System ("LACERS"). Former general counsel at the Port of Los Angeles and city attorney at the Los Angeles City Attorney's Office.

“DEMOCRACY ITSELF IS A PROCESS OF CHANGE, AND SATISFACTION AND COMPLACENCY ARE ENEMIES OF GOOD GOVERNMENT.”

By Governor Edmund G. “Pat” Brown,
addressing the inaugural meeting of the Little Hoover Commission,
April 24,1962, Sacramento, California

LITTLE HOOVER COMMISSION

Milton Marks Commission on California State Government Organization and Economy

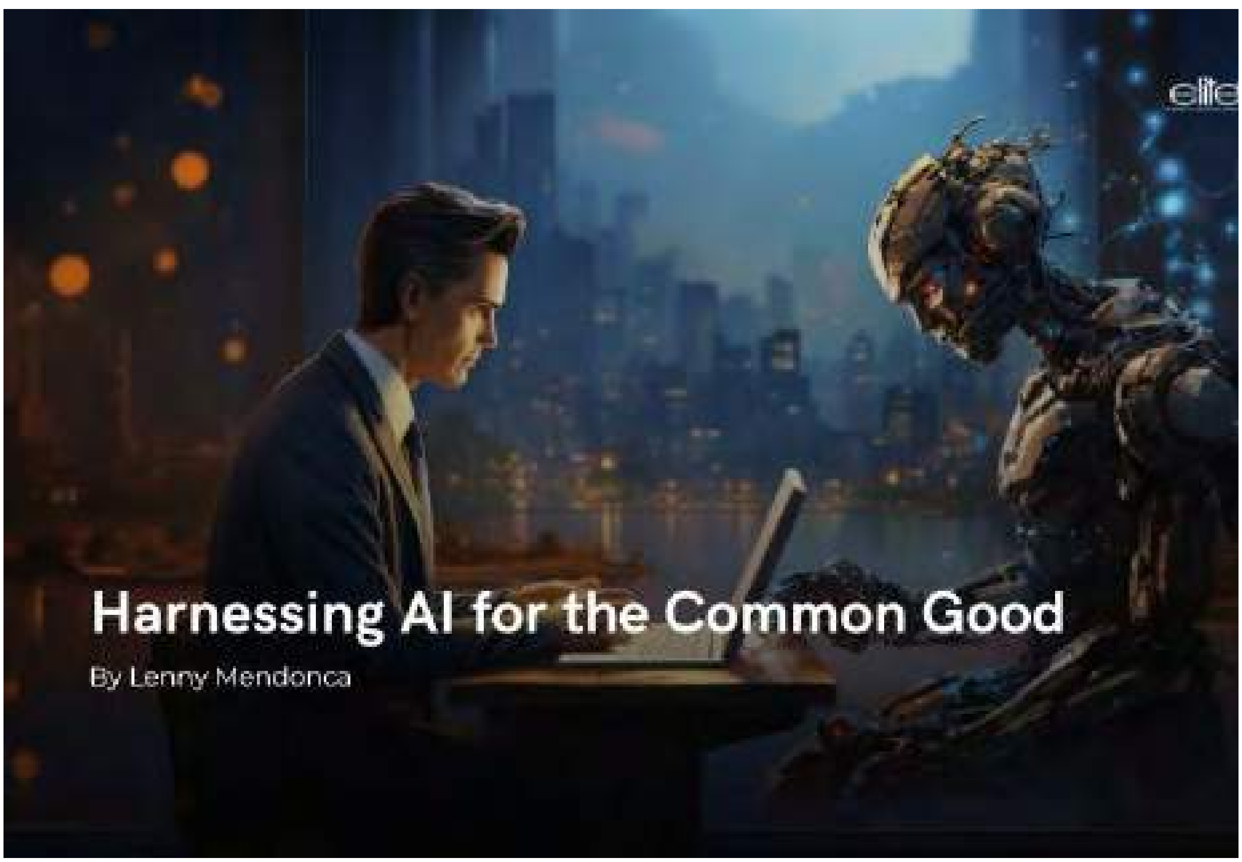
www.lhc.ca.gov

Business / Technology

Harnessing AI For The Common Good

- July 08, 2024
- by Elite + Editorial Team

- Share Post
- Share on Facebook
- Share on Twitter
- Share on Line



00:00 06:19

By Lenny Mendonca

BERKELEY – Much of the media coverage of artificial intelligence has focused on its potential negative impact. Science-fiction scenarios like the human-robot wars of the Terminator franchise and real-world controversies, such as the proliferation of deep fakes and OpenAI's unauthorized use of Scarlett Johansson's voice for its new digital assistant, have captured the public imagination, stoking fears of a dystopian future.

Fortunately, both the AI industry and US President Joe Biden's administration seem to be taking these concerns seriously. Biden's October 2023 executive order on AI, in particular, has been rightly praised for focusing on responsible innovation. By facilitating the development of crucial standards, tools, tests, and transparency requirements, it could help establish accountability, discourage illegal and dangerous applications, and guard against disinformation, privacy violations, and intellectual-property theft.

As a transformative general-purpose technology, AI is likely to have a far-reaching impact on society and the labor market. Consequently, the US government must study the technology's labor implications and act rapidly and decisively to protect workers against potential disruptions.

But, in addition to protecting workers and consumers against potential risks, the government should also play a more active role in promoting the development of AI applications that serve the public good. To achieve this, the United States could leverage its significant and growing technological lead.

Globally, roughly $50 billion in venture capital has been invested in the sector just in the past year. Consequently, the AI boom is projected to boost productivity by as much as 0.6% annually, enabling the global economy to grow despite slower workforce growth.

The US has more AI-related startups than the rest of the world combined, and the current US stock-market rally can be at least partly attributed to optimistic predictions about AI's future. With a market capitalization of $3.3 trillion, chipmaker Nvidia alone is worth more than the entire German stock market. Meanwhile, as companies like Google, Microsoft, Facebook, Apple, Tesla, and IBM spend billions of dollars integrating AI into their existing products and services, the Boston Consulting Group projects that this technological boom will account for 20% of its revenues in 2024. Even the electricity sector is growing rapidly, fueled by the enormous power needs of large language models.

AI has also made significant inroads into numerous non-technology sectors. In the health-care industry, it is used for diagnostics, personalized medicine, patient management, administrative tasks, and pharmaceutical research. The financial sector employs AI technologies for fraud detection, risk management, and personalized services. In manufacturing, these technologies enhance supply-chain optimization, equipment maintenance, and quality control, while agriculture companies integrate AI into precision farming, crop monitoring, pest control, and predictive analytics. And law firms increasingly rely on AI for document review, research, contract analysis, and compliance management.

Driven primarily by market forces with minimal government intervention, these sectors are beginning to harness AI's potential to improve efficiency, reduce costs, and deliver better products. But the government still plays a crucial role in advancing the integration of AI into essential public services like health care, transportation, and education.

While some public-policy applications will resemble private-sector uses like fraud detection, procurement, risk management, logistics, traffic management, and software development, others will require careful adaptation to meet government agencies' specific needs. These include data analytics and visualization, customer service and support, and community input, all of which necessitate greater public-private cooperation and a massive effort to recruit and train AI talent.

As is often the case in new technological arenas, California is at the forefront of global efforts to promote AI for the common good. In May, the state hosted a joint AI conference with Stanford University and the University of California, Berkeley, and launched a six-month pilot program that aims to evaluate the suitability of generative AI tools for services like customer support, highway traffic management, and public safety.

Governments can also play an important role in facilitating research and development. The US already spends $95 billion annually on defense and national-security R&D and $61 billion on health-related research – two areas that will be dramatically affected by AI. Notably, the government has vast amounts of data, both public and private, that could be harnessed for open-source applications aimed at improving traffic monitoring, weather analytics, disaster prevention and response, economic statistics, and public health.

Of course, policymakers must ensure that these data are accurate, unbiased, and used responsibly. While this will not be easy, as evidenced by the bungled rollout of the new student financial-aid forms (FAFSA), it is conceivable that individuals' information soon will be pre-filled into government forms, automatically enrolling them in certain public programs and allowing them to opt out.

Perhaps the most promising public-policy option is integration of AI into training and education programs. Recent innovations in generative AI have made personalized, one-on-one tutoring scalable and accessible, with studies showing that these tools could help bridge the gap between low- and high-skilled workers. The Khan Academy's Khanmigo, a tutoring bot for schools, is a prime example of AI's potential educational benefits.

To be sure, like most other technological advances, AI can have both positive and negative effects. Given the right policies and talent, it could fuel innovation, enhance competitiveness, boost productivity, and improve public-sector efficiency. But to unlock its full potential, the US government must ensure that AI is used to serve the common good. States like California are already leading the way.

Lenny Mendonca, Senior Partner Emeritus at McKinsey & Company, is a former chief economic and business adviser to Governor Gavin Newsom of California and chair of the California High-Speed Rail Authority.

Project Syndicate

- Tags:
- AI
- Technology
- ProjectSyndicate
- PublicGood
- Innovation
- EconomicGrowth

- Share Post
- Share on Facebook
- Share on Twitter
- Share on Line

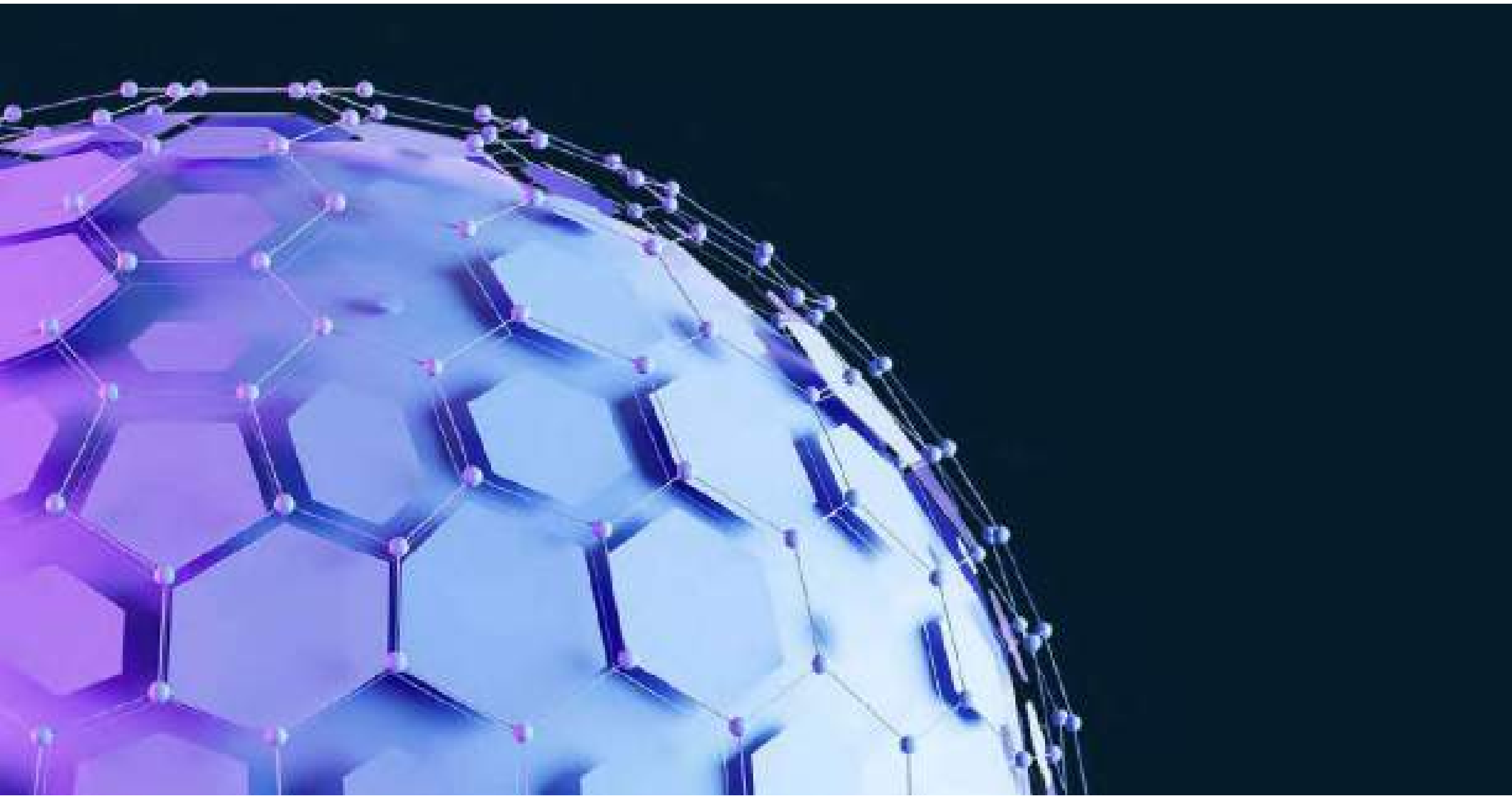
McKinsey & Company

Technology, Media & Telecommunications Practice

Sovereign AI: Building ecosystems for strategic resilience and impact

Sovereign AI is achievable only through an ecosystem effort that connects energy, compute, data, models, platforms, and applications across multiple actors.

This article is a collaborative effort by Ali Ustun, Arnaud Tournesac, Luca Bennici, and Ruben Schaubroeck, with Justin de Niese, Kaavini Takkar, Melanie Krawina, and Newfel Drahmoune, representing views from McKinsey's Technology, Media, & Telecommunications and Public Sector Practices and McKinsey Technology.



March 2026

Disclaimer:
This article is descriptive and analytical. It does not provide policy, regulatory, or national security advice, and it does not recommend specific national strategies. Examples are included to illustrate ecosystem patterns and should not be interpreted as endorsements or prescriptions.

Sovereign AI is moving from a policy debate to an economic and strategic imperative. Across governments, enterprises, and investors, leaders increasingly view the ownership of AI capabilities as central to economic competitiveness, strategic resilience, and societal trust.

Yet despite this urgency, many sovereign AI initiatives are stalling and failing to deliver their expected results. In this article, we analyze how sovereign AI efforts are being pursued and what differentiates sovereign ecosystems that successfully translate intent into scaled adoption and durable advantage. Drawing on a global survey of enterprises, providers, governments, and investors,[1] we then examine the roles different actors must play, the challenges actors face, the partnership models that consistently outperform, and a practical road map for building sovereign AI capabilities that compound over time.

Sovereign AI refers to a nation's or organization's ability to develop and control its own AI capabilities to ensure strategic independence and alignment with domestic values and laws. That said, sovereign AI does not have a single definition; rather, it is the result of the interaction between four distinct components:

— *territorial:* where data and compute physically reside

— *operational:* who manages and secures data and compute

— *technological:* who owns the underlying stack and intellectual property

— *legal:* which jurisdiction governs access and compliance

Viewed this way, sovereign AI is best thought of as a spectrum of potential solutions distributed across different tiers of sovereignty, depending on stakeholder and local circumstances (Exhibit 1).

As a result, sovereign AI represents one of the largest opportunities within AI. McKinsey estimates that 30 to 40 percent of AI spending could be influenced by sovereignty requirements. This would represent a market of some $500 billion to $600 billion globally by 2030 (Exhibit 2).

But seizing that opportunity means confronting a very specific execution challenge: Success in sovereign AI is not achieved through a single policy decision, cloud contract, or "national model" announcement. Instead, sovereignty is best thought of as an ecosystem effort that connects multiple layers—energy, compute, data, models, cloud platforms, and applications—into one coherent system, managing fragmentation across ownership, operating models, and accountability, with explicit trade-offs and deliberate choices about what truly needs to be sovereign.

[1] *Tech: Forward,* "The sovereign AI agenda: Moving from ambition to reality," McKinsey, December 18, 2025.

Exhibit 1

AI workloads can be tiered along a spectrum, based on opportunities for sovereignty.

Tier 1: Fully sovereign		Tier 2: Partially sovereign			Tier 3: Public stack	
Disconnected AI services	**Fully local AI cloud services**	**Hybrid AI stack services**	**Local entity AI services**	**Localized data and operations services**	**Public AI cloud with sovereign controls**	**Basic public AI cloud**
Fully disconnected infrastructure, with services deployed on-premises or in private cloud (eg, private-cloud offerings, solutions for defense, or government use cases)	Cloud-based offerings that are locally developed and hosted across stack (eg, local data center, with cloud platform, models, and applications developed and hosted locally)	Services offered on a hybrid local–global stack (eg, a player with local bare metal but global AI cloud stack or AI application frameworks)	Services operated as joint venture between local legal entity and global player (eg, a venture between a domestic player and a hyperscaler)	Foreign-owned services operated by local personnel in accordance with all data stored in-country (eg, a sovereign offering from a hyperscaler)	Public-cloud services that are hosted in local data centers offering enhanced protection in accordance with local regulatory controls	Public-cloud services with no guarantees on data location or legal jurisdiction

Sovereignty level

High ← → Low

Source: *Tech: Forward*, "The sovereign AI agenda: Moving from ambition to reality," McKinsey, Dec 18, 2025

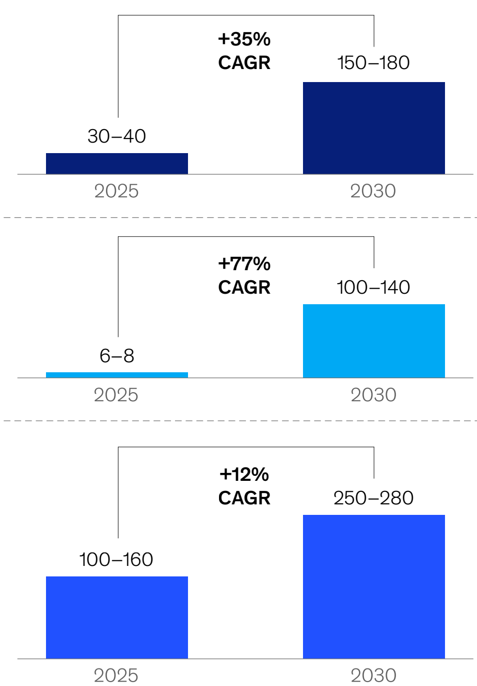
McKinsey & Company

Exhibit 2

The sovereign AI opportunity, already sizable, could expand to a $600 billion market by 2030.

Global market size, total addressable market (TAM),[1] $ billion

Applications and use cases



Models, data, and tooling

Infrastructure and compute

Key drivers

Applications and use cases: Growing local and regional regulatory and compliance requirements are leading nations to reduce dependency on global providers.

Models, data, and tooling: National and regional actors seek model independence to deliver more-relevant outcomes for local populations and to comply with data localization initiatives.

Infrastructure and compute: In response to rising geopolitical uncertainty, nations are co-investing in and building compute capacity and data centers to reduce reliance on global providers.

Total estimated demand	2025	2030	
	~150–200	~500–600	**30–40%** of AI market

[1]TAM is equivalent to total estimated demand for sovereign AI based on baseline AI numbers and sovereignty requirements, which is different from actual spend captured by sovereign AI solutions (which evolves from low percent in 2025 to high percent in 2030, differentiated based on national archetypes).
Source: *Tech: Forward*, "The sovereign AI agenda: Moving from ambition to reality," McKinsey, Dec 18, 2025

McKinsey & Company

The sovereign AI ecosystem: Moving from 'sovereign assets' to 'sovereign assets and outcomes'

For business leaders and policymakers, a useful starting point is to change the way they think about sovereign AI. Many initiatives focus on inputs—such as GPUs, data centers, cloud regions, and national model announcements—and while those inputs matter, the prize is in the long-term outcomes, such as strategic resilience, autonomy, economic value capture, and better social outcomes.

An effective sovereign ecosystem is not necessarily one in which everything is built domestically. Instead, it is one in which key control points are sovereign by design, even if other elements of the stack may remain open to partnerships, interoperability, and competition. Our analysis shows that the most effective ecosystems operationalize "minimum sufficient sovereignty" with a repeatable decision rule: Classify workloads by the importance of regulatory issues and third-party exposure, and then assign a sovereignty tier with explicit requirements for data residency, key ownership, and access controls.

Different jurisdictions are pursuing sovereign AI through distinct ecosystem archetypes. Even jurisdictions with advanced capabilities are rarely self-sufficient across all layers and often rely on external providers in at least part of the stack (particularly in hardware and advanced compute). The following are among the different approaches we are seeing:

— *End-to-end hub and frontier AI powerhouses.* In this model, private data center operators build massive AI-ready capacity to attract hyperscalers and frontier AI labs, enabling large-scale training and advanced inference ecosystems.

— *State-led with data center or cloud execution.* This approach is led by a state to keep compute, data, and model intellectual property under national control for strategic and public sector workloads. Local cloud and data center providers execute the scale-up of sovereign compute and data platforms under state requirements.

— *Model development led by research and policy.* Led by research institutions and policymakers, regulatory and state incentives steer domestic model development and compliant data access, with cloud and data center ecosystems providing trusted environments for research compute to operationalize locally developed models.

— *Industry-led and compute-hardware-driven adoption.* In this approach, data center providers and cloud players partner with local enterprises and chip ecosystem and semiconductor leaders to create AI-capable regional platforms.

— *Policy-enabled regional hubs with strong local or regional demand.* In this model, policy enables fast permitting, power access, and investment incentives, and data center operators build and aggregate AI-ready capacity to serve public sector and regional-enterprise demand and attract anchor tenants.

Even amid these differences, effective sovereign AI ecosystems tend to share a common set of characteristics:

— *A demand-led anchor and sectoral adoption.* Leaders start by clustering demand sources (such as citizen services, health outcomes, financial integrity, critical infrastructure protection, and industrial productivity). This entails an explicit shift in public sector demand: Governments align procurement, funding, and operating mandates with the sovereign AI strategy, acting as anchor customers.

— *Sovereign AI infrastructure as a key pillar enabled by foundational inputs.* Strong ecosystems typically have a deep in-country compute backbone: data centers, high-density GPU clusters, cloud platforms, subsea cables, and low-latency networks that host and run AI workloads. The compute backbone is enabled by physical resources that make infrastructure viable: reliable and affordable power, green energy, land, and water.

— *A clear sovereignty baseline and reference architecture.* Because sovereignty is multidimensional, effective ecosystems codify what must be sovereign into a reference architecture with a set of nonnegotiable control points: data classification and permitted uses; encryption and key ownership; identity and access, logging, and monitoring model risk management and evals; and incident response and legal access pathways.

— *Trusted data governance, policies, and standards.* Governance, policies, and standards constitute an enabling force that shapes speed and scale through land and power allocation, compute import and export rules, incentives, and governance frameworks that attract demand and investment.

— *A data ecosystem and a pragmatic modular model strategy.* Localization keeps data "inside," but it does not automatically make data usable. Strong sovereign ecosystems build data products and sharing mechanisms: interoperable standards and sector consortiums that increase the quantity and quality of training and fine-tuning data. With regard to AI models, ownership builds independence, but bringing the best AI models to the country often means giving up control. As a result, countries are increasingly adopting a layered strategy: leveraging global frontier models where possible and developing or fine-tuning domain and language models where sovereignty needs and value are highest.

— *Capital that matches the stack's risk profile.* Financing fuels build-out and innovation, with mechanisms spanning public incentives, venture capital, and private equity for infrastructure, start-ups, and enterprise adoption. Energy and data centers need patient, infrastructure-style capital; models and platforms need growth capital with risk tolerance; and applications and integration need venture pathways and enterprise adoption muscle. Leading ecosystems align capital instruments to each layer.

— *Local talent that is attracted or nurtured.* The AI talent pipeline is fast emerging as a scarce resource, necessitating rapid upskilling. This includes sustained investment in education, reskilling, and lifelong-learning programs that prepare workers for new roles created by AI integration.

Taken together, these seven elements describe not a checklist, but the components of a system. Effective ecosystems are those that treat sovereignty as a coordinated design problem, aligning these components so that each reinforces the others.

The roles that matter in successful ecosystems: Governments, providers, enterprises, and investors

Building a sovereign AI ecosystem requires coordination across four distinct groups: Governments shape trust, rules, and demand; providers create the underlying technology and platforms; enterprises convert infrastructure into real economic value; and investors supply the capital and risk tolerance needed to scale.

Below, we examine each role, drawing on our survey insights to highlight the specific constraints and the actions required to move from fragmented pilots to scaled outcomes.

Governments: Act as orchestrator, investor, regulator, and anchor customer
Governments have a unique ability to turn fragmented ambition into coordinated execution and thus play a central role in leading ecosystems.

Governments set the sovereignty goalposts. They define which workloads require strong sovereignty (for example, defense, sensitive citizen data, and critical infrastructure), which can use hybrid models, and which can remain largely global. They then translate those choices into actionable controls (such as data classification, auditability, and key ownership). By creating certification regimes, governments can help standardize what "trusted" means so regulated industries can adopt quickly and repeatedly.

Governments also aggregate demand to create an adoption flywheel. Public sector demand can be bundled into multiyear frameworks anchored on a small number of interoperable, at-scale providers to justify up-front investment.

Governments also catalyze supply by enabling policy and targeted investment. Policymakers can also unlock capacity by accelerating permitting and grid readiness for at-scale infrastructure developers, enabling long-term energy planning for AI loads.

Technology providers: Build capability, localize trust, and partner for legitimacy
Providers span hyperscalers, local cloud providers, neoclouds, data center operators, telecom companies, model developers, and integrators. Leading ecosystems rarely choose either hyperscalers or local providers, instead designing architecture in which each competes and collaborates at the layer where it has advantage.

Our survey results highlight a core tension that providers must navigate. While most enterprise leaders describe sovereign AI as strategically important, sovereignty alone rarely drives the decision to switch vendors—a decision that remains dominated by price, performance, and reliability (Exhibit 3). This does not signal weak demand for sovereign options. Rather, it reflects the way enterprises operationalize risk: Sovereignty matters most for a specific subset of workloads—those involving sensitive data, regulatory exposure, or critical services.

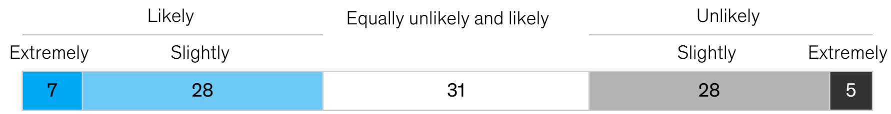
Indeed, sovereign AI offerings are perceived as being 10 to 30 percent more expensive than global alternatives (Exhibit 4). There are cases and instances in which sovereign AI players may be more advantageous than global alternatives, but those players will need to educate customers on these instances and make the case for when a premium is warranted.

For providers, the implication is straightforward. Demand for sovereign AI is real, but it is selective. Sovereignty becomes commercially relevant when it clearly reduces risk or enables deployment in regulated settings and not when it is sold as a blanket feature. In practice, this means that providers succeed when they translate ecosystem-level sovereignty requirements into offerings that are concrete and easy for enterprises to adopt at the workload level, not when asking customers to absorb complexity or uncertainty themselves.

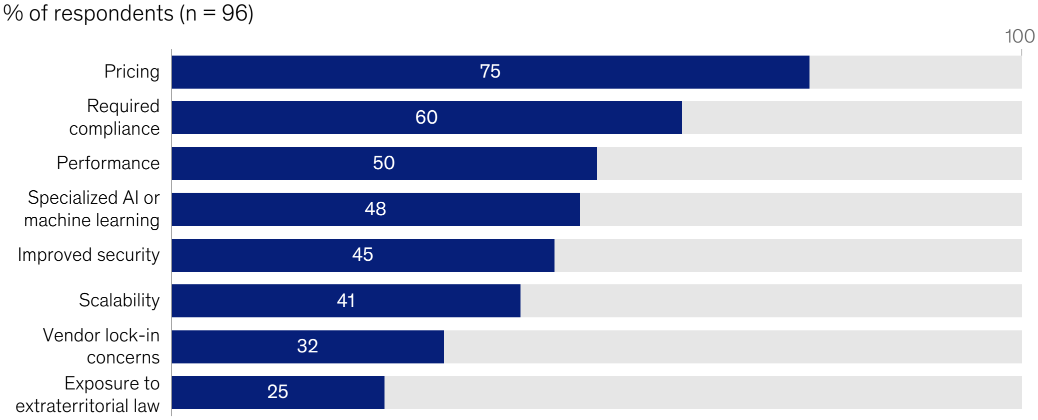
Exhibit 3

About 70 percent of enterprises are open to switching their current AI vendor for better pricing and performance.

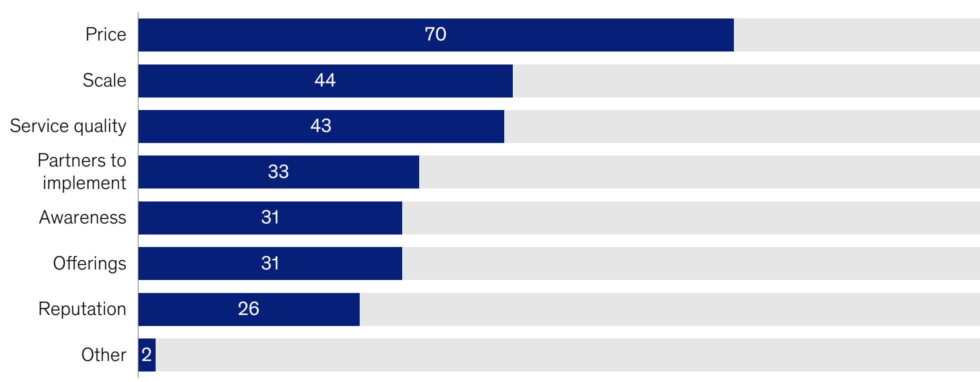
Likelihood to switch or add new vendors,[1] % of respondents (n = 96)



Factors considered when switching from current provider to a sovereign AI provider,[2]
% of respondents (n = 96)



Factors limiting use of sovereign AI provider,[3] % of respondents (n = 96)



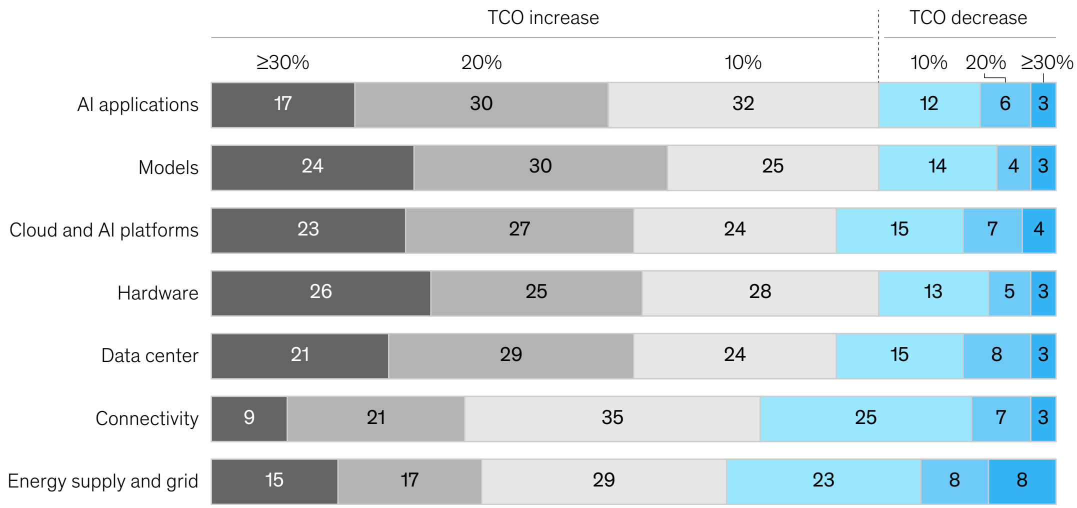
Note: Figures do not sum to 100%, because of rounding.
[1]Question: How likely are you to switch or add new vendors?
[2]Question: What are the factors you would consider when switching from your current provider to a sovereign AI provider? Respondents could select more than one answer.
[3]Question: What is limiting your use of a sovereign AI provider? Respondents could select more than one answer.
Source: McKinsey Sovereign AI CxO global survey, Dec 2025

McKinsey & Company

Exhibit 4

Sovereign AI solutions are perceived as being 10 to 30 percent more expensive across the stack.

Total cost of ownership (TCO) of sovereign solutions, compared with nonsovereign, by tech stack layer,[1] % of respondents (n = 300)



[1]Question: How does TCO compare between sovereign and nonsovereign solutions for each area?
Source: McKinsey Sovereign AI CxO global survey, Dec 2025

McKinsey & Company

Enterprises: Create demand, provide data, industrialize adoption
Enterprises are the demand engines that turn sovereignty into scaled economic value. In leading ecosystems, regulated enterprises and government-owned entities act as anchor tenants that justify ecosystem-wide investments.

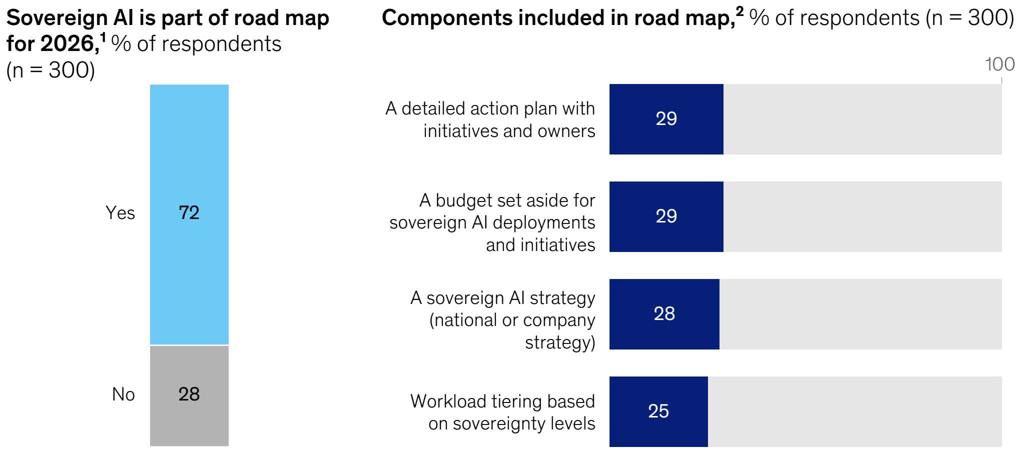
Enterprise interest in sovereign AI capabilities is now widespread, but while most have it as part of their road maps for 2026, few have a detailed strategy, action plan, budgets, and workload tiering (Exhibit 5).

This lack of operational readiness also helps explain why sovereign cloud and AI migrations typically take three to four years (Exhibit 6). These timelines are not driven primarily by technology limitations but instead reflect the organizational work required to move regulated workloads. Sovereign AI adoption is therefore a multiyear transformation, not a simple vendor switch.

At the same time, long migration timelines should not be interpreted as a lack of technical capability. Sovereign and local providers increasingly match global alternatives on service levels, especially lower in the AI stack (Exhibit 7).

Exhibit 5

Nearly three-quarters of enterprises include sovereign AI as part of their 2026 road map—though few have concrete plans in place.

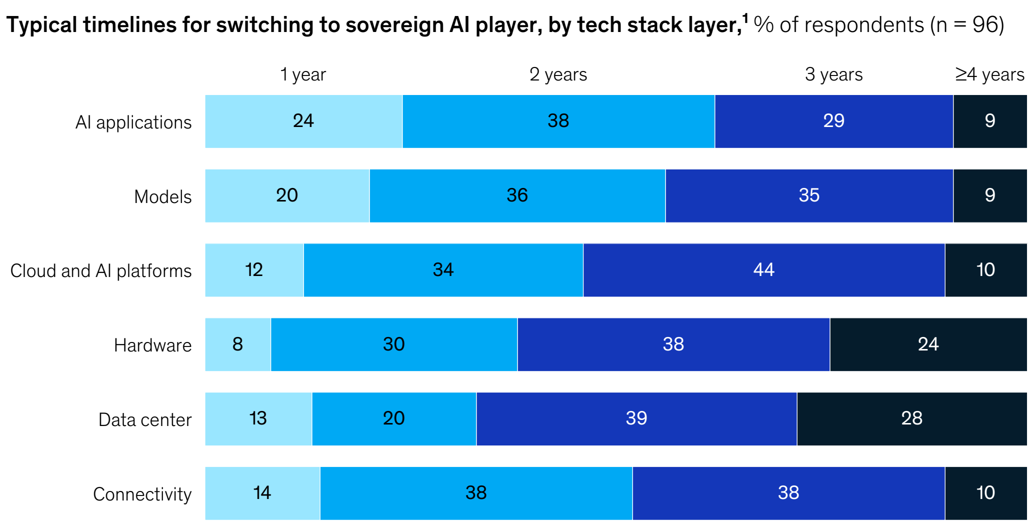


[1]Question: Is sovereign AI part of your road map for 2026?
[2]Quesiton: Which components does your company have? Respondents could select more than one answer.
Source: McKinsey Sovereign AI CxO global survey, Dec 2025

McKinsey & Company

Exhibit 6

Many enterprises expect sovereign AI migrations, especially infrastructure layers, to take three or more years.



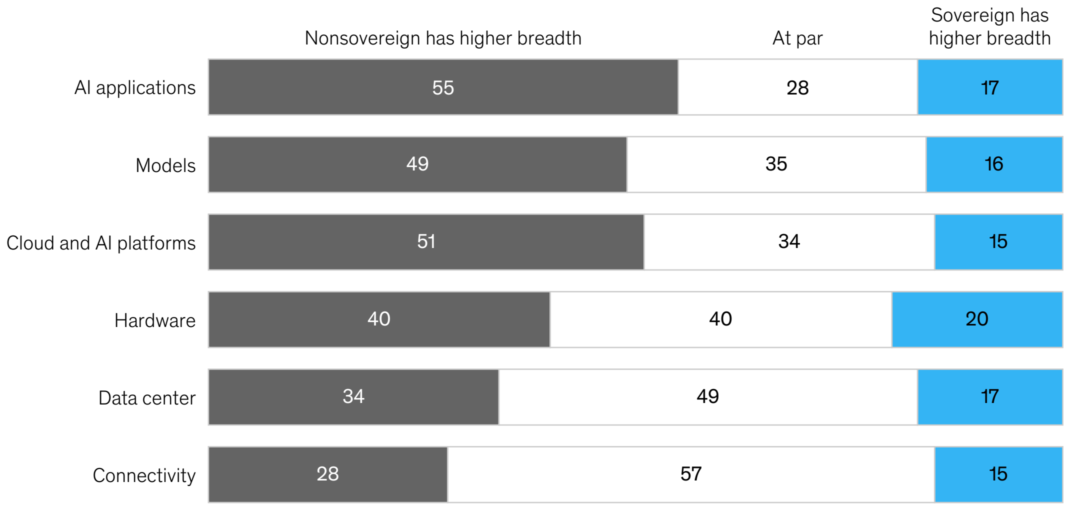
[1]Question: What are the typical timelines for switching to a sovereign AI player across the layers of the stack?
Source: McKinsey Sovereign AI CxO global survey, Dec 2025

McKinsey & Company

Exhibit 7

Enterprises generally see broader service offerings in nonsovereign AI solutions, though many infrastructure layers are viewed as comparable.

Comparison of breadth of services between sovereign and nonsovereign solutions, by tech stack layer,[1] % of respondents (n = 96)



[1]Question: How does breadth of services compare between sovereign and nonsovereign solutions for each area?
Source: McKinsey Sovereign AI CxO global survey, Dec 2025

McKinsey & Company

Sovereign AI migrations are slow not because the technology is immature but because enterprises struggle to decide where sovereignty truly matters and to adapt their operating models accordingly. As a result, providers might consider approaching sovereignty as a portfolio decision rather than as an ideology—segmenting workloads into sovereign, hybrid, and global. This avoids "all or nothing" debates and accelerates time to value while sovereignty deepens over time. Those seeking to capitalize on sovereign demand could also invest in the real bottleneck: data readiness and resolving the operating model. That means building data products with machine learning operations that can span sovereign and nonsovereign environments. Finally, they can help shape the ecosystem—joining sector consortiums, acting as early adopters and reference customers, evolving procurement, and codeveloping domain applications and models with providers and start-ups to accelerate local innovation.

Investors: Provide the capital, manage price risk, and accelerate scale

Sovereign AI is becoming a major investment theme for investors globally, especially among sovereign wealth funds, leading to an expected increase in sovereign AI investment mandates and assets under management.

Sovereign AI spans asset classes: energy, real estate, data centers, connectivity, cloud services, model development, application software, cybersecurity, and integration. Investors matter because they can help organizations bridge the "valley of uncertainty" before utilization is proven, and most expect their investments to increase across the stack, especially higher up the stack.

Across all layers, the most effective investors do two things especially well. First, they back projects with real demand and clear rules, rather than big announcements or speculative build-outs. Second, they help companies grow and exit so local innovation can scale instead of stalling.

Partnership models that outperform

Across markets, several partnership structures can move sovereign AI ecosystems from pilots to scale-up. These models succeed because they align incentives across public and private actors and reduce friction in adoption.

Sovereign AI zones with standardized controls

In this model, integrated environments combine energy, compute, secure connectivity, and compliance controls into a single operating framework. By standardizing security, data residency, and audit requirements up front, these zones reduce onboarding time and enable enterprises and providers to deploy workloads repeatedly rather than through one-off exceptions.

Examples include sovereign cloud zones offered by hyperscalers in Europe and the Middle East as well as national AI or cloud zones that bundle certified infrastructure with preapproved regulatory controls.

Demand aggregation and offtake commitments

These are multiyear frameworks that consolidate demand from the public sector and regulated industry to create predictable workloads. When paired with fast-track procurement and clear scale-up pathways, demand aggregation turns policy intent into bankable utilization.

Examples include the EuroHPC Joint Undertaking, where coordinated public demand underwrites shared supercomputing and AI capacity, with government-led frameworks that anchor early AI workloads in health, defense, or public services.

Joint operating models for sovereign environments

In this model, shared-control structures clearly define who operates infrastructure, who controls access and encryption, and how incidents and compliance are managed. These models offer a hybrid between fully state-run and fully vendor-run environments.

Examples include sovereign cloud joint ventures such as Bleu in France, which combines local operational control with global hyperscaler technology under clearly defined governance.

Model adaptation and data consortiums

Collaborative arrangements can pool data, funding, and demand to develop or fine-tune domain- and language-specific models. Governments typically set governance and evaluation standards while multiple providers compete at the application layer.

Examples include open and semi-open model initiatives such as BLOOM—a large language model created by the BigScience workshop with 176 billion parameters and trained to support 46 natural languages and 13 programming languages—as well as national language or sector-model consortiums.

Blended finance for early layers

Financing structures can combine public capital with private investment to support assets with long payback periods, such as energy, data centers, and foundational platforms. Public participation helps derisk early build-out, with a path to commercial terms as utilization scales.

Examples include public co-investment in national compute hubs and AI factories, which is often paired with long-term offtake commitments from government or regulated industries.

Across these models, the common thread is orchestration. Partnerships outperform when responsibility for aligning incentives, removing friction, and converting sovereign ambition into execution is explicit rather than assumed.

A practical road map: Three waves of ecosystem building

In practice, successful sovereign AI ecosystems tend to emerge through three overlapping waves rather than a single, linear build.

The first wave focuses on establishing the baseline and unlocking early demand. Leaders clarify which workloads truly require sovereign controls, translate those decisions into governance and procurement mechanisms, and launch a small number of lighthouse use cases large enough to justify initial investment. The goal is not completeness but credibility—creating early proof that sovereign environments can operate reliably, securely, and at scale.

The second wave concentrates on scaling shared infrastructure and data ecosystems. With demand signals in place, ecosystems expand compute and energy capacity on bankable terms, industrialize operating models, and invest in sector-specific data products and lawful data-sharing mechanisms. This is where many initiatives falter by attempting to scale infrastructure without first resolving governance, operating model, and talent constraints.

The third wave builds durable advantage and exportable capability. Ecosystems deepen specialization in selected domains, support a competitive provider landscape, and enable start-ups and integrators to scale. At this stage, trusted capabilities become not just domestic enablers but sources of regional or global differentiation.

The most common failure mode is mis-sequencing—investing heavily in shared assets before demand and governance are ready or pursuing global leadership ambitions without the data, adoption, and operating foundations required to sustain them.

Ultimately, sovereign AI is not about full-stack independence. It is an ecosystem play. Those who orchestrate coherent systems—in which sovereignty is applied deliberately at critical control points, and governments, providers, enterprises, and investors align incentives—will turn infrastructure into trusted capabilities and turn trusted capabilities into scaled outcomes.

Ali Ustun is a senior partner in McKinsey's Doha office; **Arnaud Tournesac** is a partner in the Paris office; **Luca Bennici** is a partner in the Dubai office, where **Justin De Niese** is an associate partner and **Newfel Drahmoune** is a consultant; **Ruben Schaubroeck** is a senior partner in the London office; **Kaavini Takkar** is an associate partner in the Seattle office; and **Melanie Krawina** is an associate partner in the Vienna office.

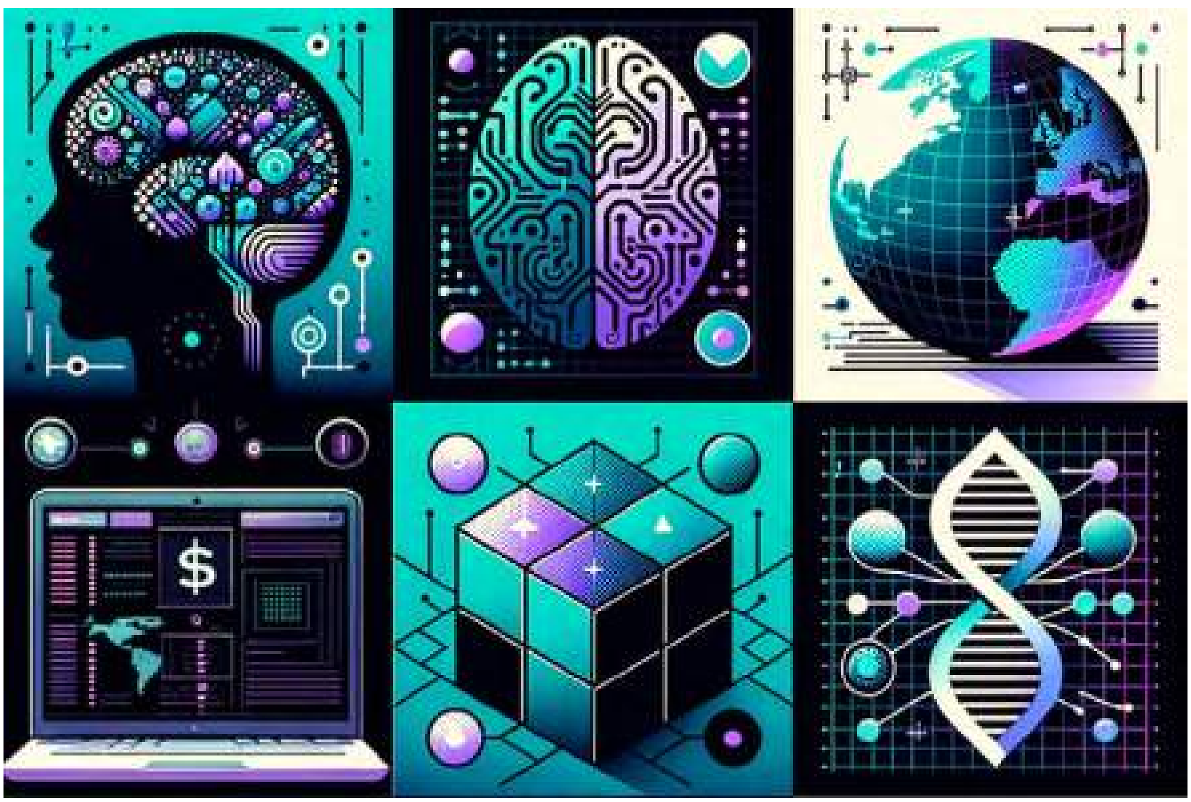
This article was edited by Larry Kanter, a senior editor in the New York office.

December 15th, 2025 | 11 min read

Artificial Intelligence

Stanford AI experts predict what will happen in 2026

The era of AI evangelism is giving way to evaluation. Stanford faculty see a coming year defined by rigor, transparency, and focus on actual utility over speculative promise.



After years of fast expansion and billion-dollar bets, 2026 may mark the moment artificial intelligence confronts its actual utility. In their predictions for the next year, Stanford faculty across computer science, medicine, law, and economics converge on a striking theme: The era of AI

Learn more about what Stanford HAI faculty expect in the new year.

AI sovereignty, global growth

James Landay, *HAI Co-Director, Professor of Computer Science, and the Anand Rajaraman and Venky Harinarayan Professor in the School of Engineering*

My biggest prediction? There will be no AGI (artificial general intelligence) this year.

But additionally, AI sovereignty will gain huge steam this year as countries try to show their independence from the AI providers and from the United States' political system. What is AI sovereignty? In one model of sovereignty, a country might build its own large LLM. In another example, a country might run someone else's LLM on their own GPUs so that they can make sure their data doesn't leave their country. The term "sovereignty" isn't well defined, and HAI is currently working on a project to help people understand these different models and provide some analysis.

We obviously have seen in 2025 a lot of investments in huge data centers around the world, whether it was UAE in May or South Korea in the fall. We probably will see organizations like Nvidia and OpenAI making tours of other countries, which is related to this AI sovereignty and how they can have an advantage. We'll see these continued AI data center investments in 2026. But at some point, you can't tie up all the money in the world on this one thing. It seems like a very speculative bubble.

" AI sovereignty will gain huge steam this year as countries

Stanford HAI Denning Co-Director

Also, in 2026 we'll hear more companies say that AI hasn't yet shown productivity increases, except in certain target areas like programming and call centers. We'll hear about a lot of failed AI projects. Now, do people learn from those failures and figure out where to apply AI appropriately in ways that they're more successful this year? Maybe.

I think we're going to see some new custom UI AI, things beyond just the chabot or even this web browser that OpenAI released. I think this new year we'll see previews of those things, if not real products.

Also, asymptoting. We have huge models, but we've seen better models that are smaller than the huge ones. It seems like we have reached some amount of peak data, both because we're running out of data and also because the data is poor quality. I expect there will be a lot more effort on curating really good datasets that are smaller and creating models that perform better on smaller data.

Obviously, there were a lot of AI video advances that happened in 2025, but they weren't very good. In one of my classes, though, a student team used AI to make a video that normally would have actors and special locations. It had some issues, but it was actually pretty good. I think that means the video tools have finally gotten good enough that we'll see real uses, and we'll see that take off in the new year. Relatedly, we'll see a lot more copyright issues.

Opening the black box is science's next mandate

Russ Altman*, the Kenneth Fong Professor in the School of Engineering and Professor of Bioengineering, of Genetics, of Medicine, and of Biomedical Data Science, and Stanford HAI Senior Fellow*

I see enormous potential in foundation models to unlock discoveries and predictions in science and medicine. Right now, there are several approaches to building these models. For example, developers can build one massive "early fusion" model incorporating all data types, or they can construct a "late fusion" model where they build separate models for each modality and integrate them afterward. I see some profound questions around this. Will separately built models work together well? What happens when some data is bad? If you have a late fusion model, where, say, you have combined several models including DNA, RNA, and proteins, you could just rebuild the

> **“I expect more focus on the archeology of the high-performing neural nets.”**
>
> **Russ Altman**
> HAI Senior Fellow

Also, people are discovering the amazing power of AI in scientific research. But in science, you need more than just an accurate prediction; you have to have insight on how the model got to that prediction. In science labs, there’s increasing focus not on the output of the model, but on the internal neural network that’s leading to the performance, the attention map of which data is paying attention to which other data.

In 2026, I expect more focus on the archeology of the high-performing neural nets. We just published a paper on this called “**Paying attention to attention in proteins**.” I’m seeing people using things called sparse autoencoders on the deep network to try to identify the features in the data that are driving performance. In science, there’s an absolute mandate to open AI’s black box, and I’m starting to see us open that box.

Now looking at AI and the business of health, this past year we saw a ton of business investment in startups using AI for medicine. A typical hospital is inundated with interest from startups that want to sell them a solution for X. Each of those individual solutions is not unreasonable, but in aggregate, they’re like a tsunami of noise coming at an executive. This year I expect we’ll start developing ways to evaluate the impact of an AI system, its technical features, its training population, how to implement it, how efficient or disruptive it is for staff, its ROI on hospital workflow, patient happiness, quality of decisions. All of that is a process that we’re just now wrapping our arms around here at Stanford, led in part by Stanford Hospital Chief Data Scientist Nigam Shah. But then we need to make it available in less technically savvy or well-resourced

reasoning

__Julian Nyarko__, Professor of Law and Stanford HAI Associate Director

I predict that two themes could define the year in the domain of AI for the legal services sector. First, rigor and ROI. Firms and courts might stop asking "Can it write?" and instead start asking "How well, on what, and at what risk?" I expect more standardized, domain-specific evaluations to become table stakes by tying model performance to tangible legal outcomes such as accuracy, citation integrity, privilege exposure, and turnaround time. There could also be a stronger focus on efficiency gains inside real workflows (document management, billing, and knowledge systems) rather than in controlled, artificial scenarios. Second, AI will take on harder work. Beyond intake and first drafts, we already begin seeing a shift toward systems that tackle, for instance, multi-document reasoning: synthesizing facts, mapping arguments, and surfacing counter-authority with provenance. This shift demands new frameworks for measurement – such as LLM-as-judge and pairwise preference ranking – to evaluate complex legal tasks at scale. Emerging benchmarks like GDPval, built around these ideas, could steer development roadmaps toward higher-order tasks.

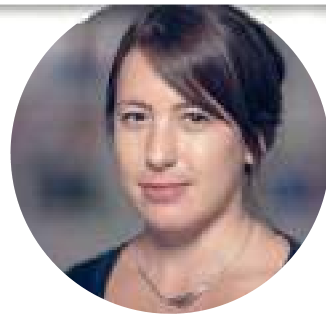
Deflating the AI bubble

__Angèle Christin__, Associate Professor of Communication and Stanford HAI Senior Fellow

The billboards in San Francisco say it all: AI everywhere!!! For everything!!! All the time!!! The slightly manic tone of these advertisements gives a sense of the hopes – and immense investments – placed in generative AI and AI agents.

So far, financial markets and big tech companies have doubled down on AI, spending massive amounts of money and human capital, and building gargantuan computing infrastructures to sustain AI growth and development. Yet already there are signs that AI may not accomplish everything we hope it will. There are also hints that AI, in some cases, can misdirect, deskill, and harm people. And there is data showing that the current buildout of AI comes with tremendous environmental costs.

I expect that we will see more realism about what we can expect from AI. AI is a fantastic tool for some tasks and processes; it is a problematic one for others (hello, students generating final essays without doing the readings!). In many cases, the impact of AI is likely to be moderate: some efficiency and creativity gain here, some extra labor and tedium there. I am particularly excited to



> “We will see more realism about what we can expect from AI.”
>
> **Angèle Christin**
> HAI Senior Fellow

A ‘ChatGPT moment’ for AI in medicine

__Curtis Langlotz__, Professor of Radiology, of Medicine, and of Biomedical Data Science, Senior Associate Vice Provost for Research, and Stanford HAI Senior Fellow

Until recently, developing medical AI models was extremely expensive, requiring training data labeled by well-paid medical experts (for example, labeling a mammogram as either benign or malignant). New self-supervised machine learning methods, now widely used by the developers of commercial chatbots, don’t require labels and have dramatically reduced the cost of medical AI model training.

Medical AI researchers have been slower to assemble the massive datasets needed to capitalize on self-supervision because of the need to preserve the privacy of patient data. But self-supervised learning from somewhat smaller datasets has shown promise in **radiology**, **pathology**, **ophthalmology**, **dermatology**, **oncology**, **cardiology**, and **many other areas of biomedicine**.

Many of us will remember the magic moment when we discovered the incredible capabilities of chatbots trained with self-supervision. We will soon see a similar “ChatGPT moment” for AI in medicine, when AI models are trained on massive high-quality healthcare data rivaling the scale of data used to train chatbots. These new biomedical **foundation models** will boost the accuracy of medical AI systems and will enable new tools that diagnose rare and uncommon diseases for which

time

Erik Brynjolfsson, Digital Economy Lab Director, Jerry Yang and Akiko Yamazaki Professor, and Stanford HAI and SIEPR Senior Fellow

In 2026, arguments about AI's economic impact will finally give way to careful measurement. We'll see the emergence of high-frequency "AI economic dashboards" that track, at the task and occupation level, where AI is boosting productivity, displacing workers, or creating new roles. Using payroll, platform, and usage data, these tools will function like real-time national accounts. In our "**Canaries in the Coal Mine**" work with ADP, we already see early-career workers in AI-exposed occupations experiencing weaker employment and earnings outcomes; in 2026, similar indicators will be updated monthly, not years later. Executives will check AI exposure metrics daily alongside revenue dashboards, and policymakers will use them to target training, safety nets, and innovation policy. The debate will shift from whether AI matters to how quickly its effects are diffusing, who is being left behind, and which complementary investments best turn AI capability into broad-based prosperity.

> **"Arguments about AI's economic impact will finally give way to careful measurement."**
>
> **Erik Brynjolfsson**
> HAI Senior Fellow

GenAI attempts to bypass the enterprise

Nigam Shah, Professor of Medicine and of Biomedical Data Science, and Chief Data Scientist for applications that are made available for "free" to end users. Consider, for example, efforts such as literature summaries by **OpenEvidence** and on-demand answers to clinical questions by **AtroposHealth**.

On the technology side, we will see a rise in generative transformers that have the potential to forecast diagnoses, treatment response, or disease progression without needing any task-specific labels.

Given this rise in available solutions, the need for patients to know the basis on which AI "help" is being provided will become crucial (see **my prior commentary** on this). The ability for researchers to keep up with technology developments via good benchmarking will be stretched thin, even if it is widely recognized to be important. And we will see a rise in solutions that empower patients to have agency in their own care (e.g., **this example** involving cancer treatment).

> **"We're at an important point of reflection to think about what we really want from AI."**
>
> **Diyi Yang**
> Assistant Professor of Computer Science

Advancing human-AI interaction for long-term benefits

Diyi Yang, *Assistant Professor of Computer Science* point of reflection to think about what we really want from AI and how to make it happen.

Personally, I would like to see more work around designing human-centered AI systems that are not only technically capable, but also meaningfully connected to how people think, interact, and collaborate. This requires going beyond optimizing AI systems for short-term engagement or satisfaction and instead prioritizing how human-AI interactions shape users' long-term development and well-being. We need AI systems that augment human capabilities, and this needs to be built into the AI development process from the start, not treated as a post-hoc alignment problem.

This story was **originally published** by the Stanford Institute for Human-Centered Artificial Intelligence.

Writer

Shana Lynch

Campus unit

Stanford Institute for Human-Centered Artificial Intelligence

Related topics

Artificial Intelligence

Share this story

Copy link

February 19th, 2026 | 7 min read

Artificial Intelligence

AI+Education Summit addresses learning and equity

Educators and researchers at the fourth annual summit explored how generative AI is reshaping assessments and impacting student creativity and human connection.

The 2026 AI+Education Summit, hosted by the Stanford Institute for Human-Centered AI and the Stanford Accelerator for Learning, took place Feb. 11 on the Stanford campus. | Ryan Zhang

In brief

marginalized communities, and offered strategies for teaching AI literacy.

Experts also encouraged educators to prioritize human connections over technology in student support.

Last week, Stanford Institute for Human-Centered AI and the Stanford Accelerator for Learning convened educators, researchers, technologists, policy experts, and more for the fourth annual AI+Education Summit. The day featured keynotes and panel discussions on the challenges and opportunities facing schools, teachers, and students as AI transforms the learning experience.

At the summit, several themes emerged: AI has created an assessment crisis – student projects no longer indicate a strong learning process; schools are awash with too many AI products and need better evaluations and sustainable adoption models; AI's benefits aren't equitable; AI literacy is a non-negotiable; human connection is irreplaceable.

Read a few of the highlights from the Feb. 11, 2026 event, and **watch the full conference on YouTube**.

AI's inequitable impact

AI amplifies whatever educational foundation already exists, said Wendy Kopp, founder of Teach for All. In mission-driven schools with strong pedagogy, AI becomes a powerful tool for teachers and learners. But without a strong pedagogy and guidelines, the technology becomes a distraction.

Wendy Kopp, founder of Teach for All, asked: “How do we make sure that it’s our teachers and our educators, especially those working with the most marginalized students, who are at the forefront of driving how we utilize AI?” | Ryan Zhang

Miriam Rivera, of Ulu Ventures, said a critical distinction emerges between consumption and creation of AI. In well-resourced schools, she said, students often learn to create with technology (3D printing, coding), while in less-resourced schools, students merely consume it.

Both panelists said that equity-focused teachers and students from marginalized communities must be at the forefront of designing AI applications, not just receiving them.

Dennis Wall, a Stanford School of Medicine professor, illustrated one way this might look. His team is developing a gamified framework to support children struggling with social communication skills. His lab is co-designing these resources with the teachers, therapists, and parents who will use them, ensuring the tools are accessible, engaging, and informed.

AI literacy is a must-have

Education has long assumed that strong products (homework, summative tests, problem sets) indicate strong learning processes, said **Mehran Sahami**, a Stanford School of Engineering professor. AI has broken this assumption. Students can now generate impressive products without engaging in meaningful learning. This directs educators to focus on assessing and supporting the actual learning process rather than just evaluating end products.

Mehran Sahami, the Tencent Chair of the Computer Science Department and James and Ellenor Chesebrough Professor in the School of Engineering, noted: “Our challenge with generative AI is not to consider it as a tool, but consider it as a topic, and build a curriculum in which we teach how to use it to foster things like creativity and deeper educational outcomes.” | Ryan Zhang

More so, we can’t treat AI solely as a tool. Students need a systemic curriculum on AI. Sahami proposed a progression: Introduce what AI is; teach about hallucinations and bias; show how to verify AI outputs; teach advanced techniques like prompting. Without this structured approach, students teach themselves – and 70-80% use AI to short-circuit learning rather than enhance it.

Mike Taubman, a teacher at North Star Academy in Newark, N.J., developed an “AI driver’s license” curriculum that maps the adolescent rite of passage of getting a driver’s license onto AI literacy. The goal is to put students in the driver’s seat, not the passenger seat, when it comes to AI. The four-part curriculum includes choosing a destination (students learn to ask want they want of AI); learning how to drive (see how these tools work and what it means to prompt, develop agentic workflows, etc.); opening the hood (understand their limitations and risks); and defining the rules of the road (decide what AI should and shouldn’t do).

Understand AI’s learning harms

Guilherme Lichand, assistant professor at Stanford Graduate School of Education, studied AI’s impact on creativity for middle school students in Brazil. He compared AI assistance with guardrails (if students ask for 10 words, the AI would give only 3) against no assistance across creativity tasks.

Students with AI assistance performed better on the task while they had the tool. But when assistance was removed within the same test, the advantage disappeared – suggesting no immediate positive transfer.

While that finding isn’t surprising, he said, the results from a follow-up creative task were more concerning:

> Students who never had AI performed best.

than their initial advantage.

This wasn't just about missing the tool – students had less fun and began believing AI was more creative than they were, he said, suggesting AI damaged their creative self-concept.

A "too many pilots" problem

Today we have no shortage of AI products, said Stanford Graduate School of Business Professor and HAI Senior Fellow **Susan Athey**, but we lack effective implementation and adoption. Schools and districts are slow to adopt new tools because of historical software lock-in and the opportunity costs of training teachers on systems that may fail.

"The bottleneck is we have too many pilots actually, and still not enough implementations that are actually effective," said Susan Athey, the Economics of Technology Professor at Stanford Graduate School of Business. | Ryan Zhang

Athey also noted a "teaching to the test" problem for developers. If teachers spend more time on an interface, does that mean it's good and they're deeply engaged, or does that mean it's terrible simulations that could be developed by universities and philanthropy to create robust measurement infrastructure the whole sector can use.

Never replace real relationships

Nearly half of all generative AI users are under 25, said Amanda Bickerstaff, CEO of AI for Education; that's over 300 million active monthly users of ChatGPT alone who are under 25. Students use AI more for mental health and well-being – seeking connection, support, and understanding – than for schoolwork. Bickerstaff warned about cognitive offloading, mental health offloading, and even "belief off-loading," where AI fundamentally shapes how people think, with just four or five chatbot makers having outsized influence on billions of users.

Because of that, she said, we must equip people with knowledge, skills, and mindsets to understand when and how to use AI and, crucially, when not to use it.

The most vulnerable, according to new research from **Pilyoung Kim**, visiting scholar at the Stanford Accelerator for Learning and a professor of psychology and director of the Center for Brain, AI, and Child (BAIC) at the University of Denver, are young people lacking human connections. She asked over 260 middle school students and their parents to compare and share preferences between two chatbot conversation styles: A "best friend" that was highly relational and would respond with comments like, "That must be so upsetting. Your ideas matter so much. I'm always here to listen," and a more transparent version that set boundaries and reminded the user that it was an AI.

Pilyoung Kim, visiting scholar at the Stanford Accelerator for Learning, said that young people with unmet social needs may be more drawn to an AI that provides social connections, making them vulnerable to rely on relationships that are not real. | Ryan Zhang

More adolescents preferred the relational AI, and even more than half of parents chose the relational AI for their teens, reasoning it would be more effective at supporting issues their children might not share with them directly.

But more importantly, children who chose the relational AI were also more likely to report feeling stressed or anxious, and they reported a lower family relationship quality.

“If they have more unmet social needs, it is possible that they’re more drawn to an AI that provides social connections,” Kim said. “That might put them in more vulnerable positions to overly rely on a relationship that is not real.”

She emphasized the common thread of the day: AI should never replace human connection.

The AI+Education Summit is co-hosted by the Stanford Accelerator for Learning and the Stanford Institute for Human-Centered Artificial Intelligence (HAI). **Watch the summit videos on YouTube.**

This story was **originally published** by the Stanford Institute for Human-Centered Artificial Intelligence.

Writer

Shana Lynch

Campus units

Stanford Institute for Human-Centered Artificial Intelligence
Stanford Accelerator for Learning

Education

Share this story

Copy link

Subscribe to Stanford Report

News, insights and events delivered to your inbox each weekday morning.

Email address

Sign up

Research Matters

Groundbreaking innovations that begin in Stanford labs flow freely into private industry to improve human well-being, fuel the economy, and strengthen American competitiveness.

Learn more about research at Stanford

Popular stories

1. Watching a lifetime in motion reveals the architecture of aging
2. Stanford scholars train AI to better augment human creativity
3. College leaders reflect on the future of higher education
4. Why we can’t stop clicking on rage bait
5. What your phone knows could help scientists understand your health

View all

Events

As AI reshapes work, what should workers do next?

Event

Science & Engineering

Stanford scholars train AI to better augment human creativity

Research

Stories to keep you informed and inspired, from one of the world's leading research and teaching institutions.

March 10th, 2026 | 4 min read

Science & Engineering

Stanford scholars train AI to better augment human creativity

Visual artists working with text-to-image tools face frustrating barriers. Computer science, cognitive psychology, and education scholars think they can help.

Getty Images

In brief

The research team is developing open-source AI tools, creating [illegible] to guide model outputs and enhance visual storytelling.

These innovations could transform various fields, empowering creators of all skill levels to express their ideas seamlessly and effectively.

The conversation around AI and art generally swings between two extremes: A flood of AI slop or the total automation of creative work. The more desirable approach may be an AI that behaves as a useful collaborator.

But thus far, visual artists working with text-to-image tools confront frustrating basic hurdles in their abilities to direct AI. Ask an AI to create an image of a house? Not too difficult. Direct it to make the house red, with four front-facing windows, a chimney, and ivy covering the left side? Good luck.

Stanford computer science, cognitive psychology, and education scholars believe they can help AI better augment human creativity by teaching models and people to communicate ideas with each other. With funding from a Stanford Institute for Human-Centered AI (HAI) **Hoffman-Yee Research Grant**, the scholars are developing a shared conceptual grounding for humans to collaborate with generative AI on production-quality visual content ranging from illustrations to diagrams to animations.

“While the models seem amazing, they are terrible collaborators,” says **Maneesh Agrawala**, professor of computer science at Stanford and a co-principal investigator for the project. “Creators have no way of knowing what the AI will produce when given a certain text prompt. If you ask for a suburban single-family home, it generates a modern duplex.”

Authoring original content requires having opinions and constantly making choices, Agrawala explains. Humans and AI need a shared set of concepts so the nuance doesn’t get lost in translation.

Deciphering the human creative process

The Stanford team is approaching this problem from two directions. First, the scholars are running experiments to better understand how people collaborate to create visual content. They have conducted several studies of people performing creative tasks to analyze through chat logs and sketches how the participants communicate as they work together.

everyone talks or draws the same way, but they still expect to be understood."

Building AI tools that understand creators

Second, the team is building open-source AI tools to apply the lessons learned about human creative communication. For example, **ControlNet** teaches text-to-image diffusion models about spatial composition, using two separate features, blocking and detailing, to mirror how artists begin with a rough sketch and then complete the detail of a drawing. Today's models struggle to capture the idea of a pose or how objects should be arranged in a scene. With this tool, creators can guide models to a layout that matches their vision.

> **"The impact of a shared conceptual grounding between humans and AI promises to yield new applications in diverse fields."**
>
> **Maneesh Agrawala**
> Professor of Computer Science

Another tool called **FramePack** enables creators to generate 3D videos from a text prompt for multi-scene storytelling. This tool teaches models to prioritize scenes based on their importance to the overall story, similar to the way a human would work on the project.

A third innovation explores the power of neuro-symbolic AI, which combines neural networks with reasoning capabilities to increase transparency and overcome the limitations of "black box" AI. Using these principles, the team has developed a visual scene coding language that works from a natural language text prompt to produce lines of code, which are executed and rendered to create a 3D scene. Human creators can stay in the loop to inspect or edit the code and prompt the AI to update its program at any time.

Reimagining education content

The impact of a shared conceptual grounding between humans and AI promises to yield new applications in diverse fields, including design, simulation, animation, robotics, and education, says Agrawala. The research team is currently working with gaming platform Roblox to enable players to generate unique 3D objects from text prompts while imposing game restrictions (so, for using a combination of natural language, example content, code snippets, and other modalities.

"We're serious about equipping the broader creative community with the tools they need to communicate with AI effectively," Fan says.

Want to learn more? **Watch this research team** discuss the latest findings during the recent Hoffman Yee Symposium at Stanford HAI.

This story was **originally published** by Stanford HAI.

Writer

Nikki Goth Itoi

Campus unit

Stanford Institute for Human-Centered Artificial Intelligence

Related topics

Science & Engineering
Artificial Intelligence

Share this story

Copy link

Sign up

Research Matters

Groundbreaking innovations that begin in Stanford labs flow freely into private industry to improve human well-being, fuel the economy, and strengthen American competitiveness.

Learn more about research at Stanford

Popular stories

1. Watching a lifetime in motion reveals the architecture of aging
2. College leaders reflect on the future of higher education
3. Stanford scholars train AI to better augment human creativity
4. Why we can't stop clicking on rage bait
5. What your phone knows could help scientists understand your health

Read next

View all

materials

News

detail

News

Stories to keep you informed and inspired, from one of the world's leading research and teaching institutions.

March 13th, 2026 | 4 min read

Events

As AI reshapes work, what should workers do next?

At the 2026 SIEPR Economic Summit, economists examine the latest labor data and offer guidance for navigating the changes ahead.

Erik Brynjolfsson, Chad Jones, and Erika MacEntarfer discuss the promise and perils of AI. | Ryan Zhang

With fears of an AI-driven jobs apocalypse rising, workers – especially younger ones – could use some advice right about now.

First, there were reassuring words from Erika McEntarfer, until recently the head of the Bureau of Labor Statistics and now the Tad and Dianne Taube Policy Fellow at the Stanford Institute for Economic Policy Research (SIEPR). The bottom isn't falling out of the job market, she said, but there are signs of weakness.

Stanford economists discuss AI and the economy at the 2026 SIEPR Economic Summit: SIEPR's Neale Mahoney, Stanford Digital Economy Lab's Erik Brynjolfsson, Stanford GSB Professor Chad Jones, and SIEPR's Erika McEntarfer. | Image: Ryan Zhang; video: courtesy SIEPR

"Unemployment is edging up for those most AI-exposed occupations," McEntarfer said, "but much more slowly than it is rising for everyone else." The slowdown in hiring over the last 18 months has mostly impacted manual labor jobs, she said.

The bigger fear is what may come next. Erik Brynjolfsson is a SIEPR senior fellow and director of Stanford's Digital Economy Lab, whose co-authored research has identified a steady slide in hiring of workers whose jobs are most vulnerable to AI's advancing capabilities – software engineering and call-center customer support, for example. Speaking at the summit, before a standing-room audience of leaders in business, government, and academia, Brynjolfsson said that if unemployment rates for these workers continue to slide and if the falloff in hiring spreads to more jobs, "we're going to see much more striking effects" of AI on the labor market overall.

> **“Unemployment is edging up for those most AI-exposed occupations. But much more slowly than it is rising for everyone else.”**
>
> **Erika McEntarfer**
> The Tad and Dianne Taube Policy Fellow at SIEPR

What’s urgently needed now, but lacking, are government policies aimed at creating a more flexible economy – one where there’s better real-time data and sufficient training programs to help workers gain new skills. “I feel like we’re kind of flying blind [and] not putting in place the infrastructure we need,” Brynjolfsson said.

Charles “Chad” Jones, a SIEPR senior fellow and the STANCO 25 Professor of Economics at Stanford Graduate School of Business, echoed these concerns. “I think a world where growth explodes is very likely a world of great inequality,” he said. A future in which AI, instead of posing an existential threat to humans, drives economic growth “is a world of incredible abundance, and that incredible abundance should let us solve problems” like extreme inequality.

So what can young workers be doing today to prepare for tomorrow’s unknowns, asked moderator Neale Mahoney, the Trione Director at SIEPR and the TG Wijaya Professor of Economics at the Stanford School of Humanities and Sciences.

Speakers offered some ideas:

Use AI to learn, not just automate. Brynjolfsson said research shows that employment is falling among workers who use AI to automate tasks, but growing for those adopting AI to learn new skills.

Identify where humans have the advantage. AI executes, but humans define problems and evaluate solutions. Identify the “weak links” – or bottlenecks – in a series of tasks needed to get a job done that can’t be automated, and focus on those as your added value, Jones said.

Don’t forget to show up. “Never underestimate just how valuable [being a worker who can] get stuff done for your boss is, regardless of your skillset,” McEntarfer said.

This story was **originally published** by Stanford Institute for Economic Policy Research.

Writer

Krysten Crawford

Campus unit

Stanford Institute for Economic Policy Research

Related topics

Events
Economics
Artificial Intelligence

Share this story

Copy link

Groundbreaking innovations that begin in Stanford labs flow freely into private industry to improve human well-being, fuel the economy, and strengthen American competitiveness.

Learn more about research at Stanford

Popular stories

1. Watching a lifetime in motion reveals the architecture of aging
2. Stanford scholars train AI to better augment human creativity
3. College leaders reflect on the future of higher education
4. Why we can’t stop clicking on rage bait
5. What your phone knows could help scientists understand your health

Read next

View all

Events

College leaders reflect on the future of higher education

Event

Events

Family Weekend can hold special meaning for first-gen students

News

Stories to keep you informed and inspired, from one of the world's leading research and teaching institutions.

McKinsey & Company

The telco reinvention: How AI can fuel value creation

March 2026

Contents

Insights

Issue briefs

Insights

McKinsey & Company

The critical bets on the future of telco value creation

After a challenging decade, telecom operators may finally be poised to spur a new era of growth. But it will require embracing fundamental change and making strategic moves in four key domains.

This article is a collaborative effort by Andrea Travasoni, Dev Patel, Theodor Vendrig, and Tiago Palma, with Alima Ziman and Marco Labianca, representing views from McKinsey's Technology, Media & Telecommunications Practice.

March 2026

Telecom operators (telcos) have played a critical role in laying the foundation for the massive, technology-driven economic and financial gains of the 21st century. Yet for most of that period, the industry hasn't managed to share in much of the spoils. While established tech titans and emerging start-ups alike were profiting from successive generations of wired and wireless infrastructure, the telcos that deployed and operated those networks found themselves weighed down by the crushing capital burden as well as intense new competition. That initial period from the late 2000s to 2015 put great pressure on the industry's revenues and margins, leaving it with an ROIC that fell below the cost of capital.

The industry has spent much of the last decade trying to spur a recovery, but its track record has been mixed. Despite making modest progress by embracing selective consolidation, more disciplined and agile operating models, and adjacent businesses or services, it's often seemed the best the industry could do over the past decade has been to tread water. The market value of telcos has done just that, as the persistent challenges of the prior five to ten years continued to take their toll. Regulatory constraints, a legacy business model, shifting customer expectations, and technological disruption have helped saddle telcos with both TSR and enterprise value multiples that have stayed flat at around 5 to 7 times for the past ten years, even as the tech sector's have continued to grow to 15 or higher.

However, as the industry turns the page on the first quarter of the 21st century there are some signs that telcos may finally be poised to embark on a new era of healthy growth. While the industry experienced notably low TSR growth over the last two decades—at 29 percent versus 235 percent for all sectors globally—it has stabilized, growing 28 percent since the start of 2024, firmly on par with the overall global market (Exhibit 1).[1]

At the same time, in most regions, the longstanding imbalance between revenue growth and investment growth is beginning to narrow. From 2019 to 2021, for instance, investment capital (IC) at North American telcos grew at a CAGR of 1.8 percent while revenues grew just 0.4 percent; by contrast, from 2021 to 2024, IC declined at a CAGR of 0.4 percent compared with revenues' annual decline of 1.2 percent. The industry is showing a more hesitant, wait-and-see perspective on capital outlays, fueled in part by the fact that many costly 5G and fiber rollouts are close to completion. In a survey of 152 telco executives last year, only a third said they expected investment capital growth to accelerate over the coming three to five years, a 50 percent decline from the prior year's outlook.[2]

Still, driving healthy, long-term growth for telcos will require more than taking a sharper approach to infrastructure investments or reaping productivity improvements. It will take the industry embracing fundamental change and making strategic moves in four key domains:

— **AI-native transformation.** Comprehensive, AI-driven reinvention of the end-to-end processes and operating model across all functions requires the capability to partner strategically with hyperscalers, system integrators, and solution and change management specialists. At the same time telcos will need to harness the technology and data to help automate workflows and break down legacy silos and fragmented systems.

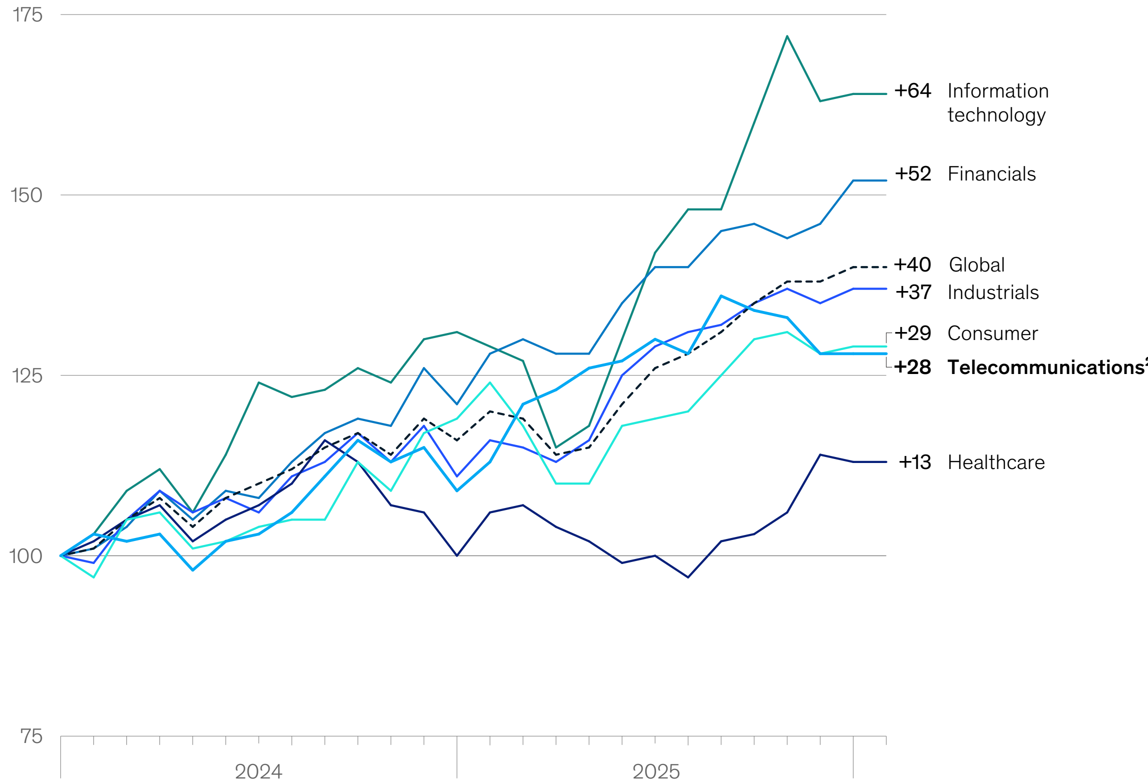
— **Growth beyond the core.** Achieving the goal of healthy, long-term growth requires using AI to redefine consumer and enterprise value propositions through hyper-personalized, automated engagement, both with core offerings as well as an expanded, diverse portfolio of low-cost adjacent products and services.

[1] Industry TSR growth as reflected in the MSCI Total Shareholder Yield index.
[2] McKinsey Telco CxO Survey was conducted online in February 2025, with a sample of 152 respondents representing Africa, Asia, Australia, Europe, Latin America, the Middle East, North America, and Oceania.

Exhibit 1

After more than a decade of lagging most sectors, telcos' market performance has more recently begun to approach global averages.

MSCI total shareholder return,[1] index (Jan 2024 = 100)



[1]Based on last sale price on Jan 30, 2026 of MSCI All Country World Index (ACWI) and sub-sector indices.
[2]Based on MSCI ACWI Select Telecommunication Services Screened 35/20 Capped Index ($).
Source: Capital IQ

McKinsey & Company

— **Infrastructure innovation.** As sovereign AI and data infrastructure become strategic differentiators in select markets and ownership models evolve to unlock capital and increase asset utilization, rethinking and upgrading legacy systems and technologies is critical.

— **Market shaping.** In an industry becoming increasingly commoditized, carefully navigating market dynamics and regulatory environments is more essential than ever; structural transformation such as the delayering of infrastructure or carveouts of non-core assets and continued consolidation remain critical, but only if pursued as part of a broader holistic strategy and transformation.

If telcos successfully execute a number of moves across these four domains, they could increase the industry's lagging ROIC rates and lay the foundation for renewed top- and bottom-line performance, which could in turn lead to more favorable market valuations. Over the last five years, fewer than one in five large telcos across the globe has achieved above-industry-average growth in both revenue and profits, and those were the only players rewarded with enterprise value (EV) growth, delivering around 11 percent annual EV growth on average.[3] Players that fail to make these significant shifts could face growing competitive pressure in an AI-driven environment. This article explores what it takes for telcos to avoid that fate.

Four strategic moves key for a new era of telco growth

The range of strategic moves telcos can make to reignite growth is broad, spanning four primary domains, but the pace of adoption varies widely throughout the industry, and not all moves carry the same importance (Exhibit 2).

AI-native transformation

Like many other industries, telcos haven't yet seen much top- or bottom-line impact from their experiments with gen AI across various parts of their organization. As with some of their previous technological initiatives, telcos have suffered from a siloed, piecemeal approach to incorporating gen AI or more recently agentic AI. They have tended to favor narrow use-case implementations or cost-reduction pilots over a comprehensive redesign of end-to-end domains, which have so far yielded little financial upside.

Industry leaders increasingly recognize that capturing AI's full potential requires a more holistic transformation, supported by two critical enablers.[4] First, data: access to a broad and integrated data foundation is essential to redesign processes end-to-end, and prevents data from becoming the source of a new generation of invisible silos. Second, the operating model: scaling AI requires cross-functional teams, rapid test-and-learn cycles, and continuous improvement embedded into day-to-day execution.

Though they remain relatively early in their AI journeys, with very few initiatives fully implemented, many telcos have started to make progress. Sixty-one percent of telco executives in a recent McKinsey gen AI–focused survey said they are focused on scaling gen AI use cases across functions and less on isolated pilots. And, 47 percent reported experiencing some impact from those efforts, up 20 percentage points from a couple of years ago.[5] That impact is limited so far, with AI driving less than two percent of total revenues for more than half the operators we surveyed that have tested AI use cases. But the industry remains optimistic about its potential. Fully 64 percent of leaders said they expected the technology to contribute more than five percent of revenues in the near future, and 40 percent anticipated AI-driven cost reductions to exceed 10 percent once the technology is fully scaled across their organizations.[6]

As they attempt to achieve those goals, telcos will need to adapt to the fact that data is the core competitive asset in an AI environment. That gives even greater urgency to the industry's pressing need to standardize its fragmented systems, break down silos, and make structured data easily accessible at all levels of the company. At the same time, standing out from the crowd could become even more challenging. The rise of AI-powered agents risks further commoditizing the telco market, with all players offering similar efficiency and service quality, a situation that could force telcos to be even more innovative and creative in communicating a distinctive value proposition.

[3] Based on a sample of 166 publicly traded, large telecom operators, including tower companies, across Africa, Asia, Europe, the Middle East, Latin America, and North America.
[4] These two enablers are part of *Rewired: The McKinsey Guide to Outcompeting in the Age of Digital and AI*, McKinsey's broader, comprehensive digital and AI transformation framework. The second edition of *Rewired* will be available April 14, 2026. For more information, see, Eric Lamarre, Kate Smaje, and Rodney Zemmel, "Rewired to Outcompete," *McKinsey Quarterly*, June 20, 2023.
[5] McKinsey Gen AI Telco CxO Survey, December 2025 (n = 49).
[6] McKinsey Telco CxO Survey, January 2026 (n = 125).

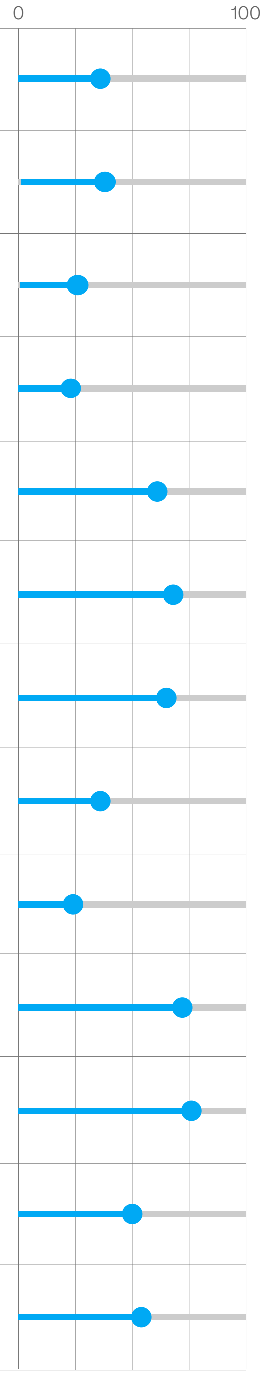
Exhibit 2

The strategic moves telcos can make to reignite growth vary by level of importance and implementation stage.

Telco value creation strategic bets by domain

Domain	Strategic bet	Overview	Relevance,[1] % of respondents
AI-native transformation	AI-driven cost efficiency	Triggering significant cost reduction by harnessing AI to redesign processes across all operations and support functions	58
	Zero-touch operations	Using AI and automation to enable proactive, autonomous, and faster operations with minimal human intervention across the enterprise	34
	AI-powered customer life cycle engagement	Driving highly tailored engagement across the entire customer journey, from brand building and conversion to retention and upsell	46
	AI-driven value proposition	Delivering personalized experiences through intelligent, value-added services, such as vendor-agnostic AI assistants	48
Growth beyond the core	Vertical ecosystem expansion	Expanding beyond core connectivity by bundling standard telecom services with adjacent offerings (eg, security, entertainment, insurance, energy)	33
	ICT[3] products and services	Becoming trusted digital transformation partners with end-to-end enterprise tech solutions that span cloud, cybersecurity, and IT managed services	30
	Network APIs	Fostering the development of new 5G-related products and services by working together to roll out standardized, global API platforms	18
	Private 5G networks	Monetizing 5G infrastructure through vertical-specific enterprise offerings in such areas as manufacturing, logistics, and public infrastructure	26
Infrastructure innovation	AI infrastructure	Positioning telcos as trusted providers of AI-capable infrastructure (including sovereign AI) to support national priorities and enable enterprise-scale AI deployment	55
	Network excellence	Incorporating AI and automation to provide intelligent and resilient networks as network equality reemerges as a key differentiator	40
	Infrastructure and IT modernization	Updating legacy networks and IT systems with cloud-native, real-time architectures to reduce complexity, cut costs, and unlock AI-enabled service innovation at scale	46
Market shaping	Delayering	Carving out high-potential telco units into independent entities to enable specialized governance, faster scaling, and stronger performance	6
	Strategic M&A	Focusing on industry consolidation to benefit from enhanced economies of scale and value proposition differentiation	58

Implementation is ongoing or fully live,[2] % of respondents



[1]Question: Which of the strategic bets listed are most relevant to you (please select the 5 most relevant)?
[2]Question: How far is your company in the implementation journey for your selected 5 most relevant bets? Please indicate level and status of implementation. Implementation figures represent the share of respondents who identified the given strategic bet as one of their 5 most relevant and said their implementations are either "ongoing" or "fully live."
[3]Information and communications technology.
Source: McKinsey Telco CxO Survey, Jan 2026 (n = 125)

McKinsey & Company

Overcoming such risks and succeeding at an AI-native transformation will, above all, require telcos to embrace foundational change in the following areas:

— **AI-driven cost efficiency.** Arguably one of the most foundational elements of an AI-native transformation is to drive significant cost reduction by harnessing AI to redesign processes across all operations and support functions. It makes sense then that this area attracted the most consensus in our most recent telco leader survey, with close to 60 percent of executives choosing it as one of their top strategic priorities.[7] Telstra, the leading Australian telco, achieved AU $122 million in cost savings in just the first year of its program to embed AI across all its corporate processes and business workflows. As part of what it dubbed Project T25, Telstra shifted from relying on a roster of 18 different tech vendors to establishing strategic joint ventures with just two, improving capability integration and governance in the process. Another telco developed a gen-AI powered copilot for its financial planning and analytics team that could handle much of the previously manual and time-consuming data collection and research activities, allowing employees to focus on higher value, more strategic analyses, and saving several millions in operating expenses.

— **Zero-touch operations.** Using AI and automation to minimize human intervention is another key piece of the puzzle. Reaching this goal of proactive, autonomous operations has the potential to radically reshape back-office and support functions, but it is no small feat. It requires, among many key enablers, cross-domain data flows that seamlessly link IT, network, and operations with real-time, cross-domain data flows. BT's procurement unit (BT Sourced) uses AI sourcing and analytics tools to automate and quicken tendering, supplier onboarding, and spend analysis, improving decision-making in the process.

— **AI-powered customer life cycle engagement.** Providing tailored engagement for both B2C and B2B customers is increasingly table stakes for any successful telco, and leveraging AI is essential to reaching that goal. If telcos can incorporate the technology into every stage of the customer journey, from brand building and conversion to retention and upsell, they can spur sizeable improvements in revenue (5 to 8 percent), upsell (50 percent), and churn reduction (30 percent). One telco call center struggling with periods of undercapacity and overcapacity, for example, was able to shave its customer wait times from 3 minutes to 30 seconds with no budget increase by using AI to better predict call traffic patterns and adjust staffing as a result. Another telco has used gen AI tools to cut by 70 percent the time it takes to develop and activate a new personalized marketing campaign or product offering.

— **AI-driven value proposition.** A key strategic bet that forms the core of an AI-native transformation is tapping the technology to provide hyper-personalized experiences and services. Such innovative offerings, which in some cases have shown the potential to increase B2C ARPU by as much as 5 to 10 percent, could range from vendor-agnostic AI assistants that help users deal with fragmented LLMs, to localized edge hosting of LLMs closer to the end user. Several telcos have already gone down this route: SK Telekom, for instance, rolled out an AI-powered personal assistant as an integrated part of its mobile plans that helps in such areas as shopping and travel.

This type of bold, multifaceted effort depends on several factors that can make the difference between realizing ambitions and encountering significant setbacks. These key enablers include:

— **Change management.** AI implementation ultimately comes down to helping a group of people, often after years of ingrained routines, to fundamentally reshape how they work day to day. That seismic shift requires time, committed and visible leadership, a comprehensive redesign of core processes and incentives, and a significant investment in upskilling. In a recent gen AI–focused survey of telco top executives, more than

[7] McKinsey Telco CxO Survey, January 2026 (n = 125).

three-quarters identified slow adoption because of weak change management as the biggest obstacle to scaling AI impact.[8]

— **AI-ready systems.** These include systems that integrate and structure data from customer journeys in real time. They also include modular gen AI tools for content creation and targeting, and cross-functional, agile teams to enable personalization sprints and rapidly roll out and refine new AI campaigns.

— **Cross-functional product squads.** These include teams from the product, engineering, and go-to-market functions, and edge-ready infrastructure that can host LLMs locally, securely, and cost-effectively.

— **Ability to form, manage, and scale strategic partnerships.** These need to be made with a broader set of players complementing telcos in the AI and tech ecosystem, including hyperscalers, system integrators, specialist solution players, digital infrastructure specialists, and impact partners.

In those markets where telcos happen to be part of larger diversified conglomerates (such as those in many parts of Asia), building these enablers often presents platforming opportunities to commercialize capabilities both internally and externally.

Growth beyond the core

As traditional connectivity markets mature and competitive intensity continues to rise, telcos face mounting pressure to unlock new sources of sustainable growth. Even data, once the industry's primary value creation engine, is increasingly commoditizing. Amid this challenging environment, telcos seeking a new commercial growth path are already testing and scaling new business models that extend well beyond core connectivity.

Telcos happen to be well-positioned for this pursuit. The sector's position at the center of the digital economy affords it real-time access to data generated across networks, devices, locations, and usage patterns, and at a scale and granularity matched by few other sectors. This unique data advantage, combined with trusted customer relationships and ubiquitous infrastructure, enables telcos to shift from selling discrete products to delivering integrated, value-added services.

By responsibly and securely leveraging network and customer insights, operators can create differentiated, customized offerings in both B2C and B2B markets. These range from personalized consumer experiences (including energy, content, insurance, security, omnichannel retail and other offerings) built around a vertical ecosystem to industryspecific enterprise solutions that enhance security, efficiency, and performance. Leading players are already translating this shift into action by expanding into adjacent digital verticals, end-to-end ICT services, network APIs, and private 5G networks. In the process, they are converting experimentation into scalable growth, unlocking new revenue pools, and building a more resilient, long-term value creation model.

— **Vertical ecosystems expansion.** Expanding beyond core connectivity to offer adjacent services that unlock new revenue pools is a critical component for telcos looking to drive a new era of healthy growth. These offerings could revolve around everything from security and entertainment to insurance, energy, and fintech, tapping the industry's reach and infrastructure to dramatically expand the scope of customer relationships. Forward-thinking telcos are already leading the way in this area, shifting from a mindset of selling products into an ecosystem of varied services that can widen the value proposition. Canada's Telus now gets a quarter of its revenues from B2B digital verticals for healthcare, agriculture, digital process and back office solutions, part of a diversification strategy that has helped the company's TSR outperform the country's broader telco sector.

[8] McKinsey Gen AI Telco CxO Survey, December 2025 (n = 49).

Several European telcos are adding additional services to their portfolio. Spanish operator MasOrange, for instance, has focused on expanding into adjacent categories such as energy, insurance, security, and health/telemedicine to serve its customer base with value-added services and to help increase ARPU and reduce churn, all while reinforcing its core connectivity value proposition. By leveraging a dynamic, TechCo operating model of independent units and agile platforms and scaling through partnerships and other flexible collaboration models, MasOrange has built a number of growing new businesses. These have included an app-controlled electricity-and-gas service, which it sold last year for 90 million euros; an insurance business (mobile, home, health, and payment protection) with a cumulative ambition of more than 7.5 million policies and over €1.5 billion in premium volume; a cobranded home security business (with ADT) that has reduced churn; and B2B focused offerings, including a cybersecurity and IoT offering that are supporting ARPU uplift and profitability through higher-value enterprise services. Italy's WindTre has enjoyed similar initial success by rolling out insurance and energy products, using upskilling and realigned incentives to enable its store clerks to help drive interest and sales of a significant volume of policies or agreements.

— **ICT products and services.** As more enterprises seek out third-party support in optimizing their rapidly evolving patchwork of technologies and systems, telcos have a lucrative opportunity to play a bigger role in digital transformation. By expanding from core B2B offerings into end-to-end solutions that span cloud, cybersecurity, and IT managed services, the industry can grow B2B ICT into a dynamic business that accounts for as much as 20 percent of total telco revenue. One European telco used that model to grow its share of the domestic cloud market, upgrading its infrastructure, upskilling employees, and striking key partnership on the way to more than doubling its cloud-related revenues and earnings in four years.

— **Network API.** The effective monetization of 5G has been a significant challenge for telcos since the wireless technology's debut several years ago. By working more closely together to roll out standardized, global API platforms, the industry can spur the creation of new 5G-related products and services that leverage features such as speed on demand, low-latency connections, speed tiering, and edge compute discovery. A trio of Brazilian operators recently partnered on an API initiative to help financial institutions improve digital security and antifinancial fraud efforts. Working with the industry's GSMA Open Gateway Initiative, TIM, Claro, and Vivo developed three customized APIs focused on continuous mobile number verification, SIM card swap checking (to prevent account takeover attacks), and instant device location and validation. The services are already reaching close to 150 million customers and early results have been promising, not only generating $9 million in incremental revenue from the SimSwap API alone but also delivering tangible consumer benefits.

— **Private 5G networks.** Another vehicle for increasing telcos' return on 5G is customized, vertical-specific enterprise architecture to enable advanced use cases. Many operators (including T-Mobile, O2, Vodafone, and KPN) are already launching such solutions, from smart hospitals or connected transportation to augmented retail experiences. By targeting areas such as logistics, manufacturing, and the public sector, such specialized networks could fuel incremental 5G monetization opportunities over time.

Achieving true commercial excellence requires a significant appetite for change as well as a number of specific enabling factors, including the following:

— embracing a new type of partner ecosystem and striking deals with vertical leaders to help drive the development of new offerings beyond the core; create an internal integration platform to ease the process of bundling, activation, and management of non-core services

— laying the operational and technical foundation to support ICT-driven partnerships, especially alliances with hyperscalers and provides of cybersecurity and software; that means developing a modular platform to manage full-stack solutions as well as a specialized salesforce skilled in catering to demanding enterprises and the integrated delivery and support systems that those customers expect

- establishing common application layer standards to minimize developer friction, and integrating network APIs with hyperscaler platforms to help expand the reach of new offerings

- obtaining reliable, enterprise-grade wireless spectrum access to back up private network service guarantees and design scalable solutions, often in partnership with industrial players or hyperscalers, and which combine hardware, software, and services targeting use cases in specific sectors or verticals

Infrastructure innovation

As AI reshapes economic competitiveness, digital infrastructure is becoming an even more valuable strategic asset. For telecom operators, this shift opens a path to move beyond the role of capital-intensive utility toward a more differentiated position in the digital economy. The combination of rising AI workloads, growing concerns around network resilience and security, and renewed emphasis on performance at the network edge is creating a set of concrete, near-term opportunities for telcos, particularly in areas where their assets, capabilities, and trusted role intersect with emerging demand.

Governments and enterprises are increasingly looking for secure, local, and reliable infrastructure to support AI-driven use cases, while at the same time expecting networks to deliver higher performance, greater reliability, and lower unit costs. Capturing this opportunity will require operators to think holistically about infrastructure, both their own internal systems and the wider external environment in which they compete. They'll have to focus not just on what they build, but how they operate it, modernize it, and turn it into a true source of differentiation and returns. The key areas include:

- **AI Infrastructure:** With AI fast becoming a private- and public-sector imperative, telcos have an opportunity to play a valuable role as a reliable provider of the enabling infrastructure, from cloud and edge to secure connectivity. This is particularly the case in the growing number of markets where governments (and the businesses they regulate) are looking to increase their technological resilience, or sovereignty, often backed by national AI or data center strategies that include mandates and sizeable public funding. Industry leaders recognize this opportunity as one of their most promising, with 55 percent of those we surveyed identifying AI infrastructure as a top strategic bet.[9] Numerous telcos are already actively exploring this potentially fertile ground, lured by the prospect of increasing their return on invested capital by as much as 50 percent and seizing a new chance to compete with established hyperscalers. In practice, telcos' advantage relative to hyperscalers lies less in scale of compute or model development and more in trusted local presence, regulatory alignment, and control of national infrastructure assets such as edge locations, secure connectivity, and data-center footprint. As a result, sovereign AI initiatives are most often pursued through partnership-based models, with telcos providing the infrastructure and compliance layer, and hyperscalers contributing cloud platforms, GPU-based compute, and foundation models. From Spain (Telefonica) and the UK (BT) to Saudi Arabia (STC) and Southeast Asia (Singtel), telcos are planning or launching AI-ready networks, data center platforms, and other cloud or edge-based products to serve government and enterprise customers seeking out secure, locally-housed (aka sovereign) AI infrastructure with models that are trained by local teams on local datasets.

- **Network excellence:** As AI workloads and digital services move closer to the edge, network quality is reemerging as a major differentiator. By incorporating AI and automation across the entire life cycle—from planning, construction, and deployment to fault management and inventory—operators can provide intelligent, high-performing, and resilient networks that be a key driver of value creation rather than just a technical asset. In certain instances, such an approach has shown the potential to reduce network operating expenses by up to 30 percent and total capital expenses by as much as 15 percent. For instance,

[9] McKinsey Telco CxO Survey, January 2026 (n = 125).

one North American telco used AI tools to improve the safety, efficiency, and customer satisfaction of its frontline technician force by unlocking 10 to 20 percent additional capacity and enabling up to a 30 percent reduction in overall field-force requirements.

— **Infrastructure and IT modernization:** Monetizing the past decade-plus of capital investments into 5G, fiber, and other technologies is as much about updating operators' legacy network infrastructure and IT systems as it is about devising new products and services. If operators can't successfully upgrade their own internal technology capabilities, they are much less likely to be able to effectively take advantage of cutting-edge innovations like AI, or to market new offerings tied to the most advanced network solutions. By embracing fiber rollouts, automation, and cloud-ready architecture that is both modular and API driven, forward-looking telcos can lay a foundation that could reduce IT costs by as much as 30 percent and boost ARPU by 10 to 15 percent; previous McKinsey research has shown that operators in the top quartile of IT maturity, on average, grow revenues three percentage points more year over year compared with their peers, and they do so at a higher rate of profitability (roughly 15.5 percent net operating profit after tax versus 14 percent).[10] Several European operators (such as Norway's Telenor and Spain's Telefonica) have already seen encouraging results from their decommissioning of copper networks, including double-digit decreases in energy consumption and fault rates, and 7 to 14 percent increases in customer satisfaction.

Successfully leveraging infrastructure as a value creation opportunity requires a number of key capabilities and approaches, including the following:

— **Compute power.** To seize the AI infrastructure opportunity, telcos will need to ensure they have sufficient access to compute capacity, either through GPUaaS or necloud providers, or through partnerships with hyperscalers. Their infrastructure will also have to be both AI- and edge-ready, able to satisfy power, cooling, latency, and location requirements of such advanced workloads. Maximizing the shared hardware will also be critical, such that it can simultaneously run more traditional network functions along with the AI inference demands via AI-RAN technology.

— **Unified data flow.** Optimizing the telco network can't happen without a comprehensive, unified data flow to fuel automated analytics. To achieve that level of transparency and understanding, operators must aggregate real-time telemetry from all the different network domains (RAN, core, transport). An autonomous, AI-based platform is equally important, enabling operations including planning, fault detection, incident resolution, and energy optimization with little to no human involvement. At the same time, a fully AI-literate network workforce of engineers and planners is essential for integrating AI tools into operations at the outset and laying the foundation for increased automation.

— **Planning and oversight.** A successful infrastructure and IT modernization campaign takes more than just an embrace of more advanced technologies. On both internal and external fronts, careful planning and oversight is needed for such significant changes. That means developing thorough, end-to-end technical and operational plans for the organization's journey from legacy to modern systems, as well as proactive, comprehensive management of customer migrations, including targeted outreach, guided onboarding, and incentive models.

Market shaping

Over the past two decades, telecom operators have faced a steadily tightening set of constraints. Capital intensity remains high, pricing pressure is persistent, and returns have struggled to keep pace with those in adjacent digital sectors. Investors, meanwhile, have become less bought in to complex, vertically integrated models that can

[10] "Transforming telecom tech: How IT excellence drives innovation and cost efficiency," McKinsey, February 28, 2025.

be difficult to manage, slow to adapt, and heavily influenced by evolving regulatory environments. All of this is playing out as customers' expectations for reliability, speed, and seamless digital experiences continue to rise.

Against this backdrop, many operators are recognizing that incremental improvement alone is unlikely to change the trajectory of the increasingly commoditized industry.

Traditional levers—cost discipline, selective network investment, and product refreshes—remain necessary but insufficient to the task at hand. More leadership teams are stepping back to ask more fundamental questions about structure, focus, and scale: which parts of the business truly benefit from being integrated? where can specialization unlock value? and how can capital be deployed more effectively?

Two market-oriented strategic responses are emerging most clearly from this reassessment: delayering and strategic M&A. Each represents a distinct way of addressing structural challenges in the telco value chain, and each has the potential to materially improve performance. At the same time, both require significant organizational change and management attention, making clear strategic intent and disciplined execution essential. Two actions in particular are required:

— **Delayering.** The challenges telcos have faced in the past decade or two, particularly with investors, have convinced some to make a dramatic break from the traditional integrated ownership model. To attempt to improve focus, innovation, and valuations, a small but growing number of operators is choosing to delayer or separate into two or more independent corporate entities, typically one built around network and infrastructure ("NetCo" or "InfraCo") and the other(s) centered on products and services ("ServCo"). While the "NetCo" can take advantage of its new status as a carrier-neutral wholesale access provider, the "ServCo" can further specialize into distinct units based on type of customer, offering, business model, investment horizon, or skills. In many cases, the establishment of a separate NetCo helps to spur consolidation in the telecom infrastructure market and greater economies of scale.

  The rewards of such a radical step can be significant—some operators that have taken it have realized as much as a ten percentage point uplift in EBITDA margins and a ten times increase in enterprise value multiple, but such positive impact is far from guaranteed. Delayering can take as much as two or three years and sizeable amounts of capital to pull off, and some industry observers think such a split can hamper each separate entity's ability to present a seamless, omnichannel customer experience, and hence requires significant change management. Still, several carriers have already experienced positive returns through delayering: one of the earliest such moves about a decade ago, by Czech carrier O2/CETIN, produced a 27 percent total market cap increase just one year after the split. More recently, carriers such as Etisalat (based in the UAE), and MTN (based in South Africa) have found success delayering into four or five different separate units or platforms.

— **Strategic M&A.** The more traditional, market-driven maneuver that telcos have looked to as a possible solution to their recent woes is consolidation. Such an approach can be successful in harnessing economies of scale, value proposition differentiation, and access to new markets, with the potential to achieve synergies amounting to more than 15 percent of combined capital expenditure and 30 percent of operating expenditure (and it can be especially useful in in-country consolidation plays). The 2024 combination of Orange Spain and MasMovil is expected to realize synergies of €490 million after four years, with the new company serving nearly 50 percent of the country's fixed and mobile subscribers and close to a third of the industry revenues; likewise, the 2025 pairing of Vodafone UK and Three UK will create the country's largest mobile operator, serving nearly 40 percent of subscribers as well as telco revenues, with an anticipated £700 million in synergies by the fifth year. More industry leaders are contemplating such moves; close to 60 percent of respondents to our most recent survey in January identified M&A as one of their top five strategic priorities going forward, a level of consensus that was only matched by AI-driven cost efficiency.[11] But combining forces is by no means a panacea for value creation. Unless it's carried out thoughtfully, with significant cost transformation and disciplined, post-merger execution, it can be a damaging waste of time and resources.

Market-driven organizational and strategic moves that generate sustainable, incremental value are easier said than done; reaching that ambitious goal takes many key factors, such as the following:

— **Operational clarity.** Delayering a traditionally integrated operator in such a way that the separated parts are greater than the whole requires a number of key enablers: the clear delineation of the different assets and activities, including the decoupling of underlying IT systems, taking into account value pools, synergies, interdependencies and compliance; as part of this, it's important to carefully draft commercial and service-level agreements between the newly distinct entities, the NetCo and ServCo, to maintain adequate continuity; and finally, the holistic redesign of each entity's operating models and processes to spur focused execution.

— **Clear assets structure.** To increase the odds that strategic M&A will ultimately be viewed positively in hindsight, operators will want to have their assets in question structured with clear financial, legal, and operational separation between business units, which help to boost valuations and simplify deal-making. It's equally important to be organizationally prepared for the aftermath of any transaction; appropriate teams and tools should already be in place to speed post-deal integration and reduce potential disruption.

Finding a place within an AI-driven economy

As telecommunications leaders contemplate their industry's place in the emerging AI-driven economy, they have valid reasons to feel cautiously optimistic about their ability to successfully reinvent themselves for this new era. Recent improvements in financial performance, more disciplined capital allocation, the near completion of costly network buildouts, and customers' evolving expectations have created a valuable opportunity for the sector to move beyond the challenging past two decades into a healthy era of sustainable growth.

There is, however, no single playbook or prescriptive formula for turning that opportunity into significant enterprise value. Market structures, competitive dynamics, starting positions, and policy regimes, among other factors, mean that successful telcos will invariably end up taking distinct paths, and the scale and pace of impact will vary significantly.

What the strongest performing operators have already shown is that they share a common mindset, if not a common strategy. Rather than betting on one narrow initiative, they have made progress across several of the critical broad strategic areas outlined in this article. Some have focused more on embedding AI at the core of their operating models or redefining commercial engagement, others on innovating with infrastructure or actively shaping markets rather than reacting to them. Companies such as MasOrange, Singtel, Indosat, Telus, and Deutsche Telekom illustrate this diversity: each has emphasized a combination of strategic moves and approaches, yet all have shaped transformation with a portfolio of actions rather than a single lever.

[11] McKinsey Telco CxO Survey, January 2026 (n = 125).

Regardless of the specific path, any brand of large-scale telco reinvention depends on a small number of critical enablers. This starts with visible, sustained leadership from the top—setting a clear strategic direction and aligning the organization behind it. Equally important is disciplined investment in data and modular IT, which form the foundation for multiple strategic bets. And finally, it requires an execution-driven culture that favors experimentation and learning by doing, with leaders willing to move from plans to action quickly and make course corrections as warranted.

For telco leaders, the overarching lesson is straightforward. Value creation may not require following a uniform template, but it does demand focus, coherence, and the willingness to act across multiple fronts. Telcos that launch a handful of bold, context specific moves have a chance to materially improve returns and regain investor confidence; in the process, they can capture more of the value their networks enable and reestablish themselves as growth platforms at the heart of the AI economy. Those that hesitate, or pursue siloed initiatives without a broader strategy, risk repeating their recent, growth-challenged history, relegated to low-return, utility-like roles as the pace of dynamic, AI-enabled change accelerates across every industry.

Andrea Travasoni is a senior partner in McKinsey's Milan office, where **Marco Labianca** is an associate partner. **Dev Patel** is a senior partner in the Chicago office, **Theodor Vendrig** is a partner in the Oslo office, and **Tiago Palma** is a partner in the Madrid office. **Alima Ziman** is an associate partner in the Almaty office.

The authors wish to thank Azamat Bakytbekov and Blanca Alcala for their contributions to this article.

This article was edited by Daniel Eisenberg, an executive editor in the New York office.

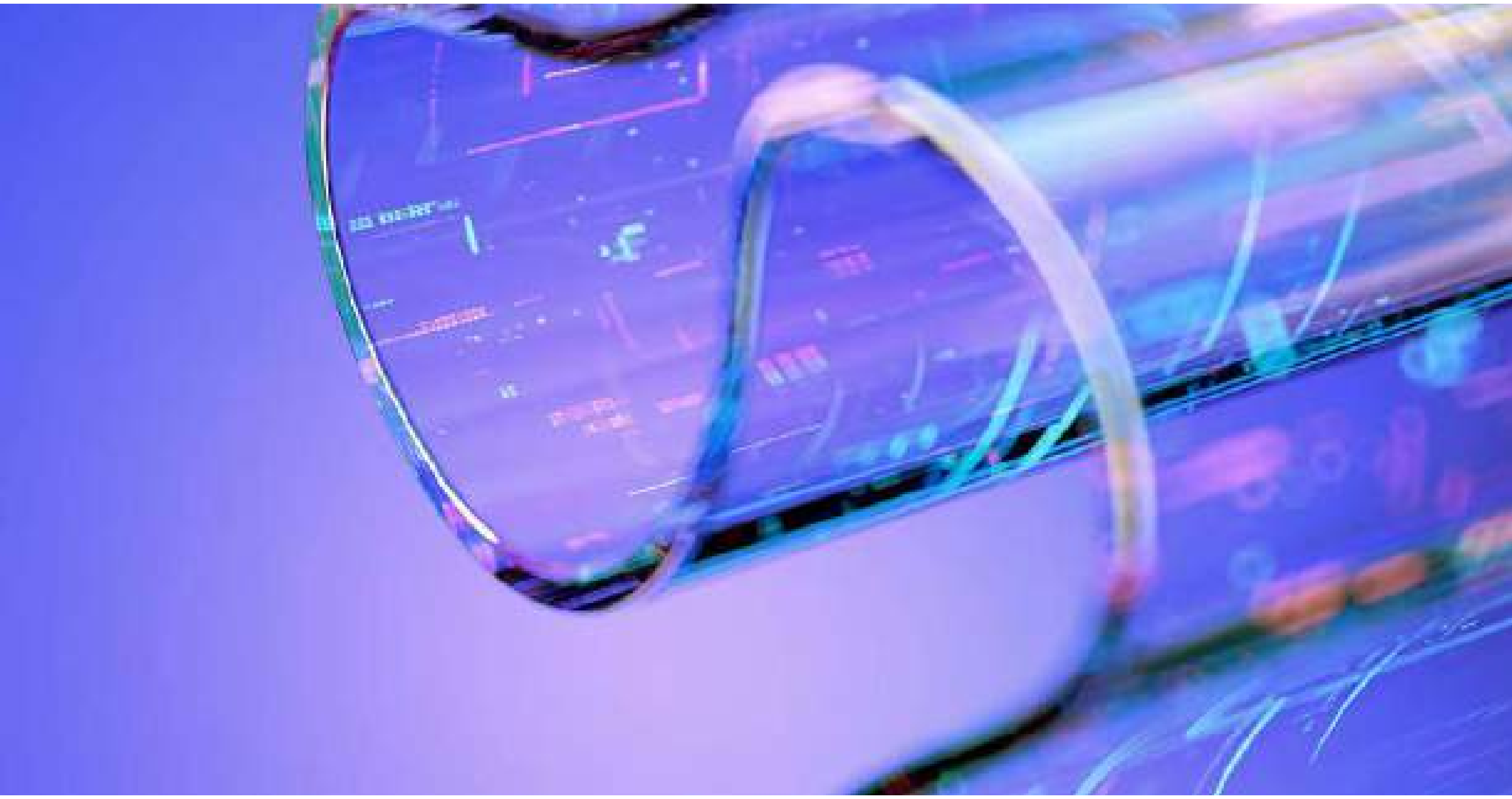
McKinsey & Company

Technology, Media & Telecommunications Practice

Telcos' AI inflection point: What leaders do to capture value

Agentic AI offers telcos a potential opportunity to reorient their recent trajectory. The question is whether they can redesign their organizations enough to make the most of it.

This article is a collaborative effort by Benjamim Vieira, Guilherme Cruz, Ignacio Ferrero, and Tomás Lajous, with Borja Belda, Daniela Mendoza, and Manuel Palacios, representing views from McKinsey's Technology, Media & Telecommunications Practice and QuantumBlack, AI by McKinsey.



March 2026

For more than a decade, telecom operators (telcos) have pursued successive waves of digital transformation to offset slowing growth, rising capital intensity, and limited product differentiation. Most delivered incremental gains, but not the type of overarching structural change that the industry requires at a time when the pace of innovation, competition, and challenging secular issues show no signs of slowing down.

Agentic AI may be the first technology capable of altering that trajectory. Unlike earlier automation tools that improved individual tasks, AI agents have the potential to reshape entire workflows, even helping make operational decisions and coordinate work across functions. This shifts AI from being a powerful productivity tool to a full-scale execution layer that can fundamentally change how telcos design operations, deploy capital, and create value.

The question for telco leaders is no longer so much about where exactly to apply AI, but how to redesign the enterprise to operate with and optimize it.

Over the past few years, much of the sector has learned the hard way that there is no quick way to make that happen. But even as most telcos have failed to gain real value from their growing AI investments, rolling out too many fragmented or duplicative AI use cases and pilots, a smaller number of forward-thinking players are showing how the rapid emergence of agentic AI can be a paradigm shift, offering a rare opportunity to reorient the industry's prospects.

The telco leaders in harnessing AI to generate significant impact have already demonstrated it requires a sustained, long-term approach that includes CEO-led sponsorship, disciplined organizational transformation and change management, and a clear focus on end-to-end processes rather than individual tasks. This article, based on a recent survey[1] of top telco executives and our experience working with clients on AI deployments, provides industry peers with a road map to similar success, with the possibility to increase both ROIC and EBITDA margins by as much as ten percentage points within five years.[2] It lays out several key elements of a coherent agentic AI operating model that is necessary for telcos to capture that scale of new value and growth over the coming years and highlights the current state of telco AI adoption and impact for both leading, early-moving telcos and the rest of the industry.

The state of AI for telcos

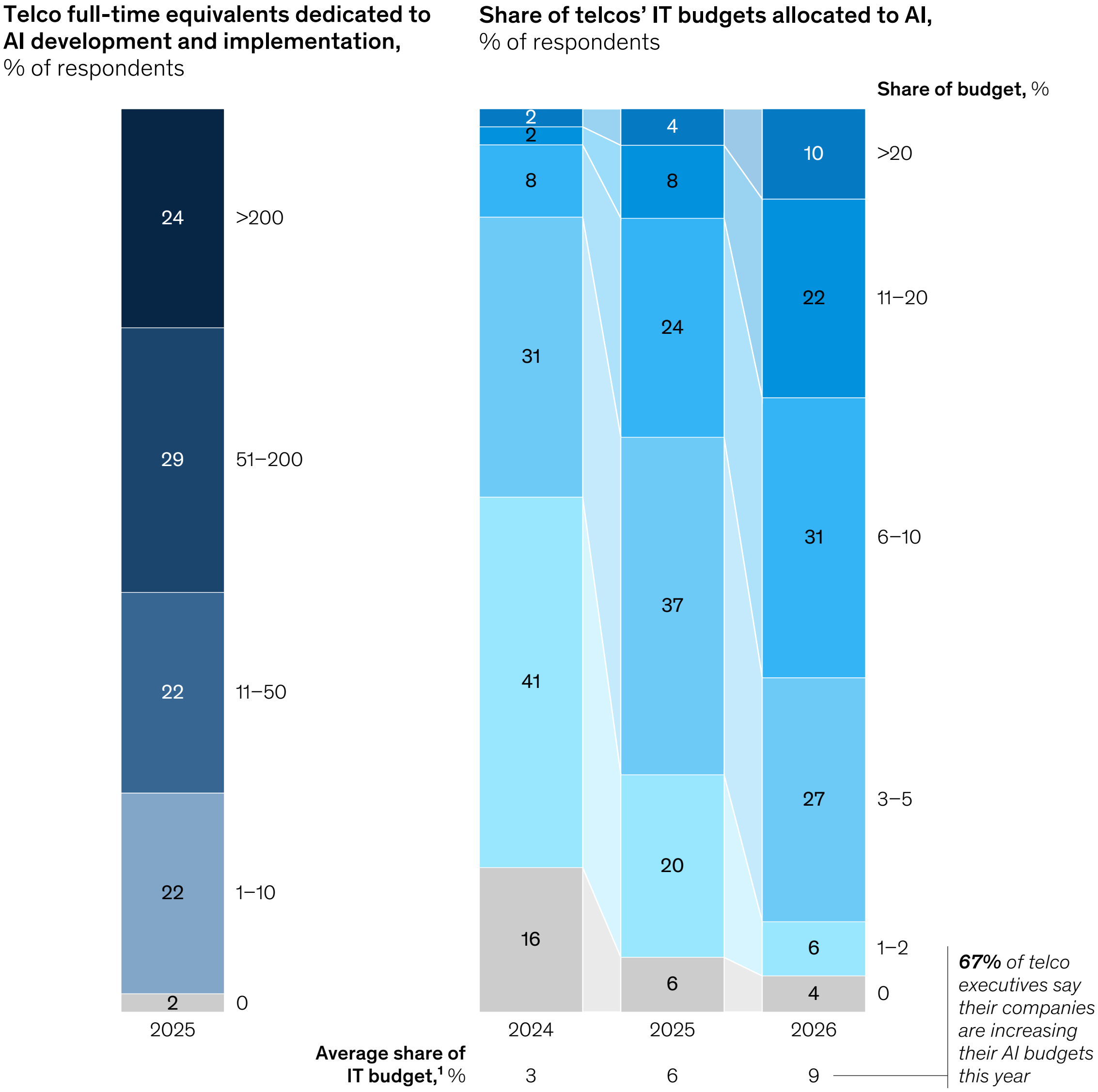
While telco executives increasingly recognize the importance of AI to their future, relatively few have seen sizable impact from their embrace of the technology. And that embrace has grown significantly in recent years.

Just over half of the telcos we surveyed have at least 50 full time equivalents (FTEs) dedicated to AI, and fully two-thirds expect to increase the AI portion of their IT budgets this year; roughly half of that segment anticipates devoting more than 10 percent of that spending to AI, with the average industry allocation growing to 9 percent (Exhibit 1).

[1] The McKinsey Telco CxO Gen AI Survey was conducted in December 2025, with 49 respondents representing telecom operators from across North America, South America, Europe, Africa, the Middle East, and Asia.
[2] Depending on baseline performance and implementation scope.

Exhibit 1

Telcos are allocating more personnel and boosting IT budgets for AI to speed up deployment and begin to capture significant impact.



Note: Figures may not sum to 100%, because of rounding.
[1]To calculate these weighted averages, we took the midpoint of the budget % ranges (0, 1–2, 3–5, 6–10, 11–20, and >20), multiplied each by the corresponding % of respondents in the category, and summed the multiplied numbers.
Source: McKinsey Telco CxO Gen AI Survey, Dec 2025, n = 49

McKinsey & Company

Though virtually all telcos we surveyed are piloting AI, only 57 percent report scaling use cases across multiple domains, led by customer service and networks, virtually unchanged from almost a year ago. Far fewer—just 16 percent—characterize AI as the "new normal" across their organization. At the same time, only 51 percent view AI as a "blockbuster technology" that will fully transform the industry, compared to 61 percent in the prior survey.[3] Most of that shift in sentiment was toward a more sober view that the technology is "something relevant that will generate impact in the industry," which is now shared by 47 percent of respondents, up from 36 percent the previous year.

This more realistic, but still bullish, mindset, which is focused on tangible results rather than the technology's seemingly limitless capabilities, reflects the industry's initial experiences with incorporating AI. While telcos have seen meaningful cost savings in certain areas, most have yet to realize widespread productivity gains. Only 12 percent of respondents report having already captured sizable impact, with most others caught in what we call "the money step," where investments have yet to yield any real balance sheet benefits despite spurring other impacts. A majority of industry leaders, however, remain cautiously optimistic that those benefits will materialize toward the end of the decade; they view the most fertile areas as customer support and network functions, followed closely by IT, where most telcos expect to enjoy at least 10 percent cost savings over the next one to two years and close to 30 percent by 2030 (Exhibit 2).

Despite that relatively rosy long-term view, telco executives acknowledge the substantial challenges they face in scaling the impact of AI, such as moving beyond limited productivity gains from isolated use cases. More than three-quarters of those surveyed agree that immature operating models, data limitations, and lagging adoption due to ineffective change management are the top issues. When it comes to translating impact into measurable, substantial value, there is less consensus (Exhibit 3). A little over half of respondents identify employee or team adoption as the biggest hurdles, while a third point to the difficulty of implementing process changes at scale, and a quarter noted inflexible budgeting as the key obstacle.

Telcos' AI inflection point

A critical reason for telcos' relative bullishness about overcoming those challenges is the recent emergence of agentic AI. Unlike previous AI tools, which have largely enabled impressive but still incremental improvements, agents represent the possibility of an automation-driven inflection point for industry. By 2030, according to the McKinsey Global Institute, AI automation could create as much $16 billion in new economic value for the telecom sector in the United States alone; and redesigning workflows around human-agent collaboration, whether for sales reps, equipment installers, or electronic engineers, may be key to capturing it. Roughly half of the industry's workforce may have the potential to be remade with collaborative human-agent automation, more than double most other sectors. That prospect of using agents to redefine (or re-create) workflows means the industry may be able to move beyond the same outdated operational processes and models that have proved to be stumbling blocks to unlocking new value.

While much of the industry discussion around agents a year ago was largely theoretical, more than half of telco survey respondents now say they are deploying agentic use cases across at least one function. Several early movers are already showing how the technology can be leveraged to capture sizable value, as the following examples across different functions illustrate.

Customer service

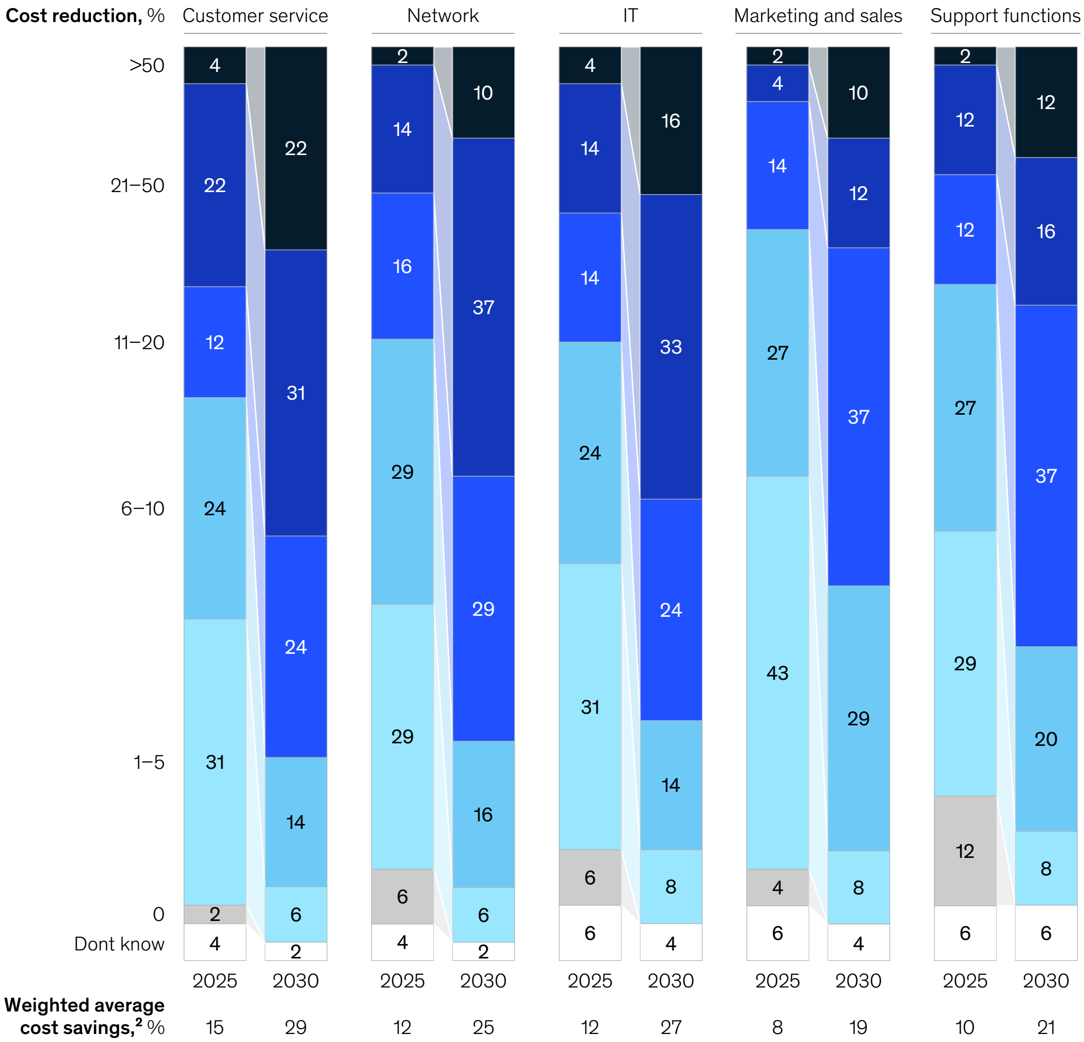
Across European telecom operators, rising customer service volumes and high call center costs have made improving efficiency in customer care a strategic priority.

[3] McKinsey Telco CxO Gen AI Survey, February 2025 (n = 52).

Exhibit 2

Telcos expect AI to help generate much greater cost savings by 2030, particularly in customer service, network, and IT domains.

Telco estimated cost reduction attributed to AI, by domain,[1] % of respondents



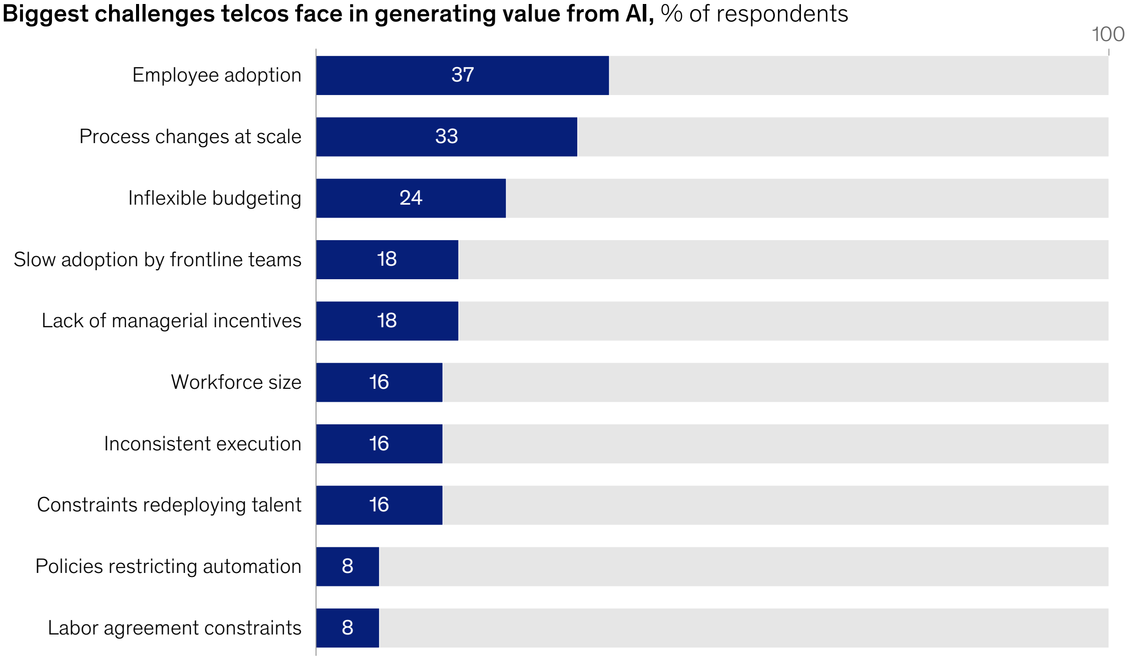
Note: Figures may not sum to 100%, because of rounding.
[1]Question: What is the expected impact (% cost reduction) attributed to AI (including gen AI and agentic AI) in different domains in [year]?
[2]To calculate these weighted averages, we took the midpoint of the cost % ranges (0, 1–5, 6–10, 11–20, 21–50, and >50), multiplied each by the corresponding percentage of respondents in the category, and summed the multiplied numbers.
Source: McKinsey Telco CxO Gen AI Survey, Dec 2025, n = 49

McKinsey & Company

Exhibit 3

Employee adoption and deploying process changes at scale are the biggest obstacles telcos say they face in generating value from AI.

Biggest challenges telcos face in generating value from AI, % of respondents



[1]Question: What are the biggest challenges your company faces in materializing value from AI? Please identify a maximum of 3.
Source: McKinsey Telco CxO Gen AI Survey, Dec 2025, n = 49

McKinsey & Company

At KPN, this has translated into developing voice-to-voice agentic AI to handle customer care interactions end to end. To build this voice-to-voice agent, the team took a three-step approach. They first used call volumes and transcripts to identify and analyze interactive voice response end points; next, they drilled down into "subintents" (such as fixed line no dial tone, defective modem or router replacement, or mobile voice/SMS roaming issues) and customer journeys to assess feasibility and agentification complexity; and then they built use cases based on impact value and development complexity and feasibility.

Leveraging this analysis, the operator prioritized six of the 18 use cases, including invoice agent, generic Q&A, and authentication for initial piloting. As a result, the operator is aiming to reduce the average handling time of calls requiring human experts, significantly reducing the run rate in overall call center spending within a year.

Network

For two leading network operators serving savvy, demanding customers in very different regions—MASORANGE in Europe and NTT DOCOMO in Asia—a deep understanding of network customer experience (CX) recently surfaced as a key competitive differentiator. This happened at a time when capital expenditure (capex) levels heightened the need for more-efficient investment decisions.

In response, each operator has used an AI for CX index, creating a daily metric that correlates the operator's network performance with customer satisfaction and risk. As a result, the operators were able to identify customers unsatisfied with the network, determine the primary cause of their subpar experience, and link those insights to capital interventions (such as site upgrades) or other new customer offerings to help improve customer retention. For example, sites with low CX index and high revenue were identified as priority sites for the operators to invest capex.

To optimize the opportunity, both operators redesigned capex planning workflows and led efforts to upskill network teams on this AI application and ensure sustained adoption of the new decisioning approach across areas and branches. With this approach in place, the operators either assessed CX along the most heavily used commuting lines (affecting an estimated 5 million to 10 million users, for example) and reprioritized their 2025–26 capex plan, or identified areas where customers were the most sensitive to network performance fluctuations and drove actions to improve the network and help enhance the loyalty and experience of those customers. (For more on NTT DOCOMO's use of the AI for CX metric, see sidebar, "NTT DOCOMO: Using AI to keep communities better connected.")

Case study

NTT DOCOMO: Using AI to keep communities better connected

In Japan's hyper-connected cities, network quality isn't an abstract metric. It's felt in fleeting moments on packed commuter trains, in dense urban corridors, during short, repeated interactions like checking a map between stations, refreshing a news app on a crowded platform, or resending a text message that lags just long enough to be noticed.

For NTT DOCOMO, those moments were becoming a warning sign.

Even as the company continued to invest heavily in network infrastructure, customer experience in some high-traffic areas was deteriorating. Traditional engineering metrics could confirm that the network was busy—but they couldn't explain why customers were increasingly frustrated.

NTT DOCOMO recognized that closing this gap required a fundamentally new way of approaching the network: one that connected technical performance to lived customer experience—and linked both directly to investment decisions. The goal wasn't incremental optimization. It was a systematic, experience-led approach to how the network is planned, operated, and improved.

Experiencing the network as a customer

To make that shift, NTT DOCOMO worked with McKinsey as its thought partner to co-develop the Customer Network Experience (CNX) index—an AI-driven metric designed to quantify network customer experience in a way that is actionable across the organization.

Built using more than 400 terabytes of network-side data, the CNX index applies advanced analytics and machine-learning models to capture how users really experience the network across locations and time periods, rather than how the network performs in theory.

The methodology was tailored to Japan's unique usage patterns. In particular, "commuting-line" features were developed to reflect the impact of repeated, short-duration connections in dense urban transit corridors—a critical driver of customer perception that conventional metrics, such as download speed, had failed to reflect.

"As networks become more complex, improving customer experience requires more than adding infrastructure," says Yoshio Umezawa, VP of R&D Innovation at NTT DOCOMO. "The CNX Index allows us to connect network performance directly to how the network is actually experienced, giving us a common, objective foundation for decision-making across the organization. AI enables us to move from reactive responses to a more systematic, experience-led approach to network management."

Beyond the analytical solution: driving real change
Crucially, CNX was never treated as a standalone analytical output. NTT DOCOMO embedded the index directly into core workflows, including network planning, preventive maintenance, and capital expenditure (CAPEX) decision-making. CNX insights were explicitly linked to site revenue, allowing teams to identify where poor experience overlapped with high economic importance. Planners could then compare interventions against experience and site revenue in a consistent, objective way.

To ensure long-term ownership, McKinsey also supported NTT DOCOMO in building internal capabilities through an AI academy that trained engineers and planners on both the fundamentals of analytics and the practical use of CNX. The two organizations also co-developed a comprehensive change-management program to drive adoption across headquarters and regions—aligning teams around shared definitions, dashboards, and decision routines.

Turning insight into better moments and better decisions
The new perspective quickly surfaced insights that had previously been hard to identify. In Tokyo and the surrounding regions, CNX analysis revealed that central urban areas often delivered lower customer experience than surrounding zones, despite having a high number of 5G sites. The issue wasn't simply coverage but had multiple causes identified by CNX. By linking the experience index to site revenue, NTT DOCOMO could pinpoint critical locations where targeted intervention would deliver disproportionate customer impact.

The result: Approximately 10 to 30 percent of planned CAPEX interventions have the potential to be reprioritized, allowing capital to be redirected toward higher-impact uses without increasing overall investment levels. Today, CNX is used by teams across headquarters and regions to support more consistent prioritization, tighter alignment between central planning and local execution, and more disciplined monitoring of outcomes.

Beyond efficiency gains, CNX gives NTT DOCOMO something more durable: a foundation for managing the network around how customers actually experience it. By embedding AI-driven insight into everyday decision making, the company is showing how advanced analytics can move from analysis to action, even as network complexity continues to rise.

"What makes this transformation powerful isn't the technology alone—it's the shift in mindset," says Takamichi Watase, a McKinsey partner based in Tokyo. "By anchoring network decisions in real customer experience, NTT DOCOMO has created a common language that aligns engineers, planners, and executives around what truly matters: how the network feels to the people who use it, every day."

Marketing and sales
Most telecom operators now use AI-augmented agencies and in-house tools to accelerate core marketing activities such as copy generation, promotion design, and campaign execution. At the same time, some operators are extending AI into frontline sales and contact center operations to unlock additional value from existing customer interactions.

The business process outsourcing (BPO) unit of leading South American telco Entel, for instance, faced a common challenge in its contact center operations: Changing customer expectations and engagement patterns were driving a decline in outbound sales, while the company continued to handle hundreds of thousands of inbound service calls each month. Yet fewer than 1 percent of these calls was systematically analyzed, limiting visibility into sales opportunities, agent performance, and evolving customer needs. This gap represented a significant value opportunity.

To unlock this value, Entel Connect, in partnership with a hyperscaler, implemented an agentic AI tool that processes and analyzes 100 percent of inbound calls daily. The solution decomposes each conversation to identify leads, flag missed opportunities, and track customer sentiment while also generating personalized coaching insights for every agent and actionable dashboards for supervisors. The tool was designed to operate at scale, processing large volumes of calls overnight at very low unit cost, while continuously refining recommendations based on real operational feedback. Over time, the contact center evolved from a reactive service operation into a proactive, data-driven sales and learning engine. Within ten weeks, inbound sales increased by 40 percent, driven by a doubling of sales attempts during service calls, with no negative impact on customer satisfaction.

Support functions

A South American operator whose finance team was spending a disproportionate amount of time on repetitive, manual tasks decided to build an agentic solution to tackle the long turnarounds of basic analyses and manual data preparation and free up capacity for higher-value work.

The team defined more than 200 recurring question archetypes across operating expenditures, capex, and profit and loss, had the solution ingest more than 15 structured and unstructured data sources, and deployed a multiagent architecture (including intent recognition, entity extraction, query generation, visualization, and guardrails) to deliver accurate, explainable answers to questions of what, why, and how. This solution automated roughly 90 percent of management report findings and produced more than 95 percent accuracy on core use cases, allowing the teams to turn their attention from manual analysis toward faster, informed decision-making.

Building a winning agentic telco

A year ago, in "Scaling the AI-native telco," we outlined the key elements for telcos to scale AI efficiently.[4] These include targeting domain- and workflow-specific transformation opportunities, building scalable, modular AI platforms, implementing adequate data foundations, and fueling adoption with change management best practices. While those overarching pillars remain valid, our recent experience in helping telcos deploy AI has highlighted some critical, more granular lessons about what it takes for telcos to build a coherent agentic operating model capable of unlocking sustainable new value and growth.

1. If AI isn't in the budget, it isn't real (like its impact)

One of the primary pitfalls of early AI initiatives across both telcos and other industries has been the failure to explicitly tie them to, and track their progress against, clearly defined, benchmarked goals. Several leading organizations from the telecom sector and beyond are already taking this approach, rigorously reviewing their budgets to pinpoint where AI can deliver measurable impact and tying it to a clear business case. That impact can be significant (Exhibit 4). In this way, companies can prioritize what drives value on both the top and bottom lines rather than pain points or challenges that they think AI could help solve. This approach gives business leaders tangible, measurable targets to achieve and be judged by. As part of this, AI should be a key component of the entire budgeting process; for instance, if a division expects a certain amount of impact from its AI initiatives in the coming year, then the company may consider adjusting budgets accordingly to reinforce accountability for delivery on the goal.

[4] "Scaling the AI-native telco," McKinsey, February 27, 2025.

Exhibit 4

Telcos have significant top- and bottom-line opportunities to use AI to improve EBITDA by up to 20 to 30 percent.

Telco today vs future telco powered by AI/gen AI (~5 year horizon), illustrative

Profit and loss (P&L) domain	What is included	Share of revenue, %: Telco today	Share of revenue, %: Future telco	Impact on P&L, pp[1]	Examples of how AI will drive impact
Revenue		100	103	+2–3	Hyper-personalized digital communications, refined CVM,[2] revamped S2S[3] (gen AI transcripts)
COGS[4] and bad debt	Cost associated with interconnect, roaming, content	32–33	31–32	–2 to –3	Optimize connectivity footprint, predictive bill shock to manage collections up front
Operations Margin[5]		67–68	71–72		
Marketing and sales	Cost from sales channels, marketing, product development for both B2C, B2B	9–10	7–8	–2 to –3	Automated sales channel calls, e-commerce reduction, store footprint optimization
Customer service	Cost from care (ie, sales to support, inbound, digital), customer value management, loyalty, and field operations	2–3	1–2	–1 to –2	Automated care channels, optimized CVM[2] (eg, predictive issue detection and proactive resolution)
SG&A[6]	Cost from HR, finance, legal, and cross-function (eg, supply chain)	6–7	5–6	–1 to –2	Automated contract management, invoicing, supplier searches, refined AI-driven HR and workforce
Network	Cost from network planning, engineering, and run (eg, network operations center)	12	10–11	–1 to –2	Self-healing agentic-driven networks, self-optimized networks (focused on network run)
IT	Cost from operation, maintenance, and development of IT systems	2–3	3–4	+0–1	Accelerated software development (automated testing, debugging, refactoring, and documentation)
EBITDA/ Total		33–34[7]	41–43	+8–10	

		Capital expenditures, %: Telco today	Capital expenditures, %: Future telco	Change, pp[1]	
Capital expenditures	Network expansion, real estate	100	92–96	–4 to –8	Optimization to increase capital expenditures ROI with impact on customer experience, real estate fine-tuning

Note: Numbers based on wireless operators with comparable characteristics (eg, ~$500 million to $1 billion in annual revenue, more than 5 million in active users). Figures may not sum, because of rounding.
[1]Percentage point. [2]Customer value management. [3]Service to sales. [4]Cost of goods sold. [5]Impact on volume of full-time equivalents (FTE) calculated based on the expected impact on P&L cost related to workforce × a correction factor to adjust for different ranges (eg, expected that agentic AI automates first low-wage workers, therefore the FTE impact is bigger in pp versus that on P&L). This impact excludes third-party spend (eg, external call centers). [6]Selling, general, and administrative. [7]EU average = 33; US average = 36.
Source: Investor reports; McKinsey Global Institute; McKinsey analysis

McKinsey & Company

2. Isolating key tasks and jobs to be done are essential for redesigning workflows

Much of AI agents' transformative power stems from how they can be leveraged to fundamentally reinvent end-to-end workflows across business functions. Yet most telcos continue to capture only a fraction of this value. Nearly three-quarters of the surveyed leaders report using AI primarily to support minimal or moderate redesigns of existing processes rather than fundamentally rethinking how work should be done in an AI-first environment. As a result, AI often accelerates legacy ways of working instead of delivering notable improvements in cost, speed, and effectiveness.

Leading operators shift the focus of redesign from processes to tasks and the accompanying skills that they require. By systematically decomposing priority value chains into their underlying activities and isolating the essential capabilities, they create a clear, detailed view of the work that is performed and what it takes to get it done. Each activity is then reassessed based on its suitability for automation, human-AI collaboration, or continued human ownership (Exhibit 5). This granular approach allows operators to redesign workflows deliberately, rather than layering AI onto inherited structures.

Exhibit 5

Reimagining telco processes from scratch by mapping existing workflows is key to capturing impact from agentic AI.

Steps for redesigning a workflow using agents

Human task · Agent task

1

Create an exhaustive list of potential workflows to automate

2

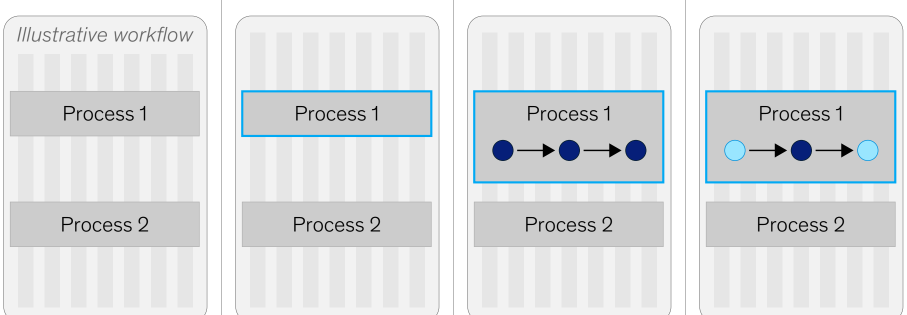
Prioritize and select key workflows to augment using agent

3

Break down workflow into subtasks and identify pain point

4

Design new workflow using agent, AI/machine learning, and humans in the loop



1
- Inventory all end-to-end workflows across priority domains
- Build a business-owned taxonomy that links each activity to cost, full-time equivalents, and outcomes

2
- Evaluate each process on value potential (cost, revenue, customer experience) and feasibility (data readiness, complexity)
- Select high-priority candidates for initial agent enablement

3
- Break target workflows into granular tasks
- Identify bottlenecks, manual activities, and non-value-add steps

4
- Redesign flow to eliminate waste and embed agentic automation (including design of process that can leverage robotic process automation and traditional AI)
- Develop proof of concept of agents, including orchestration among agent families

Critical consideraitons

Start with a common language
- A unified, business-owned process taxonomy standardizes workflow classification, revealing cross-functional overlaps and inefficiencies missed by fragmented use case lists

Attach impact to each workflow
- Linking costs and outcome metrics (revenue, customer experience) to processes quantifies value, enabling ROI-driven prioritization

Redesign, don't just automate
- Reengineer high-value workflows, embedding gen AI agents to streamline tasks, eliminate non-value-add steps, and drive productivity

McKinsey & Company

One North American operator applied this approach to its marketing and commercial operations. It started by mapping various teams' day-to-day activities to ground the transformation in existing opportunities. It then fully redesigned all major workflows (such as evergreen, always-on marketing) with AI at their core. Only after the redesigned workflows were clear were technology choices and organizational changes finalized.

This approach is critical because it can fundamentally change outcomes. By redesigning workflows from scratch (looking only at the "necessary steps"), entire stages, handoffs, and coordination layers are eliminated, with additional value unlocked. By contrast, embedding AI into existing processes mainly accelerates the status quo, just as applying AI to isolated tasks often shifts bottlenecks rather than removing them, resulting in limited productivity gains. Telcos that redesign end to end, anchored in task-level clarity and AI-first assumptions, can unlock sustained improvements in cost, cycle time, and output quality that incremental automation cannot achieve.

3. Give agents and humans clarity about their new roles

Redesigning workflows is just the first step of the organizational overhaul telcos will need to generate substantial impact from AI. It's equally important to transform their operating structures and both agent and human roles to fuel lasting gains in cost efficiency, cycle time, and overall capacity.

A critical aspect of this is explicitly defining how AI agents are embedded within teams—what work they perform end to end, how they collaborate with humans, how exceptions are escalated, and how performance is tracked and improved. Once telcos have identified the opportunities for embedding AI agents into workflows, they can turn to redefining human roles and responsibilities. Research by the McKinsey Global Institute shows that while 30 percent of all telecom workforce hours could realistically be automated by 2030, fully 91 percent of all sector jobs will require fundamental redesign and rethinking to make the most of automation. Even as traditional roles in areas such as DevOps (software development and IT operations), product marketing, customer support, and field engineering are likely to experience declines in numbers, positions like solutions sales specialists, forward-deployed engineers, product builders, and network performance analysts are just as likely to grow; entirely new roles, including AI workflow architect and AI risk manager, will also emerge and gain prominence.

Today, many responsibilities that could be automated or orchestrated by AI agents are fragmented across numerous roles, teams, and functions. Without rethinking the structure to consolidate those responsibilities into clear, accountable roles, the impact of AI is likely to remain limited. Organizing these evolving roles around employees' skills and expertise, based on a chapters model, helps create belonging and clarity regarding role expectations. As employees rely less on cross-functional coordination, team interactions become faster and more agile. Individual employee capabilities, however, will shift to include designing, supervising, and continuously improving agents.

The benefits of these moves are manifold. Clearly defining how agents are integrated reduces structural overhead by removing layers of supervision, coordination, and approvals. Shifting roles, team structures, and agent ownership allows AI to take on end-to-end work while freeing up human capacity for higher-value tasks, rather than just making those individuals marginally more productive. And transforming operating structures and roles along with workflows helps prevent humans from serving as bottlenecks in these new, AI-driven processes.

4. Change management can't be a one-off exercise

Telco leaders often assume that technical upgrades and AI pilots will eventually translate into commercial impact. Our research shows otherwise. Slow adoption, rooted in weak change management, is the primary barrier to scaling AI value. Converting AI capability into sustained productivity and margin improvement requires an operational discipline that hardwires new behaviors into daily work.

Leading operators pair a CEO-led mandate with frontline execution teams, including change managers, analytics translators, and product owners, that are embedded directly into sales and service teams. Large language models (LLMs) make it possible to measure AI usage and adoption at the individual level, enabling much finer-grained visibility into how work is done. This data foundation allows operators to track behavior with greater precision, hold supervisors accountable for reinforcing usage, and shift incentives from outcomes to inputs to drive company effectiveness and efficiency. Such proximity ensures that adoption is measured daily at the individual level, frontline supervisors are accountable for reinforcing usage through coaching and structured performance conversations, and behavioral levers such as incentives and gamification are used to normalize AI usage. Finally, change teams cannot be treated as a short-term tactic; they must remain embedded with frontline operations. Together, these practices ensure that increased AI usage translates directly into measurable improvements in sales and service performance.

At Indosat Ooredoo Hutchison (IOH), CEO Vikram Sinha has framed the company's AI transformation approach as "70 percent people, 20 percent process, and 10 percent technology," underscoring that sustained adoption depends primarily on organizational change.[5] IOH paired this vision with a scaled capability-building and change program: the full C-suite team convenes quarterly for immersive AI sessions, AI translators have been trained across every function, and employees participate in gamified recognition programs that encourage experimentation and reward impact. These efforts embed AI directly into day-to-day work and build a broad base of AI champions, creating the momentum needed to drive adoption across the organization.

Operationalizing these lessons requires concrete program design. Capability building must be role specific and reinforced through coaching incorporated into daily routines. Core workflows should be rebuilt so that AI outputs are always factored into decision-making, with clear ownership and transparent tracking of adoption. Senior leaders must model AI use, while progress is communicated frequently across the organization. Sustained engagement comes from running short, focused campaigns that combine incentives, targeted learning, and rapid iteration based on frontline feedback.

Ultimately, to turn AI into a durable source of competitive advantage and value creation, telcos need to convert change management from a one-off program into a repeatable operating rhythm.

5. AI factories model success and accelerate value capture

Telecom leaders eager to move their AI initiatives beyond a handful of "pilot purgatory" or isolated use cases often run into the same problems: fragmented efforts, high costs, slow scaling, and insufficient risk management. In practice, this often means duplicative model builds, inconsistent foundation model choices, and limited reuse of data, prompts, or agent architectures. This is even more relevant in the current market, where users increasingly leverage AI solutions in their day-to-day work (such as for code development) without clear guardrails or training, thus posing a potential risk for their organizations.

An AI factory model addresses these issues by introducing a common playbook and set of capabilities to guide how the organization delivers and scales AI responsibly. By bringing together the right process, talent, and technology, it accelerates time to value and makes it easier to redesign workflows with greater automation and consistency (Exhibit 6). Rather than creating one-off solutions for every new project, the factory model treats AI assets—platforms, tools, and reference architectures—as standardized, core capabilities that can be built once and reused. These assets can include reusable agent frameworks, shared model environments, and standardized governance controls.

[5] "McKinsey and IOH discuss cracking the AI paradox at Innovate Asia," McKinsey, November 26, 2025; Khoo Tee Tan, "Vikram Sinha on Indosat's role in shaping Indonesia's AI transformation," McKinsey, December 3, 2025.

Exhibit 6

An AI factory model gives telcos a common playbook and capabilities to successfully scale AI across the organization.

Critical elements for building an AI factory

Strategy

- Structured ideation
- AI opportunity identification
- Process redesign
- Value realization
- Portfolio strategy and governance

Tech foundations

Technology

- Monitoring
- Agentic AI platform
- Modular AI agents
- AI sandbox
- AI productivity tools
- On-premises capability

Data

- Central AI platform
- Data strategy and protocols
- AI FinOps
- AI-driven data quality
- AI orchestration
- AI data discovery
- Operational monitoring
- Unstructured data handling
- CI/CD[1] starter kit
- AI data remediation
- Experiment tracking
- AI data products
- AI LiveOps tooling

Rewiring capabilities

Operating model

- Use case intake
- AI standards
- AI control tower
- AI architecture review
- Ways of working/ pod structure

Talent

- Capability assessment
- Strategic workforce planning
- Career path
- AI-enabled HR engine
- Partner ecosystem
- Managed service provider partnerships

Scaling and activation

Change management

- AI marketing assets
- AI champion network
- Executive AI training
- Adoption model

Responsible AI

- AI principle and mandates
- Risk log tool
- Risk control
- Responsible AI training

Key attributes

- Pre-provisioned infrastructure
- End-to-end life cycle support
- Flexible model integration
- Reusable components and frameworks
- Scalable resource management
- Unified orchestration and monitoring
- Enterprise-grade security and governance

[1]Continuous integration/continuous delivery.

McKinsey & Company

Without such standardization, scaling AI across the enterprise can become very difficult. Teams tend to develop solutions independently, adopting different tools, design patterns, and operational practices, which often lead to fragmented or siloed implementations that are hard to govern and even harder to scale. Over time, this fragmentation can increase technical debt, raise inference and infrastructure costs, and create uneven performance across similar AI use cases. This inconsistency also increases complexity and introduces significant risk, particularly around security, compliance, and data governance.

The AI factory also acts as a connective layer between an organization's varied AI efforts and senior leadership, increasing transparency and helping prioritize the most valuable, cost-effective opportunities while making trade-offs explicit. By centralizing visibility into use case value and model performance, it strengthens accountability for AI impact.

When standing up an AI factory, organizations must be focused on the highest-priority use cases. This discipline determines which capabilities are established first and defines the road map for scale. The AI factory is built on a modular set of capabilities, including a structured intake process to assess and prioritize use cases, and shared, reusable components that enable teams to leverage off-the-shelf solutions.

- **Intake** provides a standardized process to review and prioritize AI initiatives, ensuring that resources focus on the highest-impact opportunities. Some telcos now deploy AI agents within the intake workflow to assess readiness, recommend development stage (proof of concept, scale, or deprioritize), and define a clear action path.

- **Shared, reusable components** reduce development time while maintaining quality and consistency. With proven assets such as propensity model blueprints and enablement tools, teams avoid reinventing solutions for each use case and accelerate delivery at scale.

In addition to these two core capabilities, an AI factory can take on some or all aspects of talent and change management. It is not a static construct but a living capability that must be built and sustained within the organization, providing a repeatable model for moving from isolated pilots to enterprise-level results.

6. Moving forward: Agents must understand the business and the data to truly unlock its power

As telcos transition toward AI-native, self-directing enterprises, the industry is approaching a structural inflection point. The next generation of agent-led operating models will likely not be limited by algorithmic capability, but by whether AI systems can truly understand the business and the network they are meant to help run. Even the most capable LLMs struggle when key concepts, policies, and dependencies are scattered across systems or left implicit.

As they work to address this, leading operators are starting to invest in knowledge frameworks (semantic data layers) to become a foundational layer of AI-native telco architecture. Those conceptual models represent the institutional knowledge "locked" in the organization. Sharing this knowledge with agents can unlock new capabilities, cut down on duplicate efforts, reduce technical complexity, and enable agents and workflows to build on one another, turning fragmented business and network data into trusted, reusable intelligence.

This approach can have multiple applications. For example, in network operations, when service quality drops, agents can analyze the network layers to narrow down root causes, identify affected customers and services, and understand effects on service-level agreements. Instead of speeding up ticket handling, agents can determine whether a single network fault is driving multiple incidents, weigh different remediation options, and guide engineers toward the best intervention. For example, a large Australian telco has accelerated service automation 30 percent by encoding engineer's knowledge on the network into agents.

One implementation approach to enable LLMs to understand enterprise data has been the use of ontologies and knowledge graphs. These explicit conceptual models of core business entities—and the relationships between them—aim to provide a shared, consistent representation of the organization's data landscape for both humans and AI agents. While powerful in theory, ontology-driven approaches can be costly and complex to design, govern, and continuously update at enterprise scale. In practice, many organizations struggle to operationalize them quickly enough to keep pace with evolving business needs.

An emerging alternative is the use of modular, context-specific AI "skills"—reusable workflows or capability layers that encapsulate business logic, decision rules, and data access patterns in a structured but lighter-weight way. Rather than requiring a fully harmonized enterprise data model up front, AI skills allow organizations to operationalize AI around well-defined tasks (for example, pricing analysis, customer segmentation, and root-cause analysis) and progressively build capability through repeatable, composable units. Tools such as Claude Skills illustrate how this approach could enable faster deployment, clearer governance, and more pragmatic codification of institutional knowledge across AI agents.

While telcos' initial efforts to incorporate AI have shown genuine promise across pilots and fragmented use cases in areas such as customer service, network, and marketing and sales, only a small number of operators have been able to capture significant value. But as those operators are showing and their peers are learning, the emergence of agentic AI may present telcos with an automation-driven inflection point, in which agents' ability to redefine entire workflows is critical to leveraging AI's ability to radically improve customer experience and operational efficiency.

The development of a coherent agentic operating model is essential to take advantage of the new agentic opportunity. This involves prioritizing the agentic redesign of some end-to-end processes, tying AI initiatives to clear financial goals and budget outcomes, defining and clarifying new roles and responsibilities for both agents and humans, building ontologies to give agents the necessary business context, and implementing comprehensive change management to drive adoption. Finally, AI factory models help bring together all these elements, creating reusable trusted data assets and processes that can be harnessed broadly across the organization to help lower unit costs and improve economics as scale increases.

Telcos that treat agentic AI as simply another layer of tooling may see marginal productivity gains and growing complexity. Those that commit to building a coherent agentic operating model will fundamentally reshape how work gets done—and, in doing so, unlock material improvements in ROIC, margins, and customer experience. As agentic AI advances, the gap between these two groups could widen. For telecom leaders, the question is increasingly less about whether AI can create value, but more focused on whether their organizations are prepared to operate differently enough to capture it.

Benjamim Vieira is a senior partner in McKinsey's Lisbon office; **Guilherme Cruz** is a partner in the New York office, where **Tomás Lajous** is a senior partner, **Borja Belda** is an associate partner, and **Manuel Palacios** is a consultant; **Ignacio Ferrero** is a partner in the Miami office; and **Daniela Mendoza** is an associate partner in the Madrid office.

The authors wish to thank Bruce Philp, Cindy Xiao, Leire Merino, Max Shackelford, and Takamichi Watase for their contributions to this article.

This article was edited by Daniel Eisenberg, an executive editor in the New York office.

Technology, Media & Telecommunications Practice

Winning B2B customers in technology and telecommunications

Our latest B2B Pulse Survey of 3,000 decision-makers shows how agentic AI is creating value beyond connectivity, raising the bar on execution, integration, and trust.

This article is a collaborative effort by Andrea Travasoni, Dev Patel, Giacomo Dolci, and Naveed Niwaz, with Bulat Gaifeev, Luca Furlani, and Olimpia Gascó, representing views from McKinsey's Technology, Media, & Telecommunications Practice.

March 2026

B2B technology and telecom growth is entering a new phase—one defined less by disruption than by rebalancing. Core connectivity remains a critical foundation, but it is no longer sufficient to define where value is created or where growth accrues. Instead, growth is increasingly shifting to higher layers of the stack, where security, intelligence, seamless engagement, and trust shape customer outcomes.

Insights from McKinsey's latest Global Technology and Telecommunications B2B Pulse Survey, based on more than 3,000 enterprise decision-makers across 11 industries and 18 countries, show that several fundamentals are holding. Overall investment intent remains positive, driven primarily by large enterprises, even as growth in core connectivity continues to lag behind higher-value domains such as security and agentic AI. Customers increasingly recognize operators' relevance beyond the network, and omnichannel engagement has become a baseline expectation across the buying journey.

At the same time, new openings for growth are emerging. Investment momentum is broadening as small and medium-size enterprises (SMEs) spend more decisively. Customer stickiness toward telecom operators is stronger than expected, though loyalty is becoming more conditional and increasingly anchored in security and trust rather than connectivity alone. Quality of experience has emerged as a decisive swing factor, shaping how customers engage and transact across channels.

Agentic AI is beginning to reset the growth equation. Adoption is accelerating across customer segments and regions, reshaping how enterprises buy, interact, and operate. As customers themselves become more AI-enabled, digital channels are moving from a supporting role to a primary growth engine—enabling larger, more complex transactions to be executed end to end and raising expectations for frictionless, outcome-driven engagement.

For telecom operators, these shifts create a dual imperative. Agentic AI offers a tangible lever to improve internal efficiency and service performance today, supported by strong customer openness to AI-enabled assistance. This makes operational deployment an immediate priority, with clear potential to reduce cost to serve and scale service quality. Over time, these capabilities may also provide a foundation to extend offerings beyond core connectivity into trusted, AI-enabled solutions. The ability to materialize these opportunities will depend less on technology access than on disciplined execution—specifically, operators' ability to deploy reliably at scale and maintain trust in an increasingly regulated, AI-driven environment.

1. Growth fundamentals are holding

While the competitive landscape is evolving, several structural dynamics from prior years remain firmly in place.

Investment outlook is positive, with core connectivity lagging behind

Survey results show that more than 60 percent of respondents plan to increase technology and telecommunications spending, driven primarily by large enterprises. This momentum reflects a broad recognition of the need to invest in innovation and remain competitive at the forefront of technology.

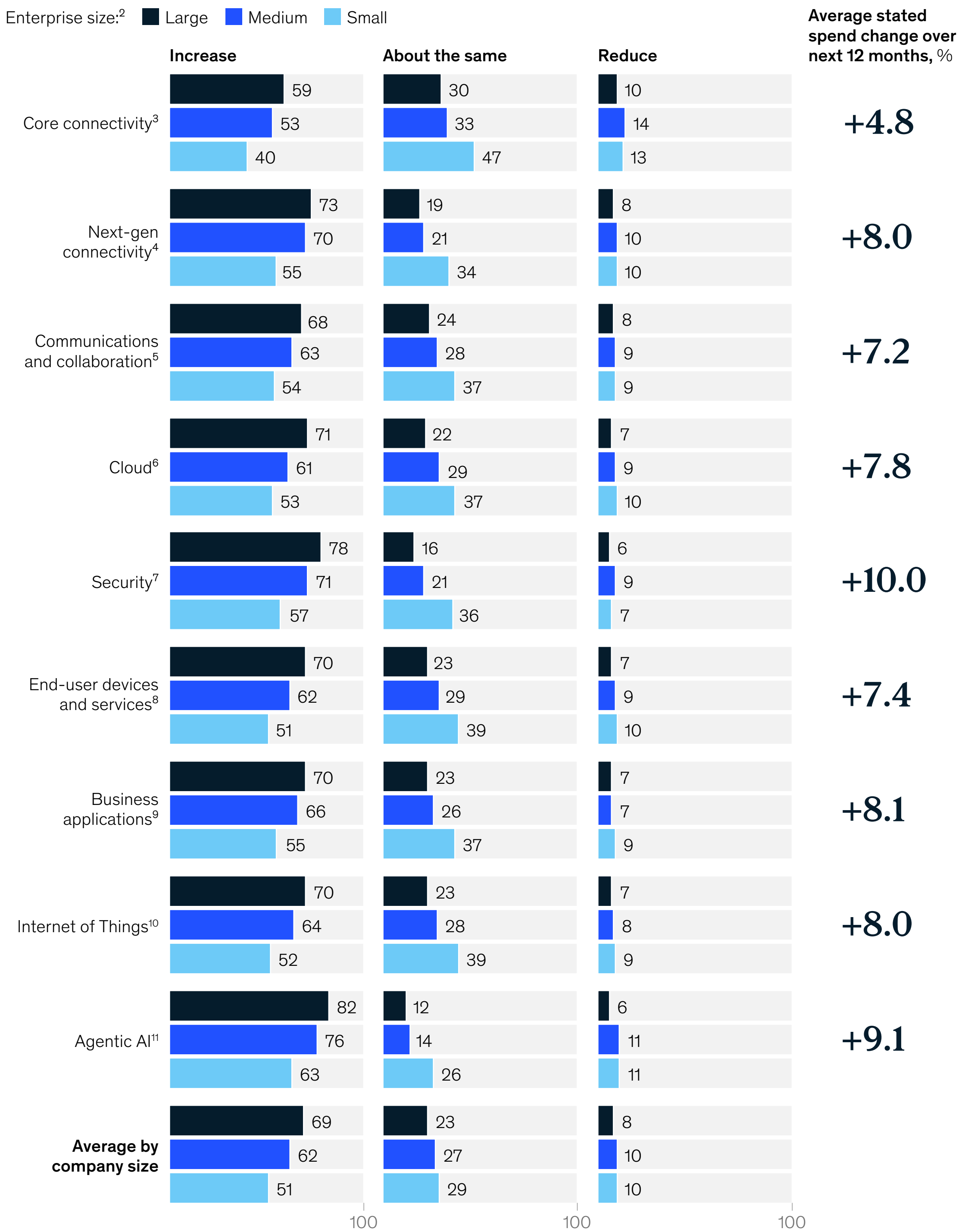
The trend holds across geographies, though optimism varies significantly. Africa (74 percent expecting growth) and Asia–Pacific (70 percent) show markedly stronger investment intent than Europe (54 percent), where macroeconomic pressures appear to be dampening spending appetite.

Growth expectations also diverge sharply by product domain. While core connectivity remains foundational, the strongest momentum has shifted to higher-value areas centered on protection, intelligence, and automation. Security and agentic AI are the fastest-growing domains (Exhibit 1), each expected to expand by more than 9 percent over the next 12 months. Core connectivity, by contrast, is projected to grow at less than 5 percent, making it the slowest-growing major category in the portfolio.

Exhibit 1

Security and agentic AI lead growth beyond the core.

Expected budget or spend evolution over next 12 months (net of inflation), by product category,[1]
% of respondents across business segments



[1]Question: How do you expect your budget or spend by product category to evolve in the next 12 months net of inflation? [2]Large = >500 employees; medium = 100–500 employees; small = <100 employees. [3]Mobile and fixed core connectivity. [4]SD-WAN solutions, 5G, private networks. [5]Collaboration software including share and sync; unified communication tools including audio and video. [6]Public, private, and hybrid cloud models, legacy on-premise, IaaS (infrastructure as a service), PaaS (platform as a service). [7]Data security, endpoint security, network security, security applications, web and email security. [8]Mobile/PC, other hardware (eg, tablets, Wi-Fi), device management, support contract/contact center. [9]On-premise/licensed software, SaaS (software as a service), internally developed software. [10]Hardware embedded with electronics, software, sensors, and connectivity, which enables these objects to connect and exchange data including all related services. [11]AI systems that can plan and execute multistep tasks and take actions.
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,214

McKinsey & Company

Operators have a right to play beyond core connectivity

B2B customers now expect operators to act as integration partners and end-to-end solution providers, placing greater emphasis on the ability to orchestrate complex, multidomain solutions rather than on network assets alone (Exhibit 2). As expectations rise, execution has become the primary constraint.

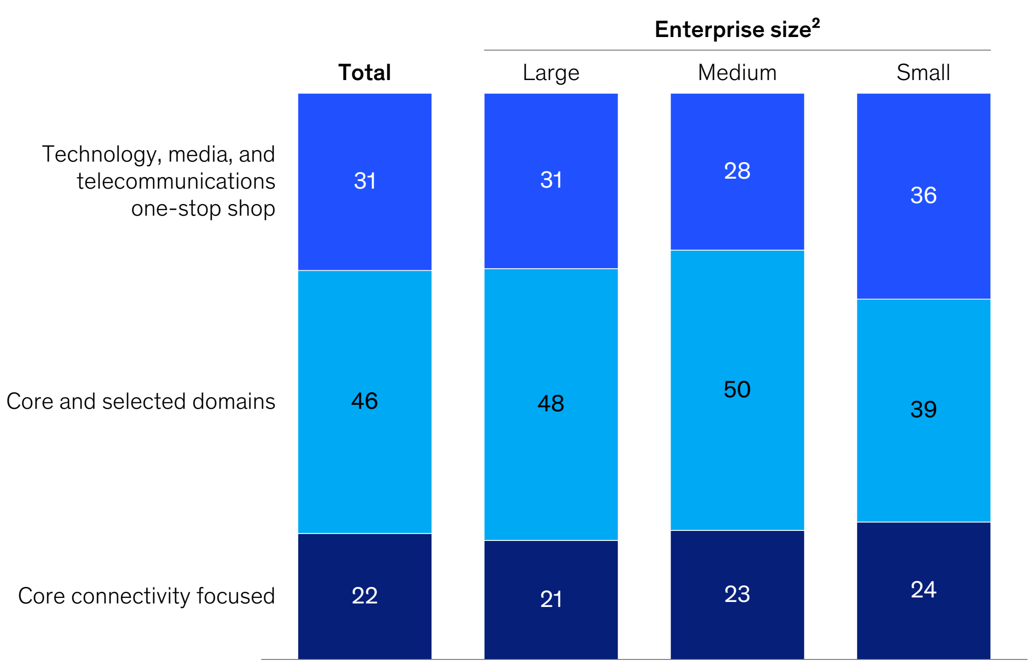
Nearly 80 percent of survey respondents recognize operators as relevant players beyond mobile and fixed connectivity, particularly in areas such as network orchestration, security, cloud provision and hosting, as well as the delivery of AI capabilities. Customer preferences, however, vary by segment. Small enterprises show the strongest interest in one-stop-shop offerings. Medium enterprises are more selective, favoring operators that concentrate on a defined set of domains. Large enterprises sit between these positions but show growing interest in advanced, integrated solutions as agentic AI adoption accelerates.

As B2B customers increasingly expect operators to act as integration partners and end-to-end solution providers, success hinges less on network assets and more on the ability to deliver complex, multidomain solutions at scale.

Exhibit 2

Operators have a right to play beyond connectivity.

Role telecom operators should play in the industry in the future,[1] % of respondents



Note: Figures may not sum to 100%, because of rounding.
[1]Question: What role should telecom operators play in the industry in the future?
[2]Large = >500 employees; medium = 100–500 employees; small = <100 employees.
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,214

McKinsey & Company

Customers want omnichannel

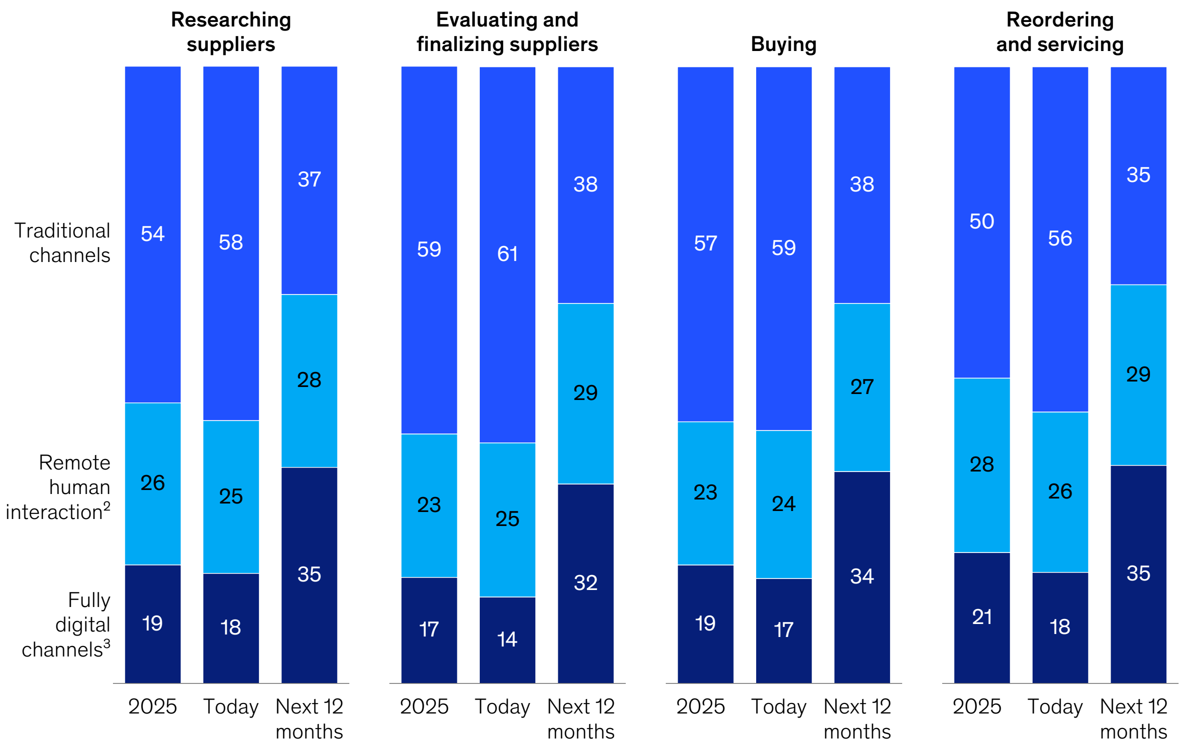
Delivering on the promise beyond core connectivity increasingly depends not only on what operators offer but on how they engage customers across channels—particularly digital ones. After a period of restraint, appetite for fully digital engagement is rebounding. Survey respondents indicate that approximately one-third of B2B customers expect to rely on digital-first interactions over the next 12 months (Exhibit 3). As a result, technology and telecom providers are being assessed less on channel availability and more on their ability to deliver seamless, consistent experiences across digital interfaces and human touchpoints.

This orchestration matters most where relationships are strongest. Nearly 60 percent of B2B survey respondents prefer hybrid engagement models that combine digital and remote human interaction, particularly in reordering and servicing journeys. Customers want the efficiency of digital channels without sacrificing continuity or trust built through existing relationships.

Exhibit 3

Digital intent outpaces usage.

Preferred communication channel with provider across the buying journey,[1] % of respondents



Note: Figures may not sum to 100%, because of rounding.
[1]Question: Please select the preferred channel of communication with your provider across the buying journey over last/next 12 months.
[2]Remote includes phone calls, video conference calls, emails, fax, etc.
[3]Fully digital includes company websites, e-commerce, chatbots, internet searches, mobile apps, etc.
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,214

McKinsey & Company

2. New trends are creating renewed openings for growth

At the same time, shifts in demand, retention, and digital behavior are opening new paths to growth.

Spending intent is rising, led by SMEs

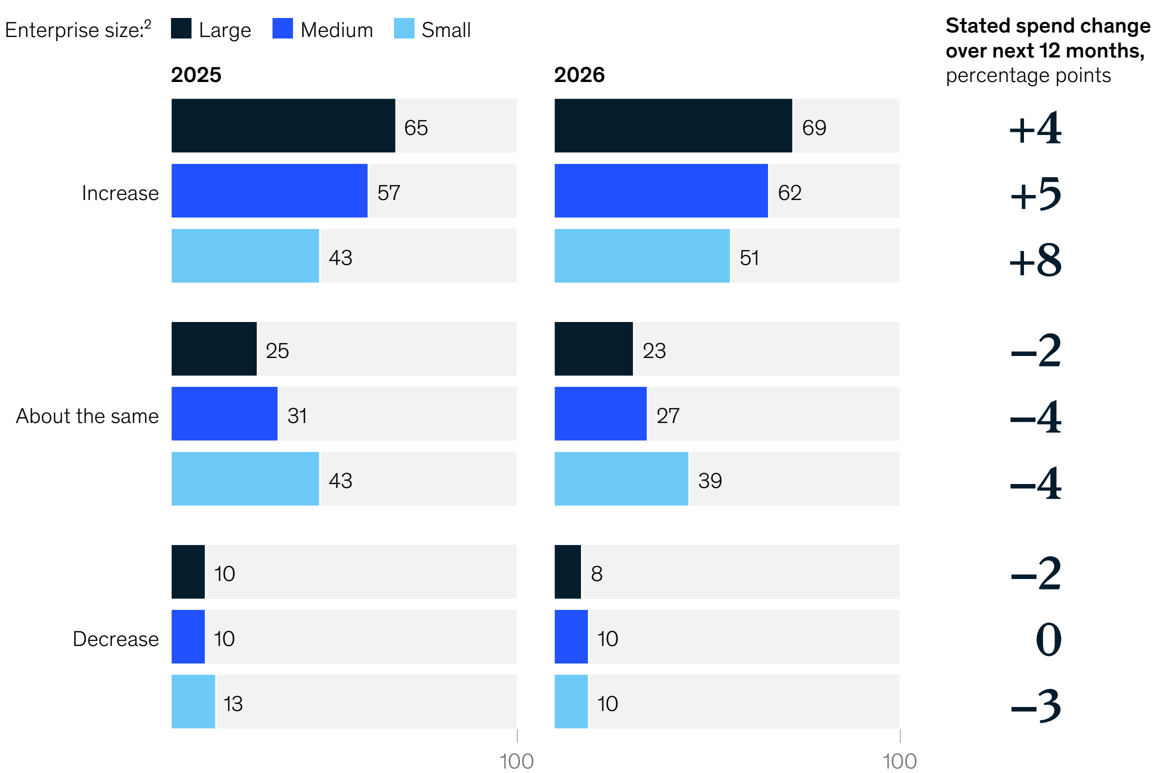
Enterprise technology budgets point to a solid growth outlook for the year ahead. Survey results indicate that more than 60 percent of companies plan to increase technology spending, driving a 7 percent year-over-year increase in overall budget-growth intent (Exhibit 4).

This renewed growth expectation is no longer driven solely by large enterprises. SMEs, often more cautious, are showing renewed confidence, with budget-growth intent among small and medium-size enterprises rising by eight percentage points year over year.

Exhibit 4

Small and medium-size enterprises show the strongest year-over-year increase.

Expected budget or spend evolution over next 12 months (net of inflation), by product category,[1]
% of respondents across business segments



[1]Question: How do you expect your budget or spend by product category to evolve in the next 12 months net of inflation?
[2]Large = >500 employees; medium = 100–500 employees; small = <100 employees.
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,214

McKinsey & Company

While large enterprises continue to account for the majority of absolute spend, the broadening of confidence marks a structural change from prior years. As SMEs reenter the market more decisively, the addressable growth base is expanding and competition for incremental demand is intensifying, particularly in high-growth domains beyond core connectivity.

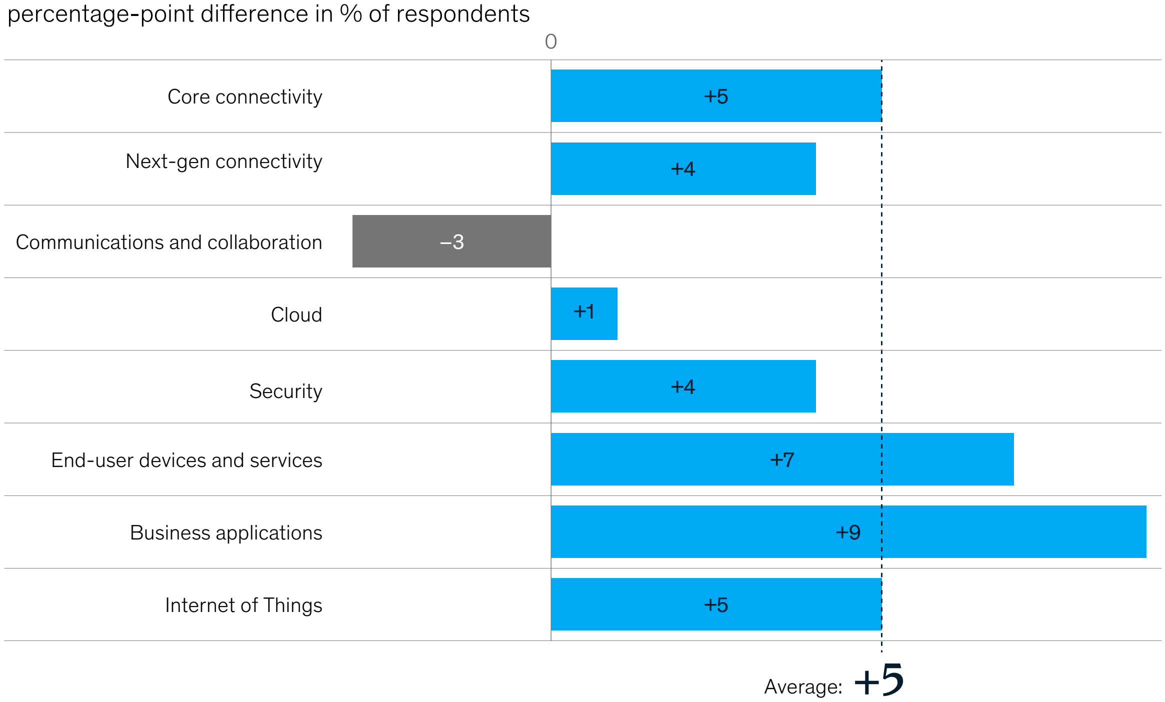
Customer stickiness is stabilizing, but loyalty is increasingly conditional
After a period of elevated churn risk, B2B customer stickiness toward telecom operators has begun to stabilize. Nearly three-quarters of surveyed customers report no intention to switch provider types over the next 12 months, up from roughly two-thirds in prior years (Exhibit 5). This shift points to a more resilient customer base and suggests that competitive pressure is translating into more selective switching rather than broad-based defection.

This resilience, however, is uneven across the portfolio. Stickiness is strongest, and improving fastest, in business applications and end-user devices and services, where customers report a seven- to nine-percentage-point increase in their likelihood to remain with telecom operators between 2024 and 2026. In these domains, operators appear to be consolidating their role as trusted delivery partners.

Exhibit 5

Telecom operator retention has improved across most product categories.

Change in share of respondents choosing telecom operators as their primary vendor,[1] 2026 vs 2024, percentage-point difference in % of respondents



[1]Question: When sourcing technology and telecommunications products, who is your primary vendor today and most likely to be in 12 months?
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,214

McKinsey & Company

By contrast, other domains remain structurally contested. Communications and collaboration, as well as cloud services, show flat or declining stickiness, reflecting intense competition and a crowded supplier landscape. In these areas, incumbency alone offers limited protection, and customer relationships remain more fluid.

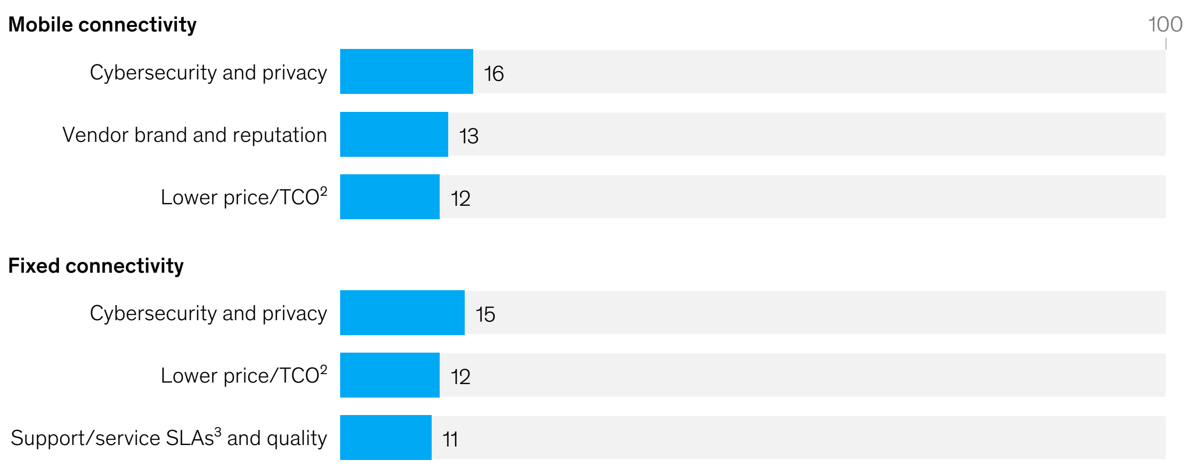
When customers do consider switching, the underlying drivers are also shifting. Among the roughly 30 percent of surveyed B2B customers that report having switched providers in the past year, cybersecurity has emerged as the leading trigger, overtaking traditional factors such as price, coverage, or service reliability (Exhibit 6). Even in relationships that are primarily connectivity-led, security considerations now rank as the number one reason for defection.

Taken together, these patterns point to a fundamental change in loyalty. While switching behavior has not yet increased materially, stable churn rates should not be mistaken for durable loyalty. Customers are no longer anchored to operators by connectivity alone; instead, loyalty is becoming increasingly conditional and must be actively earned through credible, outcome-driven security and trust capabilities. For operators that meet this expectation, cybersecurity can serve as a powerful lever to defend and expand share of wallet. For those that do not, incremental demand—and eventual erosion—is likely to accrue, especially in favor of cloud and technology providers rather than remain within the telco ecosystem.

Exhibit 6

Security now ranks as the number one reason for switching.

Top 3 reasons to switch providers per category,[1] % of times ranked as most important by switchers



[1]Question: Which of the following reasons prompted you to switch the [product category] provider? Provider switch reasons respondents selected from: lower price/TCO; product/service features and capabilities including redundancy/reliability; support/service-level agreement and quality; vendor brand and reputation; existing relationship with the provider; product innovation/industry and vertical-specific solutions; end-to-end solution and/or ease of integration with your systems; sales rep knowledge and ability to shape solutions; cybersecurity and privacy; experience in engaging across different physical and digital channels.
[2]Total cost of ownership.
[3]Service-level agreements.
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,214

McKinsey & Company

Customer experience is a swing factor for channel choice with AI-forward customers driving digital channel growth

According to the survey, B2B customers continue to prefer hybrid engagement models that combine digital channels with human interaction. At the same time, new evidence suggests that quality of experience has become the decisive swing factor in channel choice. Rather than channel availability, improved quality and reliability of digital journeys is supporting adoption growth (Exhibit 7). As digital experiences improve across the buying cycle, customers are increasingly willing to rely on digital-first interactions, particularly where journeys are intuitive, consistent, and frictionless.

Exhibit 7

Reliable digital channels drive 1.4 times higher fully digital adoption.

Share of respondents preferring fully digital channel across the customer journey, %

[1]Question: Please select the preferred channel of communication with your provider across the buying journey over last/next 12 months. (Share of respondents choosing fully digital channel.)
[2]Question: What are the challenges you encountered with digital channels across the buying journey? (Respondents selecting "No challenges" for the corresponding step of the buying journey.)
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,214

McKinsey & Company

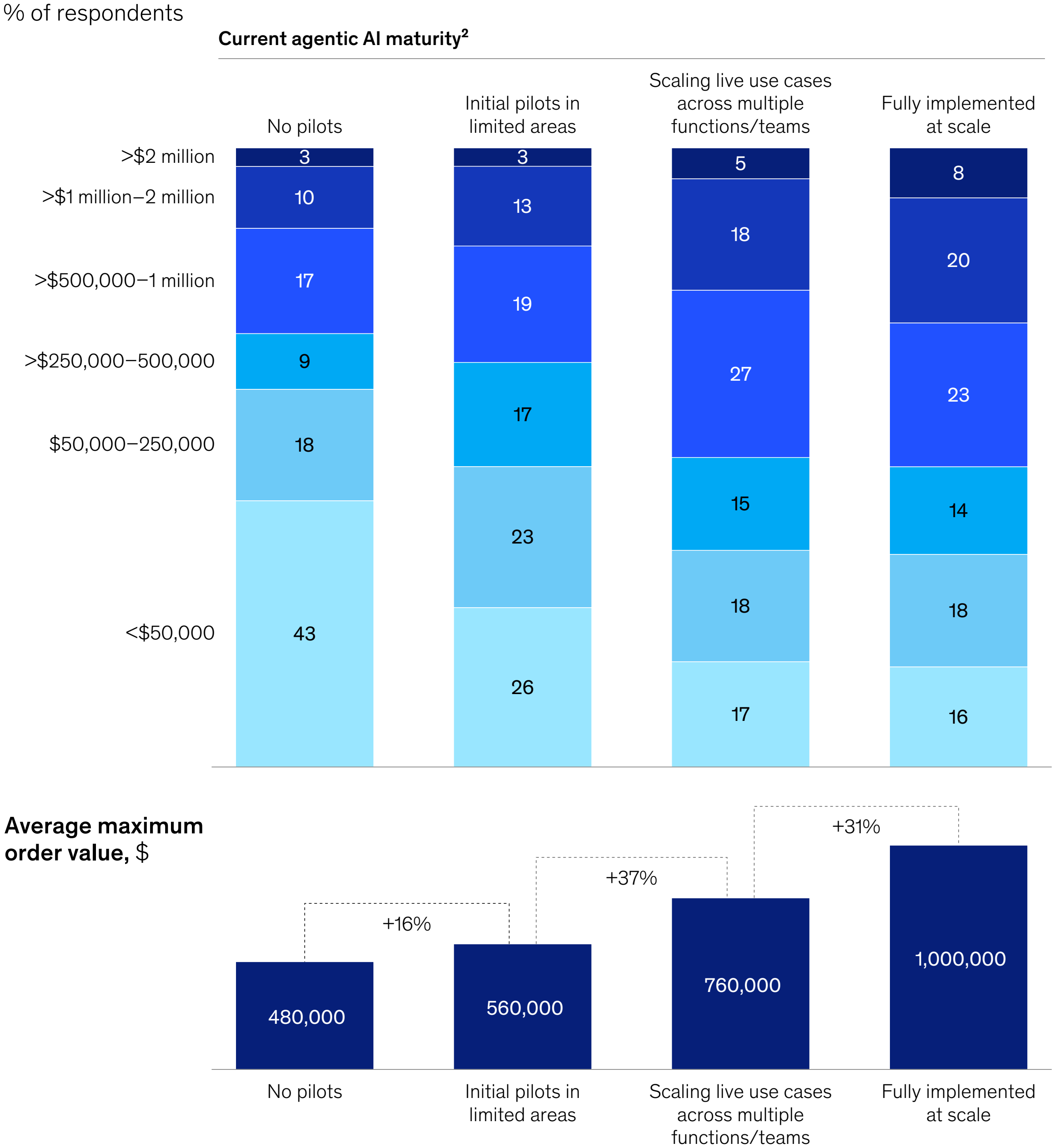
While these improvements have supported steady growth in digital usage, customer experience alone explains only part of the shift. The rapid advancement of agentic AI among B2B customers seems to be an even more powerful "accelerator," pushing digital channel adoption, as well as the size of digital transactions.

Enterprises that have fully implemented agentic AI at scale in at least one use case, or that are actively scaling live use cases across functions, show a materially higher willingness to transact through fully digital, end-to-end channels. For these AI-mature customers, the average "ticket size" for digital transactions exceeds $1 million—roughly twice the level reported by peers still in early pilots or without operationalized AI. As customers become more comfortable delegating decisions and execution to AI-enabled workflows, digital self-service seems to increasingly extend to complex, high-value purchases (Exhibit 8).

Exhibit 8

AI-forward buyers are driving digital spend.

Maximum order value respondent would purchase through end-to-end digital self-service,[1]
% of respondents



Note: Figures may not sum to 100%, because of rounding.
[1]Question: What is the maximum order value you would purchase through end-to-end digital self-service for both a new product or service?
[2]Question: With regard to agentic AI adoption, how would you best describe your company's current maturity?
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,214

McKinsey & Company

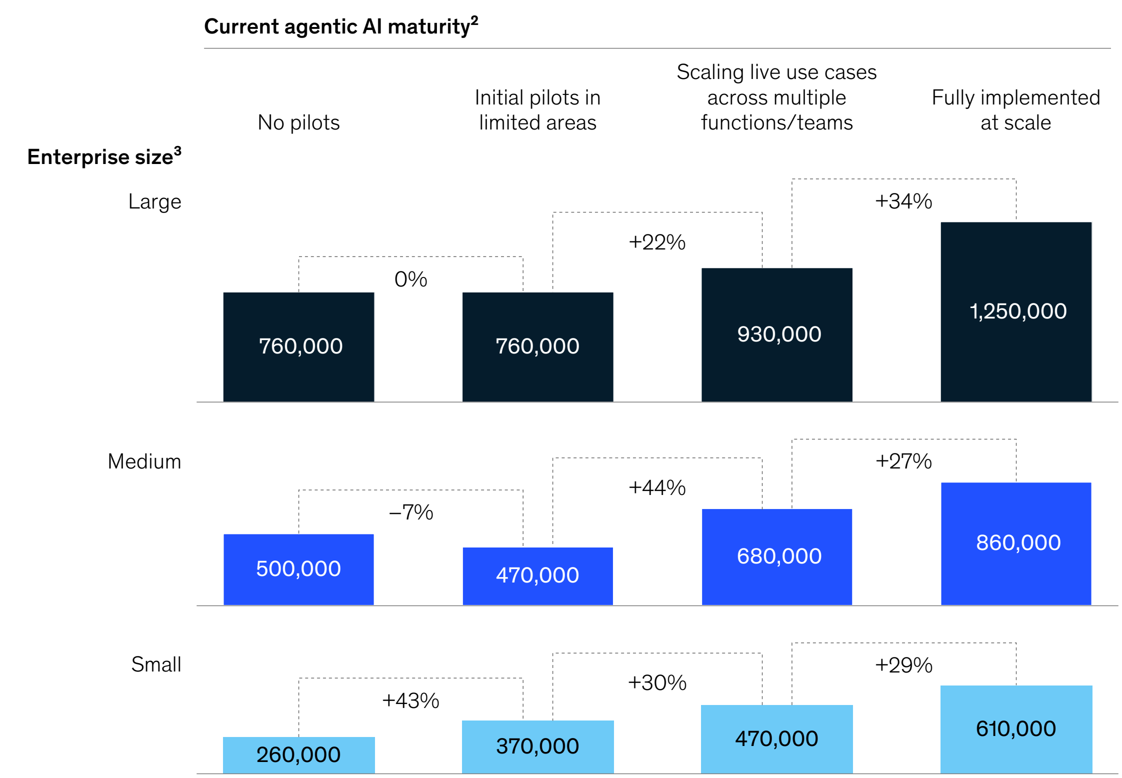
This relationship appears systematic rather than episodic. According to the survey, as customers progress along the agentic AI maturity curve—from experimentation to scaled deployment—the value they report being willing to transact digitally increases in lockstep (Exhibit 9). Each step up in AI maturity is associated with a self-reported 20 to 45 percent increase in the absolute dollar value customers are prepared to execute through operators' end-to-end digital channels.

This pattern holds across enterprise sizes but is most pronounced among large organizations. According to the survey, large enterprises with agentic AI fully implemented at scale report a willingness to transact up to one-third more value digitally than peers using AI in more limited or selective ways.

Exhibit 9

AI maturity boosts digital spend across business segments.

Maximum order value respondent would purchase through end-to-end digital self-service,[1] $



[1]Question: What is the maximum order value you would purchase through end-to-end digital self-service for both a new product or service?
[2]Question: With regard to agentic AI adoption, how would you best describe your company's current maturity?
[3]Large = >500 employees; medium = 100–500 employees; small = <100 employees.
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,214

McKinsey & Company

Taken together, these dynamics point to an impending acceleration in digital channel adoption. Improvements in customer experience have laid the foundation, but agentic AI is now changing how enterprises buy—reducing reliance on human-mediated interactions and increasing comfort with automated, end-to-end digital journeys. As agentic AI adoption continues to expand rapidly across customer segments, digital channels are likely to shift decisively from a complementary interface to a primary growth engine, enabling providers to scale sales productivity and support larger, more complex transactions online.

3. Agentic AI is resetting the growth equation

Agentic AI now sits at the center of these shifts, accelerating change across buying, operations, and delivery models.

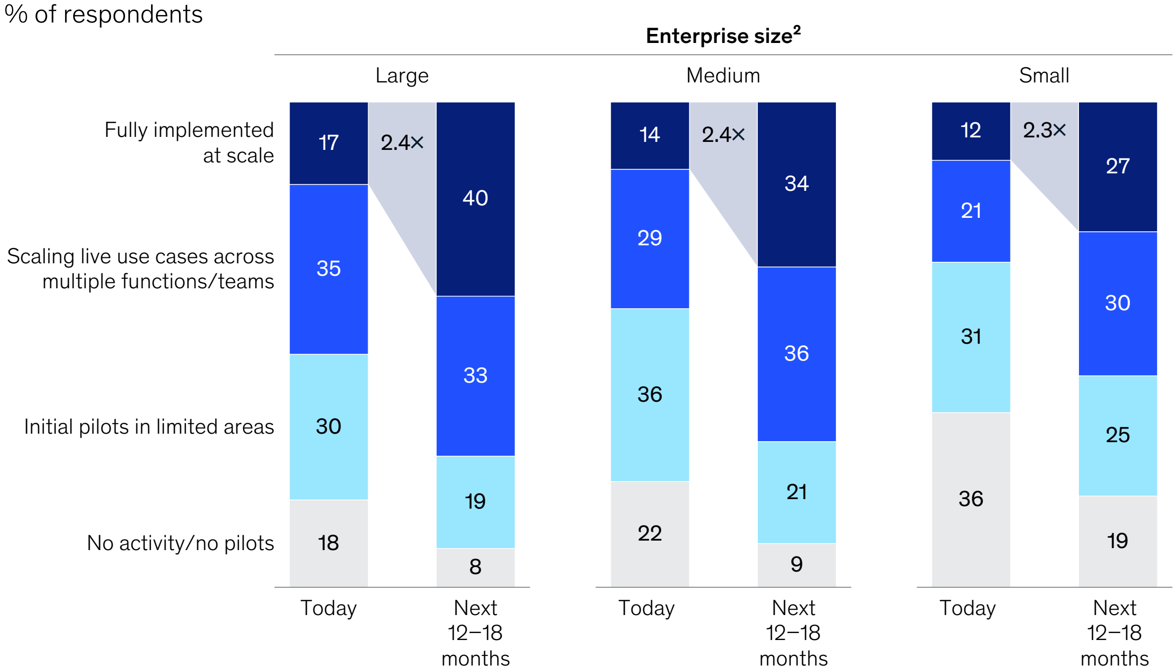
Agentic AI adoption is already widespread but not at the same pace everywhere

Roughly 80 percent of surveyed large and medium-size enterprises report having agentic AI in place in some form, ranging from early pilots to fully scaled deployments (Exhibit 10). Adoption among small enterprises is lower but still material, with approximately 65 percent of surveyed B2B customers reporting some level of implementation.

Exhibit 10

More than 80 percent of companies are expected to be active in agentic AI in the next 12 to 18 months.

Current level of agentic AI adoption maturity vs future ambition (next 12–18 months),[1]
% of respondents



Note: Figures may not sum to 100%, because of rounding.
[1]Question: With regard to agentic AI adoption, how would you best describe your company's current level of maturity and future ambition (next 12–18 months)?
[2]Large = >500 employees; medium = 100–500 employees; small = <100 employees.
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,214

McKinsey & Company

The more meaningful distinction lies in maturity. Among large and medium-size enterprises, about 40 to 50 percent report to have already moved beyond experimentation, with agentic AI either fully implemented at scale or actively scaling across multiple domains. In contrast, only about 30 percent of SMEs report to have reached a similar level of maturity.

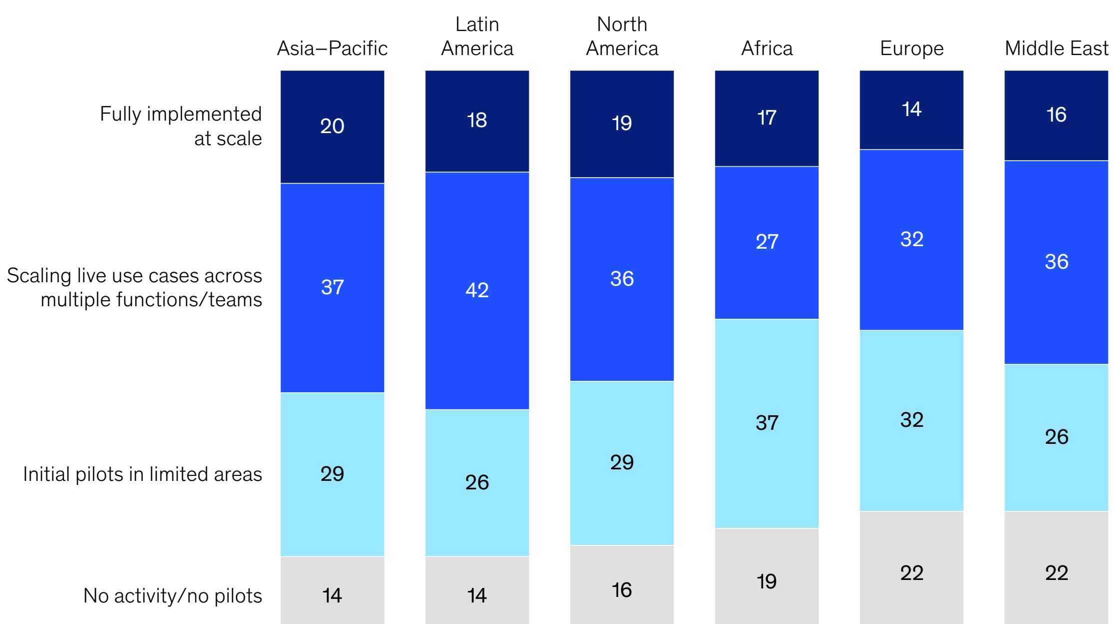
Over the next 12 months, companies across all size segments expect the share of fully implemented agentic AI solutions to increase by a factor of 2.3x to 2.4x. This signals a decisive shift from isolated use cases to enterprise-wide execution, bringing fundamental changes to operating models, decision-making, and system architectures. For B2B operators, the customer base is becoming more automated, more digital, and more capable of engaging through AI-enabled, end-to-end journeys. Portfolios, sales motions, and service models will need to evolve accordingly to remain relevant.

Across regions, momentum in agentic AI adoption remains very strong, with roughly 70 to 80 percent of surveyed large enterprises actively engaged, ranging from initial pilots to fully implemented, at-scale use cases (Exhibit 11). While this high level of activity is evident globally, meaningful regional differences are beginning to emerge.

Exhibit 11

There is strong momentum in agentic AI, with 70 to 80 percent of large enterprises actively engaged across regions.

Current level of agentic AI adoption maturity among large enterprises,[1] % of respondents



[1]Question: With regard to agentic AI adoption, how would you best describe your company's current level of maturity?
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,214

McKinsey & Company

Asia–Pacific and Latin America stand out as the most advanced regions, with more than 80 percent of large B2B companies with already initiated or scaled agentic AI. This reflects faster progression from experimentation to enterprise-wide deployment.

By contrast, Europe and the Middle East have a lower share of enterprises having implemented agentic AI fully at scale or at scale-up stage. More than 20 percent of large B2B companies report no active involvement in agentic AI, indicating a slower transition from pilots to scaled implementation.

While these gaps may appear moderate at first glance, they are strategically meaningful. Geographies further along the adoption curve are likely to experience earlier shifts in buying behavior, greater reliance on digital self-service, and rising expectations for AI-enabled engagement. As these expectations evolve, competitive dynamics will increasingly be shaped by how quickly providers adapt their capabilities to customer readiness and local market conditions, rather than by technology availability alone.

Data privacy and sovereignty are becoming non-negotiable

Rising geopolitical fragmentation and expanding regulatory scrutiny are elevating data privacy and sovereignty from compliance considerations to decisive buying criteria. Across regions and industries, surveyed B2B customers increasingly view sovereign compliance as a prerequisite for engaging with technology and telecom providers, an expectation that is intensifying as agentic AI becomes more deeply embedded in enterprise operations.

Data privacy and regulatory compliance now rank as the single most important customer concern, cited by more than half of respondents (Exhibit 12). Sensitivity is particularly acute as AI agents gain broader access to enterprise data and take on more autonomous decisioning and execution roles. At the same time, regulatory frameworks (most notably around AI governance in the European Union) are becoming more explicit and consequential, raising the stakes for customers seeking to deploy AI at scale without regulatory risk.

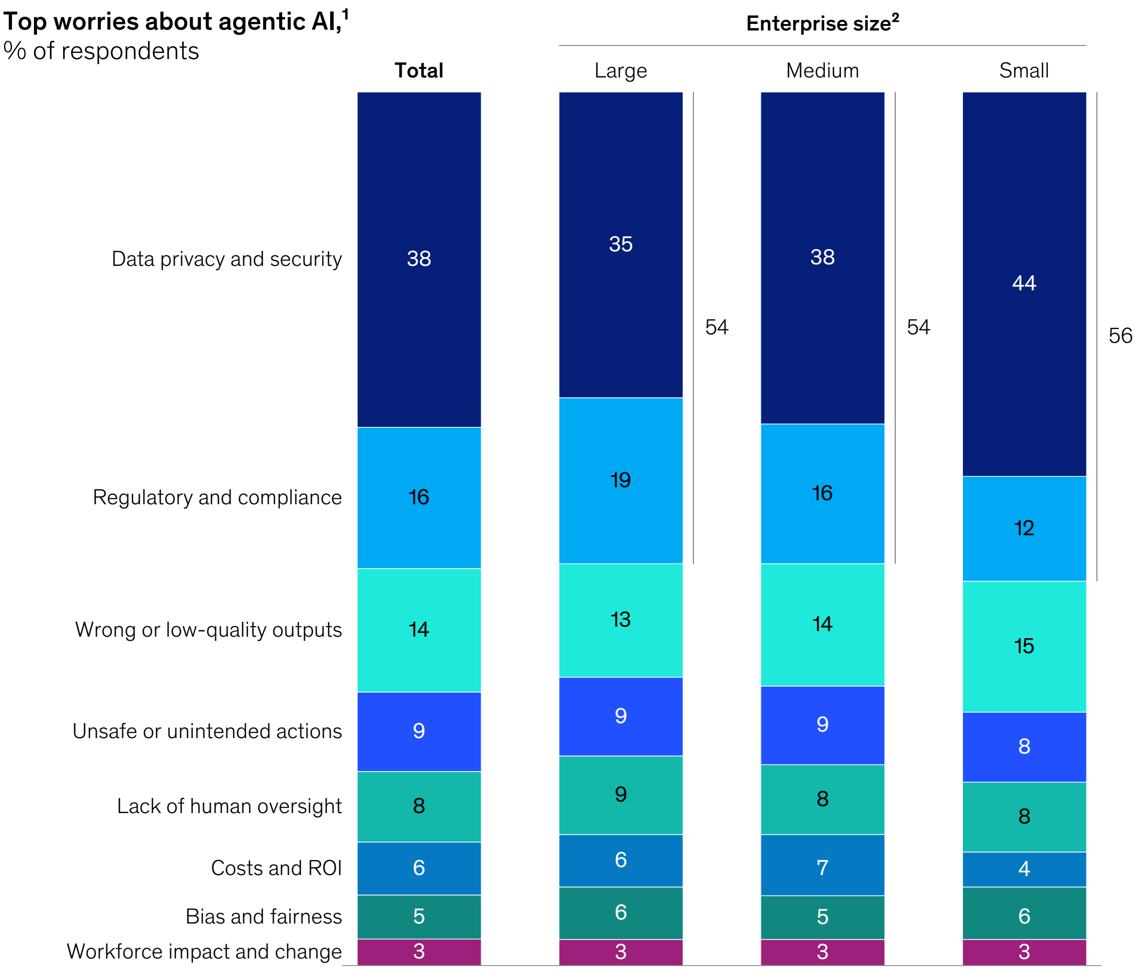
As a result, providers that fail to meet sovereignty, security, and compliance expectations are increasingly excluded from consideration, regardless of price or functionality. Trust has become a gating factor rather than a differentiator. Telecom operators enter this environment with a structural advantage, grounded in decades of experience operating under stringent regulatory regimes and managing critical infrastructure. Operators that can simplify compliance, provide transparency, and absorb regulatory complexity on behalf of customers have an opportunity to convert a baseline requirement into a source of differentiation, particularly in the context of agentic AI-enabled solutions.

Concerns are most pronounced among small enterprises, where nearly half of respondents cite data privacy and security as their top consideration, significantly higher than among large and medium-size enterprises. This gap reflects structural differences in internal capabilities—smaller organizations typically lack dedicated security and compliance resources and therefore place greater reliance on external providers to manage risk on their behalf.

Across all segments, customer preferences converge on a clear deployment mandate. Nearly two-thirds of respondents require their data to run in private cloud environments they directly control or within on-premise infrastructure (Exhibit 13). Appetite for less restrictive models remains limited, with only a small minority expressing comfort with public cloud or provider-hosted environments. Innovation and scalability matter, but they do not outweigh the need for control.

Exhibit 12

Data security and compliance dominate agentic AI concerns.



Note: Figures may not sum to 100%, because of rounding.
[1]Question: What are your biggest worries when thinking of agentic AI?
[2]Large = >500 employees; medium = 100–500 employees; small = <100 employees.
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,214

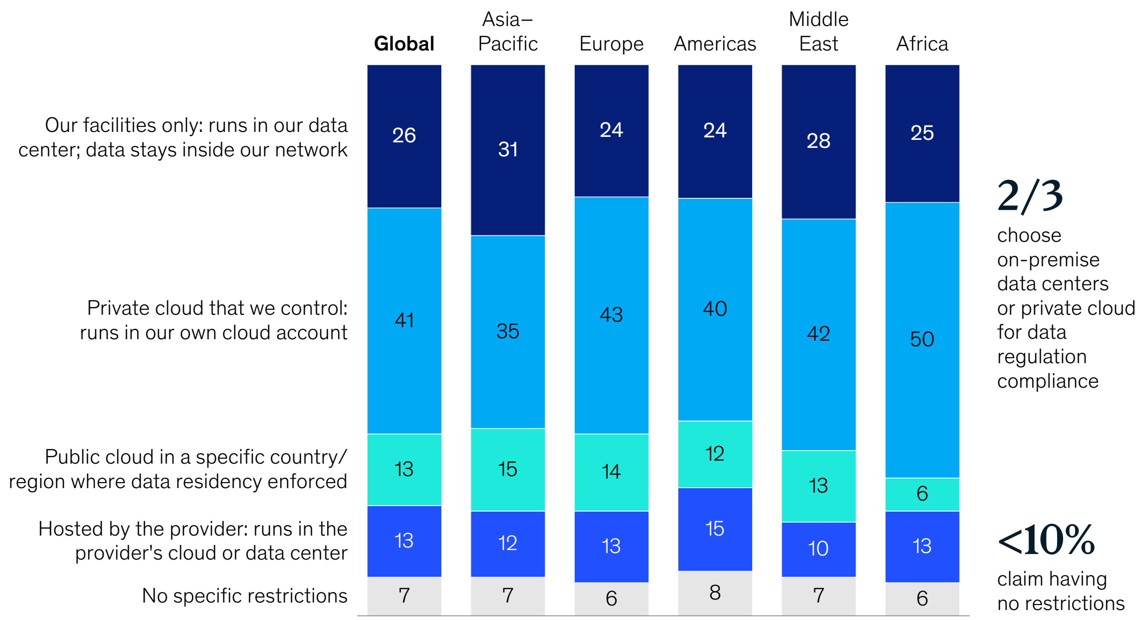
McKinsey & Company

Looking ahead, these dynamics are likely to intensify. As agentic AI adoption accelerates and enterprises delegate more data access and operational authority to autonomous systems, expectations around privacy by design, sovereign deployment, and regulatory assurance will rise in parallel. Providers that can natively support flexible deployment models and embed trust into AI-enabled offerings will be best positioned to capture growing demand. Those that cannot, risk being structurally sidelined as AI-driven buying behavior reshapes the market.

Exhibit 13

A preference exists for private cloud and on-premise data management.

Setup considered acceptable for where data is stored/processed,[1] % of respondents



Note: Figures may not sum to 100%, because of rounding.
[1]Question: Thinking about your company's data regulation and compliance policies, which setup is acceptable for where your data is stored/processed?
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,214

McKinsey & Company

Agentic AI adoption is growing, especially within customer care use cases

As agentic AI adoption moves from experimentation toward broader deployment, B2B customers are converging on a clear set of priority use cases. Customer care has emerged as the primary entry point: 80 to 90 percent of respondents report they have already implemented, or plan to implement, agentic AI in customer service processes (Exhibit 14). This consistency across industries reflects both the maturity of available solutions and the immediate, measurable impact on service efficiency and experience.

Beyond customer care, adoption is expanding into more operationally complex domains. IT and development functions, along with fraud and security, represent the next wave of high-prevalence use cases, with more than 70 percent of surveyed customers indicating active or planned implementation. Revenue-facing applications—particularly in sales and marketing—are more uneven but feature prominently in advanced industries such as technology, media, and telecommunications; retail; financial services; and energy and utilities, where competitive intensity and growth pressure increase the value of AI-enabled engagement and decision support. Taken together, these patterns suggest that agentic AI adoption is evolving toward more differentiated, industry-specific deployment paths rather than a one-size-fits-all trajectory. For B2B operators, understanding which use cases customers prioritize for optimization through agentic AI also provides a clear signal on where to focus internal efficiency efforts. By mirroring customer priorities internally, operators can accelerate capability development, prove impact at scale, and subsequently monetize these capabilities through deployment at customer sites.

Exhibit 14

Customer care leads agentic AI adoption, with more than 80 percent project adoption.

Existing and future use-case plans across B2B,[1] % of use cases selected among top 3 per industry either as already implemented or planned to be implemented within next 12–18 months

50 60 70 80 90

	Order orchestration	Field services	Billing and collections	Sales	IT/developer productivity	Marketing	Fraud and security	Customer care
Advanced industries	57	55	72	74	75	79	70	81
Energy and utilities	53	64	71	78	72	76	75	83
Financial services	50	71	71	78	70	72	80	81
Healthcare and pharmaceutical	56	59	72	73	76	73	76	80
Consumer packaged goods	61	57	73	76	77	76	72	82
Professional services	54	55	74	74	78	77	78	83
Public sector	54	63	69	61	75	66	77	84
Real estate	56	55	69	72	69	77	72	82
Retail	51	59	72	78	70	81	76	88
Technology, media, and telecommunications	52	64	68	78	82	70	81	79
Transport and logistics	54	60	68	67	66	72	78	84
Average across industries	54	61	71	74	74	74	76	82

[1]Questions: In which areas has your company already implemented agentic AI use cases? In which areas does your company plan to introduce agentic AI use cases in the next 12–18 months?
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,038

McKinsey & Company

Despite strong intent and widespread experimentation, however, at-scale implementation remains the exception rather than the norm. Among surveyed customers that have initiated agentic AI adoption, only 3 to 11 percent report full-scale deployment for any given use case (Exhibit 15). Even when including organizations that are actively scaling AI across multiple domains, adoption reaches only 7 to 21 percent. Across use cases, the largest share of respondents remains in the pilot phase, highlighting a persistent gap between ambition and execution.

Exhibit 15

Less than 11 percent of companies have fully implemented agentic AI solutions.

Current level of agentic AI adoption maturity by use case,[1]
% of respondents

0 5 10 15 20 25

Current agentic AI maturity[2]

	Fully implemented at scale	Scaling live use cases across multiple functions/teams	Initial pilots in limited areas
Customer care	11	21	23
Marketing	7	15	18
Fraud and security	9	14	15
Sales	7	14	15
IT/developer productivity	7	15	15
Billing and collections	6	14	14
Field services	6	11	11
Order orchestration	3	7	7

[1]Question: In which areas has your company already implemented agentic AI use cases?
[2]Question: With regard to agentic AI adoption, how would you best describe your company's current level of maturity?
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,308

McKinsey & Company

This execution gap is evident even in customer care, the most advanced and mature use case. While the majority of organizations are either piloting or scaling agentic AI in this domain, only around one in ten surveyed customers report having achieved full-scale deployment. Nearly half remain in early stages, with a substantial share still working to move beyond pilots and isolated team-level implementations.

Taken together, these findings point to a critical inflection point. Customer demand for agentic AI is increasingly well defined, but the ability to execute reliably at scale remains constrained. Providers that can bridge this gap—by integrating agentic AI into core processes, operating models, and governance frameworks—will be best positioned to capture value as adoption accelerates from pilots to enterprise-wide deployment.

4. Agentic AI creates immediate efficiency and a longer-term growth option for operators

For operators, this creates a dual opportunity: capture near-term operational gains while selectively building toward differentiated AI-enabled offerings.

Agentic AI improves efficiency while preserving customer trust

Agentic AI presents telecom operators with a near-term opportunity to materially improve internal efficiency. Applied to high-volume, operationally intensive interactions, agentic AI can reduce cost to serve, extend service availability, and improve responsiveness, while preserving human-led engagement where commercial stakes and relationship value are highest. Rather than replacing human interaction, agentic AI enables a more effective allocation of human effort.

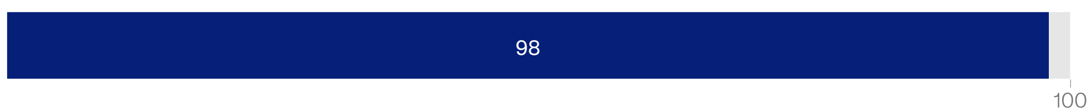
Customer acceptance provides a strong foundation for this shift. Across surveyed B2B segments, 98 percent of customers indicate openness to receiving AI-enabled assistance in at least one interaction (Exhibit 16). Acceptance is highest in service and support contexts, where speed, availability, and consistency are most critical. Between roughly half and two-thirds of respondents are open to AI-assisted 24/7 phone or chat support, and close to half would accept AI guidance when resolving technical issues consistently across enterprise sizes.

By contrast, tolerance for AI involvement is lower in commercially sensitive interactions. Only around one-third of customers are comfortable with AI support in sales and proposal activities, underscoring that trust, judgment, and negotiation remain distinctly human domains. This boundary is instructive. It suggests that the value of agentic AI lies not in full automation, but in intelligent orchestration.

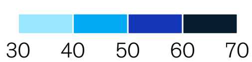
Exhibit 16

Ninety-eight percent of B2B customers are open to AI-assisted service.

Share of respondents open to be serviced with AI in at least one use case,[1] % of respondents



Areas where respondents are open to AI-enabled servicing from their connectivity provider,[1]
% of respondents



	Enterprise size[2]		
	Large	Medium	Small
24/7 answers by phone or chat: AI that replies to common questions without waiting for a human	62	56	52
Fix technical issues: AI that guides me step by step to resolve service or equipment problems	55	50	48
Self-service tasks online: AI that helps me change plans, add services, or check status	50	47	42
Prevent problems proactively: AI that monitors services and fixes issues before we notice them	50	46	41
Manage support cases: AI that opens, routes, and keeps me updated on support tickets	50	44	39
Billing and accounts: AI that explains charges and helps correct or change items	40	41	40
Sales and proposals: AI that learns my needs and drafts tailored offers/proposals	39	36	38

[1]Question: Thinking about how your connectivity provider supports you today and in the future, where would you be open to be serviced with the use of AI?
[2]Large = >500 employees; medium = 100–500 employees; small = <100 employees.
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,214

McKinsey & Company

For operators, the winning model is therefore hybrid by design. Agentic AI should be deployed to absorb volume, complexity, and variability at scale, freeing human experts to focus on high-value interactions that drive growth, differentiation, and long-term relationships. Operators that successfully rebalance this mix can unlock meaningful productivity gains in the near term, while strengthening service quality and customer trust rather than eroding it.

The provider market remains highly contested

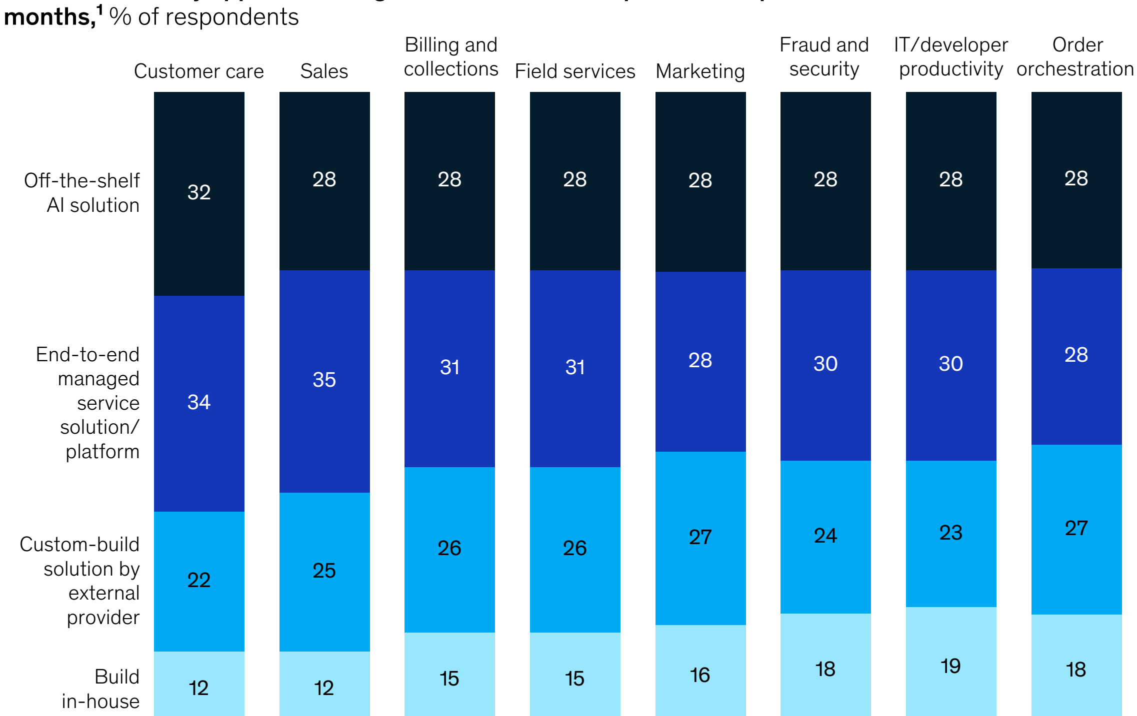
As agentic AI adoption accelerates, enterprises are making a clear build-vs.-buy choice and are increasingly opting to buy rather than develop capabilities in-house. Across use cases, the survey shows that roughly 80 percent of B2B customers prefer to source agentic AI through off-the-shelf solutions, managed services, or third-party-delivered custom builds (Exhibit 17). This preference holds consistently across major use cases, from customer care and sales to billing, marketing, and field services.

Despite this strong demand signal, the provider landscape remains fragmented. No use case requires a fundamentally distinct delivery approach, and no provider type has yet emerged as a clear winner. Customer preferences vary by application. Surveyed foundation model and AI-native providers currently lead stated preferences, particularly in customer care and sales, reflecting their early internal adoption and proven functionality. System integrators follow as a credible alternative.

Exhibit 17

Over 80 percent prefer agentic AI from external providers.

Preferred delivery approach for agentic AI use cases implemented/planned within next 12–18 months,[1] % of respondents



Note: Figures may not sum to 100%, because of rounding.
[1]Question: For each agentic AI use-case area you said your company has implemented or plans to introduce in the next 12–18 months, how would you prefer it to be provided?
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,214

McKinsey & Company

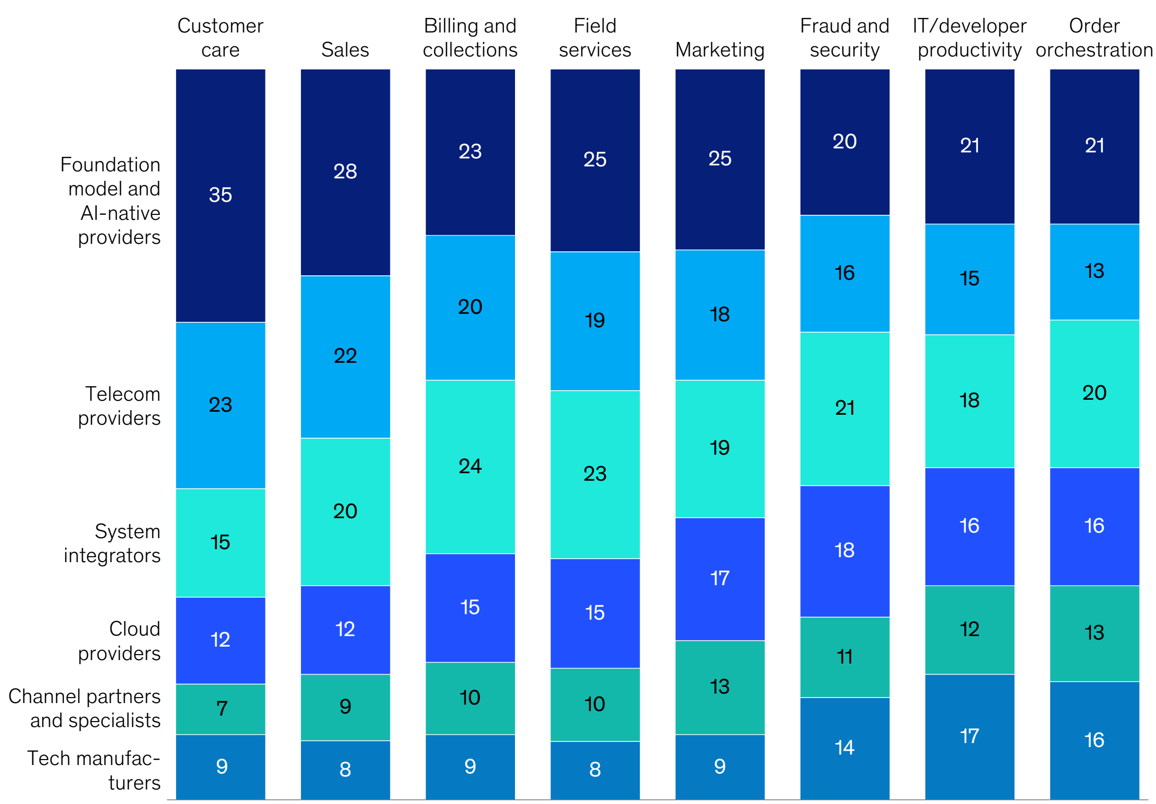
More unexpectedly, telecom operators also feature among the top three preferred provider types (for the next 12 to 18 months) across several use cases. This signal should not be read as a reflection of today's market reality; few operators currently offer a clearly defined, scaled agentic AI proposition. Rather, it appears to reflect a latent customer expectation—or even aspiration—around the role operators could play, anchored in trust, operational reliability, and deep integration into existing processes.

This lack of consolidation could be strategically significant. While customer openness to buying agentic AI solutions is clear, no operator has yet established a clearly defined, scalable agentic AI proposition with a proven go-to-market model (Exhibit 18). As a result, the opportunity for operators exists in principle, but remains early, uncertain, and largely untested. Realizing it would require rethinking value propositions, delivery models, and commercial packaging rather than extending existing offerings.

Exhibit 18

No provider type consistently dominates across agentic AI domains yet.

Preferred provider for agentic AI use cases implemented or planned within next 12–18 months,[1]
% selecting telecom as provider



Note: Figures may not sum to 100%, because of rounding.
[1]Question: For each agentic AI use-case area you said your company has implemented or plans to introduce in the next 12–18 months, who would you prefer to provide it?
Source: McKinsey analysis; McKinsey Global Technology and Telecommunications B2B Pulse Survey, Q1 2026, n = 3,214

McKinsey & Company

Customer behavior further clarifies where this opportunity is, and where it is not. When speed, standardization, and ready-to-deploy functionality are the primary decision criteria, enterprises gravitate toward AI-native, off-the-shelf offerings, which are preferred roughly twice as often as custom-built or fully managed alternatives. By contrast, integrated, managed, and domain-embedded deployments (those requiring deep process knowledge, trusted operations, privacy controls, and long-term accountability) remain less clearly served.

It is in these environments that telecom operators could selectively explore a differentiated role, particularly in domains such as customer care, sales, and billing, where they already own customer relationships and process context. Whether this opening can be translated into a scalable and repeatable business will depend less on technology access and more on disciplined execution, clear value articulation, and the ability to operate agentic AI reliably at scale in a fast-moving and still-maturing market.

Connectivity still underpins B2B telco value, but it no longer defines where growth is created. As value moves up the stack, agentic AI is accelerating both efficiency gains and new ways for operators to engage customers beyond the network. Capturing this opportunity will depend less on technology access than on execution, combining security, digital experience, and thoughtful AI–human orchestration.

Andrea Travasoni is a senior partner in McKinsey's Milan office, where **Giacomo Dolci** is a partner and **Luca Furlani** is an associate partner; **Dev Patel** is a senior partner in the Chicago office; **Naveed Niwaz** is a partner in the Copenhagen office; **Bulat Gaifeev** is a capabilities and insights specialist in the Lisbon office; and **Olimpia Gascó** is a consultant in the Geneva office.

Issue briefs

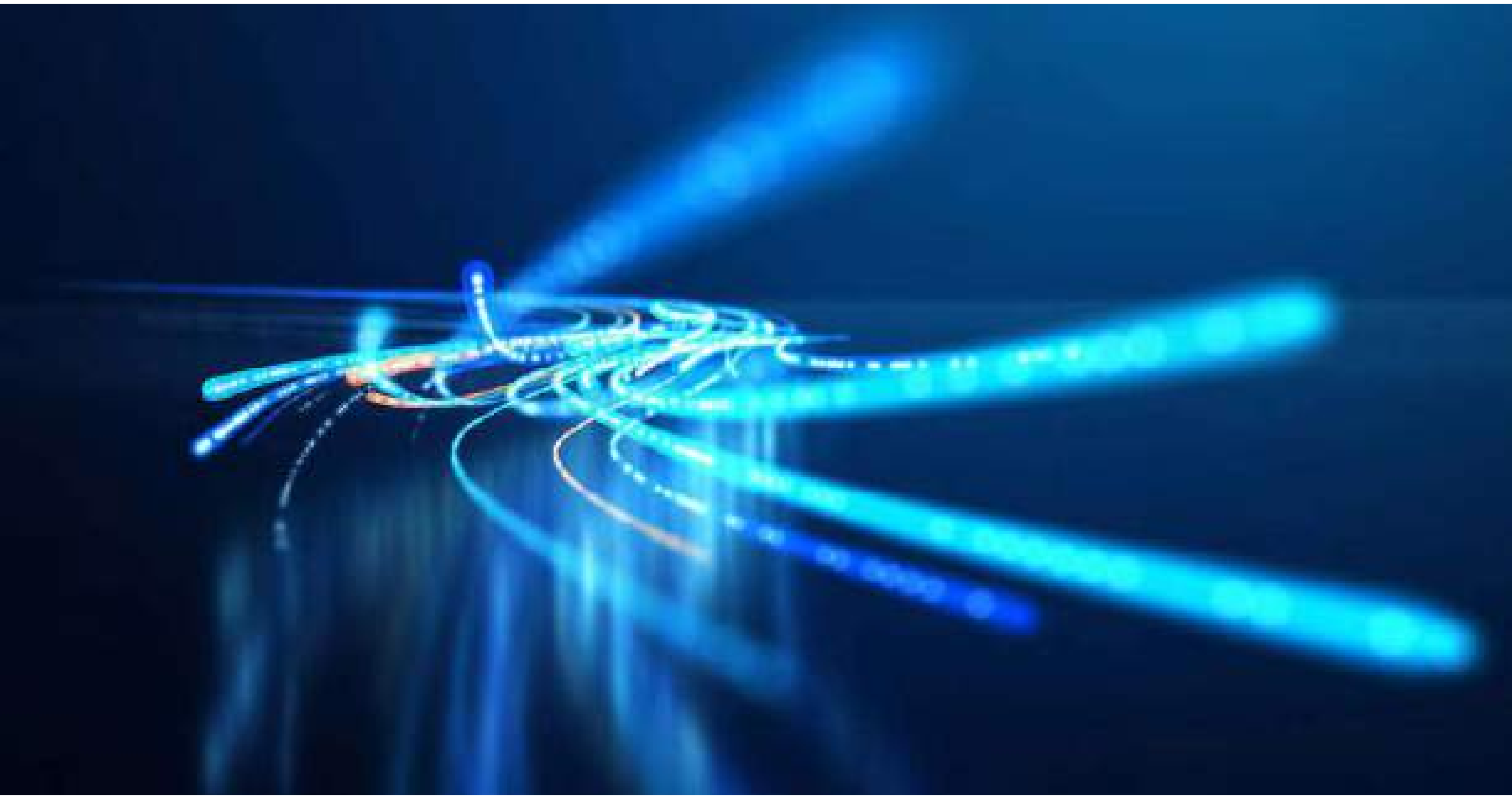
McKinsey & Company

Technology, Media & Telecommunications Practice

AI-driven telecom networks

Structural and cost pressures have leading telecom operators looking to AI to reset network economics, enable disciplined growth, and unlock new revenue streams.

This article is a collaborative effort by Gustav Grundin, Sebastian Cubela, and Tomás Lajous, with Borja Belda, Matyas Zetek, and Sebastián González, representing views from McKinsey's Technology, Media & Telecommunications Practice.



March 2026

Telecom network economics are under structural long-term pressure. Network complexity continues to rise, customer expectations for performance remain high, and cost pressures arise in the context of an expected data traffic growth deceleration in developed markets.

The rise of AI provides multiple opportunities for telcos to turn the tide. For starters, AI can be embedded in telco value propositions offering differentiated services. Operators can also participate in the emerging AI value chain by becoming the AI infrastructure backbone (as explored in an accompanying issue brief in this compendium). And telcos can leverage AI to transform their operations from the ground up.

Most directly, in an industry where improved operational efficiency has become central to the networks' value equation, AI represents a rare opportunity to reimagine the domain and reset its economics, while delivering exceptional performance. According to our latest survey of global telco top executives, the network domain will be one of the primary focus areas for AI deployments during the next two years, alongside customer care.

However, the difference between marginal gains and structural impact lies not in the technology itself but in how operators redesign processes, roles, governance, and budgeting around it. To capture these opportunities, operators need to raise their ambition and apply the same type of discipline to the network domain.

Recent progress in this field is encouraging. One of the loftiest and potentially most consequential ambitions many operators have articulated—a fully autonomous, self-optimizing, self-healing network—is no longer a distant vision, but an achievable goal in the coming years.

What's at stake?

AI's impact on networks spans both capital expenditure (capex) and operating expenditure (opex), across the following three major domains:

1) Value-led planning and simulation:
Network planning is shifting from static engineering thresholds to AI-driven, value-based optimization. To complement traditional network planning based on capacity and coverage estimates, advanced machine learning models and digital twins now simulate thousands of rollout and upgrade scenarios before capital is deployed. These simulations consider and estimate impacts on customer experience (CX), traffic evolution, customer churn and average revenue per user (ARPU), and potential competitive moves from other operators.

Operators deploying AI-based planning engines are already reporting success metrics:

— 10 to 20 percent lower greenfield rollout capex

— 15 to 35 percent lower upgrade capex

— Reallocation of 20 to 30 percent of annual capex away from low-ROI interventions

However, AI alone will not deliver results. Operators need to embed those simulation engines into their capital process and governance. Budget decisions need to be explicitly tied to the outputs, planning teams need to be trained to act on and improve model outputs, and engineering teams must provide feedback, so models consider new or unexpected restrictions.

2) AI-driven operations:
AI is also reengineering and optimizing network operations across multiple areas: In energy management, AI dynamically optimizes energy consumption by managing sleep features and detects anomalies without affecting service quality. In field operations, route optimization and automated scheduling reduce idle time and unnecessary dispatches. In maintenance, predictive models shift operators from reactive repairs to proactive interventions on critical assets.

Combined, AI-driven operational use cases can reduce total network opex by 15 to 30 percent.

To maximize value, operators need to redesign current workflows around human–AI collaboration, automating workflow steps, eliminating redundant handoffs, coordination efforts, and inefficiencies.

3) Automated repair and self-healing networks
Issue resolution and self-healing capabilities are emerging as one of the most widespread AI applications in network operations. Operators are deploying AI across the entire "issue management journey." For example:

Advanced anomaly detection and CX monitoring models identify potential faults early, even before customers are aware of a problem. Smart co-pilots and root-cause models analyze historical incidents and equipment documentation to recommend remediation steps, while dynamic matchmaking systems assign tickets to engineers with the most relevant expertise. Critical change agents automatically identify planned updates with disruption risks and design fallback or remediation plans to prevent major disruptions.

At scale, these capabilities have enabled operators to achieve 30 to 70 percent fewer troubleshooting tickets, leading to 55 to 80 percent reductions in network operations center costs, and 30 to 40 percent faster mean time to repair, alongside measurable improvements in customer experience.

What's ahead?

As operators move beyond pilots, four AI-enabled shifts are emerging:

— **Increased focus on "fixing the foundations:"** Network data has historically been fragmented and less mature than commercial data due to its complexity and volume. Scaling AI requires centralizing, governing, and standardizing network data, and defining a hybrid architecture (cloud and on-prem) to support AI use cases and workflow automations at different levels of the network.

— **Capturing institutional knowledge:** Much network know-how resides in engineers' experience. To enable network autonomy, operators are systematically codifying this expertise—through document ingestion, interviews, and structured modeling—so AI systems can infer network dependencies and remediation logic. This semantic layer becomes the foundation for self-directed decisioning.

— **Growth of network APIs and programmable connectivity:** AI-native and enterprise applications increasingly require identity, location, security, and quality-on-demand as programmable network functions rather than static connectivity. Network APIs allow operators to expose these capabilities directly to developers and enterprises, enabling differentiated, usage-based monetization. API coverage already spans around 80 percent of global mobile connections, positioning operators to participate in application-layer value creation rather than remaining transport providers.

— **Intelligent network services and cloud interconnect:** As AI workloads become more distributed, enterprises need software-defined network services to dynamically manage latency, routing, bandwidth, and regulatory constraints across cloud and hybrid environments.

What does it take to succeed?

Access to AI doesn't separate leaders from laggards; instead, it's the ability to embed AI as a continuous management capability across the network life cycle. Top performing operators take a business-led approach to AI, prioritizing use cases by value rather than technology, investing early in data foundations and digital twins, building internal capabilities beyond pilots, and partnering selectively with hyperscalers and AI specialists. Crucially, they recognize that AI is evolving at exceptional speed—models, tooling, and use cases are advancing every month.

AI is now the primary lever to reset network economics, protect margins in mature markets, enable disciplined growth in emerging markets, and unlock new, differentiated revenue streams. Operators that industrialize AI end-to-end can materially outperform peers on ROIC, EBITDA resilience, and long-term strategic optionality.

Gustav Grundin is a McKinsey partner in the Helsinki office; **Sebastian Cubela** is a partner in the Miami office; and **Tomás Lajous** is a senior partner in the New York office, where **Borja Belda** is an associate partner. **Matyas Zetek** is an associate partner in the Prague office, and **Sebastián González** is a consultant in the Connecticut office.

The authors wish to thank Abhimanyu Harlalka, Cindy Xiao, Eduardo Zamudio, and Manuel Palacios for their contributions to this article.

This article was edited by Daniel Eisenberg, an executive editor in the New York office.

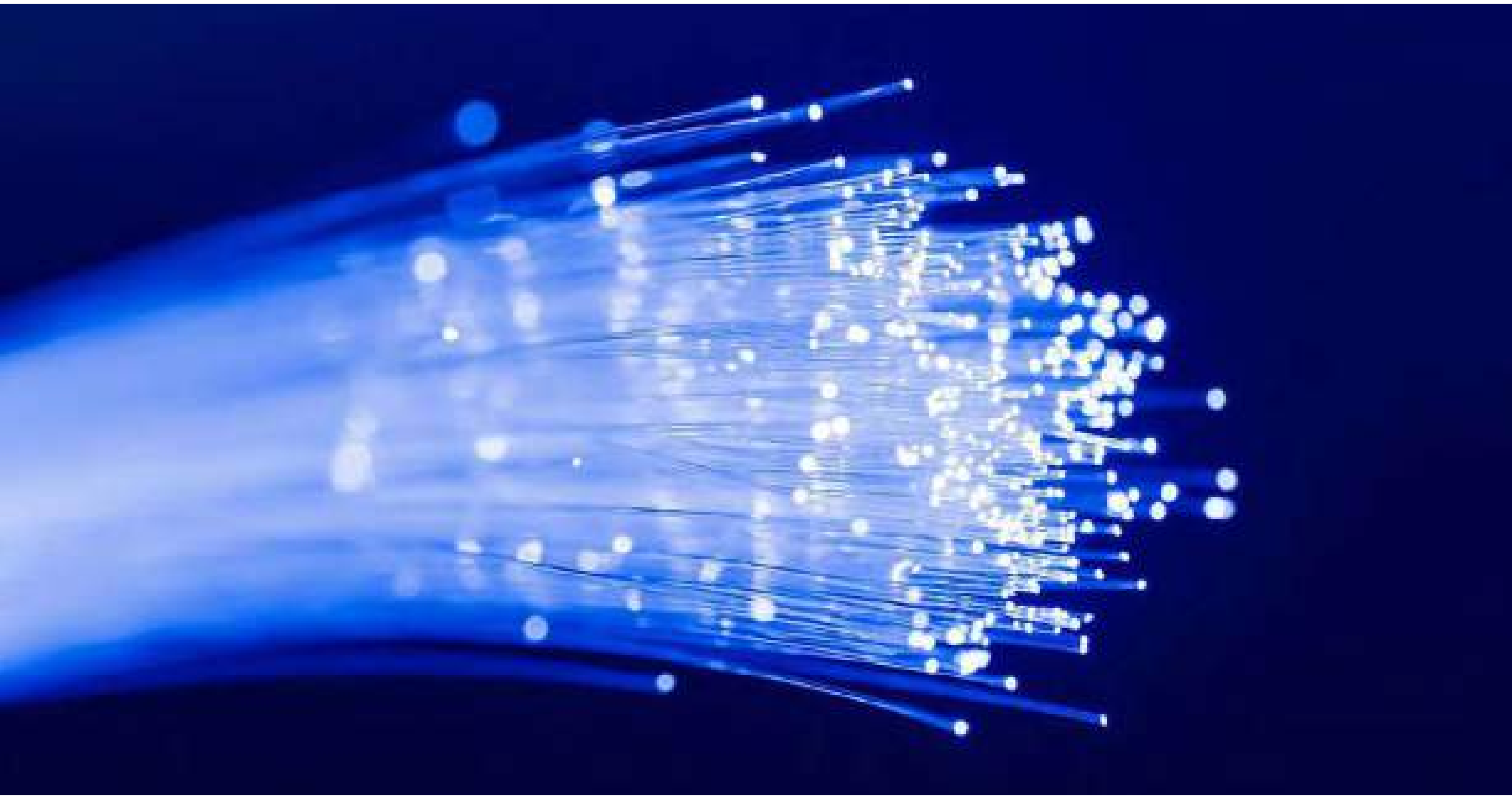
McKinsey & Company

Technology, Media & Telecommunications Practice

AI infrastructure

Playing a major role in enabling the AI era could reignite growth for telcos along the AI infrastructure value chain, utilizing assets many telcos already control.

This article is a collaborative effort by Gustav Grundin, Miguel Frade, Sebastian Cubela, and Tomás Lajous, with Borja Belda and Lorraine Salazar, representing views from McKinsey's Technology, Media & Telecommunications Practice and QuantumBlack, AI by McKinsey.



March 2026

AI's continued unprecedented growth depends heavily on accelerating investment in the physical and digital infrastructure required to train, deploy, and scale AI workloads—from data centers and fiber connectivity to intelligent networks, power, real estate, and accelerated, GPU-based compute. Historically, telcos have not captured a fair share of growth from tech disruptions that substantially increased data traffic such as the rise of video and social media consumption.

Today that dynamic is beginning to shift, giving telcos a new opportunity to play a central role in enabling the AI era. In the process, this could help reignite the industry's growth.[1]

The rapid adoption of generative and agentic AI is driving demand not only for centralized compute but also for distributed infrastructure closer to end users. Building and operating this infrastructure require assets that many telcos already have: extensive fiber networks, national footprints, space and presence at the edge, access to power, and experience managing complex, high-availability networks. The question is no longer whether telcos are relevant to AI infrastructure, but where along the value chain they can compete effectively and profitably.

What's at stake?

Between 2012 and 2025, global mobile data traffic grew by over 50 percent per year, while telecom service revenues barely increased by 1 percent annually, with hyperscalers capturing most of the economic value.

The AI wave offers telecom operators an opportunity to address that long-standing value capture challenge. Global data center demand could more than triple by 2030, reaching at least 170 gigawatts, driven largely by AI workloads. Importantly, this is not a single "bet-the-company" opportunity. The AI infrastructure value chain offers multiple entry points with different risk, capital intensity, and return profiles, allowing operators to align ambition with balance-sheet capacity and strategic intent.

Within this space, three value pools are particularly relevant for telcos:

- Fiber connectivity for new data centers in core and Tier 2 cities represents a global revenue opportunity of approximately $30 billion to $50 billion by 2030, as new facilities require high-capacity and often dark-fiber[2] connections.

- Beyond connectivity, opportunities are present in offering intelligent, software-defined network services to help enterprise customers manage AI workloads, control annual cloud data transfer fees, and meet latency and regulatory requirements. Commonly referred to as egress costs, cloud data transfer fees are estimated at $70 billion to $80 billion annually, creating scope for differentiated telco offerings focused on performance, cost efficiency, and control.

- Telcos may unlock new value by utilizing existing space, power, and cooling for distributed compute needs, including participation in the rapidly emerging GPU-as-a-service (GPUaaS) market, which is estimated to be worth $35 billion to $70 billion by 2030 (excluding hyperscalers). This opportunity is supported by accelerated compute workloads that are growing at more than 30 percent CAGR and expected to represent more than two-thirds of data center demand within five years.

---

[1] "AI infrastructure: A new growth avenue for telco operators," McKinsey, February 28, 2025.
[2] Dark fiber is unused or "unlighted" fiber optic cable.

Who are the key players and stakeholders?

The ecosystem is broad and increasingly competitive. Key stakeholders include hyperscalers and colocation providers, cloud operators, and GPUaaS players—many of which prefer dark fiber. Telcos enter this landscape as infrastructure owners and operators, often competing with—and simultaneously partnering alongside—hyperscalers. Other important players include the public sector; regulators that play a central role by shaping fiber access, data sovereignty, and licensing requirements; and key regulated industries that drive demand for sovereign AI.

Additional value chain participants include chip makers, software providers, systems integrators, and partners supporting data center retrofits and cooling solutions. Finally, energy providers and investors are increasingly influential, given power constraints and the capital intensity of AI-ready infrastructure.

What are the recent important developments?

As AI infrastructure scales and decentralizes, several developments are worth watching.

- **AI data center capacity is expanding and spreading to new geographies.** Hyperscalers and colocation providers have announced plans for more than 2,600 new data centers. Roughly one-quarter are to be located in cities with no existing data center footprint, pushing demand into new markets. Procurement preferences are also becoming clearer, with growing emphasis on dark fiber. Overall, the global footprint is expected to approach 11,000 facilities by the early 2030s.

- **Power and permitting constraints are reshaping where value is accruing.** Grid power limitations and long permitting timelines are slowing new builds, increasing the value of existing space and available power. This dynamic favors players that control "ready" assets—particularly sites with assured, contracted power in locations that can support distributed compute for inference.

- **Telco participation is shifting from pilots to execution.** Early movers among operators are signaling a transition from experimentation toward scaled delivery. Some are partnering with hyperscalers to provide dark fiber and edge compute; others are launching GPUaaS offerings or repurposing central offices or spaces with power and cooling to support AI workloads.

What are the biggest challenges?

Despite the opportunity, risks are substantial. There are still uncertainties around demand, particularly across use cases and geographies, as well as the technology and connectivity needs to enable computing at the scale required by AI providers. Competition from hyperscalers and cloud-native players is intense, with superior scale, capital, and developer ecosystems. And it's still not a given that investments can generate enough compute capacity to meet AI ambitions, while still delivering strong returns.

Commercial and competitive uncertainty also remains. Intelligent network services do not yet have a clearly established market definition, and telecom operators increasingly compete with hyperscalers and network service providers across connectivity, space and power, and GPUaaS offerings.

Capital requirements are high, especially for GPUaaS, dark fiber, and AI-ready data centers, while technology cycles are short and pricing pressure could intensify as supply expands. Many telcos lack the software, sales, and partnership capabilities required to compete beyond traditional connectivity.

What does it take to succeed?

Capturing value from AI infrastructure requires several shifts for telcos:

— Commercial capabilities including dedicated hyperscaler-focused sales teams, faster design and contracting cycles, digitized routing and asset data, and productized connectivity solutions are essential. GPU-as-a-service offerings require advisory-led selling and customer success capabilities. Close partnerships with hyperscalers are key to understand their evolving needs and new emerging technologies in this space.

— Partnering with and leveraging infrastructure investors, software providers, systems integrators, and retrofit specialists is key to lower capital needs, de-risk projects, and accelerate service delivery and facility readiness.

— Disciplined underwriting is important, including evaluation of fiber routes for multi-tenant upside and careful management of cost structures across compute, power, cooling, networking, and memory.

— Risk management for space and power, including prioritizing sites requiring limited retrofit or committing capital only after securing firm tenant demand, is critical; for edge inference, at least 500 kilowatts of power are typically required to achieve economic viability.

What should players be watching?

Telco leaders should monitor where hyperscalers and colocation providers expand into new markets and how power constraints reshape infrastructure economics. Operators have an opportunity to be early movers in the build-out of distributed compute or edge infrastructure, but much will depend on the level of demand for GPUaaS offerings and whether demand for intelligent network services picks up and can be successfully monetized. Ultimately, the biggest risk facing telcos may not be choosing the wrong path—but failing to choose at all.

Gustav Grundin is a partner in McKinsey's Helsinki office, and **Miguel Frade** is a partner in the New York office, where **Tomás Lajous** is a senior partner, and **Borja Belda** is an associate partner. **Sebastian Cubela** is a partner in the Miami office, and **Lorraine Salazar** is a senior knowledge expert in the Singapore office.

The authors wish to thank Arjita Bhan, Eduardo Zamudio, Manuel Palacios, Matyas Zetek, and Nicholas Shaw for their contributions to this article.

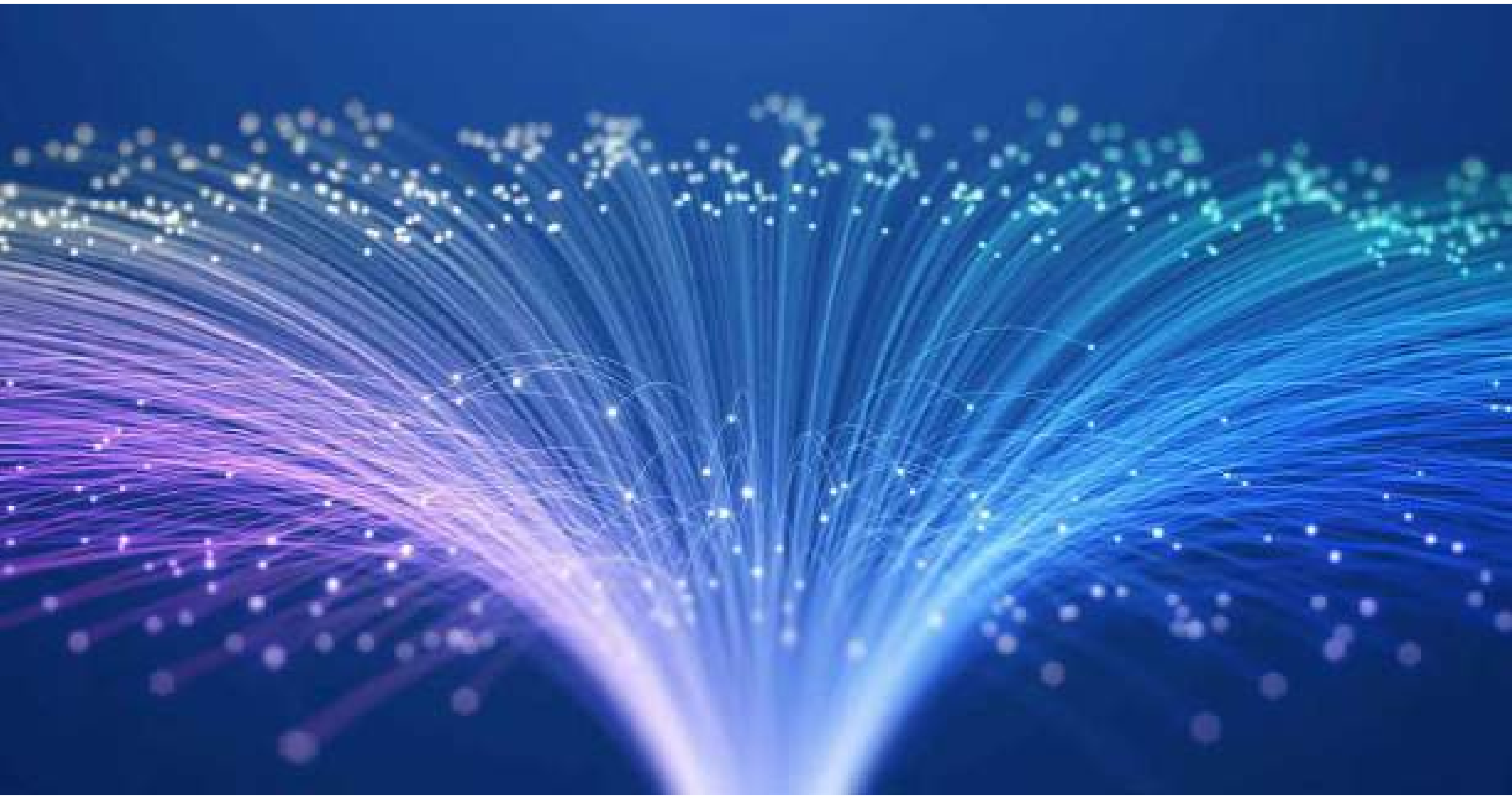
This article was edited by Daniel Eisenberg, an executive editor in the New York office.

Issue brief

B2C growth beyond the core

The beyond-the-core model offers telcos an attractive path to B2C growth as traditional growth engines are slowing. Leaders are learning how to capture the opportunity while addressing the challenges.

This article is a collaborative effort by Davide Schiavotto and Santiago Fernandez, with Louis Maus, Mate Deutsch, and Tamas Kabay, representing views from McKinsey's Technology, Media & Telecommunications Practice.



March 2026

Traditional B2C telco growth engines are increasingly constrained. Upsell opportunities face more challenges due to unlimited mobile voice and data plans, while fiber uptake is commoditizing previously premium, gigabit speed offerings. At the same time, many markets are running out of cross-sell opportunities due to fixed-mobile convergence (FMC) saturation; FMC has at least 50 percent penetration in major markets such as Belgium, Portugal, and Spain, with accompanying significant decline in linear TV viewership. Lastly, price increases are delivering decreasing returns, as less than 50 percent of headline price uplifts are typically retained once customers migrate to lower-priced, front-book offers.

As a result, telcos face a strategic question: whether to retain and maximize profitability via cost optimization levers or search for growth beyond the core. Our experience and research show that beyond-the-core value propositions can unlock new revenue streams, deepen customer engagement, and support long-term value creation; roughly 70 percent of consumers are willing to purchase or own at least one adjacent service via their telco provider.

Telcos are well positioned to capture this opportunity, given their trusted customer relationships, billing integration, extensive distribution and service footprint, and depth of proprietary assets such as data or existing partnerships. All these advantages are further accelerated by technological advancements (such as APIs and agentic AI) that make the expansion to new verticals easier than before. They are also reflected in rising telco investor expectations.

What's at stake?

The scale of the beyond-the-core opportunity extends well beyond incremental adjacency. By 2035, the integrated network economy (both B2B & B2C) is expected to represent a total global revenue pool of some $125 trillion. Within this, B2C-relevant connectivity adjacent domains such as housing (~$7.4 trillion), health (~$7.9 trillion), digital content (~$4.8 trillion), mobility (~$3.0 trillion), education (~$5.5 trillion), and B2C marketplaces (~$12.1 trillion), create meaningful headroom for telcos to expand their role in consumer value chains.

Beyond-the-core models also offer a structurally attractive financial profile. Distributor and B2B2C partnership models require limited incremental capital, leverage existing assets (retail, billing, customer care, field force), and can generate high-margin revenues while reinforcing the core business through higher engagement and reduced churn.

Importantly, investors increasingly recognize B2C and digital revenues as a driver of improved valuation multiples for integrated telcos relative to pure connectivity players.

What's happening?

Historically, beyond-the-core ventures have been difficult to execute. A previous McKinsey survey of top global telecom executives showed that a high share of new business builds failed or remained unprofitable in earlier waves of experimentation.[1] Recent evidence, however, shows that several leading telcos are now generating meaningful revenues from beyond-the-core activities. These operators derive between 5 and 40 percent of total revenues from adjacencies such as fintech, insurance, digital media, energy, and security, corresponding to annual revenues of ~€1 billion to ~€15 billion. These players demonstrate that beyond-the-core models can scale when pursued with focus, the right operating model, and strong ecosystem partnerships.

[1] "How telcos can succeed in launching new businesses beyond connectivity," McKinsey, February 25, 2022.

What does it take to succeed?

Despite growing traction, beyond-the-core growth remains challenging, with the primary barriers being organizational and operational rather than strategic. Traditional telco structures—designed for scale, reliability, and cost efficiency—often lack the speed, experimentation, and customer-centricity required to execute ecosystem plays effectively.

Leading players address these challenges through an innovative, platform-led approach with clear organizational focus. Once a telco has found the right vertical(s) relevant to its footprint, success depends on rigorous execution, including operating model (structure, process, people, technology) and link to core business P&L accountability and shared targets and incentives with core business.

If early results in these verticals are promising, longer-term, scalable success requires “bringing it together.” This entails defining the overarching customer value proposition with strong ecosystem synergies. As part of this, it’s critical to have the platform-first proposition design centered on orchestration, integration, and bundling, rather than as standalone products.

Davide Schiavotto is a senior partner in McKinsey’s Milan office; **Santiago Fernandez** is a partner in the Madrid office; **Lous Mas** is a fellow associate in the Munich office; and **Mate Deutsch** is an associate partner in the Budapest office, where **Tamas Kabay** is a research science expert.

This article was edited by Daniel Eisenberg, an executive editor in the New York office.

McKinsey & Company

Technology, Media & Telecommunications Practice

Satellite connectivity

As LEO consumer broadband and D2D connectivity push satellite capacity into the telecom mainstream, telcos must decide where to compete, partner, or ignore.

by Brooke Stokes and Javier Gil Gomez, with Karol Pyrzak

March 2026

Satellites are back in the telecom industry conversation, as the topic of former last resort connectivity is now landing on the desks and strategic agendas of telecom operators around the world. Two markets are moving from "niche" to more mainstream and relevant at scale: (1) Low-Earth-Orbit (LEO) consumer broadband and (2) direct-to-device (D2D) connectivity. Both shifts ride the same recent supply-side step change: Cheaper launches, higher cadence, mass-produced spacecraft, and falling terminal costs are turning satellite capacity into something that can be scaled and packaged like a consumer product.

Satellites won't replace terrestrial networks, but they will reset customer connectivity expectations at the "edges" of fixed and mobile, and those boundaries are expanding as LEO capacity scales. Because satellite players compete with a different economic model, they introduce new strategic uncertainty to the market and leave telcos with some key questions to grapple with: where to compete, where to partner, and where to ignore.

1. Consumer satellite broadband

As LEO capacity grows, consumer satellite broadband's target markets could expand significantly.

What's on the horizon?

Geostationary Earth orbit (GEO) broadband served remote households because terrestrial last-mile economics were poor, but latency and quality of experience limited adoption. LEO changes the experience profile. Latency is structurally far lower than GEO, enabling more demanding applications. Performance still varies with cell loading and handovers, but the addressable segment is no longer limited to the fully unconnected.

Two main adoption axes will affect take-up:

— **Strength of competing fixed alternatives:** Where fiber-to-the-home (FTTH) and modern cable are widespread, LEO will skew to rural coverage. Where fixed access is scarce, slow to build, or low quality (legacy xDSL, congested fixed wireless access [FWA]), LEO becomes a credible primary option.

— **Pricing sophistication and affordability thresholds in some markets:** LEO constellations are deployed globally, but pricing can be highly local. Providers can flex prices to maximize beam utilization, for instance, discounting in underfilled areas and selectively undercutting incumbents. That could pressure telco pricing norms. Upfront terminal costs, however, remain a material barrier and will likely cap adoption in some emerging markets.

What are the challenges?

So far, LEO broadband has been capacity-constrained above a certain household density (beams saturate and peak-time performance degrades). But sector roadmaps—including Starlink's V3 next-gen satellites—point to step-changes in capacity as larger spacecraft and higher launch cadence arrive; Amazon LEO and other entrants will add supply alongside Starlink's existing fleet. As "capacity in the sky" rises, the density constraint should ease, pulling competition from rural into suburban markets (and selected urban niches).

What could happen?

— **Capacity unlock plus local pricing.** With next-gen capacity (such as Starlink V3) and new constellations, LEO pushes into suburban markets, using so-called beam fill, reduced pricing to discount selectively and win share.

— **“Served-on-paper” broadband is vulnerable.** LEO targets households with weak real-world experience (congested FWA, copper-based infrastructure), not just rural areas.

— **ARPU compression to fill the capacity.** As satellite capacity scales and competition intensifies, increased satellite pricing erosion may anchor lower price expectations in competitive, non-fiber technologies.

2. Direct-to-device connectivity

The extent of D2D satellite connectivity market growth depends on which delivery models and commercial patterns become dominant and how the sector overcomes physical challenges.

What’s on the horizon?

D2D satellite connectivity is gaining attention not because it can replace terrestrial mobile service, but because it can cost-effectively close coverage gaps where propagation limits and site economics make towers impractical. Unlike consumer satellite broadband, D2D positions satellites as another mobile access layer, embedded in smartphones and integrated into spectrum, roaming, and plan design.

Two delivery models are emerging:

a. **Mobile network operator (MNO)-spectrum (partnered) model:** Satellites use an operator’s terrestrial mobile bands via commercial partnerships (such as Starlink–T-Mobile; AST SpaceMobile with different operators). While this has the ability to reach many existing phones with minimal user behavior change, scaling is limited by the number of partnerships and requires market-by-market deals with tight regulatory and interference coordination.

b. **Mobile-satellite-spectrum (MSS) model:** Uses dedicated mobile-satellite spectrum, typically enabled through OEM/chipset integration (such as Apple–Globalstar or Skylo). This has the advantage of clearer spectrum rights and more consistent cross-border footprint, but it depends on device penetration and multi-party ecosystem alignment (OEMs, chipsets, satellite operators).

What are the physical challenges?

Three physical constraints define near-term D2D: (1) Handset uplink: Unmodified smartphones are built for nearby towers and are capped by power, antenna gain, thermal limits, and battery; satellites can boost downlink, but not the uplink. (2) Indoor coverage remains weak because building penetration and clutter erode link margin and users often lack clear sky view. (3) Dense, urban capacity is structurally limited: satellite “cells” are large and shared, so capacity per km^2 is far lower than terrestrial dense reuse. As a result, early D2D fits low-rate messaging, safety, and intermittent voice in rural/outdoor settings; high-rate data degrades quickly as user density rises.

What could happen?

D2D go-to-market models are defined by who owns the customer/brand, who controls the user experience, and how integration is delivered (direct versus hub/platform). Four primary commercial patterns are forming:

— **MNO-led add-on/wholesale non terrestrial network (NTN):** The operator owns the customer, sells D2D as a plan add-on, and buys satellite capacity directly from the provider.

— **OEM-led feature:** The device maker owns activation and UX (free safety first, likely paid tiers later); the operator plays a limited bundling role.

— **Satellite-led retail/hybrid mobile virtual network operator (MVNO):** The LEO player owns the customer under its own mobile brand, uses terrestrial networks as MVNO/roaming for most traffic, and adds satellite as edge coverage.

— **Hub/platform-led enablement:** A third party provides a "one integration" layer (authentication/routing/charging/APIs) and can aggregate multiple satellite networks, accelerating multioperator and multicountry rollout.

There are three main scenarios for operators to plan for:

— **Safety becomes table stakes.** As emergency messaging/location standardizes, operators bundle and resell the service, with customers expecting the feature as part of premium plans.

— **NTN becomes "just another roaming layer."** Hubs normalize integration and accelerate cross-market launches.

— **Spectrum and regulation set the pace.** MNO model scales through partnerships; MSS scales with chipset penetration.

What can telcos do next?

1. **Map "underserved but paying" pockets** in fixed and mobile by micro-geography, where complaints, usage patterns, and churn signal a readiness to switch.

2. **Choose posture by segment and market.** Compete or partner (bundle/roaming) depending on market conditions and ignore altogether only where the market is genuinely saturated with high-quality fixed and mobile.

3. **Protect the experience layer.** Even when capacity is wholesale, telcos can win on packaging: plan design, billing, care, device programs, and predictable tiers.

Satellite connectivity won't replace terrestrial networks. But it will change what "good enough" looks like at the edges, and who owns the customer there.

Brooke Stokes is a partner in McKinsey's Southern California office, **Javier Gil Gomez** is a partner in the Madrid office, and **Karol Pyrzak** is a capabilities and insights specialist in the Wroclaw office.

This article was edited by Daniel Eisenberg, an executive editor in the New York office.

Technology, Media & Telecommunications Practice

Sovereign AI

Despite huge waves of momentum and large spend commitments, many sovereign AI initiatives embody intent without readiness. Sequencing is key to making sovereignty repeatable and adoption scalable.

This article is a collaborative effort by Alex Panas, Ali Ustun, Arnaud Tournesac, and Luca Bennici, with Justin De Niese, Kaavini Takkar, and Newfel Drahmoune, representing views from McKinsey's Technology, Media & Telecommunications and Public Sector Practices and QuantumBlack, AI by McKinsey.

March 2026

***Disclaimer:** This article is descriptive and analytical. It does not provide policy, regulatory, or national security advice, and it does not recommend specific national strategies. Examples are included for illustration of ecosystem patterns and should not be interpreted as endorsements or prescriptions.*

Over the last year or two, sovereign AI has moved from a niche policy concept to what many view as a strategic imperative. Business and political leaders increasingly see the capability to deploy and scale AI while retaining appropriate control across key dimensions of sovereignty as central to economic competitiveness, national security, and societal trust. But sovereignty is multi-dimensional (operational, technological, legal, and territorial). The practical goal is to apply sovereignty deliberately at the control points that drive trust, auditability, and accountability, rather than maximizing sovereignty everywhere. Yet many initiatives stall because they build "sovereign assets" (data centers, GPUs, sovereign regions, national model announcements) faster than they build the demand, governance, and trust needed for scaled adoption.

What's at stake?

Sovereign AI represents one of the largest opportunities in AI: Up to 30 to 40 percent of AI spending is influenced by sovereign requirements, comprising an estimated $500 billion to $600 billion market by 2030. But the strategic prize is broader than spend. The outcomes include (1) resilience and autonomy for critical services (reducing exposure to geopolitical shocks), (2) end-to-end risk management across data, models, and applications with controlled access and auditability, and (3) economic value capture through productivity and service-quality improvements, startup and scale-up formation, higher-value jobs, and new digital exports as ecosystems mature.

Who are the key players/stakeholders?

The sovereign AI ecosystem comprises four groups, each of which control a different constraint on progress:

— **Government (orchestrator, regulator, investor, anchor customer):** Governments set sovereignty goalposts and make trade-offs explicit (which workloads require strong sovereignty versus hybrid versus largely global). They build a "trust market" through standards, certification, and enforcement so adoption becomes repeatable, not bespoke. They catalyze supply through enabling policy and targeted investment (including permitting and energy planning). And they aggregate demand into multi-year frameworks and lighthouse programs to create an adoption flywheel and justify early investment.

— **Technology providers:** Providers span hyperscalers, regional cloud and data center operators, telecom operators, model developers, and integrators. Effective ecosystems rarely pick either hyperscalers or local providers. Instead, they design an architecture where players collaborate and compete at the layer where each has advantage, while localizing trust through sovereign controls.

— **Investors:** Investors bridge the "valley of uncertainty" before utilization is proven, funding the stack across energy, connectivity, data centers, cloud services, model development, applications, cybersecurity, and integration.

— **Enterprises:** Enterprises convert infrastructure into economic value. They decide where sovereignty truly matters, tier workloads accordingly, and redesign governance and operating models so AI can scale beyond pilots.

What are the recent important developments?

Three signals stand out in our research and survey results:

1. **Momentum and headlines.** Commitments are surging—billions toward GPUs, data centers, and sovereign regions—yet scaled outcomes lag.

2. **Demand is real but pragmatic.** More than 70 percent of enterprises see sovereign AI as strategically important, but decisions remain workload-specific and dominated by price, performance, and reliability.

3. **Execution is the constraint.** Many organizations have intent but lack the operating model and "trust fabric" to move regulated workloads quickly and repeatedly.

What are the biggest challenges?

Sovereign initiatives can fall victim to mis-sequencing: building assets before demand and governance are ready. Practically, this shows up as:

— **Intent without readiness:** While 72 percent of enterprises say sovereign AI is part of their 2026 roadmap, only 28 to 29 percent report having a sovereign AI strategy, a detailed action plan, or a budget set aside—and only 25 percent have workload tiering by sovereignty levels.

— **A perceived cost premium:** Sovereign solutions are widely perceived as 10 to 30 percent more expensive across the stack, making buyers selective about where sovereignty is worth paying for.

— **Slow transformation cycles:** Sovereign cloud and AI migrations typically take three to four years, driven less by technology constraints and more by governance redesign, risk/compliance alignment, operating model shifts, and workforce retraining.

— **Fragmentation and lock-in:** Without shared standards and interoperability, deployments become bespoke, slowing procurement and scaling.

What are the key success factors and critical enablers?

A set of key moves can make sovereignty repeatable and adoption scalable:

1. **Build a scalable trust fabric:** Codify sovereignty into standards, reference architectures, and certification regimes (including model governance expectations such as evaluation, monitoring, and documentation), so regulated industries can adopt quickly and repeatedly.

2. **Make sovereignty a portfolio decision:** Segment workloads into sovereign, hybrid, and global lanes to avoid all-or-nothing debates and focus controls where risk and exposure demand them.

3. **Design for interoperability:** Build portability and common control patterns so enterprises and startups can scale through interoperability rather than bespoke lock-in.

4. **Sequence execution in three waves:** A practical roadmap is: ***Wave 1:*** Baseline controls plus lighthouse demand; ***Wave 2:*** Scale shared infrastructure plus data ecosystems; ***Wave 3:*** Specialize in selected domains/languages and export solutions once the foundation is proven.

What should players be watching over the coming year?

Watch whether ecosystems shift from announcements to adoption mechanics: credible certification regimes and enforcement; governments bundling demand into multi-year frameworks; providers narrowing the perceived 10 to 30 percent premium via standardized, auditable controls and scale; and enterprises accelerating workload tiering plus operating model changes that shorten multi-year migrations. The leaders will be those that treat sovereign AI as an ecosystem build—sequencing demand, governance, and trust in step with infrastructure investment.

Alex Panas is a senior partner in McKinsey's Boston office; **Ali Ustun** is a senior partner in the Doha office; **Arnaud Tournesac** is a partner in the Paris office; **Luca Bennici** is a partner in the Dubai office, where **Justin De Niese** is an associate partner, and **Newfel Drahmoune** is a consultant; and **Kaavini Takkar** is an associate partner in the Seattle office.

This article was edited by Larry Kanter, a senior editor in the New York office.

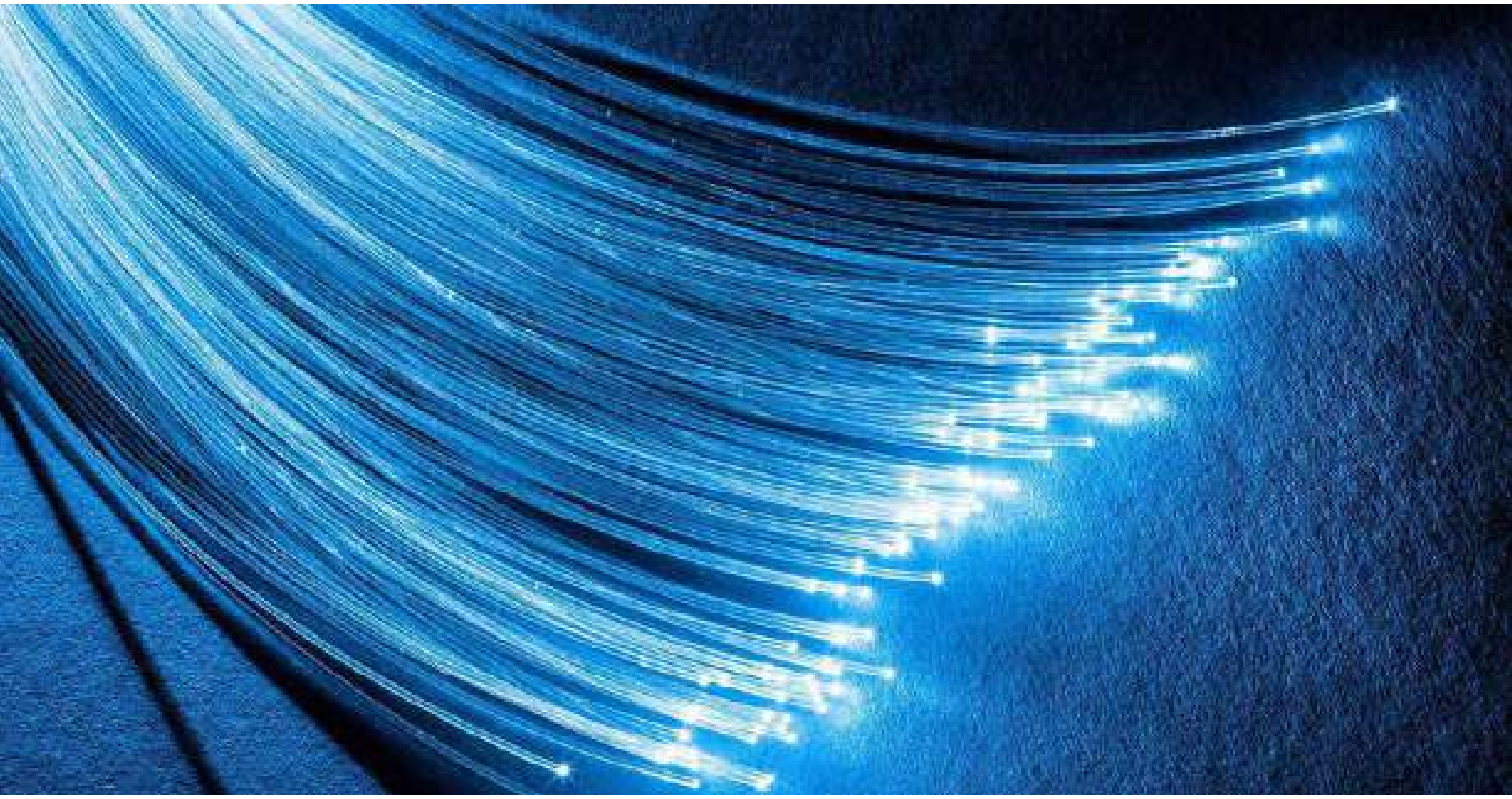
McKinsey & Company

Technology, Media & Telecommunications Practice

Telecom infrastructure

Telecom infrastructure is at an inflection point due to slower rollout and other growth challenges. Success creating value depends on whether players can transition toward a more integrated platform.

by Luca Fiandro and Nemanja Vucevic, with Lorraine Salazar and Shu Chern Lim



March 2026

Telecommunications infrastructure—anchored by towers and fiber networks—has long been a cornerstone of digital connectivity and a resilient investment class. Today both asset types face a shared inflection point: Traditional growth engines are slowing, even as demand for connectivity accelerates and diversifies, driven by data-intensive applications and new compute architectures.

Over the past two decades, telecommunications towers (which provide the physical locations for radio equipment that enables mobile connectivity) have benefited from rapid mobile adoption, network densification, and growing tenancy. Fiber networks, which form the backbone of fixed broadband, mobile backhaul, enterprise connectivity, and long-haul connectivity, have replaced copper as the future-proof wireline technology, delivering stable, infrastructure-like returns.

As coverage has expanded and markets mature, however, both towers and fiber face slower rollout and challenges to maintain utilization growth. These dynamics are pushing operators and investors to rethink how value is created, sustained, and differentiated across the infrastructure stack.

Though both remain foundational assets, the next phase of value creation will favor players that move beyond passive ownership and reposition their infrastructure as dynamic platforms at the center of the digital economy.

What's at stake?

Despite slowing deployment, infrastructure executives remain optimistic. In towers, 98 percent of executives expect long-run profitability to increase, reflecting confidence in the sector's ability to generate a second wave of value. For towers, the opportunity lies less in adding new sites and more in extracting additional value from existing assets. Future growth will depend on tenancy optimization, selective densification, and expansion into adjacent services including small cells, private networks, and edge data centers, rather than footprint expansion.

Fiber coverage has similarly reached scale—with approximately 80 percent of Europe, 70 percent of Asia–Pacific, and 60 percent of North America. Executives nevertheless expect rollout to continue, adding roughly 15 percentage points of additional coverage globally over the next five years. While rural and subsidized builds remain relevant, a new stimulus for growth is emerging: the rapid expansion of data centers, particularly in Tier 2 cities, where grid constraints and compute demand are reshaping fiber routing and investment priorities. Beyond coverage, the largest value pool lies in improving take-up, minimizing future duplication, and expanding the range of services delivered over fiber infrastructure.

Who are the key players and stakeholders?

In towers, key stakeholders include independent and telco-owned tower companies; mobile network operators as anchor tenants; infrastructure investors; and regulators governing site permitting and spectrum usage. Public authorities and enterprise customers are increasingly relevant as nontraditional tenants. More recently, there is growing interest in towers as potential locations for distributed compute supporting AI inference workloads, although this remains at an early stage.

The fiber ecosystem is broader and increasingly interconnected. It includes fiber operators, fixed and mobile service providers, infrastructure investors, hyperscalers, enterprise customers, public-sector stakeholders, and data center operators. Telecom operators remain the primary deployers of fiber infrastructure, but ownership and influence from specialized infrastructure players and investors continue to grow. Regulators play a critical role, shaping wholesale access models, fiber migration policies, and rural deployment subsidies. Adjacent participants—such as energy providers, private-network users, and edge data center customers—are becoming more important as fiber players extend their scope.

What are the recent important developments?

First, tower valuations have declined, with average transaction multiples falling from approximately 22 times EBITDA in 2022 to around 17 times in 2025, reflecting market maturity and slower organic growth. In response, attention is shifting away from pure site expansion toward active equipment, energy management, and operational efficiency.

Fiber deal multiples have likewise declined, albeit at a more modest pace, from 20 times EBITDA in 2022 to around 19 times in 2025. Fiber deal activity has also slowed, dropping at an annual rate of 20 percent between 2022 and 2025. Owners appear reluctant to sell at lower prices, leading to longer holding periods.

Third, massive investment in data centers to support AI workloads—both training and inference—is reshaping fiber demand. Constraints on grid power in major hubs are driving new subsea, long-haul, and metro fiber builds, particularly toward secondary markets and Tier-2 cities that can support incremental compute capacity.

Finally, low-Earth-orbit (LEO) satellites are emerging as a complementary technology. Executives expect LEO to capture close to 20 percent of existing fiber's addressable market, particularly in hard-to-reach areas, reinforcing a hybrid connectivity model rather than direct substitution.

What are the biggest challenges?

Both asset classes face structural constraints. Towers confront expiring or renewing anchor-tenant contracts and limited greenfield rollout opportunities, putting pressure on revenue growth. Fiber networks—particularly consumer-focused builds—struggle with low take-up rates, often near or below the roughly 30 percent threshold required for sustainable profitability. This challenge is compounded by overlapping networks and slow migration from legacy copper infrastructure.

Across both sectors, expansion into new services—such as active equipment, energy management, or enterprise connectivity—requires new capabilities and presents new risks, altering the traditionally infrastructure-like investment profile. Regulatory complexity, overbuilding, and operational inefficiencies further complicate value capture.

What does it take to succeed?

Six levers underpin successful value creation:

1. **Consolidation and optimization** to unlock scale efficiencies and avoid duplicative networks

2. **Selective new rollouts,** using more efficient designs such as small cells, "street work," or targeted expansion into Tier 2 cities to support AI infrastructure growth

3. **Extending the scope of services,** including radio access networks (RANs), energy management, and operations and maintenance for tower operators; and active services, enterprise connectivity, and cloud interconnection for fiber players

4. **Growth in adjacent areas,** such as small cells, distributed antenna systems, private networks, and edge data centers

5. **Boosting tenancy and take-up** through wholesale models, faster fiber migration, and nontraditional tenants

6. **Operational efficiency,** leveraging AI, analytics, standardized processes, and energy optimization, to reduce costs

What should leaders be watching over the coming year?

Key developments include the pace of tower and fiber consolidations; adoption of active-equipment and energy-management models; regulatory progress on fiber migration; the real impact of LEO satellites on capital allocation; and the accelerating race to build fiber infrastructure that supports AI-driven data center growth.

More broadly, success will hinge on whether infrastructure players can transition from narrowly defined, real-estate-like models toward integrated, engineering-led platforms—while preserving the stability investors expect.

Luca Fiandro is a partner in McKinsey's Milan office, **Nemanja Vucevic** is a partner in the Madrid office, **Lorraine Salazar** is a senior knowledge expert in the Singapore office, and **Shu Chern Lim** is a capabilities and insights expert in the Boston office.

The authors wish to thank Aleksander Matlok, Alex Mestre, Asif Khan, Boris Gusev, Federico Mentesana, Gerardo de Geest, Gustav Grundin, Jiaqi Zhang, Kah Man Leo, Lieven Verboven, Margherita Piccini, and Ming Shen Oon for their contributions to this article.

This article was edited by Daniel Eisenberg, an executive editor in the New York office.

McKinsey.com

@McKinsey
@McKinsey
@McKinsey