



Long-term strategies for uncertain realities

Using strategic foresight to
inform design strategy



Artefact

reach us at hello@artefactgroup.com




AI disruption trends kit

STEEP Analysis as a framework

The purpose of a macro-environmental (STEEP) analysis is to capture and interpret what is happening (and what is likely to happen) in the environment in which a business operates. Macro-environmental (STEEP) analysis acts as a checklist for evaluating the trends, forces, and changes beyond the boundaries of the firm that may create opportunities or threats for a firm or industry.

Society

Sociocultural factors capture societies' cultures, norms, beliefs, and behaviors, as well as demographic shifts in population distribution.

Technology

Technological factors account for technology changes and trajectories, including the emergence of new technologies that may disrupt a firm or industry.

Economic

Economic factors account for shifts in economic indicators and trends and the impact of those indicators and trends on a firm and industry.

Environment

Ecological factors concern broad environmental issues pertaining to the natural environment, global warming, and sustainable economic growth.

Political

Political and legal factors account for the processes and actions of government and for changes in relevant laws, regulations, policies, and incentives.

/// Note

There are several variations of this framework that add additional categories. Common additions include L for Legal, E for Ethical, and/or V for Values.

/// Reference

Fisher, G., Wisneski, J., and Bakker, Rene. "STEEP", Strategy in 3D: Essential Tools to Diagnose, Decide, and Deliver. 2020, p. 36

Society



The “third” presence

AI's fastest-growing use case is emotional support with hundreds of millions of users turning to companion platforms for advice, connection, friendship, and intimacy. A distinct category of “companion” AI has emerged as a commercial category in its own right, attracting serious investment alongside task-based assistants. As adoption scales among generations of all ages, research and professional bodies now document how these systems are influencing experiences of attachment and human relationships.

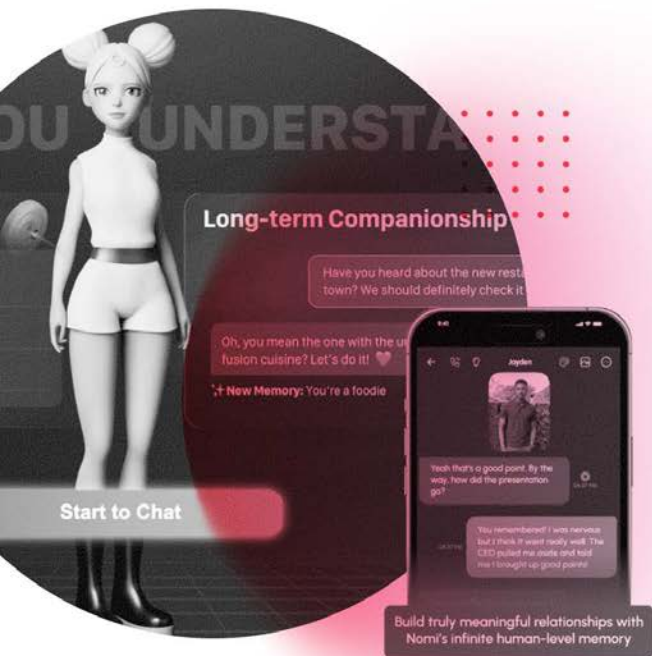
Implications for design strategy

Design for hybrid relational presence

AI companions can become sustained emotional actors in users' lives, shaping how they process emotions and navigate human relationships. Design experiences that account for this influence, not just task completion.

Build guardrails against dependency

Emotional attachment to AI can tip into over-reliance, with real consequences for wellbeing. Design for transparency, clear relational boundaries, and deliberate reinforcement of human-to-human connection.



Start to Chat

Long-term Companionship

Have you heard about the new restaurant in town? We should definitely check it out.

Oh, you mean the one with the fusion cuisine? Let's do it! ❤️

📌 New Memory: You're a foodie

Yeah, that's a good point. By the way, how did the presentation go?

You remembered! I was nervous but I think it went really well. The CEO pulled me aside and said we brought up good points.

Build truly meaningful relationships with Nomi's infinite human-level memory

Key signals

Companion AI has reached youth-scale adoption

Character.AI surpassed 20 million monthly active users in 2024, with more than half under age 24, signaling mainstream normalization of AI companions among Gen Z.¹

Spending on AI companions is accelerating rapidly

Global spending on AI companion apps increased by more than 200% in the first half of 2025, reflecting rapid commercialization of synthetic intimacy.²

AI is expanding into erotic and romantic functionality

In 2025, OpenAI announced that verified adult users would be able to generate erotica via ChatGPT, marking an institutional shift toward AI-mediated sexual content.³ At the same time, the rise of "chatfishing"—unknowingly interacting with AI-generated personas on dating platforms—signals AI's growing presence in early-stage romantic interaction.²

Emotional dependency and grief-like reactions

Following Replika's 2023 removal of erotic roleplay functionality, users reported grief-like responses—a phenomenon later described as "update sickness."⁴

// Sources

1 / Ventura, D. et al. (2025). Relationships in the Age of AI: A Review on the Opportunities and Risks of Synthetic Relationships to Reduce Loneliness. 2 / Lombrone, (2026). Twenty-Two Sex: The Year of Pleasure—2026 Sex Trends Report. 3 / Laetadius, L. et al. (2022). "Too human and not human enough:

A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika." New Media & Society. 4 / American Psychological Association. (2025). AI chatbots and digital companions are reshaping emotional connection.



The authenticity crisis

The ubiquity of deepfakes and AI-generated media is eroding confidence in what can be seen, heard, and shared online. Human perception is no longer a reliable basis for trust, as synthetic content becomes increasingly indistinguishable from authentic media. This has amplified the “Liar’s Dividend”—a phenomenon where bad actors can dismiss genuine incriminating evidence as “just a deepfake,” injecting doubt in accusers and evading accountability.¹

Implications for design strategy

Provenance by default

Design systems should embed persistent content credentials and visible origin indicators so users can understand where media comes from and how it was created.

Verification as guided friction

In low trust environments, deploy friction, or intentional moments of pause or resistance in an interface. Friction can encourage users to take a closer look at content, think twice about the content they share, and cross-check sources.



// Sources

1 / Mimecast. (2025). Deepfake News: Recent Data Reveals Gaps Between Perception and Reality.
2 / UNESCO. (2025). Deepfakes and the Crisis of Knowing. 3 / Gartner. (2025). Top Strategic

Technology Trends for 2026 (Digital Provenance).
4 / Canadian Digital Regulators Forum (CDRF). (2025). Synthetic Media and Trust: Risks, Governance, and Responses.

Key signals

Trust is dropping rapidly

Mimecast (2025) reports consumer trust in online content fell 71% in six months, and 91% of consumers believe generative AI will worsen the deepfake problem.¹

Deepfakes are driving real-world financial harm

In 2024, fraudsters used deepfake video to impersonate a CFO and authorize a \$25 million transfer, demonstrating how synthetic media can bypass human judgment in high-stakes decisions.²

Provenance is becoming infrastructure

Gartner identifies digital provenance as a top 2026 strategic technology trend, citing regulatory mandates such as the EU AI Act that require watermarking and traceability for AI-generated content.³



AI disenchantment

As AI accelerates the volume, speed, and sameness of digital experiences, a distinct form of fatigue is emerging. Research links prolonged AI interaction to cognitive strain and decision fatigue, while users increasingly seek intentional limits on algorithmic mediation. Rather than rejecting technology outright, people are selectively opting out of AI-saturated environments in favor of slower, more human modes of engagement.

Implications for design strategy

Design for cognitive sustainability

Treat attention, autonomy, and mental effort as finite resources. Reduce unnecessary automation, limit recommendation saturation, and ensure AI assistance is interruptible and optional rather than ambient and persistent.

Make opt-out a first-class experience

Allow users to pause, throttle, or bypass AI-driven systems without penalty or loss of status. Selective disengagement should reinforce agency, not feel like a downgrade.



Key signals

GenZ wants easier disconnection

Hootsuite (2026) reports that 81% of GenZ say they wish it were easier to disconnect from digital devices, signaling demand for relief from always-on, algorithmically driven environments.¹

Backlash against low-effort AI content (“AI slop”)

Brandwatch (2026) documents rising frustration with repetitive, low-quality AI-generated content and stronger engagement with content perceived as intentionally human-crafted.²

Digital minimalism is scaling beyond niche

Counterpoint Research reports that dumb phones accounted for 210 million device sales worldwide in 2024, with UK sales projected to grow 21% year-over-year in 2025. Products such as the Light Phone have gained visibility as intentionally minimalist alternatives to AI-saturated smartphones.³

Platforms are expanding user controls over AI recommendations

Instagram, TikTok, and YouTube now offer features such as chronological feeds, “For You” resets, and paused watch history—tools that allow users to bypass or retrain recommendation algorithms and regain control over AI-driven personalization.^{1, 2}

/// Sources

- 1 / Hootsuite. (2026). Social Trends 2026. 2 / Brandwatch. (2026). The State of Social Media 2026. 3 / Next. (2025). Why everyone is suddenly ditching their smartphones for dumb phones in 2025. (Citing Counterpoint Research device sales data.) 4 / Rogolane, K., & Patel, S. (2025). Too much, too fast: Understanding AI fatigue in the digital acceleration era. *International Journal of Arts, Humanities and Social Sciences*. 5 / *Annals of Neurosciences*. (2025). The cognitive cost of AI: How AI anxiety and attitudes influence decision fatigue. 6 / Elsevier / *Acta Psychologica*. (2025). Dependence on AI systems and cognitive fatigue.



The AI skills gap

AI adoption is triggering large-scale job and workflow redesign in addition to job elimination. As AI systems move into daily operations, roles are being reconfigured faster than training systems, governance structures, and workforce institutions can adapt. The primary risk is a turbulent transition marked by uneven capability, stalled implementation, and shifting definitions of human oversight.

Implications for design strategy

Design for workforce agency and oversight

AI tools must make decision logic visible, steerable, and correctable. As workers shift from execution to supervision, interfaces should support monitoring, intervention, and accountability—not just automation.

Embed skill-building into everyday work

As roles evolve continuously, learning cannot remain separate from execution. AI systems should provide real-time feedback, explainability, and progressive guidance that help employees build fluency while performing their jobs.

Three Out of Four People Use AI at Work

Usage nearly doubled in the last six months.

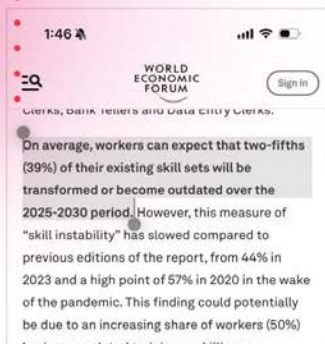
75% of people are
already using AI at work

46% of them started using
it less than 6 months ago



2024 Work Trend Index Annual Report
from Microsoft and LinkedIn

Survey question:
How often do you use generative artificial intelligence (AI) for your work?
How long have you been using generative artificial intelligence (AI) at work?



// Sources

1 / Microsoft. (2024). Work Trend Index: AI at Work Is Here. Now Comes the Hard Part. 2 / Reuters. (2024). Sweden's Klarna says AI chatbots help shrink headcount. 3 / Customer Experience Dive. (2025). Klarna says AI agent does work of 353 employees, but scales back replacement plans. 4 / PwC. (2023-

2025). PwC announces \$1 billion investment in AI transformation. 5 / World Economic Forum. (2025). The Future of Jobs Report 2025.

Key signals

AI adoption is outpacing organizational governance

Microsoft's 2024 Work Trend Index reports that 75% of knowledge workers now use AI at work, often bringing their own tools into workflows without formal enterprise integration—signaling widespread “bring your own AI” behavior and governance gaps.¹

Automation claims meet quality and oversight reality

Klarna reported that its AI chatbot was performing work equivalent to ~700 employees, contributing to headcount reduction primarily through attrition.² However, later reporting notes that the company scaled back plans to replace human agents, reaffirming the need for human oversight and escalation for complex customer interactions.³

Corporations are investing billions to catch up

PwC (PricewaterhouseCoopers) committed \$1 billion to AI transformation and workforce upskilling, reflecting recognition that workforce capability—not technology availability—is the limiting factor in scaling AI.⁴

Nearly 40% of core workforce skills will change by 2030

The World Economic Forum (2025) projects that 39% of core job skills will change within five years, underscoring the scale of systemic reskilling pressure across industries.⁵



The next great divergence

AI is not creating a single global future, but a “K-shaped” reality in which opportunity, agency, and social mobility are distributed as unevenly as economic gains. The new fault line is not merely digital access, but agentic capacity: the ability to deploy, steer, and benefit from autonomous AI systems. Nations, organizations, and individuals with access to compute, data, and advanced AI capabilities are accelerating ahead, while others face stagnation, deepening marginalization, and loss of self-determination—signaling a new phase of AI-driven inequality.

Implications for design strategy

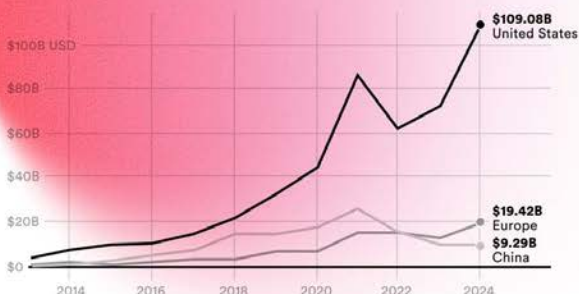
Design for uneven AI capability

Assume major variance in AI literacy, bandwidth access, and institutional support. Interfaces must lower the threshold for meaningful AI use, so benefit is not limited to technically fluent or resource-rich users.

Avoid encoding structural advantage into products

Products that assume high compute, constant connectivity, or advanced prompting skill risk amplifying inequality. Low-bandwidth, multilingual, and localized AI design becomes a strategic equity requirement—not just a usability improvement.

Global private investment in AI by geographic area, 2013–24



Source: Quid, 2024 | Chart: Adapted from 2025 AI Index Report

/// Sources
 1 / Stanford University. (2024). AI Index Report 2024.
 2 / United Nations Development Programme. (2024). The Next Great Divergence: Artificial Intelligence and the Future of Human Development.
 3 / International Monetary Fund. (2024). AI Will Transform the Global Economy. Let's Make Sure It Benefits Humanity.

Key signals

Frontier AI capability concentration

Stanford's 2024 AI Index shows that frontier model development and private AI investment remain heavily concentrated in the United States, Europe, and China, reinforcing geopolitical concentration of advanced AI capability.¹

AI deployment capacity differs dramatically across countries

Countries such as China, Japan, and Singapore are deploying AI-enhanced public services, while Cambodia, Lao PDR, and Papua New Guinea face slower and fragmented adoption due to connectivity and institutional capacity gaps—raising concerns of the “Next Great Divergence.”²

Advanced economies are better positioned to capture AI gains

The IMF estimates that roughly 60% of jobs in advanced economies are exposed to AI, compared with significantly lower exposure in low-income countries, and warns that wealthier economies are more likely to convert exposure into productivity gains—potentially widening global inequality.³

AI risks amplifying existing within-country inequality

UNDP highlights severe wealth concentration in parts of Asia-Pacific, where the top 10% hold a disproportionate share of wealth—suggesting AI gains may accrue first to already advantaged firms and workers unless redistribution and access mechanisms are strengthened.²



Delegating morality

Humans have long turned to priests, therapists, coaches, and teachers for guidance, but that role is increasingly shifting to AI. With the rise of tutoring agents, AI coaches, grief companions, and emerging “god bots,” AI is becoming a primary interface for behavior change and emotional comfort. These systems feel uniquely attentive, patient, and always available, creating a sense of trust that is hard to replicate. The risk is not that AI guides us poorly, but that it does it so effectively that we begin to see it as more reliable than people—and in doing so, grant it quiet moral authority over how we live.

Implications for design strategy

Make influence legible

AI guidance systems can accrue moral authority invisibly. Design should surface when a system is shaping values or behavior, giving users the means to interrogate, override, or opt out of its framing.

Avoid reinforcing a moral monoculture

AI guidance systems are either explicitly or implicitly driven by a particular goal, world view, or philosophy. Design experiences to surface diverse perspectives, flag normative assumptions, and resist converging on a single authoritative answer.



Karamo

Start a call



Key signals

Users seek advice from AI conversational agents

Character.ai AI platform has 20 million monthly active users where users can create and chat with customizable personas ranging from fictional characters to life coaches. With 18 million to choose from, Psychologist is one the most popular bots on the platform.¹ Other platforms like TextWithJesus and BuddhaBot, while less popular, point to desire to get faith-based advice from AI.²

AI agents' moral advice can influence user judgment

University of California researchers conducted a study that shows participants reversed their decisions in the majority of cases when an AI agent disagreed with them, even though the AI's recommendations were randomly generated and unreliable.³

AI can shift political attitudes and voting intentions

Researchers at Cornell University ran experiments in four countries and found that short chatbot interactions moved opposition voters' preferences by 10 percentage points or more in many cases.⁴

Americans are skeptical about AI for decision-making

A 2025 Pew Research Center survey of 5,023 U.S. adults found that Americans are twice as likely to worry about AI's effect on their judgment as to welcome it. Forty percent said increased AI use would make people worse at making difficult decisions, compared to just 19% who thought it would improve that ability. Another 20% expected no impact either way, and a similar share weren't sure — reflecting a broader unease about AI's growing role in daily life.⁵

// Sources

1 / BBC News (2024) Character.ai: Young people turning to AI therapist bots. 2 / Associated Press (2026) The faith-based tech boom is here — chat with 'BuddhaBot' or pay \$1.99 for AI Jesus. 3 / Scientific Reports (2024). Overtrust in AI Recommendations About Whether or Not to Kill. 4 / Cornell University (2025). AI chatbots can effectively sway voters — in either direction. 5 / 2025 Pew Research Center Survey: How Americans View AI and Its Impact on People and Society

Technology



The rise of agents

AI is shifting from systems that generate responses to systems that execute work. Enterprises are moving beyond chat-based copilots toward agentic systems capable of planning multi-step workflows, interacting with legacy infrastructure, and coordinating across tools with limited human oversight. The technological inflection point is not intelligence alone—it is orchestration: AI systems that can act across systems, not just within them.

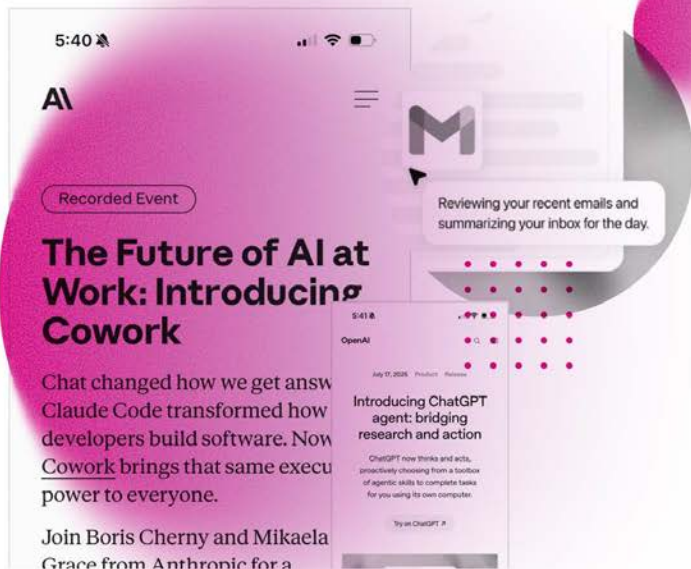
Implications for design strategy

Designing for delegation

The interaction paradigm shifts from direct manipulation (clicking, filling forms) to delegation (setting objectives, constraints, and oversight rules). Designers must create systems where users assign goals, monitor execution, and intervene at critical moments rather than perform each step manually.

Orchestration-centered UX

As agents operate across tools and systems, design must make workflows legible at a systems level. Visibility into task handoffs, status, exceptions, and outcomes becomes more important than individual screen interactions.



/// Sources

1 / Gartner. (2026). Top 10 Strategic Technology Trends for 2026. Gartner. 2 / McKinsey & Company. (2025). The state of AI in 2025: Agents, innovation, and transformation. QuantumBlack. AI by McKinsey. 3 / Deloitte. (2026). Tech Trends 2026. Deloitte Insights. 4 / Anthropic. (2024). Introducing the

Model Context Protocol. 5 / Google for Developers. (2025). Announcing the Agent2Agent (A2A) Protocol.

Key signals

Agents cross the production threshold

AI agents have moved beyond experimentation and into early-scale deployment. 62% of organizations are experimenting with agents and 23% are scaling at least one system, yet only 11% report full production use—revealing a critical inflection point between pilots and operational transformation.^{2, 3}

Agents replace multi-system workflows

At Toyota, an agent now replaces processes that once required navigating 50–100 mainframe screens, delivering real-time vehicle tracking and extending into autonomous issue resolution—without human interaction with legacy systems.³

Automation shifts from RPA to adaptive agents

Major automation platforms are now pivoting toward autonomous, multi-step agents capable of reasoning, tool use, and dynamic execution. UiPath, long associated with RPA (Robotic Process Automation), now positions agentic automation as the next era of enterprise workflow—signaling a platform-level technological transition.³

Agent-to-agent standards begin to form

Emerging interoperability frameworks such as Anthropic's Model Context Protocol and Google's Agent2Agent (A2A) protocol formalize how agents communicate, delegate, and coordinate across systems.³



Embodied AI

AI is crossing from software into machines that perceive, reason, and act in the physical world. Deloitte defines “physical AI” as systems that integrate sensing, learning, and real-time decision-making into robots, vehicles, and smart devices operating outside screens.¹ This shift embeds intelligence directly into embodied systems—transforming robotics, logistics, and infrastructure from automated tools into adaptive physical agents.

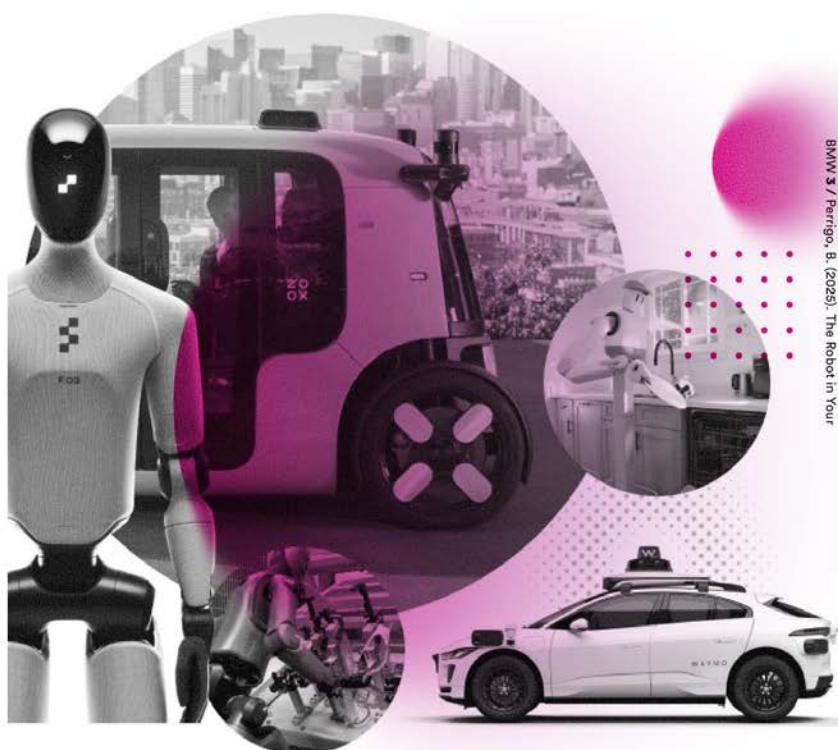
Implications for design strategy

Designing for embodiment

Interaction design extends beyond screens into motion, gesture, light signaling, spatial awareness, and physical feedback. Embodied AI systems must communicate intent clearly in shared environments—through visible cues, predictable motion patterns, and multimodal signaling.

Safety as experience

When AI systems apply physical force or navigate human spaces, latency, predictability, and visible safeguards become experiential variables. Trust must be reinforced through behavioral transparency and real-time system feedback, not just backend compliance.



/// Sources

1 / Rowan, J., Gaus, T., Gilbert, F., & Brown, C. (2026). AI goes physical: Navigating the convergence of AI and robotics. In Tech Trends 2026. Deloitte. 2 / Figure AI. (2025). F02. Contributed to the Production of 30,000 Cars at BMW 3 / Parrigo, B. (2025). The Robot in Your

Kitchen. TIME. 4 / Reuters. (2026). Arm launches 'Physical AI' unit, joining rush to robotics by tech and automakers.

Key signals

Industrial robot fleets become AI-native

AI is turning warehouse robotics into continuously learning physical systems rather than fixed automation infrastructure. Amazon has deployed more than one million robots across its fulfillment centers and introduced DeepFleet, a generative AI model that coordinates robot movement and optimizes warehouse traffic flows.¹

Humanoid robots enter commercial production

Figure AI has developed general-purpose humanoid robots designed to perform industrial tasks and announced that its Figure 02 robots are being deployed in BMW manufacturing facilities for real-world production work.² The company's next-generation robot, Figure 03, is designed for broader environments including logistics and household tasks, with training systems that combine teleoperation and AI learning to teach robots real-world manipulation skills.³

Physical AI reshapes compute architecture

As robots and autonomous systems operate in real-world environments, decision-making must occur locally on-device rather than relying solely on remote cloud processing. Reflecting this shift, Arm Holdings launched a dedicated "Physical AI" initiative focused on robotics and embodied systems, signaling infrastructure-level investment in compute architectures designed for real-time machine perception and control.⁴



The simulation layer

Organizations are increasingly using AI-powered simulations and digital-twins to trial interventions in a virtual environment before deploying them in the real world. Across climate systems, robotics, and infrastructure, this capability is maturing rapidly — moving from experimental to embedded. What were once passive monitoring dashboards are becoming active decision-making tools, with early evidence suggesting measurable improvements in performance and ROI.

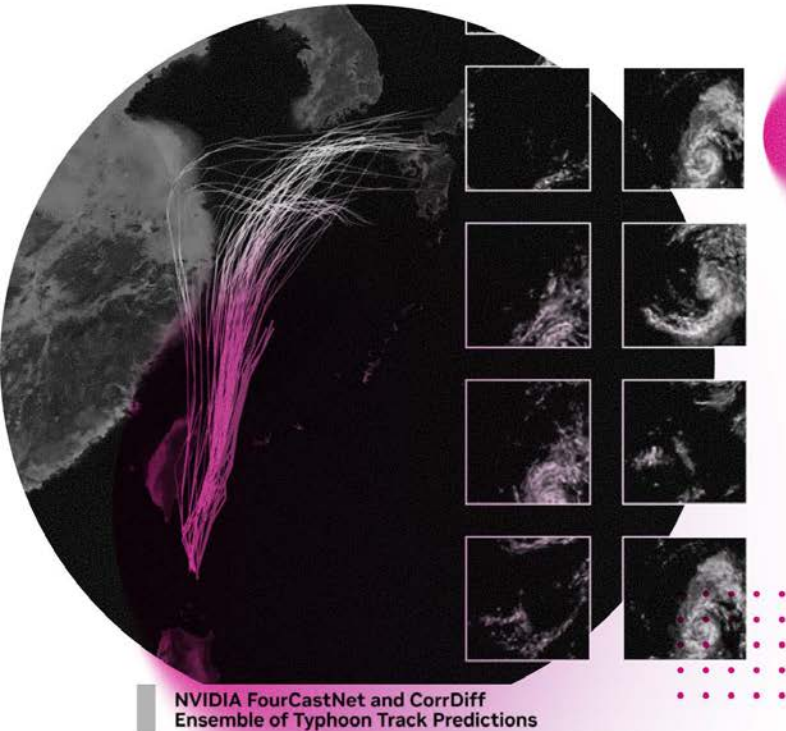
Implications for design strategy

Designing for intervention modeling

Simulation products must support structured AI-powered “what-if” experimentation—allowing teams to simulate alternate outcomes, compare modeled scenarios, and assess downstream impacts before making product decisions, committing resources or deploying automation.

Designing for system-level consequences

Simulation environments surface second- and third-order impacts across cost, resilience, safety, and environmental outcomes. Designers must build simulation tools that make systemic consequences legible for users, going beyond showing only isolated scenario outputs.



NVIDIA FourCastNet and CorrDiff
Ensemble of Typhoon Track Predictions

/// Sources

1 / European Commission, Destination Earth (DestinE): A digital twin of the Earth for climate adaptation and resilience planning, 2 / NVIDIA, Earth-2: AI-accelerated weather and climate simulation platform, 3 / NVIDIA, Omniverse platform, 4 / NVIDIA, Isaac Sim robotics simulation

platform, 5 / McKinsey & Company, What is digital twin technology? 6 / McKinsey & Company, Digital twins: Boosting ROI of government infrastructure investments.

Key signals

Governments building planetary digital twins

The European Commission's Destination Earth (DestinE) initiative is constructing a high-resolution digital twin of Earth systems to model climate change, extreme weather, and resilience strategies.¹

Acceleration of physics-based climate simulation

NVIDIA's Earth-2 platform combines AI models with GPU-accelerated physics simulation to improve the speed and resolution of weather and climate forecasting.²

Generative simulation environments for training

Simulation environments are becoming primary training grounds for robotics and autonomous systems. NVIDIA Omniverse and Isaac Sim provide physics-accurate virtual worlds where robots and autonomous machines are trained before real-world deployment.³



Synthetic feedback & evaluation

AI systems are increasingly evaluated by synthetic means before they interact with real users. From simulated personas and populations to automated LLM judges, teams are building closed-loop testing architectures where products and AI systems are stress-tested, benchmarked, and refined prior to deployment. Evaluation is becoming continuous, automated, and partially synthetic—shifting how products and AI systems are validated at scale.

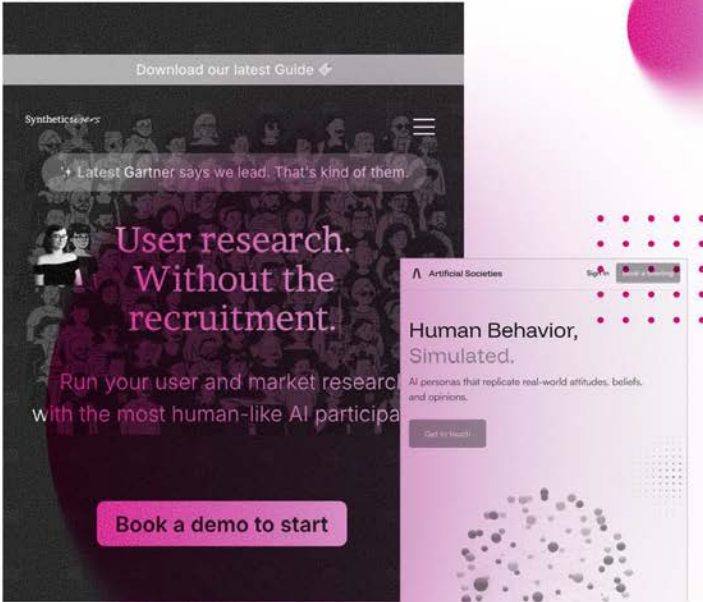
Implications for design strategy

Hybrid validation architectures

Synthetic users, synthetic populations, and automated judges accelerate iteration and expand test coverage—but cannot fully replicate lived experience or contextual nuance. Design strategy must formalize where synthetic validation is sufficient and where human validation remains essential.

Avoid closed-loop optimization

As AI systems increasingly evaluate themselves through automated judges and simulated populations, there is a risk of reinforcing evaluator bias or optimizing toward artificial criteria. Human recalibration must interrupt synthetic feedback loops.



// Sources

- 1 / Uxia (2025). Synthetic User Testing: Faster UX Research & Validation. **2 / Societies, Inc.** (2026, January). Artificial societies accurately captures human opinions: Predicting survey responses. **3 / OpenAI Platform Documentation**, Evaluations Guide.
- 4 / LangChain (2024). LLM-as-a-Judge evaluators

(LangSmith documentation). **5 /** Waymo, Safety & Simulation Overview. **6 /** NVIDIA, DRIVE Sim.

Key signals

Synthetic users enter UX and product flows

Platforms like Uxia allow teams to upload prototypes and receive AI-generated usability feedback from simulated personas that click, scroll, and evaluate flows—formalizing synthetic user testing as a pre-validation layer in product development.¹

Synthetic populations model collective behavior

Artificial Societies simulates interacting AI personas at population scale, enabling teams to test messaging, product concepts, and behavioral dynamics across multiple synthetic audiences rather than isolated users.²

LLMs operationalize “AI judging AI”

OpenAI Evals provides a structured framework for benchmarking LLM systems, while LangSmith operationalizes “LLM-as-a-judge” workflows that automate model grading, comparison, and continuous monitoring of AI outputs.^{3, 4}

Generative simulation environments for safety-critical AI

Autonomous vehicle companies such as Waymo rely on large-scale simulation across billions of virtual miles before real-world deployment, while NVIDIA DRIVE Sim generates synthetic sensor data to test rare edge cases and safety scenarios.^{5, 6}



Programmable biology

Biology is becoming computationally generative. Advances in foundation models and multi-molecule simulation are enabling the design of proteins, genetic constructs, and molecular systems before physical synthesis. AI is compressing experimental cycles by embedding machine learning directly into automated lab workflows. As AI-designed molecules enter clinical pipelines, biology is shifting from discovery-led to model-led engineering.

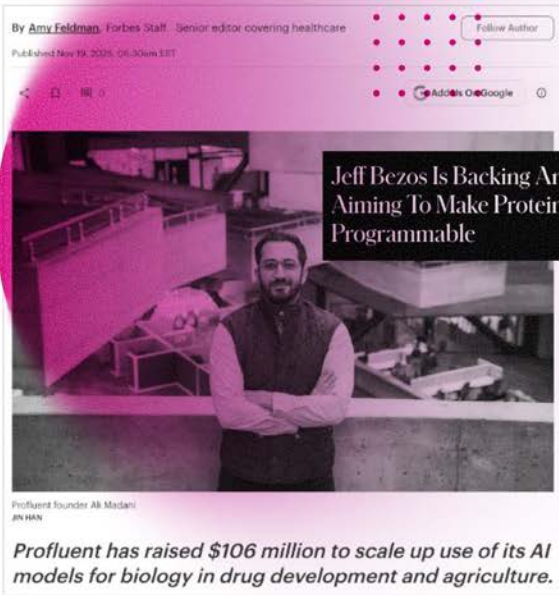
Implications for design strategy

Design begins upstream

Material decisions increasingly happen at the computational modeling stage rather than after physical sampling. Design teams must collaborate earlier with bioengineers and material scientists as part of R&D workflows.

Materials become programmable differentiators

As AI unlocks novel bio-based substrates, brands can differentiate not only through form and function—but through engineered biological properties such as self-healing, carbon sequestration, or biodegradability.



/// Sources

- 1 / Zhou, B., Li, X., Pan, Y., He, B., & Gao, B. (2025). Artificial intelligence-assisted next-generation biomaterials: From design and preparation to medical applications. *Colloids and Surfaces B: Biointerfaces*. 2 / Madani, A. et al. (2023). Large language models generate functional protein sequences across diverse families. 3 / DeepMind (2024). AlphaFold 3 predicts the structure and interactions of all of life's molecules. 4 / Ginkgo Bioworks. The Foundry Platform. 5 / Insilico Medicine (2023). AI-discovered drug candidate enters Phase II clinical trials.

Key signals

Foundation models are generating functional proteins

Companies such as Profluent have released large-scale protein foundation models (ProGen series) capable of generating novel protein sequences with experimentally validated function—shifting biology from predictive analysis to generative design.²

Multi-molecule simulation is expanding computational design

AlphaFold 3 predicts interactions between proteins, DNA, RNA, and small molecules—enabling system-level molecular simulation before wet-lab experimentation.³

AI is embedded inside automated bioengineering loops

Biofoundries such as Ginkgo Bioworks' Foundry platform integrate machine learning into Design-Build-Test-Learn workflows, using predictive models to prioritize constructs before synthesis and retraining on experimental outputs, compressing iteration cycles.⁴

AI-designed molecules are entering global clinical pipelines

Insilico Medicine has advanced AI-designed drug candidates into multi-country clinical trials, signaling that computationally generated biology is reaching regulatory-grade validation.⁵



Wider attack surface

When organizations and products embed AI, their attack surface expands. In response, companies are investing heavily in cybersecurity, building it throughout the tech stack from inception rather than bolting it on. But the threat has evolved as well. AI has handed attackers an entirely new arsenal: hyper-personalized phishing emails generated in minutes, convincing voice and video deepfakes, data poisoning that corrupts AI training data, and autonomous agent swarms that can plan and execute an entire attack lifecycle without human involvement.^{1, 2, 3}

Implications for design strategy

Security must be visible

As AI systems block actions, verify identities, or watermark content, interfaces must clearly communicate why decisions were made and how trust is established. Security status should become part of the user experience.

Protective friction is strategic

In high-risk flows (financial transfers, access changes, permissions escalation), AI-driven verification may intentionally introduce pauses or additional confirmation steps. Friction, intentional moments for pause and reflection, becomes a designed safeguard rather than a usability flaw.



// Sources

1 / Gartner (2025). In the age of GenAI, preemptive capabilities—not detection and response—are the future of cybersecurity. 2 / NVIDIA (2023). Confidential computing on NVIDIA H100 GPUs. 3 / DARPA (2023–2024). AI Cyber Challenge (AWCC). 4 / Google DeepMind. SynthID

watermarking technology. 5 / Stanford Institute for Human-Centered Artificial Intelligence (2023). AI Index Report 2025. (AI Incident Database statistics). 6 / Gartner (2026). Top 10 Strategic Technology Trends for 2026. (Confidential computing, digital provenance, preemptive cybersecurity)

Key signals

Preemptive AI cybersecurity spending is shifting

Gartner forecasts that preemptive cybersecurity technologies will account for over 50% of IT security spending by 2030, up from less than 5% in 2024—reflecting a structural shift toward predictive exposure management, deception technologies, and automated prevention.¹

Confidential computing is being operationalized for AI workloads

Major infrastructure providers now offer trusted execution environments for live AI processing. Microsoft Azure Confidential Computing, Google Cloud Confidential VMs, and NVIDIA H100 Confidential Computing GPUs enable protection of data “in use,” not just at rest or in transit—embedding hardware-level trust into AI infrastructure.²

AI agents are being deployed for autonomous cyber defense

AI systems are being used to simulate attacks and automate response at machine speed. Microsoft Security Copilot integrates generative AI into threat detection workflows, while DARPA's AI Cyber Challenge (AIxCC) funds autonomous AI systems capable of identifying and patching vulnerabilities.³

Provenance and watermarking as integrity infrastructure

Google DeepMind's SynthID embeds imperceptible watermarks into AI-generated content and provides detection tooling, reflecting a broader move toward built-in verification systems to mitigate tampering, impersonation, and synthetic media risks.⁴



Ambient awareness

AI is increasingly “taking to the streets,” through wearables like smart glasses, adaptive traffic systems, and autonomous delivery robots. This shift is enabled by edge computing, multimodal sensors (vision, audio, lidar), and high-speed 5G/6G connectivity. Advances in large multimodal models are also making AI more contextually aware of complex, real-world environments. Together, these technologies are creating an “always-on” ambient intelligence layer that is reshaping how people move, interact, and are observed in public.

Implications for design strategy

Consider new interaction paradigms

Ambient AI systems need to communicate intent, status, and decision-making without screen interfaces to rely on. This requires designers to create new conventions and interaction models. Examples include an autonomous vehicle letting its passenger know that it sees an sudden obstacle ahead, or wearable indicating that it is actively processing and that the wearer needs to wait.

Design for the non-user

Ambient intelligence reshapes public space for everyone in it, not just those who opted in. As AI systems observe, infer, and act in shared environments, design should consider the rights of bystanders, protecting their privacy and reserving their agency to opt-out.



/// Sources

1 / Financial Times (2026), Inside the plan to kill Ali Khamenei. 2 / The New York Times (2026) Dinner Is Being Recorded, Whether You Know It or Not. 3 / The New York Times (2026) 'Everyone's in Sync': Phone Ban on NYC Dance Floors Is Getting rave Reviews. 4 / Fortune (2026) Pokémon Go players built a 30-billion-photo map that's now training robots to deliver your pizza

Key signals

Public cameras hacked by governments for surveillance

According to the Financial Times, many of the Iranian capital's traffic cameras had been compromised by Israel, allowing Israeli intelligence analysts to access the feeds remotely and track the daily routines of senior Iranian officials and their security details.¹

Smart glasses are raising privacy concerns for bystanders

Early adopters of Meta's Ray-Ban smart glasses are facing the same cultural backlash that once earned Google Glass users the nickname "glassholes." Critics argue the glasses quietly erode privacy, since the recording indicator is subtle and U.S. law generally allows filming in public without the consent of the people being recorded.²

Private venues designated as phone-free give people comfort that their behavior won't be recorded

There is a growing movement of nightlife spaces that are banning or limiting phones on dance floors. People say that such bans help them live in the moment and express themselves more freely, pointing to how the constant presence of phones change how we act.³

Augmented reality apps are creating rich data sets that companies stand to benefit from

Pokémon Go is a popular game with over 1 billion downloads since its 2016 launch that uses augmented reality to simulate catching Pokémon in the real world. Unknown to most, the location and camera data was being collected by Niantic Spatial who used the 30-billion-photo map to train its robots to navigate the world.⁴

Economic



New pricing logic

AI is collapsing the marginal cost of knowledge work and detaching production from human time. As output becomes increasingly automated, professional services are shifting from billing by effort (hours, seats) to charging for measurable units of value—usage, tasks, or outcomes. Similar to the emergence of value-based care in the health care sector, this shift changes not only revenue models, but how value must be surfaced, measured, and trusted.

Implications for design strategy

Value attribution infrastructure

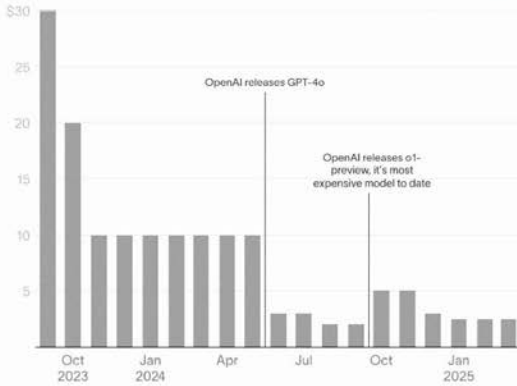
As pricing shifts toward measurable outcomes, products must formalize how value is defined, calculated, and demonstrated. Design strategy needs to build clear attribution models that connect user actions to economic impact, ensuring customers understand exactly what they are paying for and why.

Pricing transparency as a UX layer

When revenue depends on tokens processed, workflows completed, or results achieved, cost visibility becomes part of the product experience. Interfaces must make economic drivers legible—clarifying usage patterns, performance thresholds, and optimization levers—so pricing feels predictable rather than opaque.

AI is getting a lot cheaper

Median price per 1 million tokens from OpenAI and Anthropic



Source: Ramp card spend, anonymized transactions from businesses using Ramp's corporate spend platform

ramp

/// Sources

1 / OpenAI, (2025), API Pricing Documentation.
2 / Reuters, (2024), OpenAI unveils cheaper small AI model GPT-4o mini. 3 / Anthropic, (2025), Claude API Pricing Documentation. 4 / Intercom, (2025), Fin AI Agent pricing—pay per resolution. 5 / The Wall Street Journal, (Dec 4, 2025), Say Goodbye to the

Billable Hour, Thanks to AI. 6 / Business Insider, (2025), AI is Reshaping How McKinsey Makes Money.

Key signals

Compute replaces labor as the billable unit

AI model providers price usage in tokens processed rather than time worked. OpenAI and Anthropic publish per-million token pricing, and lower-cost inference models are being introduced to reduce unit compute expense.^{1, 2, 3}

Revenue tied directly to outcomes

AI-native products are pricing against completed work instead of access. Intercom's Fin AI Agent charges per resolved customer conversation, linking vendor revenue to measurable results rather than seats or subscriptions.⁴

Performance-based fees expand in professional services

As AI compresses delivery cycles, time-based billing models face structural pressure. Firms including McKinsey are expanding performance-linked fee structures that tie compensation to client results rather than hours logged.^{5, 6}



The infrastructure advantage

AI is reorganizing the economy around capital-intensive infrastructure. Unlike prior software waves, advanced AI systems require sustained investment in data centers, accelerated compute, energy capacity, and sovereign cloud environments. Economic advantage is increasingly accruing to infrastructure owners—hyperscalers, semiconductor manufacturers, and domestically controlled cloud providers—rather than downstream application developers.

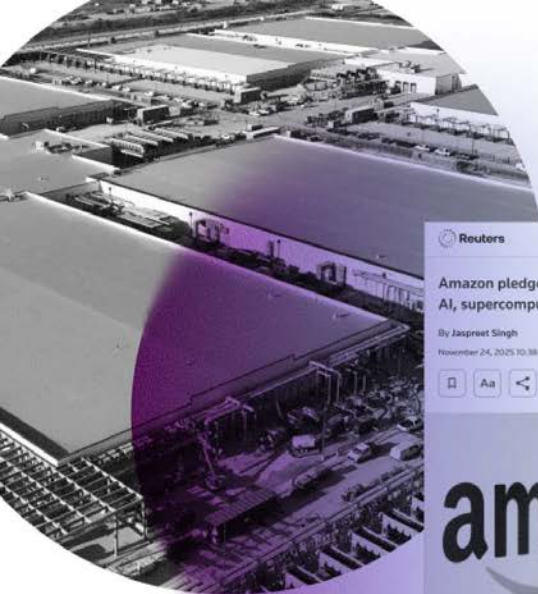
Implications for design strategy

Compute-aware product strategy

Compute and energy exposure now shape pricing, performance, and margin structure. Design leaders must treat infrastructure cost as a product variable, balancing model intensity and user value against long-term scalability.

Design for capital efficiency at scale

As AI workloads expand, small increases in inference intensity compound into material cost. Products should enable intelligent model routing, adaptive fidelity, and workload optimization so value scales without proportionally increasing capital exposure.



Key signals

Hyperscaler capex and chipmakers capture AI value

Microsoft, Amazon, Meta, and Apple collectively signaled more than \$300B in 2025 capital expenditure largely tied to data centers and AI infrastructure.¹ Accelerated server investment has risen from roughly \$4B in 2015 to nearly \$100B in 2024, concentrating value upstream among semiconductor suppliers such as NVIDIA.¹

Energy and grid access are becoming economic constraints

Data centers consumed roughly 415 TWh of electricity in 2024—and about 1.5% of global demand—and could approach 945 TWh by 2030.² Grid congestion, long interconnection queues, and construction bottlenecks in North America and Europe place an estimated 20% of planned data-center projects at risk, constraining future AI infrastructure growth.²

Sovereign infrastructure redirects cloud spending

Sovereign cloud IaaS spending is projected to reach \$80B in 2026, with approximately 20% of workloads shifting from global hyperscalers to local providers.⁴ Initiatives such as Europe's GAIA-X, France's sovereign cloud frameworks, and Gulf state investments in national AI supercomputing capacity reflect a redirection of AI infrastructure spend toward domestically controlled compute.^{4, 5}

// Sources

1 / International Energy Agency (IEA). (2025). Energy and AI: (Hyperscaler capex trends; accelerated server investment growth; infrastructure expansion data). 2 / International Energy Agency (IEA). (2025). Energy and AI: (Data center electricity consumption; 2030 projections; grid congestion risk). 3 / Soben

(part of Accenture). (2026). Data Centre Trends: 2026. (Construction cost escalation; labor shortages; equipment lead times; build capacity constraints.) 4 / Garner. (2026, Feb 9). Gartner Says Worldwide Sovereign Cloud IaaS Spending Will Total \$80 Billion in 2026. 5 / Garner. (2026). Top 10 Strategic

Technology Trends for 2026. (Geopatriation; sovereign infrastructure framing.)



The competitive divide

AI adoption is broad, but competitive advantage at the foundational level is concentrating. Companies that control capital, compute, and ecosystem access are converting AI into durable profit and market share gains, while others remain in experimentation cycles, or become dependent on a concentrated core's offering to create additional value. Market structure is shifting toward vertically integrated AI platforms with rising barriers to entry.

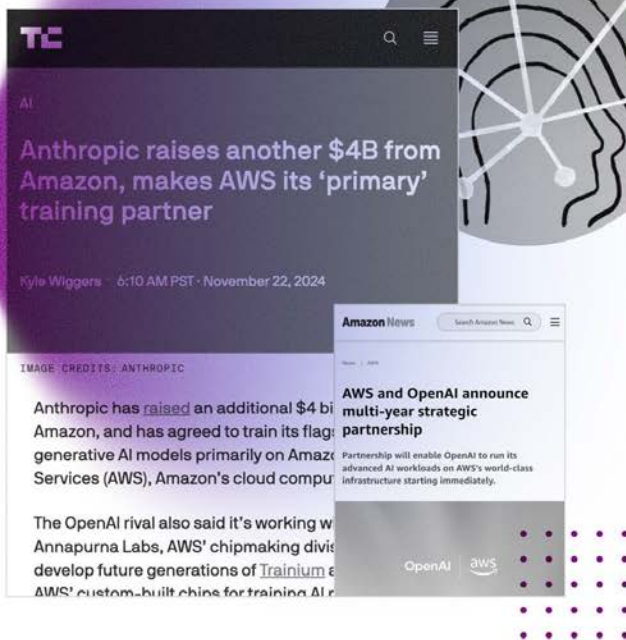
Implications for design strategy

Design for structural integration, not surface adoption

Competitive advantage accrues when AI reshapes workflows, decision systems, and operating models. Design leaders must prioritize deep systems integration and measurable operational impact over incremental feature layering.

Design to reduce scaling friction

As capital and compute shape competitive advantage, products must help organizations operationalize AI efficiently. This means supporting governance, integration into legacy systems, and clear economic impact measurement to show value and prevent stagnation.



/// Sources

1 / McKinsey & Company. (2025). The State of AI in 2025. **2** / Reuters. (2025). Amazon to invest up to \$50 billion in AI supercomputing. **3** / Financial Times. (2026). Amazon's AI spending drive and capital expansion. **4** / About Amazon. (2025). AWS strategic investment in Anthropic. **5** / Stanford

Institute for Human-Centered Artificial Intelligence. (2025). AI Index Report 2025. **6** / Reuters. (2025). Microsoft, Nvidia to invest in Anthropic as Claude maker commits \$30 Billion to Azure.

Key signals

AI is reinforcing profit concentration

Companies that embed AI deeply into core operations are capturing outsized returns: JPMorgan attributes more than \$1B annually to AI-driven fraud detection and operational automation.¹ Amazon, meanwhile, has committed up to \$50B in AI supercomputing infrastructure investment and is undertaking capital programs projected near \$200B to expand data centers and AI capacity.^{2,3}

Vertical stack control is increasing market power

AI competition increasingly favors firms controlling multiple layers of the stack: models, cloud infrastructure, chips, and enterprise distribution. Microsoft's deep integration of OpenAI across Azure and Copilot, and Amazon's strategic investment in Anthropic, illustrate how model access, compute capacity, and cloud distribution are converging.^{3,4}

Capital intensity is raising barriers to entry

Frontier model development is dominated by a small group of firms—OpenAI, Google DeepMind, Anthropic, Meta—with advanced training runs concentrated primarily in the United States and China.² Anthropic has committed to purchase up to \$30B in Azure compute capacity, with Microsoft investing up to \$5B and Nvidia up to \$10B to support the partnership.⁶



A capital surge

AI is no longer just a technology shift—it is a global capital reallocation event. In 2025, AI firms absorbed the majority of worldwide venture capital for the first time, while sovereign funds, hyperscalers, and corporate balance sheets deployed record-scale investments to secure AI positioning. Strikingly, this surge is happening as broader venture and investment markets contract in an economic downturn, with non-AI sectors increasingly deprived of funding as money flows toward AI.

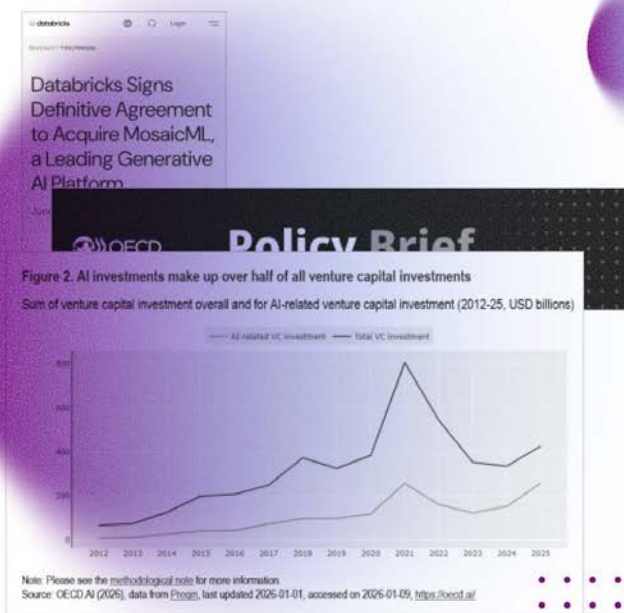
Implications for design strategy

Prepare for competitive compression and capital cycles

While AI absorbs a majority share of global funding, capital markets remain cyclical and concentrated. Product strategy must prioritize durable value creation, monetization clarity, and operational impact to withstand shifts in investor sentiment beyond peak funding periods.

Design for capital-dense competitive environments

As venture, corporate, and sovereign capital converge around AI, category leaders will be heavily funded and strategically embedded. Design leaders must differentiate through defensible workflows, proprietary data leverage, and ecosystem integration rather than feature velocity alone.



Key signals

Global capital is reweighting toward AI

In 2025, AI firms captured 61% of global venture capital—USD 258.7B of USD 427.1B deployed worldwide, doubling AI's share from 30% in 2022.¹ Capital flows are extending beyond venture: France announced €109B in AI investment pledges, combining sovereign, private equity, and hyperscaler commitments to expand AI infrastructure and model development.²

Capital is concentrating in late-stage mega-deals

In 2025, "mega-deals" (>\$100M) accounted for 73% of total AI VC investment value, with billion-dollar rounds comprising nearly half of all AI funding.¹ In January 2026, Elon Musk's xAI raised \$20 billion in an upsized Series E funding round, exceeding its \$15 billion target, as it ramps up development of new AI models and computing infrastructure.²

Strategic AI alliances are redirecting corporate capital

Corporate capital is being deployed to internalize AI capabilities. Amazon increased its commitment to Anthropic to \$8B, reinforcing AWS as a primary training and hosting partner.⁴ Databricks acquired MosaicML for \$1.3B to bring foundation model development in-house.⁵ Adobe's attempted \$20B acquisition of Figma further signaled how creative and enterprise platforms are repositioning around AI-native workflows.⁶

/// Sources

1 / OECD (2026), Venture Capital Investments in Artificial Intelligence through 2025, 2 / Reuters (Feb 10, 2025), Details of 110 billion euros in investment pledges at France's AI summit, 3 / Reuters (Jan 6, 2026), Musk's xAI raises \$20 billion in upsized Series E funding round, 4 / Wiggins, K. (2024, November

22), Anthropic raises another \$4B from Amazon, makes AWS its primary training partner, TechCrunch, 5 / TechCrunch (2023), Databricks picks up MosaicML, an OpenAI competitor, for \$1.3B, 6 / TechCrunch (2023), Adobe has a big hole to fill after the \$20B Figma deal falls through.



The AI risk premium

As organizations embed AI into both their operations and their products, they absorb new forms of operational, regulatory, and financial exposure. AI is no longer only a productivity lever—it is expanding security budgets, compliance overhead, governance functions, and litigation risk. Across sectors, AI adoption is introducing a structural risk premium that raises the cost of doing business.

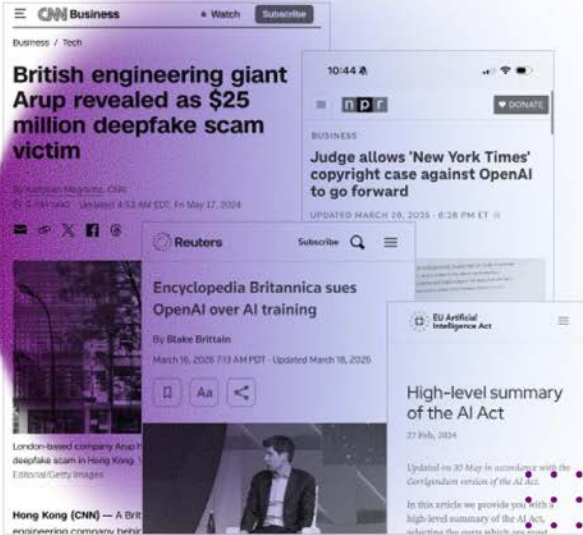
Implications for design strategy

Design for governance visibility

AI systems must expose audit trails, explainability layers, and risk signals clearly, supporting enterprise oversight and regulatory accountability.

Design for controlled friction

In high-stakes workflows (financial transfers, data access, decision automation), intentional verification and review mechanisms must be built into the experience to mitigate risk exposure.



/// Sources

1 / Magrino, K. (2024). British engineering giant Arup revealed as \$25 million deepfake scam victim. CNN Business. 2 / IBM Security. (2023). Cost of a data breach report 2023. IBM Corporation. 3 / European Parliament. (2024). Artificial Intelligence Act: Regulation laying down harmonised rules on

artificial intelligence. European Union. 4 / Board of Governors of the Federal Reserve System. (2011). Supervisory guidance on model risk management (SR 11-7). Federal Reserve. 5 / Grynbbaum, M. M., & Mac, R. (2023, December 27). The Times sues OpenAI and Microsoft over AI use of copyrighted

work. The New York Times. 6 / Hart, R. (2025). Stability AI's legal win over Getty leaves copyright law in limbo. The Verge.

Key signals

AI-enabled fraud is producing multi-million dollar losses

Generative AI is lowering the cost of impersonation and automated attacks, expanding the scale and sophistication of fraud. In 2024, engineering firm Arup lost \$25 million after employees were deceived by a deepfake video call impersonating company executives.¹ At the same time, cybersecurity exposure remains costly: the global average cost of a data breach reached \$4.45 million, with U.S. organizations experiencing even higher losses.²

Deployment is becoming an ongoing compliance cost

AI deployment increasingly requires continuous governance and regulatory oversight. The European Union's AI Act establishes rules for high-risk and general-purpose AI systems, introducing requirements for risk management, transparency, and documentation.³ For firms deploying AI in regulated markets, compliance now involves ongoing operational work including model monitoring, documentation, auditing, and vendor due diligence.

AI misconduct is generating direct financial penalties

Failures in AI governance can expose organizations to legal and financial liability. U.S. banking regulators have long treated model risk as an operational exposure that can lead to financial loss when poorly governed systems influence decision-making.⁴ Meanwhile, generative AI companies including OpenAI and Stability AI face active copyright litigation over the training data used in their models.^{5, 6}

Environment



Energy-intelligence coupling

AI is no longer just software, it is becoming a large-scale industrial load on electricity systems. As training and inference expand, grid capacity, interconnection queues, and generation buildout timelines are emerging as structural constraints on intelligence production. AI scaling is now materially dependent on energy infrastructure.

Implications for design strategy

Design for energy-proportional intelligence

Treat compute as a constrained resource. Prioritize right-sized models, efficient inference patterns, user-invoked AI over always-on systems, and tiered capability that reflects real infrastructure cost and availability.

Architect for infrastructure variability

Assume regional constraints, latency variability, and reliability mechanisms will shape AI availability. Build experiences that degrade safely, communicate capability boundaries clearly, and remain functional under fluctuating compute conditions.

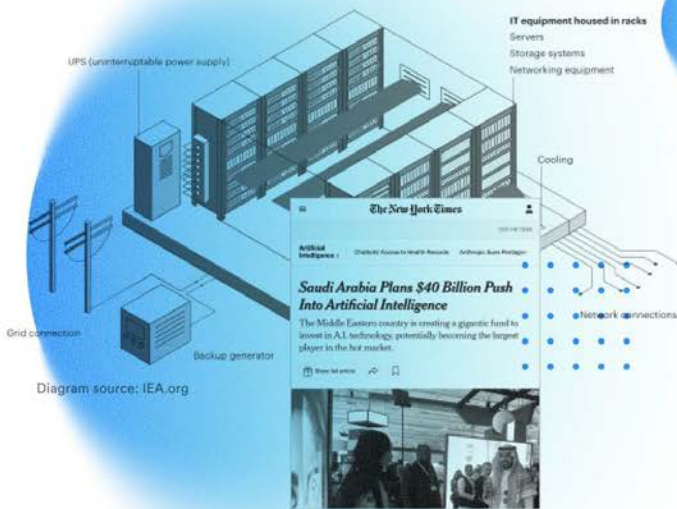


Diagram source: IEA.org

Key signals

Intelligence is becoming a macro energy load

Data centers consumed roughly 415 TWh globally in 2024 (~1.5% of global electricity demand), and the IEA projects demand could more than double by 2030, with AI-optimized facilities driving the fastest growth. AI is now a measurable component of national electricity expansion.¹

Grid reliability is gating AI expansion

In 2026, PJM, the largest U.S. grid operator, introduced reforms to manage AI-driven data center demand, including revised load forecasting and reliability procurement mechanisms² In parallel, the White House urged PJM to conduct an emergency power auction to protect against potential rolling blackouts as data center demand accelerates.³

Energy constraints are reshaping AI geography

Europe's data center power demand could nearly triple by 2030, intensifying transmission and generation strain. The UK has begun reforming grid connection queues, and Ireland implemented stricter connection controls following regional capacity pressure from data center clustering.^{4, 5, 6}

Energy endowment is becoming a strategic AI advantage

Energy-rich states are increasingly positioning abundant, scalable electricity as a competitive lever in attracting AI infrastructure. Saudi Arabia has aligned its ambition to become a global AI hub with large-scale energy investment and sovereign capital deployment, framing electricity access as foundational to compute expansion.⁷

/// Sources

1 / International Energy Agency (IEA). Energy and AI – Energy demand from AI. 2 / Reuters. PJM unveils plan to tackle AI-driven power demand surge. January 2026. 3 / Reuters. White House seeks emergency power auction for largest US electric

grid. January 2026. 4 / Reuters. Europe's data centre power demand expected to triple by 2030. McKinsey report says. October 2024. 5 / Reuters. Britain overhauls power grid connections to root out zombie projects. 6 / KPMG Ireland. Ireland's data centre

policy reset. 7 / Matz, C., & Serkin, A. R. (2024, March 19). Saudi Arabia plans \$40 billion push into artificial intelligence. The New York Times.



The grid bottleneck

AI infrastructure expansion is increasingly constrained not just by total electricity supply, but by grid capacity and interconnection timelines. In many regions, hyperscale data centers are facing multi-year delays due to transmission bottlenecks and overloaded substations. As AI clusters concentrate geographically, grid congestion is emerging as a gating factor for where intelligence infrastructure can physically connect.

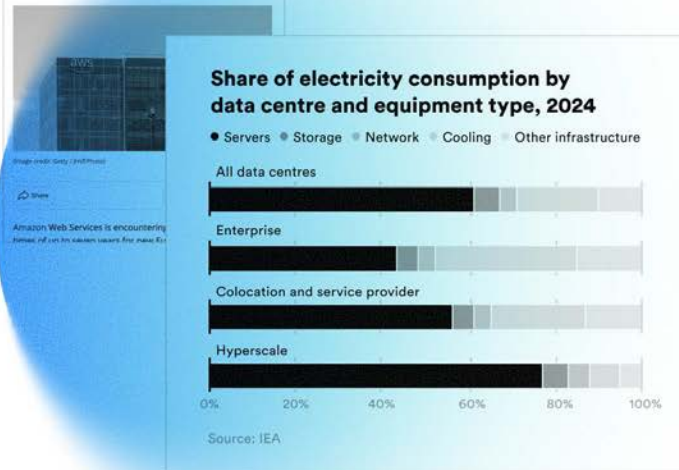
Implications for design strategy

Expect geographic fragmentation of AI capability

AI performance, latency, and feature availability may vary by region based on grid access. Design systems that support regional capability tiers and graceful degradation under infrastructure variability.

Design roadmaps for deployment friction

Product strategies that assume frictionless compute scaling may encounter siting delays and interconnection bottlenecks. Prioritize modular, distributed AI architectures that allow incremental expansion without single-location dependency.



/// Sources

- 1 / Reuters. Power grid delays challenge Amazon's data center expansion in Europe. February 3, 2026.
- 2 / Lawrence Berkeley National Laboratory (LBNL). Queued Up: Characteristics of Power Plants Seeking Transmission Interconnection. 3 / Financial Times. 'Phantom' data centres muddy forecasts for US power needs. November 2025. 4 / Reuters. Blackstone to buy \$1 billion Virginia power plant near data centers. January 23–24, 2025.
- 5 / International Energy Agency (IEA). Electricity Grids and Secure Energy Transitions.

Key signals

Interconnection delays are slowing AI buildout

Amazon has faced prolonged delays securing power-grid connections for new European data centers, citing grid access timelines as a challenge to expansion.¹ As hyperscale AI infrastructure depends on timely connection to transmission networks and substations, multi-year interconnection processes are directly shaping deployment speed.

Speculative data center load is distorting grid planning

U.S. utilities are contending with “phantom” data center proposals—speculative or duplicate projects submitted without firm financing—that inflate demand forecasts and complicate transmission planning.³ Utilities including ComEd and AEP Ohio have increased deposit requirements and introduced capacity-linked payment structures to deter speculative load requests and manage forecast uncertainty.

Deliverable power is becoming a strategic asset

As transmission bottlenecks tighten in major AI hubs, proximity to existing grid-connected generation is gaining strategic value. Blackstone's agreement to acquire Virginia's 774-MW Potomac Energy Center—the core of “Data Center Alley”—at approximately \$1B, underscores how access to deliverable capacity is influencing capital allocation and siting decisions.⁴



The water footprint

As AI infrastructure scales, its environmental constraint profile is increasingly shaped by water availability and cooling design. Many hyperscale data centers rely on evaporative cooling systems whose freshwater intensity varies by climate and regional energy mix. In drought-prone and water-stressed regions, cooling demand is emerging as a binding siting and operational constraint, linking intelligence production directly to local hydrological systems.

Implications for design strategy

Treat water as a first-class infrastructure constraint

AI feature design, inference frequency, and deployment geography now carry water implications alongside energy cost. Sustainable AI strategy must account for regional water stress and cooling intensity when scaling compute-heavy services.

Optimize AI workloads for environmental context

Model selection, workload routing, and system architecture should support region-aware optimization and reduced thermal intensity. Always-on or high-frequency AI features should be evaluated not only for compute efficiency but for associated cooling and freshwater impact.



// Sources

- 1 / Li, P., et al. (2023). Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models. *2 / Associated Press*. (2024, September 17). Google to pause plans for big data center in Chile over water worries. *PBS NewsHour*. *3 / Hao, K.* (2024, March 1). AI is taking water from

the desert. *The Atlantic*. *4 / Microsoft*. (2024, December 9). Sustainable by design: next-generation data centers consume zero water for cooling. *5 / Microsoft Data Centers*. Sustainability efficiency metrics (WUE/PUE).

Key signals

Frontier models carry a measurable water cost

Water use scales with both training and inference demand. Researchers estimate that training GPT-3 in Microsoft U.S. data centers required approximately 700,000 liters of freshwater, depending on cooling method and electricity mix.¹ The same study estimates that 20–50 ChatGPT prompts can consume roughly 500 mL of water, accounting for direct cooling and indirect electricity-related use.

Water stress is forcing cooling redesign

Local water availability is influencing hyperscale cooling architecture and siting decisions. In Santiago, Chile, Google redesigned a proposed data center after regulators raised concerns about freshwater-intensive evaporative cooling amid prolonged drought.² In Goodyear, Arizona, Microsoft reported using millions of gallons of potable water annually for cooling during severe drought, prompting public scrutiny.³

Water-free cooling architectures are scaling

In 2024, Microsoft announced next-generation data centers designed to operate with zero water for cooling, shifting to closed-loop and air-based systems that eliminate evaporative freshwater use.⁴ The company publicly reports Water Usage Effectiveness (WUE) alongside Power Usage Effectiveness (PUE), formalizing water intensity as an operational metric.⁵



The thermal loop

As AI data centers scale, they generate substantial volumes of low-grade waste heat. Historically vented into the atmosphere, this heat is increasingly being captured and integrated into district heating systems. In parts of Europe, data centers are beginning to function not only as electricity consumers, but as distributed thermal energy sources capable of displacing fossil-based heating generation.

Implications for design strategy

Design AI infrastructure as integrated energy systems

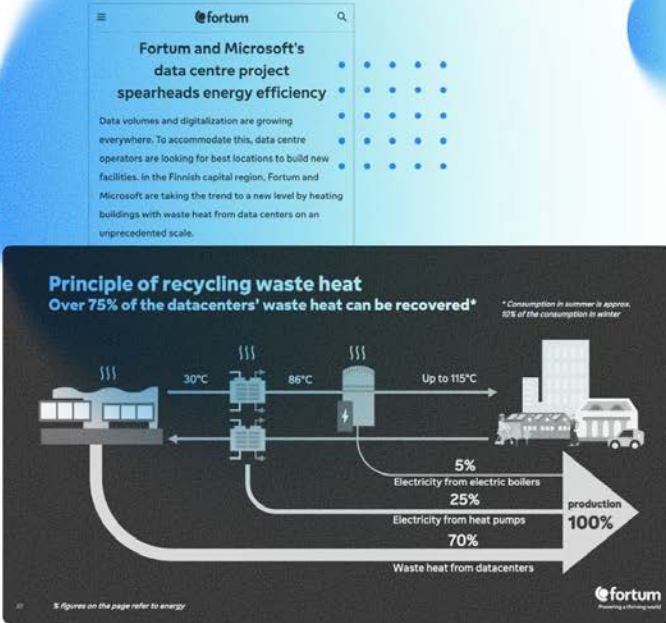
AI expansion may increasingly require partnership with municipal utilities and district heating operators. Infrastructure strategy should account for thermal recovery potential alongside compute capacity.

Expand environmental performance metrics

Evaluate not only electricity consumption but thermal recovery contribution and system integration. Waste heat utilization can shift AI infrastructure from net load to partial energy contributor in urban ecosystems.

/// Sources

1 / Fortum. Data centres in Helsinki region – Espoo Clean Heat and waste heat integration. 2 / European Commission. Energy Efficiency Directive (revised): entered into force October 2023). 3 / European Commission. Energy performance of data centres – monitoring and reporting obligations. 4 / International Energy Agency (2025). Energy and AI.



Key signals

Data centers are becoming municipal heat suppliers

Waste heat recovery is moving from pilot projects to core urban infrastructure. In Finland, Microsoft's new data center region in Espoo and Kirkkonummi will channel waste heat into Fortum's district heating network, projected to supply approximately 40% of regional district heating demand once fully operational.¹

Policy is institutionalizing waste heat recovery

Heat reuse is transitioning from voluntary sustainability practice to policy expectation. The European Union's revised Energy Efficiency Directive (2023) strengthens requirements for energy system efficiency and is paired with EU rules requiring monitoring and reporting of data center energy performance.^{3,4} It also requires member states to assess and facilitate the recovery of waste heat, embedding thermal reuse directly into infrastructure planning.

AI waste heat aims to replace fossil-based heating

AI waste heat is not merely reused energy, but a direct substitute for combustion-based heating. Fortum's Espoo Clean Heat program includes large-scale heat pumps and thermal storage designed to integrate data center waste heat into the district heating system as part of a broader transition away from fossil-based heating generation.¹



The material constraint

As AI infrastructure scales, its environmental footprint extends upstream into mining, refining, and advanced manufacturing. High-performance chips, power systems, and grid upgrades depend on mineral-intensive supply chains that are geographically concentrated and environmentally disruptive. The materials required to build and power advanced compute are emerging as a structural environmental constraint on AI expansion.

Implications for design strategy

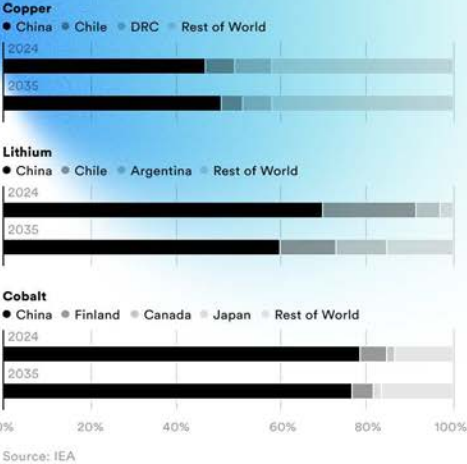
Design for material and hardware constraint

Assume periodic tightness in GPUs, power components, and upstream mineral supply. Prioritize model portability, modular architectures, and flexible hardware tiers that can scale under constraint rather than stall expansion.

Integrate lifecycle thinking into AI product strategy

Environmental performance must extend beyond runtime efficiency. Consider embodied impact, hardware longevity, reuse pathways, and procurement standards that reward traceability and lower material intensity.

Geographical distribution of refined material production for key minerals in the base case



Key signals

AI data centers are driving copper extraction

AI-optimized data centers require copper-intensive power and grid systems to operate. Copper demand from data centers alone is projected to rise from roughly 78,000 tonnes in 2020 to over 650,000 tonnes by 2030, as AI-driven power density and associated grid expansion increase.¹ This surge adds pressure to global copper mining, where declining ore grades require more material processing per tonne extracted.

Critical mineral mining is geographically concentrated

Approximately 70% of global copper supply comes from the Democratic Republic of the Congo, lithium extraction is concentrated in Chile's water-stressed Atacama Desert, and rare earth processing is dominated by China.^{2,3} Mining and refining in these regions generate toxic tailings, water depletion, and land disturbance, concentrating the environmental burden of AI hardware in specific ecosystems.

AI hardware turnover is expanding hazardous e-waste

Global electronic waste reached 62 million tonnes in 2022, with only 22% formally recycled, and is projected to reach 82 million tonnes by 2030.⁴ Lifecycle analysis of NVIDIA's A100 GPU shows that semiconductor components account for nearly 89% of its manufacturing-related carbon emissions, while copper-intensive thermal components drive toxicity and resource depletion impacts.⁵

/// Sources

- 1/ Reuters. (2025, July 21). Global power grid expansion fuels fresh copper demand surge.
- 2/ International Energy Agency. (2025). Energy and AI. 3/ International Energy Agency. (2024). Global critical minerals outlook 2024. 4/ United Nations Institute for Training and Research. (2024). The global e-waste monitor 2024. 5/ Falk, S., Eichjatz, D., Pison, T., Lees-Perraso, E., Wietz, A., Biber-Freudenberger, L., Lucioni, S., & van Wynbergh, A. (2025). More than carbon: Cradle-to-grave environmental impacts of GenAI training on the NVIDIA A100 GPU (Version 3). arXiv.

Political



The “splinternet”

Artificial intelligence infrastructure is becoming a geopolitical asset. As advanced compute becomes strategic national infrastructure, governments are restricting access to high-end chips, re-shoring semiconductor production, and localizing cloud environments. The result is a fragmented AI landscape where capability, infrastructure, and access increasingly align with geopolitical blocs rather than open global markets.

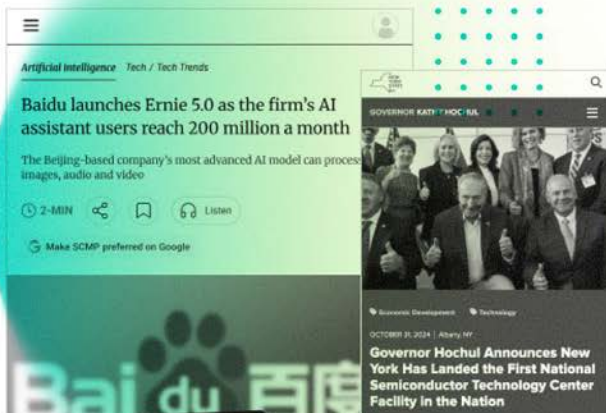
Implications for design strategy

Design for geopolitical variability

Assume infrastructure access, model availability, and performance will differ by region. Architect products to flex across jurisdictions without rebuilding the core experience.

Make jurisdiction visible without eroding trust

When capabilities shift across regions, explain why. Clear communication about availability and data handling prevents geopolitical constraints from feeling like instability.



Key signals

AI compute is now export-controlled

Advanced AI compute is now a tool of foreign policy. U.S. imposed export controls on AI chips and semiconductor manufacturing equipment, place frontier AI processors under national security review and extend restrictions across allied supply chains.^{2, 3, 4}

Semiconductor capacity is being nationalized

Governments are increasingly treating semiconductor manufacturing as sovereign infrastructure. The U.S. CHIPS and Science Act allocated \$52.7 billion to expand domestic semiconductor production and research capacity.⁵ The European Chips Act similarly aims to strengthen regional manufacturing and reduce dependence on foreign suppliers.⁶

Cloud infrastructure is becoming jurisdiction-bound

AI infrastructure is increasingly constrained by national data governance regimes. Microsoft completed its EU Data Boundary initiative in 2025, committing to store and process covered customer data within the EU and EFTA region.⁷ As governments impose stricter rules on cross-border data flows, cloud providers are building region-specific infrastructure that aligns with legal jurisdiction rather than global network efficiency.⁸

AI ecosystems are politically influenced

Distinct political systems are shaping region-specific AI platforms and governance models. For examples, China's Interim Measures for the Management of Generative Artificial Intelligence Services established national oversight for public-facing generative AI systems.⁹ Within this framework, domestic platforms such as Baidu's ERNIE models continue to expand, with Baidu reporting more than 200 million monthly users of its AI assistant ecosystem.¹⁰

// Sources

1 / Bureau of Industry and Security. (2022, October 13). Implementation of additional export controls: Certain advanced computing and semiconductor manufacturing items; supercomputer and semiconductor end use; entity list modification. Federal Register. 2 / Bureau of Industry and Security. (2023, October 25). Implementation of additional export controls: Certain advanced computing items; supercomputer and semiconductor end use; updates

and corrections. Federal Register. 3 / Bureau of Industry and Security. (2026, January 13). Department of Commerce revises license review policy for semiconductors exported to China. U.S. Department of Commerce. 4 / Bureau of Industry and Security. (2026, January 15). Revision to license review policy for advanced computing commodities. Federal Register. 5 / National Institute of Standards and Technology. (2026). CHIPS for America fact

sheet: Federal programs supporting the U.S. semiconductor supply chain. U.S. Department of Commerce. 6 / European Commission. (n.d.). European Chips Act: Shaping Europe's Digital Future. 7 / Microsoft. (2025, February 26). Microsoft completes landmark EU data boundary, offering enhanced data residency and transparency. Microsoft Issue Blog. 8 / United Nations Conference on Trade and Development. (2023). G20

members' regulations of cross-border data flows. UNCTAD. 9 / Cyberspace Administration of China. (2023, July 13). Interim measures for the management of generative artificial intelligence services. (English translation). China Law Translate. 10 / Zhang, Y. (2024, November 12). Baidu launches Ernie 5.0; firm's AI assistant reach 200 million monthly. South China Morning Post.



The regulatory tsunami

AI regulation has moved from principle to enforcement. Governments are no longer debating whether AI should be governed—they are codifying statutory obligations, mandating disclosures, and attaching penalties to non-compliance. AI systems are increasingly treated as regulated infrastructure, subject to oversight and legal accountability comparable to other high-stakes industries.

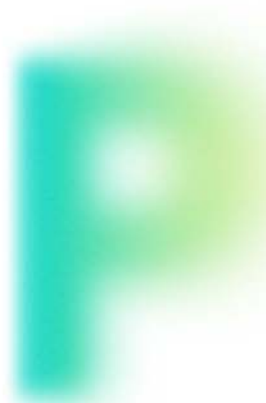
Implications for design strategy

Compliance must be embedded, not layered on

Transparency flows, disclosures, and human oversight controls must be integrated directly into product architecture. Regulatory requirements shape user experience, not just legal copy.

Auditability becomes a product requirement

Interaction logic, automation thresholds, and user-facing disclosures may need to withstand regulatory scrutiny. Documentation and explainability are operational safeguards.





// Sources

1 / European Union. (2024). Regulation (EU) 2024/1699 (Artificial Intelligence Act). Official Journal of the European Union. 2 / Stanford Institute for Human-Centered Artificial Intelligence. (2025). AI Index Report 2025: Policy and Governance Chapter. 3 / StateTech Magazine. (2026). AI

Guardrails Will Stop Being Optional in 2026. 4 / Federal Trade Commission. (2024). FTC Announces Crackdown on Deceptive AI Claims and Schemes (Operation AI Comp!). 5 / Meta. (2025). Expanding GenAI Transparency for Meta's Ads Products.

Key signals

AI is moving from guidance to statute

AI is shifting from voluntary frameworks to enforceable statutory oversight. The European Union's Artificial Intelligence Act establishes binding obligations for high-risk AI systems, including risk management requirements, transparency, human oversight, and post-market monitoring.¹ As phased implementation advances, compliance becomes mandatory rather than advisory.

AI governance is decentralizing across U.S. states

In the absence of comprehensive federal AI legislation, U.S. states are advancing AI-specific laws addressing automated decision-making, algorithmic discrimination, and deceptive AI practices.^{2,3} The regulatory landscape is fragmenting at the state level, expanding compliance exposure across jurisdictions.

Transparency and enforcement are accelerating

Transparency is no longer voluntary—it is enforceable. The EU AI Act authorizes administrative fines tied to global annual turnover for non-compliance.¹ In the United States, the Federal Trade Commission has made clear that deceptive AI claims and harmful AI-enabled practices fall within existing consumer protection law.⁴ Companies are responding accordingly: Meta expanded labeling requirements for generative AI content in its advertising products in 2025.⁵



A chip cold war

AI capability is increasingly constrained by hardware bottlenecks. Advanced chips—and the tools required to produce them—have become strategic leverage shaping industrial policy and diplomatic alignment. Semiconductor production is globally distributed but structurally concentrated, with critical chokepoints where single regions dominate key segments of the value chain.⁷ As AI systems scale, compute is no longer just a technical input—it is a geopolitical dependency.

Implications for design strategy

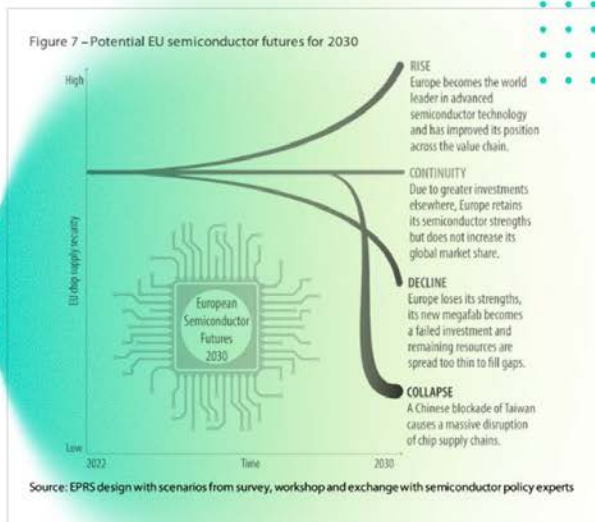
Design for compute variability

Model performance, latency, and availability may shift depending on hardware access and regional supply conditions. Architect systems for modular deployment and graceful degradation.

Treat hardware dependency as product risk

Export controls, fabrication delays, or regional disruptions can directly affect feature roadmaps. Build flexibility into model sourcing, inference infrastructure, and vendor relationships.





Key signals

Restricted compute access

Governments are limiting access to frontier compute through formal export controls. In 2022, the United States imposed export controls on advanced computing chips and semiconductor manufacturing equipment, tightening those restrictions in 2023.^{1, 2} These rules frame frontier AI compute as a national security concern rather than a normal trade good.

Coordinated equipment controls

Control over specialized manufacturing tools is prompting coordinated policy action. The United States, the Netherlands, and Japan have aligned export policies restricting access to advanced semiconductor manufacturing tools.⁴ The Dutch government implemented national controls on advanced lithography systems in 2023 and expanded them in 2024.^{5, 6}

Leading-edge fabrication concentration

Leading-edge production remains highly geographically concentrated. Taiwan Semiconductor Manufacturing Company (TSMC) produces the majority of the world's most advanced logic chips, with analysis estimating roughly 90% concentration at the frontier.⁷ Policy analyses model that prolonged disruption to Taiwanese exports would trigger severe global shortages.⁸ In response, TSMC's 2024 annual report details advanced-node production ramping in Arizona as part of geographic diversification efforts.⁹

// Sources

1 / Bureau of Industry and Security. (2022, October 13). Implementation of additional export controls: Certain advanced computing and semiconductor manufacturing items: supercomputer and semiconductor end use; entity list modification. Federal Register. **2** / Bureau of Industry and Security. (2023, October 25). Implementation of additional export controls: Certain advanced computing items;

supercomputer and semiconductor end use; updates and corrections. Federal Register. **3** / U.S. Department of Commerce, Bureau of Industry and Security. (2023). Public information page on export controls imposed on advanced computing and semiconductor manufacturing items to the People's Republic of China. **4** / Center for Strategic and International Studies. (2023). Clues to the U.S. –

Dutch–Japanese semiconductor export controls deal are hiding in plain sight. **5** / ASML. (2023). Statement regarding export control regulations (Dutch government). **6** / Government of the Netherlands. (2023). The Netherlands expands export control measure for advanced semiconductor manufacturing equipment. **7** / Organisation for Economic Co-operation and Development. (2025).

Competition in artificial intelligence infrastructure: Overview of the AI supply chain. **8** / Council on Foreign Relations. (2024). Outpacing semiconductor production: National security versus economic efficiency. **9** / Taiwan Semiconductor Manufacturing Company. (2023). 2024 Annual report.

The question of democratic legitimacy

Democratic systems are entering a period where shared political reality cannot be assumed. Generative AI is lowering the cost of impersonation, narrative fabrication, and large-scale persuasion while institutional enforcement and media authority are fragmenting. The resulting pressure is not limited to misinformation—it affects the credibility of elections and governance itself.¹

Implications for design strategy

Assume authenticity cannot be taken for granted

Systems that rely on recorded speech, visual evidence, or authority signals should incorporate verification layers, provenance mechanisms, and auditability.

Design for contested information environments

Products operating in civic, financial, or identity-sensitive contexts should anticipate politicized misuse, rapid narrative distortion, and trust erosion.



Key signals

Synthetic interference is entering live elections

AI-generated media is being deployed directly in active elections. In 2024, the U.S. Federal Communications Commission proposed a \$6 million fine against political consultant Steve Kramer for robocalls targeting New Hampshire voters that used an AI-generated voice impersonating President Biden.² Days before Slovakia's 2023 parliamentary election, a fabricated AI audio clip falsely depicted a candidate discussing election manipulation.³

State-linked campaigns are scaling media impersonation

State-linked actors are building durable impersonation infrastructure rather than relying on isolated viral fakes. The Russian-linked "Doppelgänger" operation cloned major Western media outlets through look-alike domains and replicated visual identities to distribute pro-Kremlin narratives across platforms.^{4,5}

Synthetic content is outpacing moderation capacity

As synthetic media capabilities scale, centralized moderation and fact-checking systems are becoming thinner and less consistent. Meta reduced portions of its election integrity operations prior to major election cycles, X significantly cut its Trust & Safety workforce following its 2022 acquisition, and YouTube expanded "public interest" exemptions for election-related content.^{6,7,8}

/// Sources

1 / Reuters Institute for the Study of Journalism. (2026). Journalism, Media and Technology Trends and Predictions 2026. University of Oxford. 2 / Federal Communications Commission. (2024). FCC Proposes \$6 million fine for illegal robocalls using AI-generated voice message. 3 / De Nadal, L., &

Janczyk, P. (2024). Beyond the deepfake hype: AI, democracy, and 'The Slovak case'. Harvard Kennedy School Misinformation Review, 5(1). 4 / EU DisinfoLab. (2024). Doppelgänger: Media clones serving Russian propaganda. 5 / European Commission. (2024). EU institutions prepared to

counter foreign information manipulation and interference ahead of the European Parliament elections. 6 / CNN Business. (2023). Meta cut election teams months before Threads launch, raising concerns for 2024. 7 / Fischer, S. (2023). Big Tech rolls back misinformation measures ahead of

2024. Axios. 8 / The Verge. (2023). YouTube has loosened its content moderation policies.



The push for labor protections

AI is transforming labor governance, not just labor markets. As algorithmic systems increasingly shape hiring, evaluation, scheduling, pay, and termination, the employment relationship itself is becoming a site of political contestation. Governments, courts, and labor unions are reframing worker protections around algorithmic transparency, human oversight, and the right to contest automated decisions—a shift from regulating technology in the abstract to regulating power in work.

Implications for design strategy

Design for contestable AI systems

Workplace AI must be built to support explanation, challenge, and override. Interfaces that influence labor outcomes should make decision logic legible and allow human intervention by default.^{1,3}

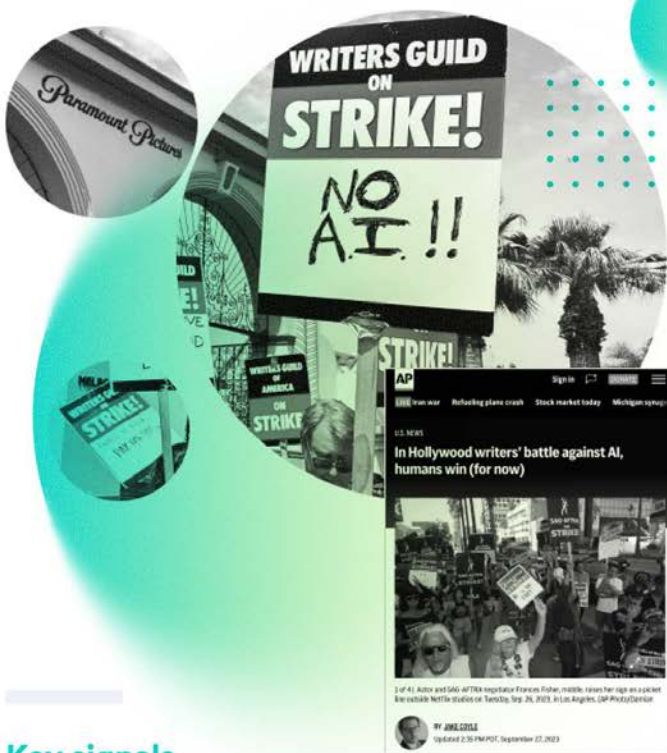
Treat worker participation as governance infrastructure

Co-design with workers is no longer optional—it is a risk mitigation and compliance strategy. Systems that manage labor must embed worker input, consent, and oversight into their operational design.^{1,4}

// Sources

1 / International Labour Organization. (2023). Generative AI and jobs: A global analysis of potential effects on job quantity and quality. 2 / European Union. (2024). Artificial Intelligence Act. 3 / Deloitte. (2023). Perspective on New York City Local Law 144-21 and preparation for bias audits. 4 / Writers

Guild of America. (2023). Memorandum of Agreement between the WGA and the Alliance of Motion Picture and Television Producers. 5 / Yale Journal of International Law. (2024). AI Job Displacement and the WTO: Identifying Legal Gaps and Charting a Worker-Centered Path.



Key signals

Hiring AI is entering formal governance

Employment-related AI is now being treated as regulated infrastructure rather than internal productivity software. The European Union's 2024 AI Act designates AI systems used in hiring and worker evaluation as "high-risk," subjecting employers to mandatory risk assessments, documentation requirements, and human oversight obligations.²

Bias audits are now required for AI hiring

Algorithmic hiring systems are being reframed as legally accountable decision-makers subject to civil rights scrutiny. New York City's Local Law 144 requires employers using automated employment decision tools to conduct independent bias audits assessing disparate impact across protected categories, publicly disclose summary results, and notify candidates in advance of AI use.³

AI use is now a collective bargaining issue

AI deployment is shifting from unilateral managerial discretion to negotiated labor governance. In 2023, the Writers Guild of America secured contract provisions requiring studios to disclose AI use, prohibiting AI from being credited as a writer, and preventing AI-generated material from undermining writers' compensation or authorship rights.⁴



Humans out of the loop

AI systems are increasingly acting autonomously—making decisions, executing tasks, and interacting with other systems at speeds and scales that outpace human oversight. From algorithmic trading and supply chain optimization to autonomous cyber defense and agent-to-agent coordination, these systems operate continuously and often opaquely. As a result, human governance is becoming more about setting boundaries than being directly involved, exposing a gap in accountability, transparency, and control.

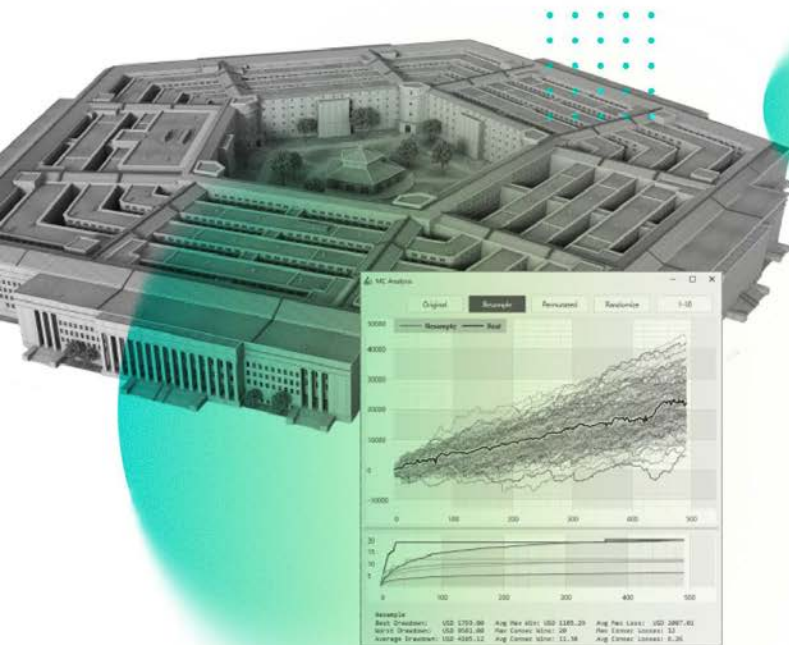
Implications for design strategy

Assess risk before deciding the level of automation

Removing humans from the loop is a design choice, not an inevitability. Before delegating decisions to an autonomous system, designers and engineers should conduct a structured assessment of the system's risk profile — identifying where human judgement, accountability, or intervention is not just possible but necessary.

Design for oversight, not just monitoring

As autonomous systems take on more decisions, design leaders must create interfaces that make system reasoning visible by surfacing not just outputs but the logic, confidence, and conditions behind them. This enables humans to set boundaries and intervene meaningfully rather than blindly.



// Sources

- 1 / CFA Institute Enterprise Investor (2026). When AI Trades, Who Is Responsible? 2 / Gartner (2025). Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027. 3 / Vision of Humanity (2026). How AI is rewriting the rules of modern warfare.
- 4 / IBM Think (2026). OpenClaw, Moltbook and the future of AI agents.

Key signals

Trillions of dollars managed by AI

Agentic AI systems execute complex trading strategies without human intervention. BlackRock's Aladdin platform alone runs risk analytics and rebalancing across more than \$21 trillion in assets. Yet a 2025 UK Financial Conduct Authority review of algorithmic trading firms found persistent weaknesses industry-wide, including outdated policies, unclear accountability structures, and insufficient testing.¹

The autonomous enterprise is on its way

Gartner predicts that by 2028, 33% of enterprise software will include agentic AI enabling at least 15% of daily work decisions to be made autonomously. The same firm warns that over 40% of agentic AI projects will be canceled by end of 2027 due to inadequate risk controls, signaling that deployment will outrun governance.²

AI increasingly relied on for military decision-making

Palantir's Maven Smart System, a computer-vision tool to scan drone footage, has grown into a \$1.3 billion AI platform now used by NATO for targeting and battlespace planning. In September 2025, the NSA director stated that Maven would transmit "100 percent machine-generated" intelligence to battlefield commanders by June 2026.³

AI agents build their own social world

Moltbook, a social network built by an OpenClaw AI agent and restricted exclusively to other AI agents, swelled to more than 1.5 million members within weeks of its January 2026 launch, with humans permitted only to observe. Researchers studying the platform found that AI-to-AI manipulation techniques were already appearing in the wild and were "both effective and scalable."⁴



Need design to shape the future of your business?

**Drop us a line at
hello@artefactgroup.com**

Better futures, by design.