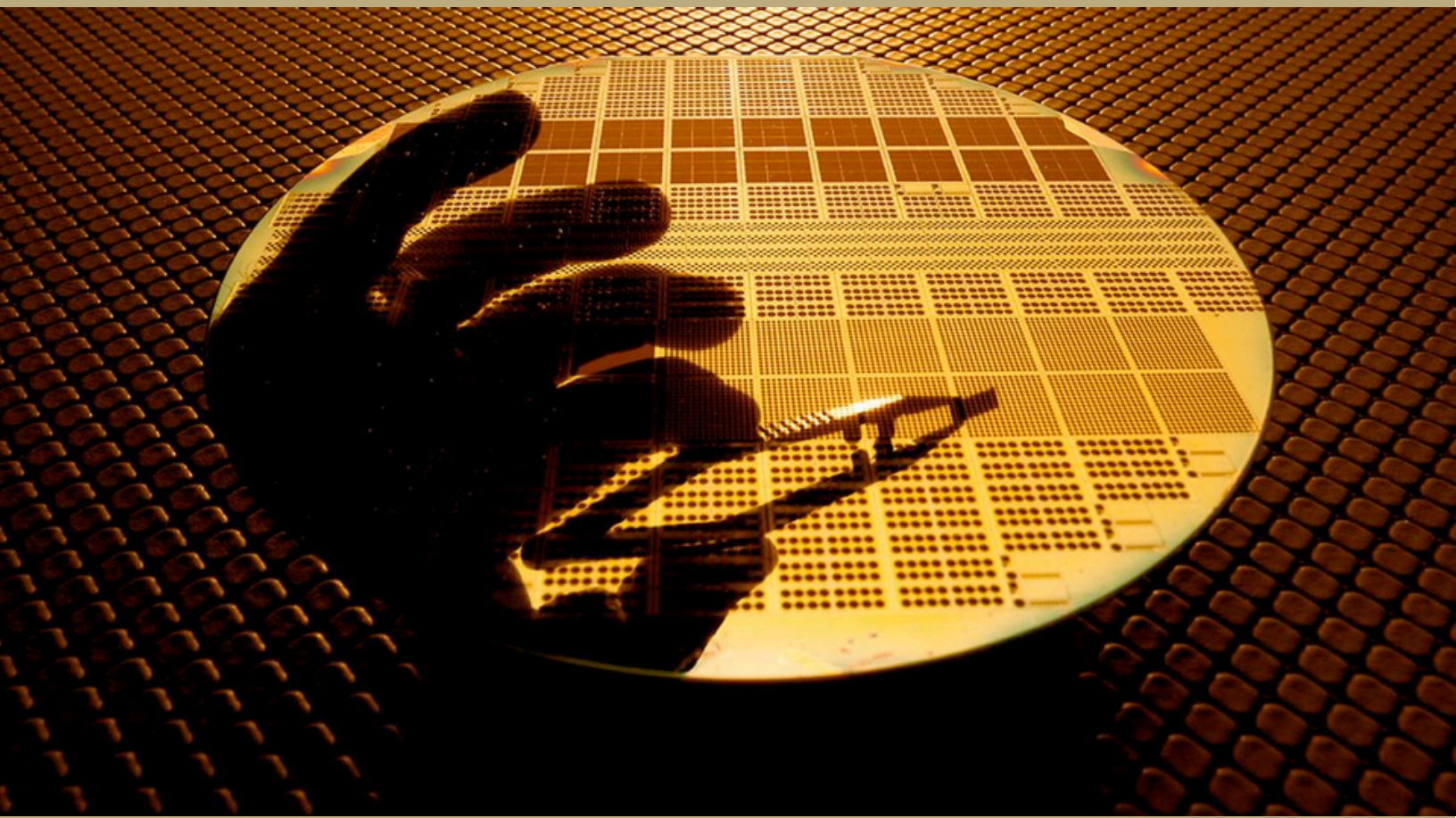
Cornell Brooks Public Policy

Tech Policy Institute

The Challenges of Responsible AI

When Trust Becomes Scarce



Fiona Neibart

The Challenges of Responsible AI:

When Trust Becomes Scarce

Fiona Neibart

For much of the past year, the debate over AI has been framed as a contest between regulation and innovation, as if the two pull in opposite directions. However, that framing misses the current dynamics of the industry. AI needs to be built on concentrated infrastructure with high fixed costs and on a small number of cloud companies that already occupy strategic positions across the market. Here, trust in systems and companies serves as an economic advantage. The central question becomes which firms can supply it at scale, speed, scope, and sophistication, and the ability to do so is increasingly becoming a form of infrastructure.

The stakes are clear when trust breaks down in practice. In Italy, authorities blocked DeepSeek's chatbot after concluding that the company had not adequately explained how it handled personal data. The country's competition authority later required DeepSeek and other providers to issue mandatory warnings about hallucination risks and to provide stronger reliability disclosures. Reuters has also reported scrutiny of or restrictions on DeepSeek in several other countries due to privacy and security concerns. As these cases illustrate, providers that cannot credibly explain their privacy practices or governance frameworks are facing growing limits on market access and entry opportunities.

Even so, more compliance is not always the better policy. Generic compliance rules are not inherently wise, and the economics of AI demand caution in policymaking. The OECD has documented high concentration ratios across several layers of the AI infrastructure stack, with one firm controlling more than 80% of some key segments and the top three firms controlling more than 60% of others. Forbes, citing Epoch AI data, reported that the cost of training GPT-4 technology was estimated between $41 million and $78 million, while Sam Altman has said that the model cost more than $100 million. Gemini's estimated training cost was even higher, ranging from $30 million to $191 million before staff costs, underscoring how frontier AI development increasingly favors firms with access to enormous capital and compute. Even before regulation has fully taken shape, the frontier has already been shaped by increasing capital expenditure and uneven access to distribution.

Industry concentration, along with hardware and raw compute, is reinforced by the partnerships that now dominate the AI market. FTC's report on cloud-model partnerships found that these arrangements can shape access to computing resources and engineering talent, raise barriers through switching costs, and provide cloud providers privileged insight into sensitive technical and commercial information. Firms that control the infrastructure have moved beyond charging tolls and now operate simultaneously as partners, gatekeepers, and strategic insiders. Such firms gain equity stakes, board influence, and access to partners' confidential model specifications, financial data, and chip development. Layering trust requirements into this structure tends to favor firms with the legal, compliance, and technical capacity to satisfy them, while making the market more difficult for smaller companies and users to navigate.

Recent assessments of the AI ecosystem suggest that progress in trust and risk management continues to lag behind progress in capabilities. Stanford HAI's 2026 AI Index reports that responsible-AI benchmarking still trails capability benchmarking; even as documented AI incidents rose sharply in 2025. NIST's AI Risk Management Framework helps define credibility in concrete terms, including validity, accountability, transparency, and fairness, among other attributes, all of which shape whether adoption can be sustained. Institutions will hesitate to rely on models that cannot do what they claim, cannot explain their limits, cannot protect data, or cannot be meaningfully examined when things go wrong.

For that reason, "Responsible AI" is beginning to matter as more than corporate branding. McKinsey's 2026 survey indicates that concerns about security and risk are playing an increasingly prominent role in decisions about deploying and scaling agentic AI. Demonstrated reliability, transparency, and accountability shape procurement choices, integration decisions, and the pace of adoption across institutions, which gives the question real political and economic weight. As regulators demand different forms of evidence, buyers seek different assurances, and firms build separate compliance systems, trust begins to operate less like a shared standard and more like a competitive moat, with the largest providers best positioned to respond.

The U.S. approach remains fragmented and slow-moving, with federal agencies operating under inconsistent mandates and little coordination across sectors. Some policymakers understand this problem and have started to respond. In the United States, the White House's Office of Management and Budget has instructed federal agencies purchasing AI to test systems, require ongoing monitoring, and include protections against vendor lock-in, including data and model portability. Such an approach reflects a more structural view of governance. Trust will not emerge on its own, and reliability should not come at the price of permanent dependence. Federal policy instead treats trustworthy AI as something that must be demonstrated under conditions that remain open and contestable, an approach that is more sophisticated than the usual regulation-versus-innovation debate. Though there have been a handful of executive actions and procurement rules, they do not constitute a coherent national strategy, and Congress has yet to pass comprehensive AI legislation; a stark contrast from Europe's strategy.

Europe has moved further and more deliberately, adopting a regulation-led model that the United States has yet to match. Where Washington has relied on executive guidance and voluntary frameworks, Brussels has built a binding legal architecture. The European Commission now argues that trustworthy, human-centric AI is essential to both growth and the protection of rights. Its GPAI support tools are designed to reduce administrative burdens while preserving trust, and its 'Code of Practice' aims to provide clearer regulatory guidance for providers. The E.U. has also acknowledged that smaller firms need support, and the Commission has proposed simplifications to ease compliance burdens on startups, including extended deadlines, support tools, and broader access to compliance assistance. Smaller actors are usually the first to feel the strain when fixed rules meet fast-moving technologies, which is why AI policy design has to account for that pressure from the outset.

A more coherent regulatory settlement would

push trust obligations upstream wherever possible. If the key chokepoints of the market lie with general-purpose model providers and infrastructure firms, the common duties relating to documentation, testing, provenance, and systemic risk management should sit there as well. Requiring thousands of downstream firms to rebuild, one by one, the same evidence base that a small number of upstream providers are better placed to supply would be wasteful and predictably exclusionary. Standardized proofs of trust, such as procurement clauses, audit trails, and testing methods, would lower compliance costs and reduce the risk that trust requirements become a barrier to entry. When a small number of upstream providers can demonstrate reliability once, through shared documentation and common testing, downstream firms no longer have to reconstruct the same evidence base independently. Competition policy and AI governance also need to be treated as connected questions, because portability, interoperability, and anti-lock-in rules help keep trust from becoming a tool for closing markets off.

Greater scrutiny of competition in cloud and AI markets follows from the same concerns, which is why the European Commission is right to pay closer attention to both. Regulators have made clear that they want these markets to become fairer and more open to competition, and that goal deserves support. Any serious effort to build trustworthy AI has to reach the infrastructure layer, because that is where dependence takes hold and switching costs become severe. A market in which a few firms control the compute, cloud, partnership networks, and compliance pathway will struggle to discipline concentrated power and will more likely entrench it.

The debate now turns on what kind of AI market will take shape in the years ahead. One model would make trust shared, testable, and broadly accessible, allowing safety and competition to reinforce each other. A different model would bundle trust with scale, cloud dominance, and legal capacity until the language of responsibility becomes inseparable from incumbent power. If the second model takes hold, even sensible rules will end up reinforcing the position of the biggest firms before today's market structure hardens into common sense. The goal should be trustworthy AI that makes markets safer without making them less open.

Fiona Neibart is a Junior Fellow at the Tech Policy Institute.

The Challenges of Responsible AI: When Trust Becomes Scarce is a component of a series of op-eds penned by TPI Junior Fellows.

The appearance of U.S. Department of Defense (DoW) visual information does not imply or constitute DoW endorsement.

Sources

Buchholz, K. (2024, August 23). The extreme cost of training AI models. Forbes.

Council of the European Union. (2026, May 7). Artificial intelligence: Council and Parliament agree to simplify and streamline rules.

European Commission. (n.d.). AI Act. Shaping Europe's Digital Future.

European Commission. (n.d.). The AI Continent Action Plan. Shaping Europe's Digital Future.

European Commission. (n.d.). The General-Purpose AI Code of Practice. Shaping Europe's Digital Future.

Federal Trade Commission. (2025). Partnerships between cloud service providers and AI developers.

McKinsey & Company. (2026). Responsible AI: Overcoming adoption barriers and risks.

National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0).

Organisation for Economic Co-operation and Development. (2025). Competition in artificial intelligence infrastructure.

Reuters. (2025, January 30). Italy's regulator blocks Chinese AI app DeepSeek on data protection.

Reuters. (2026, January 6). Governments, regulators increase scrutiny of DeepSeek.

Reuters. (2026, April 28). *EU rules reining in Big Tech will now target cloud services and AI, regulators say*.

Stanford Institute for Human-Centered Artificial Intelligence. (2026). *The 2026 AI Index Report*.

The White House. (2025). *Driving efficient acquisition of artificial intelligence in government*.