Asking The Wrong Questions

Part 2: AI Risks

By Ben Phillippi

The Wrong Risk

The loudest debate about AI is about whether it will become sentient, whether a super-intelligence will develop its own goals and decide that humanity is in the way. This scenario dominates headlines, fuels Hollywood, and consumes an enormous share of the policy conversation. It is also, I believe, a distraction from the risks that actually matter.

The case for a rogue AI requires a long chain of assumptions. It demands unbounded autonomy, recursive self-improvement without bottlenecks, and a system that somehow develops independent motivation to act against human interests. Each assumption has to hold for the next one to matter, and the model is a house of cards.[1]

Michael Pollan makes a simpler point. Sentience requires feeling, and feeling requires vulnerability. A body that can be hurt, the ability to suffer, and perhaps mortality. AI has none of these. Whatever a chatbot reports about its inner experience is, as Pollan puts it, weightless. This is not consciousness.[2]

Arvind Narayanan and Sayash Kapoor, authors of "AI Snake Oil," go further. They argue that the preconditions for a truly autonomous, self-directed AI are empirically unlikely to materialize. AI systems do not have independent motivation and they do not want things. The danger is not that AI will decide to harm us, the danger is that people will use AI to harm each other.[3]

That reframing matters, because while we argue about sentient machines, real risks are compounding. Trust in information is eroding, economic displacement is accelerating without a plan, and institutional frameworks are weakening under pressure they were not designed to handle. These are not speculative. They are happening now.

In Part 1 of this series, I argued that AI will restructure work rather than eliminate it and that the opportunity ahead is significant. I believe that. But I also believe the restructuring will not be painless, and we are not preparing for the scale of disruption it will create. The risks are real, they are just not the risks dominating the conversation.

AI as Normal Technology

There is a more useful way to think about AI risk. Arvind Narayanan and Sayash Kapoor call it "AI as Normal Technology." The word "normal" is misleading. They compare AI to electricity in the paper's second sentence. What they mean is that AI, like every other powerful general purpose technology before it, should be understood through the frameworks of diffusion, adoption, and institutional response, not treated as an unprecedented exception to how technology reshapes society.[4]

The core insight is that there is a long causal chain between AI capability and real-world impact. Benefits and risks are realized when AI is deployed and adopted at scale, not when a model posts an impressive benchmark score. GPT-4 scored in the top 10 percent on the bar exam, but that tells us almost nothing about practicing law, because the exam tests exactly what language models are good at and ignores the judgment, creativity, and context that define the profession.[5] This gap between capability and utility shows up everywhere, and it means we are consistently overestimating how quickly AI will transform industries.

Adoption is also slower than headlines suggest. Generative AI reached 40 percent of US adults within two years, but that translated to less than 4 percent of actual work hours.[6] Even with instant digital distribution, the real speed limit on AI's impact is the speed at which people, organizations, and institutions can change. Narayanan uses a vivid analogy. Amtrak bought trains that can go 160 miles per hour, but the track infrastructure averages 65. Most AI productivity claims are about upgrading the trains while ignoring the track.[7]

In safety-critical areas like medicine, criminal justice, and autonomous vehicles, diffusion lags decades behind capability by design. Regulation and institutional caution enforce this, and they should. This slowness is not a failure. It is how societies absorb powerful technologies without being destroyed by them.

This is actually good news. It means there is time for institutions to regulate, for workers to adapt, and for organizations to make deliberate choices about integration. But "slow" does not mean "harmless," and "normal" does not mean "comfortable." The Industrial Revolution was a normal technology transition. It still made life significantly worse for workers across two generations before wages finally rose.[8]

Ethan Mollick, a Wharton professor and author of "Co-Intelligence," captures what this looks like in practice. He calls it "many little apocalypses." Not one dramatic event, but educated, well-paid, creative workers facing displacement for the first time. Organizations losing agency because they are not making deliberate choices about AI. The quiet erosion of human decision-making autonomy. His key insight is that if you do not actively decide how to integrate AI, that choice gets made for you.[9]

I find the Normal Technology framework most convincing. AI is powerful, transformative, and worth taking seriously. But it is not magic and it is not a separate species. Understanding it through history gives us better tools than treating it as unprecedented.

The normal technology frame gives us reason for measured confidence. But there is a specific pattern of risk within that frame that deserves much more attention than it is getting.

The Accumulating Storm

The AI risk debate has been captive to one model of catastrophe. A sudden, dramatic event where a superintelligence seizes control and humanity is eliminated. We have already discussed why that scenario is a house of cards. But Atoosa Kasirzadeh, a Carnegie Mellon professor of philosophy and AI, argues that in focusing on it, we are missing a more plausible and more dangerous pathway to catastrophe.[10]

She calls it accumulative risk. In this model, existential damage arrives not through a single event but through the gradual accumulation of individually manageable disruptions that compound over time. AI-driven erosion of trust in media and information. Economic instability from rapid labor displacement. Decline of democratic institutions through surveillance and algorithmic manipulation. Cyber-vulnerabilities proliferating through increasingly interconnected systems. None of these, on their own, end civilization. But each one weakens societal resilience, and together they create feedback loops and cascading failures across systems that depend on each other. Eventually a triggering event, which might itself be modest, causes damage that cannot be reversed because the structural supports had already quietly given way.

This is the boiling frog applied to civilization. The water gets warmer so gradually that by the time anyone notices, it may be too late to respond.

This framing matters because it bridges a divide that has paralyzed the AI community for years. The "Safety" camp worries about future superintelligence. The "Ethics" camp worries about present-day harms like bias, misinformation, and surveillance. These two groups have largely talked past each other, sometimes with real hostility. Kasirzadeh's argument reframes the relationship. The present-day harms the Ethics camp worries about are not a separate, lesser category of problem. They are potentially the early stages of systemic failure, if they are allowed to accumulate and compound.

This is where the Normal Technology framework meets its most serious challenge. Narayanan and Kapoor count on institutions to manage AI's integration and I find their reasoning persuasive. But what if AI is gradually weakening the very institutions it depends on to manage its impact? Eroding trust in the information those institutions need to function. Displacing the workers those institutions were built to protect. Concentrating power in ways that make institutional oversight harder to enforce. That tension, between trusting institutions to manage AI and worrying that AI is degrading those same institutions, is the most important question none of these frameworks fully resolves.

This is the risk that concerns me most. Not a rogue AI with its own agenda, but the slow erosion of the systems our society depends on to function. And the erosion is not happening equally.

Who Benefits?

AI is built on the combined knowledge of all humanity. Every book, every article, every conversation posted online contributed to training these models. But the economic benefits of this collective contribution are accruing in an outsized way to a handful of the largest companies in the world.

This is not new. In "Power and Progress," Nobel laureate Daron Acemoglu and Simon Johnson draw on a thousand years of technology history to demonstrate that the benefits of powerful technologies have never automatically trickled down. They required political struggle, countervailing institutions, and deliberate redirection.[11] The printing press benefited elites for over a century before it empowered broader literacy. The Industrial Revolution, as we have already discussed, made life significantly worse for workers across two generations before the gains were shared. The pattern is consistent and the lesson is clear. Powerful technology creates enormous value, but that value flows to whoever controls it unless society actively intervenes.

Acemoglu asks a question that none of the dominant AI frameworks adequately address. AI for whom? He argues that the current trajectory of AI development is optimized for automation, replacing workers, rather than augmentation, making workers more capable. This is not an accident. It reflects the incentive structures of the companies building it.[12]

Thomas Friedman documented how globalization created enormous aggregate wealth while devastating specific communities. The political backlash from displaced populations has reshaped democracies worldwide. AI is following the same pattern at a faster pace, and we have even less infrastructure to manage the transition.[13] As a society, we have almost no resources in place to help people through this disruption. No retraining programs at scale. No income support frameworks designed for this kind of shift. No serious policy conversation about how to distribute the gains from a technology built on everyone's knowledge.

The companies that own the most capable AI models are consolidating power in ways that make competition harder and oversight more difficult. AI models that are owned and controlled by a single company, with no visibility into how they work, are becoming infrastructure that businesses depend on for daily operations. These "closed models" are dependencies that are difficult to reverse, and they concentrate power in exactly the ways Acemoglu warns about.[14]

Further, when the economic gains of AI are captured by a handful of massive corporations, it actively accelerates the accumulative risks we discussed earlier. A concentration of wealth inevitably leads to a concentration of power, which weakens the democratic institutions we are relying on to govern this technology.

In Part 1, I made the case that the pie is growing. I believe that. But a growing pie that is captured almost entirely by those who were already at the table is not the outcome any of us should accept. This is not just a policy problem. It is a leadership problem. Every executive making decisions about AI deployment is making choices about who benefits and who does not.

We know what the real risks are. The question is whether we will organize around them.

What We Should Actually Be Doing

The wrong question is whether AI will become sentient and destroy us. The right question is whether we will address the accumulating disruptions that are already weakening the systems our society depends on. That means building adaptive institutions that can respond to whichever future materializes. It means creating real transition support for displaced workers at the scale of the disruption. And it means developing governance frameworks that address the concentration of power and ensure the gains from this technology are broadly shared. But governance cannot come at the cost of innovation. We should be supporting open models and open knowledge, not reinforcing the closed dependencies described above. The faster this technology advances in the hands of many, the less likely it is to be controlled by a few.

AI capabilities are advancing faster than governance can adapt. This is not a reason to panic, but it is a reason to act with urgency and intention. Every hour spent debating sentient machines is an hour not spent building the frameworks that could actually protect people.

For those of us in a position of leadership, the choices we make about how AI is deployed in our organizations are not just business decisions. They are decisions about what kind of society we are building. In Part 1, I asked us to rethink how work gets done. In this article, I am asking us to take seriously what happens to people and institutions during that transition. Both questions matter. Neither is being asked enough.

> Start with why. Why are we building this? Who does it serve? What kind of future are we creating? These are the questions worth asking. The answers will determine whether this technology lifts all of us or only a few.

Citations

1. AI 2027 Project. "AI 2027: A Scenario." https://ai-2027.com/
2. Pollan, M. (2026). A World Appears: A Journey into Consciousness. Penguin Press. See also NPR interview: https://www.npr.org/2026/02/19/nx-s1-5713514/michael-pollan-ai-consciousness-a-world-appears
3. Narayanan, A. & Kapoor, S. (2024). AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference. Princeton University Press. See also "AI as Normal Technology" (2025): https://www.normaltech.ai/
4. Narayanan, A. & Kapoor, S. (2025). "AI as Normal Technology." Knight First Amendment Institute. https://knightcolumbia.org/content/ai-as-normal-technology
5. Narayanan, A. & Kapoor, S. (2025). "AI as Normal Technology," Section on Benchmarks and Construct Validity.
6. Narayanan, A. & Kapoor, S. (2025). "AI as Normal Technology," Section on Diffusion and Adoption Speed.
7. Narayanan, A. (2025). Substack note on productivity and bottlenecks. https://substack.com/@aisnakeoil/note/c-189219135
8. Acemoglu, D. & Johnson, S. (2023). Power and Progress: Our Thousand-Year Struggle Over Technology and Prosperity. PublicAffairs.
9. Mollick, E. (2024). Co-Intelligence: Living and Working with AI. Portfolio/Penguin. See also "We're Focusing on the Wrong Kind of AI Apocalypse," TIME. https://time.com/6961559/ethan-mollick-ai-apocalypse-essay/
10. Kasirzadeh, A. (2025). "Two Types of AI Existential Risk: Decisive and Accumulative." Philosophical Studies. https://arxiv.org/html/2401.07836v3
11. Acemoglu, D. & Johnson, S. (2023). Power and Progress.
12. Acemoglu, D. (2024). "The Simple Macroeconomics of AI." NBER Working Paper 32487. https://www.nber.org/papers/w32487
13. Friedman, T. L. (2005). The World Is Flat: A Brief History of the Twenty-first Century. Farrar, Straus and Giroux.
14. Acemoglu, D. (2025). "Will We Squander the AI Opportunity?" Project Syndicate. https://www.project-syndicate.org/commentary/ai-on-a-socially-harmful-path-by-daron-acemoglu-2025-02