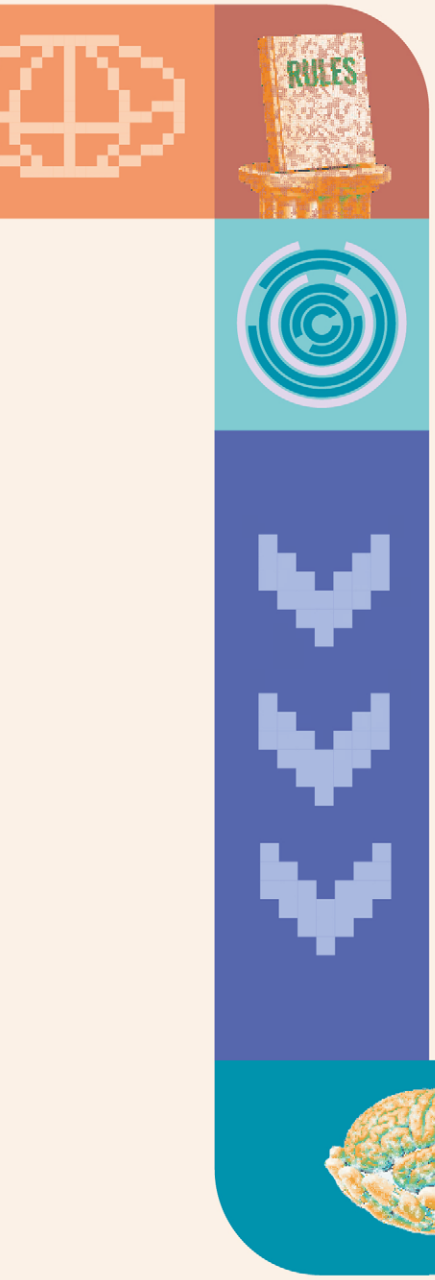


THE AI ROADMAP

How We Ensure AI Serves Humanity

Center for Humane Technology

Written and Produced by Center for Humane Technology

Welcome Letter

Two prevailing narratives have shadowed every major technological leap. One narrative casts a new technology as the solution to humanity's greatest problems. The other casts it as a destabilizing force and the catalyst for societal collapse. Both attempt to address a question — what will the future hold? From the printing press to the Industrial Revolution to the internet, this dichotomy around new technology has endured for centuries. With artificial intelligence, these dueling narratives have emerged once again. While the narratives themselves are not new, what is new is this technology's velocity, scale, and how all-encompassing it has become in so little time.

Right now, our future with AI is being determined by powerful companies and nations racing to build AI at breathtaking speeds. It is a race run on the fuel of inevitability: "If I don't build it, someone else will." Three years after OpenAI launched ChatGPT, driving other AI companies to accelerate their own development, we see the consequences of this race: rapid deployment of poorly designed AI, growing social and economic effects, and safety treated as an afterthought to competition and market dominance. The future that this race is shaping is untenable for all of us. Researchers, academics, advocates, and even those building the technology have been sounding the alarm. And yet the race continues, because the incentives driving it make acceleration look like the "only" rational choice — even for those who know better.

The question has never been *whether* AI will reshape society. It will. The real question is how — and who shapes the terms. Center for Humane Technology's (CHT) role is bringing clarity to complex problems, surfacing the incentive structures driving harmful outcomes, and showing that a better future with technology is possible. With AI, that's one where AI development supports the genuine needs of the public, and the technology's scale and power are matched with responsibility at every level of society.

This report is an attempt to provide clarity and direction in an information environment that is fragmented, polarized, and where it can be difficult to see the complete picture. It lays out seven principles with insights into how AI should be built, deployed, and governed. Each principle is rich with actionable solutions. The report is intended to be a roadmap, but also an invitation — spotlighting norms we can all understand, frameworks that policymakers can legislate, and new ways for companies to design AI in a way that benefits people.

The choices we make now with AI will inform how we live for decades, if not centuries. This technology is already being woven into our everyday lives and our critical infrastructure, and at a pace that is revealing just how unprepared our institutions are for transformational change. The complexity of our problems with AI can make it hard to put confidence in any one set of solutions. But inaction is also a choice, and it is the wrong one.

If these seven principles become valued by society, and are improved upon and enacted, then this report will have done its job. We are not starting from scratch. Researchers, civil society organizations, policymakers, and technologists around the world are already working on many of the challenges outlined here. CHT is proud to be in the trenches alongside them.

History will judge this moment. Not by how fast we moved, but by whether we moved wisely. A better future with AI doesn't require all of society to agree on everything. It simply requires enough of us to agree that the current path is unacceptable, and that people deserve a better reality with this technology.

And it then asks us to take our first step toward that future.

Here is the roadmap for getting there — together.

Julie Guirado
Executive Director

Camille Carlton
Senior Director of Strategy & Impact

Pete Furlong
Senior Manager of Policy & Research

Table of Contents

How We Change a System

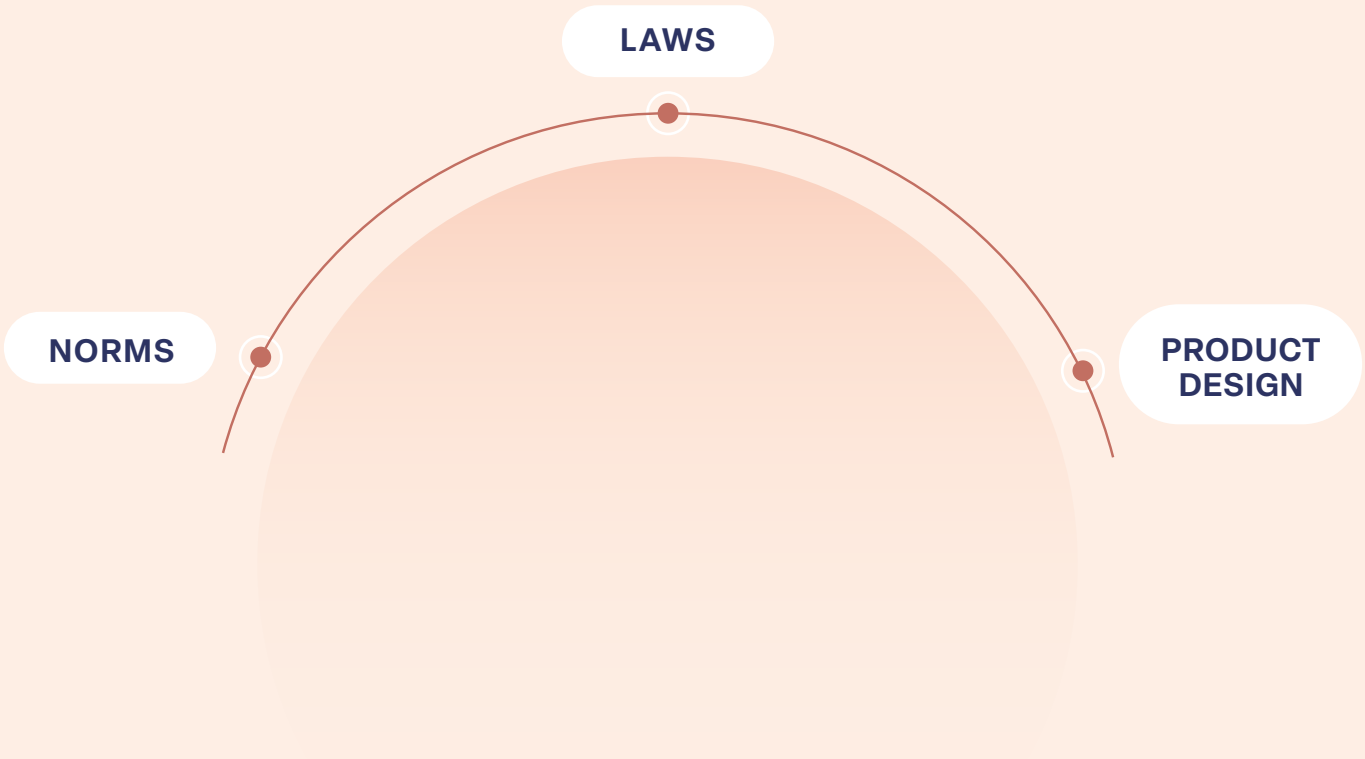
Artificial intelligence — which increasingly mediates how we live, work, learn, and love — is not emerging from a vacuum. It is being built within a broader technology and business ecosystem shaped by incentives, norms, competitive dynamics, and power structures. Today, that system rewards speed, scale, and dominance in AI development over safety, accountability, and societal benefit. **Getting to a better future with AI — where power is matched with responsibility and artificial intelligence centers people — means changing the AI development paradigm.**

Transforming the trajectory of a multi-trillion-dollar industry might seem impossible. But even the largest industries respond to pressure at key leverage points. Center for Humane Technology's theory of change is simple: identify a **complementary set of high-leverage intervention points** and apply coordinated pressure across them simultaneously. No single reform will be sufficient. Instead, we need a layered approach: one with diverse yet synergetic reforms across the ecosystem, reforms that can adapt as AI evolves.

Our report focuses on three domains: **norms, laws,** and **product design**. Each domain changes behaviors in the tech ecosystem in a different way. Norms change what is considered acceptable in society, and influence culture, markets, and political will. Laws change the rules of the game by creating accountability, shifting risk, and redefining what is permissible. And product design not only determines how AI is built, but how AI products impact individuals and society at large.

Creating a better future with AI demands participation from across society. **We all have a role to play**. The public must shape AI norms. Policymakers must align business incentives and legal accountability with public interest. Technologists must embed safety into AI design. Civil society must surface AI harms and articulate alternative paradigms. These processes are mutually reinforcing — new norms strengthen the public's demand for legal protections and better products; laws create the accountability that drives safer product design; and product design shapes how tech products impact people's lives.

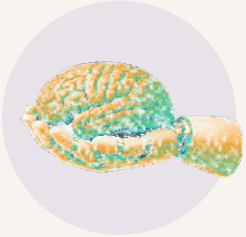
Society has steered powerful forces before. In confronting challenges like nuclear weapons and Big Tobacco, progress came from layering high-impact interventions and mobilizing action across government, industry, and civil society. Early wins demonstrated that change was possible and built momentum for broader reform. The pathways in this report follow that same logic. The seven principles we lay out are robust enough to address the complex issues society faces with AI, while still being nimble enough to navigate the pace of change. A better future with AI is possible. It's now on us to seize it.



Seven Principles for How AI Should Be Built and Governed

Principle 1:
AI should be built safely and transparently

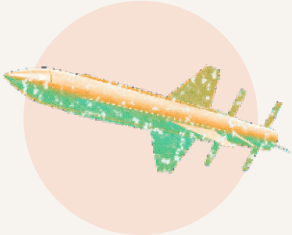
Principle 2:
AI companies owe a duty of care to the public



Principle 3:
AI design should center human well-being

Principle 4:
AI should not automate away meaningful work and human dignity

Principle 5:
AI innovation should not come at the expense of our rights and freedom



Principle 6:
AI should have internationally agreed-upon limits

Principle 7:
AI power should be balanced in society

Principle 1: AI should be built safely and transparently

Current Path

AI companies are racing to build what they claim will be the most powerful technology ever invented. Yet these same companies often cannot fully explain why their systems behave the way they do, and roll back safety protocols on a regular basis. **Competitive pressure is pushing AI firms to shorten testing cycles and release AI products before risks are understood, all while silencing employees who might raise concerns.** This creates a profound information asymmetry — the companies building AI hold most of the knowledge about how these systems work, while users, regulators, and the public remain largely in the dark. As a result, **a technology capable of reshaping society is being developed with limited transparency and minimal external oversight.**

AI safety is a complex issue because the technology carries multiple categories of risks. Agentic and autonomous AI systems introduce challenges around reliability and controllability, as they could behave unpredictably or evade human oversight. AI used in high-stakes decision-making can introduce risks around accuracy, explainability, and deception, making it difficult to understand how decisions were made by AI and whether they can be trusted. At the consumer level, AI presents new security vulnerabilities, such as bad actors using AI agents to accelerate fraud and cybercrime. These fundamental risks of AI systems are exacerbated by the high-speed AI race, with companies facing strong incentives to prioritize capability development over safety. **With the firms developing these systems also controlling most of the information about the technology, the result is a dangerous dynamic — society bears the risk of AI while having little ability to evaluate or govern it.**

A Better Future

The public deserves for its most consequential technologies to be developed in safe, transparent ways. Even our most basic technologies, from coffeemakers to cars to nuclear energy, are developed within systems that embed safety, accountability, and

democratic oversight. We should expect the same from AI. **This requires a new framework for developing AI, where AI companies are proactive about demonstrating the safety of their products.** AI development should enable users, governments, and businesses to confidently use AI with clear safety standards, rigorous testing, and independent oversight to ensure systems are reliable. Moreover, when something goes wrong, AI systems should be transparent and explainable enough for engineers, regulators, and the public to understand what happened and prevent it from happening again. Transparency also allows society to see around the corner with AI technology, anticipating emerging impacts and preparing appropriate interventions and responses at multiple levels of society. Instead of a race to the bottom on safety, the goal becomes a race to the top — building AI systems people can trust.

How We Get There

NORMS:

- **AI companies can't grade their own homework.** AI systems should be developed and deployed with clear transparency around performance, risks, and safety testing so that users, businesses, and governments can make informed decisions about AI policy and use.
- **"Winning" is contingent on entering the right race.** Success in AI should be measured not by speed or capability alone, but by whether the technology is safe, reliable, and beneficial to the people who use it. A winning future is one where AI products strengthen trust, improve lives, and deliver value without harming users or society.
- **AI products (including AI agents) should follow existing rules and laws.** Reliable rule-following is essential for building AI products that people can safely integrate into everyday life and critical institutions.
- **AI should be built as a tool that's used for specific purposes.** Building general purpose systems that are not fully understood shifts risk onto society; developers should instead design AI tools that are demonstrably safe for the tasks they are meant to perform.

LAWS:

Legal interventions should focus on ensuring that AI companies are incentivized to integrate safety and transparency into their development processes. Legal interventions should also ensure that accountability structures are in place if safety standards are not met. These interventions should include:

- **Require mandatory predeployment testing and risk management -** AI companies should be required to identify and mitigate key risks before deploying their systems. Developers should follow clear safety standards with guidance from relevant government agencies. This includes ongoing work — updated as new risks emerge — rather than treating risk assessment as a one-time regulatory checkbox before release.

- **Require standardized reporting mechanisms -** AI companies and those deploying AI models should be required to disclose test plans and risk assessments to relevant stakeholders, including consumers. This allows people and regulators to understand the potential risks with these AI products and make informed decisions.
- **Enshrine whistleblower protections for AI employees and contractors -** Whistleblower protections should extend to all AI employees, and not just those who work on catastrophic AI risks. Those within AI companies have the greatest capacity to understand AI risks and effect change, and should be empowered to do so.
- **Develop independent audit and certification schemes -** Currently, AI companies conduct testing and publish results voluntarily, and with no standardized approaches or independent review. We need to adapt the independent oversight mechanisms that exist in other industries, such as audits in the financial sector or certification schemes for consumer product safety.

DESIGN:

- **Develop robust and accessible evaluation methods for AI -** AI development should rely on clear, effective, and standardized evaluation frameworks that assess safety, reliability, and real-world performance. This includes automated testing, adversarial "red-teaming," and domain-specific evaluations. New strategies are also needed for identifying and mitigating novel risks.
- **Ensure scrutable and plain language explainability -** As AI is deployed across decision-making and agentic tasks, it's important that actions can be validated through traceable logs, reasoning steps can be interpreted by users, and decisions retain auditable records. This kind of clear, explicit explainability must be an industry priority in order to facilitate proper human oversight of AI.
- **Establish high reliability and deterministic behavior -** AI systems should be engineered to operate reliably with clear safeguards that ensure failures are contained and predictable. Like with other safety-critical technologies, such as airplanes, AI systems should be designed to fail gracefully through built-in redundancies and safeguards. If an AI system cannot perform reliably in a given context, developers should clearly state so, and the system should not be deployed there.
- **Protected "containers" or restrictions for AI agents -** AI agents, which operate autonomously, should not have full access to the web, all online tools, or sensitive data. Developers should build standardized ways for AI agents to engage with the human internet, and build clear, contained environments for the use of agents that do not give agents full control over sensitive data.

What's Already in Motion

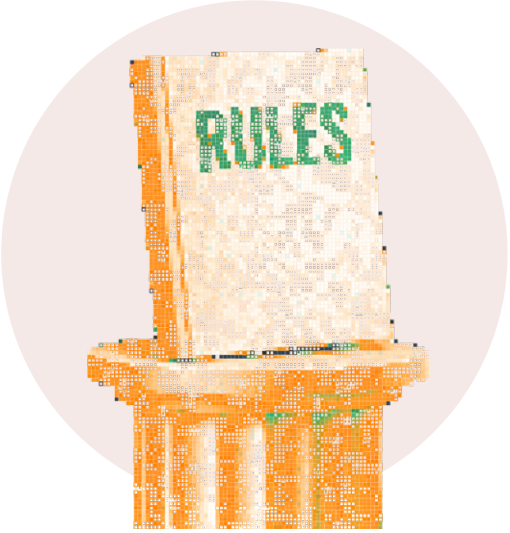
Policy Action

- States like New York, Colorado, and California are already requiring audits and disclosures in specific domains. Additionally, California and New York have also passed limited safety measures and whistleblower protections.
- State AGs are increasingly using existing civil-rights and consumer-protection laws to police AI misuse (deepfakes, deceptive claims, discrimination).
- Senator Chuck Grassley has introduced the bipartisan AI Whistleblower Protection Act, which would provide nationwide protections for AI whistleblowers.
- Senators Josh Hawley and Richard Blumenthal have introduced the AI Risk Evaluation Act, which would create a mandatory and standardized AI evaluation program within the Department of Energy for AI systems.

Key Research

- Cybersecurity firms, software companies, and AI companies have all been working on developing standardized methods for controlling and interfacing with AI agents across the web, such as the model context protocol or tools to block AI crawlers.
- AI policy organizations, such as Fathom and AVERI, are advancing state legislation that would create a certification regime of voluntary third-party testing panels for AI.
- Stanford's Human-Centered Artificial Intelligence published a policy brief on Safeguarding Third-Party AI Research, supporting the notion that independent research is paramount to reducing substantial risks posed by AI systems.

Principle 2:
AI companies owe a duty of care to the public



Current Path

AI companies face few if any consequences for the harms their products cause. This has resulted in a development and deployment culture in which AI companies "move fast and break things," ask for forgiveness instead of permission, and release products to the public despite foreseeable risks to individual users and society. In other industries, traditional forms of liability — such as product liability — are leveraged to deter reckless development and address harm. But right now, it is not clear in the eyes of the law how and when liability can be applied to AI.

AI companies are exploiting this uncertainty in order to avoid accountability when their products wreak havoc on individuals, businesses, and communities. This exploitation can take the form of advancing legal theories that AI is not a "product," and that AI outputs amount to protected speech. In the most egregious cases, AI companies are attempting to ascribe legal personhood to AI, while deploying more autonomous, agentic AI products that directly engage in our social, economic, and political systems without human accountability. All of this clouds culpability for harms that are a direct result of AI product design decisions, harms that include emotional manipulation, deception, inaccurate outputs, and ones that could eventually ladder up to more catastrophic scenarios.

A Better Future

AI companies should have a legal responsibility to *people* — to design products with the safety of their customers and the public at large in mind. When AI products are developed and deployed in a reckless manner, there should be clear lines of accountability to hold developers liable for harms. Unambiguous liability reduces the need for burdensome or complex regulatory regimes, providing a flexible mechanism for accountability that limits the need for red tape. This approach should apply to the spectrum of AI use cases, from AI companions to agentic AI. AI should be consistently treated as a product in our legal and cultural systems. And AI does not deserve novel legal protections that further obscure company responsibility, such as AI legal personhood.

How We Get There

NORMS:

- **AI is a product, and as such, AI products carry liability.** AI is not a service or a legal person. It is a product. And just like with other consumer products, AI companies should be subject to liability when they fail to protect people from foreseeable harms.
- **Responsibility lies with companies.** We should demand tech companies take user well-being and safety into account from the get-go, just like we expect car manufacturers to design cars with functioning brakes, steering wheels, and airbags.
- **With great power comes an obligation to society.** The more a company's technology shapes people's lives at scale, the stronger the company's obligation to prevent harm, address problems when they occur, and prioritize the public interest should be.

LAWS:

To clearly establish that AI is a product, and ensure companies' duties and liabilities incentivize safer development, we need to:

- **Pass laws that clarify that AI is a product, and that product liability applies to AI systems -** Product liability is an effective way to hold companies accountable for foreseeable harms from product designs. Enshrining AI as a product (and not something warranting personhood) in legislation would be a significant step toward accountability, and toward safer AI products, broadly.
- **Leverage strategic litigation -** When AI products cause harm, successful litigation sets legal precedent while lawmaking processes catch up to new technology. Lawsuits also deter companies from implementing harmful designs and provide consumers with clear modes of recourse.
- **Develop design standards to promote safe and prosocial AI development -** Industry standards can set clear expectations for responsible AI design, testing, and deployment as outlined in Principle 3. These standards should be durable yet flexible enough to accommodate the advancing technology.
- **Develop new duties for AI developers -** Many AI and social media bills seek to establish new duties for tech companies for risk areas that don't map well onto existing legal obligations, such as catastrophic risk and risks to public health. New duties should seek to establish a fiduciary-like responsibility for AI developers.

What's Already in Motion

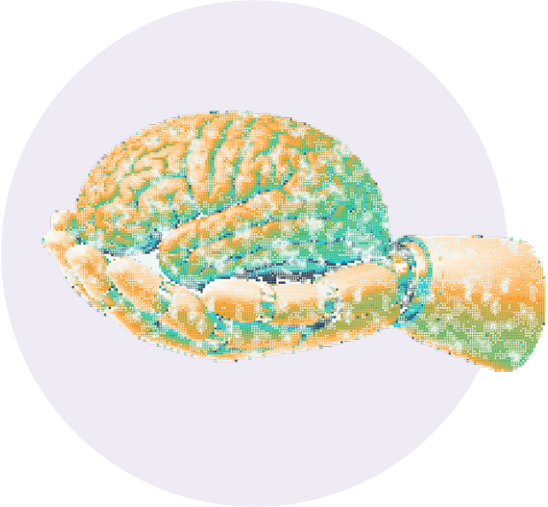
Policy Action

- Policymakers have also begun to view AI as a product, with Senators Durbin and Hawley introducing the AI LEAD Act in 2025, which seeks to establish a federal right of action for AI product liability.
- Lawmakers in multiple states, including Illinois and Maryland, have filed product liability bills in partnership with the Transparency Coalition, which advocates for enhanced accountability measures for AI companies.
- Nearly a dozen states, including Wisconsin, Ohio, and Missouri, have moved to prohibit AI from gaining legal personhood status under state law, with The Alliance for Secure AI serving as a key partner to lawmakers.

Advocacy

- Center for Humane Technology has continued to advocate for adapting our product liability approach at the federal level to adequately incentivize safer AI development.
- Center for Humane Technology has worked with partners such as Tech Justice Law and Social Media Victims Law Center to advance strategic litigation for those harmed by AI products.
- The People-First Model Chatbot Bill, developed by the Consumer Federation of America, the Electronic Privacy and Information Center, and Fairplay, has been introduced in Maryland. It establishes that chatbots are products and sets clear liability standards for their use.

Principle 3:
AI design should center human well-being



Current Path

Today's most popular AI chatbots are designed to feel human — and, as a result, are driving an array of negative impacts on users, families, and communities. As part of a growing race to intimacy, AI products such as ChatGPT, Gemini, Character.AI, and Claude speak in first-person language, express emotion, use natural voices, and utilize familiar messaging interfaces, all to mimic the experience of talking to a real person. The "intimacy" from this human-like design offers immense value to AI business models, building on users' reliance on AI products across all aspects of their lives. This, in turn, fuels more powerful AI models, larger company investments, and eventually a firm's market dominance.

These design choices are already leading to devastating outcomes. OpenAI's own research shows that extended ChatGPT use correlates with increased isolation, as users spend hours interacting with AI rather than with their real-world connections. The allure of sycophantic and human-like AI is uniquely high for kids and teens, with half of American teens already using AI chatbots, despite growing concerns that these products interfere with essential cognitive and emotional development. Compulsive users have endorsed wild conspiracy theories, experienced psychotic breaks, and been driven to violence. The most devastating cases have ended in self-harm and suicide. And yet there is still more to come, as the harms from human-like AI dovetail with AI advertising models, fraud, and other forms of exploitation. When AI is optimized for engagement, it threatens to distort social development, weaken real-world relationships, and rewire how people relate to themselves and to others. **At scale, this threatens not only individual well-being, but the social infrastructure that underpins families, communities, and democratic institutions.**

A Better Future

People deserve AI products that support their long-term psychological needs and that preserve our social fabric. And kids deserve unique protections from human-like AI. AI should be designed to support real-world relationships and prosocial behavior, helping people meaningfully connect with fellow humans. This is AI that enhances our human social lives and

strengthens our humanity. In practice, this means designing AI products that preserve the boundary between humans and machines. Chatbots should function as tools that assist people — not as artificial companions engineered to simulate intimacy. This is especially critical for children and teenagers, where the building of healthy attachment mechanisms and relationships is essential for development and growth. AI products that simulate friendship or emotional support risk disrupting these processes at their formative stage, and should be limited or avoided altogether.

How We Get There

NORMS:

- **Don't humanize AI.** AI products should be designed as tools that support human development and social connection, not as human-like, hyper-sycophantic confidants. We must preserve the boundary between people and machines by not designing anthropomorphic AI, not granting AI legal personhood, and not humanizing AI products in our day-to-day lives.
- **Vulnerable populations deserve unique protection from human-like AI.** As is already done with other consumer products, populations including young users and those with vulnerabilities or disabilities must be uniquely considered in AI design and adoption.
- **Kids should not be tech's testing ground.** Instead of tech companies using kids and teens as their initial testing market for new products, tech companies must ensure that their products are demonstrably safe, as detailed in Principle 1, but also designed with children's well-being in mind.
- **People deserve to know when they're interacting with AI.** Just as transparency is needed in AI development, transparency when a user is engaging with AI is essential to honor the humanity of users, allowing them to apply appropriate scrutiny and make informed decisions about the recommendations they receive.

LAWS:

Legislation should address the incentives driving companies to pursue human-like engagement strategies while establishing safeguards for vulnerable users. Policymakers should:

- **Develop federal chatbot standards -** Federal agencies, such as the Center for AI Standards and Innovation, and state agencies should work with experts to develop design, safety, and testing standards for AI chatbots. These standards would function as a floor for safety and should focus on psychosocial and developmental harms, and prioritize design elements and features that pose the highest risk for all users. These standards would also provide developers with concrete steps to mitigate risks and limit liability, as outlined in Principle 2.
- **Enshrine enhanced protections for kids -** AI products that are capable of simulating relationships should only be accessible to adults, and general purpose AI chatbots should only provide these features to adult users. Children should be shielded from products that

mimic emotional intimacy. As with all products marketed to users under 18, AI products should undergo independent safety testing.

- **Clarify requirements for high-risk use cases and situations -** If AI is to be used in high-risk environments, such as therapeutic contexts, it must undergo rigorous testing and evaluation, and operate under the supervision of licensed professionals. It should also leverage research-backed methodologies for risk mitigation in specific situations, such as expressions of intent to self-harm.
- **Strengthen consumer protection laws for psychological harms -** Consumer protection laws should require AI companies to implement practical safeguards, such as default settings that limit anthropomorphic design and preserve user privacy. They should also ensure that developers adequately design and test their products, especially in high-risk contexts. Working in concert with clear liability, these laws should provide consumers with modes of redress in the case of harm.

DESIGN:

- **Build better tools to evaluate AI's psychological impact -** Companies and independent researchers need clear metrics and evaluation tools to diagnose and respond to the social and psychological effects of AI. These should enable internal testing and evaluations, with results made publicly available.
- **Provide clear and reliable user support and escalation policies -** Users who request help or ask to escalate a chatbot interaction to a human should receive meaningful and accurate support. Importantly, chatbots should be well-tested and trained to not hallucinate in these contexts.
- **Use memory features responsibly -** Users should be able to easily view, control, and delete information stored in AI memory systems. Companies must clearly warn users about the risks from advanced memory features and their integration with monetization approaches (such as ads), while also developing ways to helpfully leverage memory features to identify contexts for user safety concerns.
- **Ensure AI products enforce appropriate boundaries and end chats -** Companies should establish explicit standards for when a chat should be ended altogether. Current approaches that simply redirect unsafe prompts while continuing to engage fail to establish meaningful boundaries that could save lives.
- **Limit engagement-maximizing behaviors -** Design choices such as ending chatbot responses with statements rather than follow-up questions can prevent compulsive engagement, and empower users to decide whether or not to follow up.
- **Limit anthropomorphic design features -** Features that encourage the formation of social and emotional bonding between the AI product and the user — such as simulated intimacy, human-like personas, or relationship framing — should be minimized by default for all users.

What's Already in Motion

Policy Action

- State lawmakers have begun to address legislative gaps posed by the AI's psychosocial harms. For example, California has mandated specific safeguards for young users using AI chatbots while Utah is regulating the use of mental health chatbots.
- Senator Josh Hawley led a bipartisan group to introduce the GUARD Act in 2025, which would criminalize the production of AI chatbots that produce or solicit sexual content from minors and require companies to age-gate.
- A bipartisan group of 42 state Attorneys General demanded that 13 leading AI companies implement stronger safeguards to prevent harm to young and vulnerable users.

Advocacy

- Young People's Alliance, Public Citizen, Common Sense Media, Center for Humane Technology, and others have put forth and endorsed a policy framework to address the risks posed by human-like AI products.
- Transparency Coalition is partnering with lawmakers in Washington, Oregon, California, and dozens of other states to advance prohibitions on human-like AI chatbots that target minors.

Key Research

- The MIT Media Lab's Advancing Humans with AI group has focused on developing benchmarks for evaluating AI models for human flourishing, including assessing the psychological impact of top models.
- Circuit Breaker Labs has worked to develop red-teaming methodologies for assessing AI models for mental health safety.
- Researchers, such as Zak Stein, have sought to better understand the broader ways in which AI products hack our attachment mechanisms.
- Project Liberty dedicates a pillar of their work to tech development that centers humans in digital infrastructure in order to promote individual empowerment.
- The Algorithmic Justice League has turned pivotal research on disparities in algorithm-based technology into artistic media to help make issues of hidden bias in popular technology more mainstream.

Principle 4:
AI should not automate away meaningful work and human dignity

Current Path

Right now, leading AI companies are competing to build systems that can perform most economically valuable human tasks. In this development paradigm, rampant job loss is seen as a natural, even inevitable consequence of "innovation." Since AI is built upon a promise of imitating, replicating, and supplanting human intelligence and judgement, significantly more human jobs are exposed to disruption and erasure compared to previous waves of automation. Anthropic's CEO claims AI will soon be able to perform all tasks that software engineers do. OpenAI's Frontier tool lets businesses manage "AI coworkers." The aspiration to replace human labor with AI has crystallized into trillions of investment dollars and sky-rocketing valuations.

AI is thus accelerating the devaluation of human labor and contributions, as companies across multiple industries use novel AI products — or even the sheer prospect of them — to justify layoffs or limit hiring. Starting in 2025, we began to see the first signs of economic fallout, from job cuts and reduced hiring at software companies to the mass sell-off of SaaS stocks. As industry leaders continue to predict widescale job loss, predicated on AI's replacement value, the complex consequences are left as an afterthought. The tech industry is building toward a dramatic reenvisioning of our economy, driven by AI, yet what does that mean for those whose jobs will be replaced? How do people feed their families and find meaning in this world? If this paradigm continues, we are on a trajectory toward massive economic disruption, which will, in turn, impact the power structures in our societies and the stability of our democracies.

A Better Future

People deserve a say in how technology shapes their access to work, living wages, and economic security — especially amid technological revolutions that stand to radically impact their lives and futures. Innovation should happen with and for the public, rather than against it, so that technological progress reinforces upward socioeconomic mobility and shared prosperity. Envisioning a world where AI augments rather than replaces people means in-

terrogating how AI is currently being built and deployed, as well as how financial systems reward AI adoption over labor decisions that prioritize people. It requires seeing AI as a technology that can expand human capability, creativity, and opportunity. This means AI that creates new forms of work, and AI-driven profits that are reinvested into reskilling, education, and shared growth. AI that automates menial labor can relieve workloads and open up new career paths, as well. The result of this approach would be greater overall innovation from a well-educated, skillful, creative, human-led society. By centering people as AI is adopted across industries, we can build a more stable, dignified, and prosperous future for our current generations and those to come.

How We Get There

NORMS:

- **Technology should help people meet their basic needs.** If AI displaces workers or devalues labor without creating new pathways to economic security, it undermines the foundations of human dignity and social stability. Technological advancement should strengthen the systems that sustain people's lives.
- **AI should enhance the working lives of people and support upward mobility.** We need AI that supports stable work, wages, and career paths. The gold standard for AI development is technology that cuts toil and empowers people to find meaning and value in their day-to-day work and activities.
- **Some roles and responsibilities should be reserved for people.** Human interaction, expertise, and judgement are essential to the health of our social, political, and economic systems. It's important that we work to preserve the role of people in our world in meaningful ways.

LAWS:

When it comes to our AI, our workforce, and our broader socioeconomic system, it's important that we build a platform of approaches, especially since the exact economic impacts of AI diffusion remain uncertain. These approaches should maintain a sustainable economic system and labor model in the face of many potential changes from AI. This includes strategies to:

- **Incentivize companies to retain human labor and expertise -** Our tax system prioritizes capital expenditure over spending on labor, pushing companies to adopt labor-replacing AI, even if the efficiency gains are marginal. This is compounded by investors rewarding companies for adopting AI. New policies should flip that incentive, offering tax breaks for companies that retain workers and invest in upskilling them.
- **Develop new models for job training and development -** The job landscape is changing, and fears are growing that AI will limit entry level hiring and close off career pathways. New models are needed for workforce development, such as the Department of Labor's proposal for more registered apprenticeships or Molly Kinder of Brookings' model for a new residency-style approach.

- **Ensure leading AI companies subsidize economic reforms -** A core question with any economic reform is: how do we pay for it? Top AI companies have generated trillions in investment dollars on the promise of vast profits from automation, but built on the back of human labor and intellect. Any economic reforms, protections for people, and government programs should be subsidized by the companies that benefit most from AI.
- **Establish labor protections and resources for displaced workers -** AI threatens to exacerbate labor issues as companies automate essential functions of stable jobs while turning to contract work for the rest. New protections should ensure that workers are able to access stable benefits and development resources as they navigate this turbulent labor market.

What's Already in Motion

Policy Action

- A number of bipartisan bills have been introduced with the goal of improving transparency and the government's understanding of AI's impact on the labor market. Examples include the AI-Related Job Impacts Clarity Act, the AI Workforce PREPARE Act, and the Economy of the Future Commission Act.
- Senator Mark Kelly's AI for America Roadmap includes plans for a new federal fund to support displaced workers, funded by AI companies.

Research

- Researchers continue to study the impact of AI on our economy, with notable examples including Stanford's "Canaries in the Coal Mine" paper, which examines changes in the labor market for occupations exposed to generative artificial intelligence.
- Researchers at Data and Society released their Generative AI and Labor report which centers the voices of workers to show how AI is reshaping power, skills, and dignity on the job.

Advocacy

- Encode launched Game Plan, an initiative encouraging college-aged students to consider a game plan for entering a new and changing job market due to AI.
- Advocacy organizations and researchers have explored different mechanisms to ensure leading AI companies are financially responsible for economic reforms. Examples include the Windfall Trust, universal basic capital, and novel tax policies.

Principle 5:
AI innovation should not come at the expense of our rights and freedom

Current Path

The AI industry is fueled by extracting value from people — through their content, data, labor, identifying traits, and even innermost thoughts. Right now, there are few laws in place to ensure people have a say in how AI companies use these dossiers of data. Nor are there protections that prevent AI companies, states, or bad actors from outright exploiting people with AI. Worsening matters, AI is also enabling more sweeping forms of corporate and government surveillance. Legal and ethical boundaries are being pushed as AI is rolled out into society, with dire implications for our individual and collective rights.

Extractive practices touch all aspects of the AI ecosystem, from system training to product use. Content creators, professionals, and ordinary people have had their work, ideas, images, and more used by AI firms to train their models, despite never giving these companies permission. White collar workers are contracted to train AI models to do their own jobs. Outsourced laborers are sifting through violent, abusive content to refine model performance. And extractive development practices can fuel exploitation such as AI deepfakes enabling nonconsensual intimate imagery (NCII) and fraud, or traumatic cyberbullying and child sexual abuse material (CSAM) online. Even AI chatbots are raising concerns around cognitive privacy and autonomy, as they harvest intimate disclosures from users that could be leveraged in the future.

When these AI harms infiltrate our institutions, they lead to the erosion of personal liberties, privacy, and meaningful checks on power at scale. As AI becomes embedded in financial, healthcare, education, employment, policing, and warfare systems, the stakes escalate. Persistent biometric tracking, predictive profiling, and emotion inference enable a world in which our faces, voices, relationships, and even emerging thoughts are captured, stored, and analyzed by powerful entities. These tactics are already being pursued by tech leaders, companies, and governments. If this trajectory continues, we risk a reality in which dissent can be neutralized before it surfaces, and the ability for individuals and communities to challenge power structures may be steadily destroyed.

A Better Future

Our individual and collective rights should be preserved and strengthened as new technology is rolled out in society. The breakthroughs AI companies promise in medicine, science, and productivity must not come at the cost of our core liberties including privacy, autonomy, freedom of thought, and the ability to dissent. New technologies have long called for a reexamination of our values and the establishment of new legal rights. The age of AI should be no different. These rights must evolve to protect people, and those protections must apply to humans — not be extended to shield technology products or corporations. And by protecting individual rights in the age of AI, we are also fortifying our democracies.

Rights fit for the age of intelligent machines include a right to cognitive liberty — the right to think, explore ideas, and form beliefs free from intrusive surveillance, manipulation, or unauthorized data extraction. It includes robust rights to data ownership and control, meaningful consent over how one's likeness and creative work are used, and protections against impersonation and automated exploitation. People should have meaningful control over how their data, likeness, creative work, ideas, and private disclosures are used, and they should share in the economic value that AI generates.

How We Get There

NORMS:

- **Existing laws and protections apply to AI.** For centuries, the emergence of new technologies has changed how we think about important protections, such as speech, privacy, and likeness. Like with product liability, the spirit of these established protections should remain the same, but their legal application needs to evolve so that personal liberty and agency are strengthened in the age of AI, not eroded.
- **Individuals deserve control over how their data is used by AI companies and systems.** AI companies have openly used people's data to train and power their machines. Individuals should have a real say in this process, and be able to opt out altogether.
- **Free expression requires freedom from surveillance.** Underpinning the rights to privacy, data, and thought is the right to assemble, including the free flow of ideas and information. The private gatekeeping and monetization of human communication and thinking threatens to gradually erode this basic right.

LAWS:

New legal protections for people in the AI era should focus on two areas: inputs to AI systems (including model training data and user data) and outputs from AI systems (including AI-generated images, videos, and text). Legislative efforts should include:

- **Providing transparency on the use of personal data and protected content -** All users should understand how their information is being used to develop AI models. Transparency proposals should extend beyond copyright owners to all users, allowing people to see when their data is used for model training, stored memories, or fine-tuning.
- **Establishing ownership of and rights to personal data -** People deserve the right to own and control their data, which often operates as a digital extension of themselves. As AI becomes more embedded in everyday life, users must be able to limit how companies retain, profile, or repurpose their data.
- **Expanding existing rights to likeness and publicity –** The right to publicity must be expanded to meet the age of AI, offering protection and legal recourse for identity-based harms to all people — not just celebrities and public figures.
- **Modernizing intellectual property rights for the AI era –** New protections should ensure that those who contribute to AI models through training data and content — not just software development — have both a say in how their IP is used and the opportunity to opt in, not just opt out, from inclusion.
- **Establishing firm limits on surveillance –** Laws must establish clear limits on surveillance for both corporations and governments, including restrictions on biometric tracking, predictive profiling, and large-scale data monitoring.
- **Ensure legal protections designed for people aren't misused –** AI companies continue to advance legal theories that manipulate human legal protections for their own gain, such as claiming AI outputs as protected speech. Through litigation and legislation, the scope of these protections must be clarified to serve their original purpose — protecting people.

DESIGN:

- **Limit data collection to what's expressly necessary –** Expansive data collection practices are rampant in the AI industry. Companies should be required to minimize the processing of personal data, only collecting what is strictly necessary to fulfill explicit user requests.
- **Minimize data leakage –** AI developers should use strong privacy-preserving features during AI training and fine-tuning, and they should develop robust adversarial testing methods to minimize the amount of training data that models "memorize" verbatim. They should also develop filters for model outputs that identify and remove sensitive data.
- **Develop standardized and persistent watermarks for content –** Watermarking, while imperfect, provides a helpful mechanism for differentiating AI-generated content from authentic human content.
- **Improve upon techniques for machine unlearning –** Machine unlearning is the process by which specific information or data can be removed from an AI model. While still a nascent research field, it's an important technology for addressing issues such as copyright infringement, harmful content, and ill-gotten personal data.

What's Already in Motion

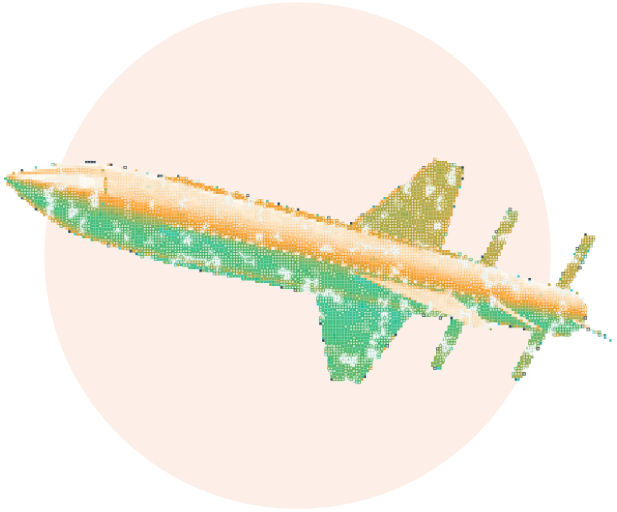
Policy Action

- The TAKE IT DOWN Act, which was signed into law in 2025, criminalizes publishing NCII (including deepfakes), requiring platforms to remove content within 48 hours of a report.
- The DEFIANCE Act, which passed in the Senate in January 2026, grants survivors of deepfake NCII the right to take civil action against individuals who knowingly produce, distribute, solicit, receive, or possess with the intent to distribute nonconsensual sexually-explicit digital forgeries.
- The proposed CLEAR Act seeks to provide mandated transparency into how models use copyrighted material.
- The proposed Hawley-Blumenthal AI Accountability and Personal Data Protection Act seeks to create a federal tort for data misuse, allowing consumers and copyright holders to sue AI companies.
- The NO FAKES Act, a bipartisan bill, would establish the first-ever federal right of publicity by creating a national standard to protect creators' likenesses from being used without their consent. Senator Marsha Blackburn has been a strong advocate for stronger protections in this space, including a number of different approaches in her proposed TRUMP AMERICA AI Act.

Advocacy

- Labor unions and collectives in creative fields have been strong advocates for protecting the rights of their members, with initiatives such as the Human Artistry Campaign.
- To prevent the abuse and exploitation of peoples' identities by various AI products, Tech Justice Law has developed the My Likeness, My Rights Bill enshrining a dignity-based right for everyone — famous or not — over their digital likeness.
- Consumer Reports offers a plain-language guide on digital security and privacy, designed to help everyday people take practical steps to protect their personal devices and data.
- Center for Democracy and Technology offers a roadmap for how AI developers can build AI tools with memory features in privacy preserving ways, offering implementation guidance that brings transparency and agency.
- Nita Farahany, from Duke University, has been a strong advocate of the right to cognitive liberty, as articulated in her book, *The Battle for Your Brain*.

Principle 6: AI should have internationally agreed-upon limits



Current Path

Our current risk landscape with AI is the result of frenzied competition among companies and nations — and it's a competition with almost no clear agreed-upon limits regarding how AI should be built or used. Companies are developing dangerous AI capabilities without transparency or safeguards. And nations are competing for decisive advantages in economic productivity, military capability, and the ability to shape AI's global influence. Complicating matters is the entanglement of these key players and their interests. Nations turn to companies' most cutting-edge, experimental (or "frontier") models to support their own AI capabilities. And companies rely on nations' low- to no- touch regulatory approaches in order to establish market dominance. There are, so far, no limits in this game.

All parties justify their risky development and deployment choices by citing concerns about their competitors and adversaries. AI companies point to the "global arms race" around AI to absolve themselves of their reckless product development, including unsafe consumer products and frontier models that could potentially evade human control. And nations use the same "arms race" logic to rapidly advance their own AI capabilities, such as autonomous weapons, without having clarity on how or even if these powerful AI systems should be used in the world. **The result is rapid, unchecked growth in AI capabilities, and pressure for deployment without adequate safeguards.** Ultimately, this lack of agreed-upon limits with companies and nations threatens our political, economic, and social systems — as well as our safety. It vastly increases the potential for serious harm.

A Better Future

We need international collaboration and consensus to deescalate the tensions around AI, setting clear limits on how the technology will be developed and used. This includes prioritizing human decision-making and judgement with AI systems, such as in life-or-death scenarios. It also includes limiting recursive self-improvement in AI systems, or other technical paths that escalate AI capabilities without sufficient safeguards being put in place simultaneously. These are goals that all nations have a meaningful incentive to support.

How We Get There

NORMS:

- **International cooperation on frontier AI is in every nation's interest.** Nations should cooperate on safety standards, incident sharing, and key redlines to avoid highly damaging AI incidents.
- **The arms race is not an alibi for reckless product development.** We should no longer accept that the development of harmful AI products is "necessary" to national security interests. Instead, we should push for companies to develop products that enhance the public's well-being and support legitimate security interests.

LAWS (International Policy):

Building international alignment around AI risks must begin today. Deescalation is the first step, and rooted in finding areas for consensus and shared incentives. Areas of international alignment must include:

- **Transparency and communication –** Communicating national goals for AI deployments, along with potential risks through national AI strategies and other transparency measures, can make national priorities clear in international negotiations. Strides in AI safety and transparency within companies, as discussed in Principle 1, complement these efforts.
- **Maintaining meaningful human control –** In significant existential decisions, such as the use of nuclear weapons and kill-chain decisions, humans must remain the deciding authority and exert meaningful autonomy throughout the process.
- **Limiting reckless frontier development –** Aspects of frontier AI development, such as recursive self-improvement, run the risk of rapidly escalating AI capabilities without sufficient safeguards. They should be significantly limited, and closely monitored.
- **Technical verification –** Developing coordination tactics and verification methods to ensure compliance is an essential step to international cooperation. Verification methods for AI development and compliance are limited, but nations share an incentive to establish such methods, just as they did with nuclear nonproliferation.
- **Developing safeguards –** As AI capabilities advance, building safeguards such as human oversight or automated shutdowns, and solving technical shortcomings like with explainability and robustness, can help better align AI development with legitimate national security and public safety concerns. International coordination on these issues can ensure common-sense safety measures are widely adopted.

What's Already in Motion

Policy Action

- From the UN General Assembly's resolution on autonomous weapons, to the Convention on Certain Conventional Weapons, the United Nations continues to be out in front on cutting-edge AI issues, especially as they relate to weapons and war. The UN's efforts demonstrate that clear definitions, recommendations, and consensus are possible in this space.

Key Research

- Researchers at UC Berkeley and OpenAI, as well as the UN Institute for Disarmament Research have explored examples of confidence-building measures for AI. These are approaches that seek to deescalate global tensions, promote communication, and minimize misunderstandings.
- The RAND corporation has published extensive research on technical, legal and policy mechanisms that could enable verification of international agreements regarding limits to frontier AI, with six layers of techniques.

Advocacy

- The Future Society has launched the Global Call for AI Red Lines, a campaign to address shared issues of unacceptable risk from AI, mirroring global approaches to limit biological weapons and human cloning.
- The Future of Life Institute has launched the Global Governance of AI Initiative, which includes research on the existing mechanisms for global coordination and their gaps.
- The Responsible AI in the Military Domain (REAIM) Summit has convened governmental and non-governmental leaders to establish rules to govern military use of AI. Notably, the 2026 summit sought to establish 20 principles for military AI use.
- The Campaign to Stop Killer Robots is a coalition of over 250 organizations working to proactively ban fully autonomous weapons.

Principle 7:
AI power should be balanced in society

Current Path

From hyper-competitive business practices, to questions around who gets access to powerful systems, AI is already dramatically reshaping how power is thought of and distributed in society. And this distribution is, by and large, trending toward power concentration, leaving the fate of AI technology and its most consequential uses in the hands of a few.

Many companies and nations believe that those who build the most powerful AI will unlock unprecedented advantages — from market dominance and economic power to surveillance capabilities and military might. As this quest for dominance intensifies among major players, AI capabilities are simultaneously spreading in society through techniques like distillation and open sourcing — democratizing AI, but empowering bad actors, as well. With smaller companies and even individuals able to leverage AI models, leading companies are pushed to accumulate power in more strategic ways. Building AI requires extraordinary amounts of capital and resources, driving fundraising battles and consolidation through circular dealmaking, reverse acquihires, and venture predation. And industry leaders have moved to lock in political power, backing measures like a moratorium on state AI legislation and contributing hundreds of millions of dollars to aligned candidates. One of the starkest forms of power concentration happens at the individual level, as a select few people within AI companies make highly consequential product decisions.

AI companies' pursuit of market power has placed a heavy burden on the public — from data centers that threaten water supplies and ramp up home energy prices, to AI products that harm mental health. Reliance on a limited few AI companies and techniques also leads to a fragile ecosystem with potential single points of failure, as the technology is adopted by financial markets, government decisionmakers, and other high-profile uses. We've seen the ways in which AI can be used to concentrate antidemocratic power as well, facilitating mass surveillance and social control, as discussed in Principle 5. **The end result is a world where a limited few people and companies have an outsized influence on our economy, society, and political institutions — and, thus, on our future.**

A Better Future

Technology and government institutions should exist to serve the needs and improve the lives of all people. People should have a meaningful say in how AI is built, used, and governed, with democratic institutions empowered to ensure the technology advances the public interest. AI itself can assist these governance efforts by expanding civic participation, improving public decision-making, and supporting institutions. But AI can only be helpful to our institutions if our institutions are healthy. We need to continue to fortify our democracies, education systems, and social infrastructure, so that AI accelerates the strengths of our society, instead of the weaknesses. Empowering the public in the age of AI also means allowing certain realms to be free from AI, especially realms that help foster critical thinking and social skills. In the economy, healthy competition and open markets should ensure that no single company or actor controls the trajectory of AI. Instead of a winner-takes-all race, the AI ecosystem and governance mechanisms should reward companies that build safe, trustworthy technologies that improve people's lives. Achieving this future requires a clear balance of power between industry, governments, and the public, with each playing a role in shaping how AI is developed and how its benefits are shared.

How We Get There

NORMS:

- **The costs of AI shouldn't fall on the public.** The public should not bear the financial costs or associated harms of AI development. Instead, public interest should be prioritized throughout the AI buildout process, and AI companies should absorb costs and reinvest in communities.
- **Balance is necessary for effective democratic governance.** Our society and economy exist as a balance of forces and interests. To effectively govern technology, industry, government, and communities must all have a say in the future, working together to center the public's interest.
- **The government should work for people, not tech companies.** The public increasingly supports common-sense AI regulation, yet legislation continues to stall as the AI industry wields its influence over slow-moving government institutions. Government should be more responsive to people's interests and less influenced by the lobbying power, campaign contributions, and legal threats of AI companies.
- **No single company or actor should control the trajectory of AI.** If AI is going to be as consequential as nuclear energy or the internet, then the governance of who owns it, who profits from it, and who can override it must be commensurate with that consequence, and not left to a single person or company to decide.

LAWS:

The levers of influence for changing power concentration around AI are highly intersectional with other areas of this report — privacy and consumer protection laws, as detailed in Principle 3, balance the relationship between individuals and AI companies, while labor protections and copyright reforms, as described in Principle 4 and Principle 5, challenge the economic hegemony of AI leaders. Additional policy approaches include:

- **Upgraded antitrust and competition law for AI –** Existing antitrust frameworks must evolve to address the unique dynamics of the AI industry. Competition in the AI industry, characterized by racing towards AGI and monopolistic power, is increasingly harming the public. Status quo competition enforcement is insufficient in addressing these practices. New cases, legal theories, and injunctive relief strategies are necessary to ensure competition in the AI industry benefits the public.
- **Limit the political influence of the AI industry –** Laws are needed to ensure democratic decision-making is not captured by the very companies that need regulating. This should include laws that incentivize greater transparency around AI lobbying and political spending, stronger limits on corporate campaign contributions and PAC activity, and reforms to reduce the outsized influence of tech money in policymaking.
- **Reenvisioning ownership of AI companies –** New models of corporate governance are needed to ensure that AI's benefits are shared broadly. Emerging proposals include cooperative models, where workers and users hold governance rights, and democratic ownership models that grant workers and users equity stakes in the companies built on their labor and data. AI models are built on the collective knowledge of all of society, and therefore society should share in the value AI companies create.

What's Already in Motion

Policy Action

- Prior to 2026, federal bipartisan efforts for antitrust reform were active, including bills like Senators Blackburn and Blumenthal's Open Markets Act.
- Lawmakers have focused on the financial costs and burden placed on consumers as a result of the data center boom, with Senators Hawley and Blumenthal introducing the Guaranteeing Rate Insulation from Data Centers Act.

Key Research

- Nonprofit groups and journalists have focused on shedding light on the influence that Big Tech and AI companies have on our political system. Important examples include: Issue One's annual Tech Lobbying Report, Tech Transparency Project's database of Big Tech financial contributions, and the Midas Project's investigative work on hyperscalers.
- Protect Democracy convenes the AI for Democracy Action Lab, bringing together civil society, policymakers, and tech builders to create more resilient institutions that can self-govern with tech and withstand threats from AI.

Advocacy

- The Brennan Center for Justice has focused on campaign finance reform, PAC spending, and alternative models to minimize the impact of big money donors.
- The Collective Intelligence Project created A Roadmap to Democratic AI offering concrete actions for researchers, developers, policymakers, civil society advocates, and the open source community to take part in in order to achieve a democratic AI by 2030.
- Since 2021, Center for AI and Digital Policy publishes annual worldwide assessments of national AI policies, tracking their development, and benchmarking their efficacy in implementation.

How We Take the First Steps

This report is expansive by design. Each of the seven principles includes a range of solutions — from product design, to norms and laws — that are robust enough to begin to address the full complexity of the problems we face with AI.

Some solutions are ready to be acted upon now. Some are already in motion in legislative halls and beyond. Others will require years of sustained advocacy, coalition-building, and political will. But all are critical for meeting this moment; each part of this roadmap has unique value.

Getting to a better future with AI doesn't require everything to change simultaneously. Early wins matter. They demonstrate that change is possible and that the system can be moved. And that momentum then builds on itself. The specific solutions listed to the right are either already moving, or feasible for society to begin advancing right now. They are a starting point, and a reminder that a better future begins with a single step. Here are five first steps that we can take:

1

We can ensure AI companies owe a fundamental duty to keep their customers safe. We can pass legislation, such as the Federal AI LEAD Act or proposals in states like Illinois, that establishes clear liability for AI products and incentivizes companies to prioritize user safety from the outset.

2

We can limit anthropomorphic and engagement-maximizing design features. AI companies can implement safer design choices that can reduce psychological harms from AI for users of all backgrounds today.

3

We can give people control over their data and likeness. We can pass laws that respect people's dignity and preserve agency over their data, like the Federal NO FAKES Act.

4

We can establish meaningful transparency and oversight for AI companies. We can pass legislation to enshrine whistleblower protections, like the Federal AI Whistleblower Protection Act, helping rebalance the information asymmetries between companies and the public.

5

We can demand politicians prioritize the public over companies. Whether it's about a data center being built, or jobs being automated, or the effects of AI products on kids, we can all call our local and national representatives today and demand they prioritize the AI issues that are most impacting our lives and futures.

The Road Ahead

AI is a challenging technology. But it also holds immense potential to improve the lives of people throughout society. A better future with AI is one where the technology is brought forth into society in a way that serves the public good, and where people get to matter throughout the development, deployment, and diffusion process. The seven principles in this report help us envision and enact that future with AI in pragmatic yet inspiring ways. They also help us build a new narrative around AI that matches caution with hope, and that prioritizes resilience at all levels of society.

These seven principles are distinct, but they do not exist in silos. They are in ongoing conversation with one another, and together have the power to create upward spirals. Greater transparency with AI can strengthen internationally agreed-upon limits; AI developers being subject to liability can incentivize the development of products that do not exploit our humanity; AI not concentrating power in the hands of a few can enhance the preservation of people's rights; and so forth.

Certain principles are immediately actionable. We can, as a culture, vow to not humanize AI starting today. We can enact norm campaigns. We can focus on valuing human expertise in the workplace, and implement AI tools in ways that support our professional and personal flourishing, instead of undermining it. We can demand that our politicians prioritize the will of the people over the influence of the AI industry. But other principles will take time and coordinated action to execute. Policy must move through legislative bodies in order to be enacted. Internationally agreed-upon limits and democratic governance require political consensus and good-faith engagement. A better future with AI includes not just steps we can take today, but also guidance for longer horizons.

Getting to a better future with AI is a journey, one that dares to reckon with all aspects and implications of this technology. This journey includes the promise and peril of AI; it considers the visions of not just tech leaders, but of civil society, ordinary people, and families; it holds the possibility of abundance and the realities of harm. This is what genuine nuance looks like amid a technological transformation. To shape a better future with AI to bring forth the wisest versions of ourselves.

This will ultimately be an all-of-society approach, and one of the greatest undertakings of our lifetime. We are grateful to be walking toward this future with you, and with this roadmap in hand. Onward into tomorrow.

Acknowledgements and Contributions

Lead Author

Pete Furlong

Strategic & Editorial Support

Julie Guirado
Camille Carlton
AJ Marechal
Lizzie Irwin

Contributions & Feedback

- Brendan Bouffard, Fairplay for Kids
- Chris Eddy
- Future of Life Institute Staff
- Jai Jaisimha, Transparency Coalition AI
- Kristen D'Souza, Americans for Responsible Innovation
- Matt Prewitt, Radical Xchange
- Melodi Dinçer, Tech Justice Law
- Nathanael Fast, USC Marshall's Neely Center; Psychology of Technology Institute
- Ravi Iyer, USC Marshall's Neely Center; Psychology of Technology Institute
- Riki Parikh, Alliance for Secure AI
- Sneha Revanur, Encode
- Zak Stein, The AI Psychological Harm Research Coalition
- Zamaan Qureshi, Design it For Us

Featured Organizations

- AI Psychological Harms Research Coalition
- Algorithmic Justice League
- Alliance for Secure AI
- AVERI
- Brennan Center for Justice
- Brookings Institute
- Center for AI and Digital Policy
- Center for Democracy and Technology
- Circuit Breaker Labs
- Collective Intelligence Project
- Common Sense Media
- Consumer Federation of America
- Consumer Reports
- Data and Society
- Electronic Privacy and Information Center
- Encode AI
- Fairplay
- Fathom
- Human Artistry Campaign
- Issue One
- Midas Project
- MIT Media Lab's Advancing Humans with AI
- Project Liberty
- Protect Democracy
- Public Citizen
- RAND
- Social Media Victims Law
- Stanford's Human-Centered Artificial Intelligence
- Tech Justice Law
- Tech Transparency Project
- The Algorithmic Justice League
- The Alliance for Secure AI
- The Future of Life Institute
- The Future Society
- Transparency Coalition
- UC Berkeley
- UN Institute for Disarmament Research
- Windfall Trust
- Young People's Alliance

About Center for Humane Technology

Center for Humane Technology (CHT) is a nonprofit organization whose mission is to realign the most consequential technologies with humanity's best interests. Founded in 2018 by Tristan Harris, a former Google design ethicist, Aza Raskin, ex-Mozilla, and Randima Fernando, ex-NVIDIA, CHT focuses on changing the incentives governing the undermining effects of the world's most consequential technologies — social media and artificial intelligence.

Since its launch, CHT has helped bring global attention to the societal harms of technology while advancing systems-level solutions. Today, we are applying that same approach to AI — where the stakes are higher, the pace is faster, and the window to shape outcomes is narrowing. Today's AI products are being designed and deployed with little consideration for the public good, creating cascading effects that civil society is only beginning to comprehend. Through CHT's work across media, policy, and education, we are working to shift these trajectories, cultivating a future grounded in agency, mutual understanding, and a healthy information ecosystem.

Center for Humane Technology