CHAPTER 9

PRINCIPLES FOR MATURE AI GOVERNANCE

DESIGNING RULES FOR REALITY, NOT MYTH

Having identified the four myths and their costly misconceptions, we can now develop the antidote. The solution is not a single law but a new governing philosophy, built on three core principles: Accountability, Proportionality, and Durability. This chapter outlines each principle, synthesizing the antidotes to our myths into a practical blueprint for mature governance.

Principle #1: Accountability as Antidote to Magic and Sin

The first principle, accountability, is the direct antidote to the myths of Magic and Sin. The Magic myth creates an accountability vacuum by framing AI as an unknowable "black box," while the Sin myth uses the algorithm as a "digital scapegoat" to absorb blame for human failings. Accountability rejects both fictions. It

embraces a democratic realism that holds individuals and institutions accountable for the tools they create and deploy.

The Mental Shift: AI Tools, Not Gods

The ritual of the scapegoat is, above all, a ritual of relief. By casting our sins onto the algorithm, we absolve ourselves of the messy business of accountability. But this relief is a dangerous fiction.[1] The alternative is to reject the technological determinism that fuels the myths and insist on a simple truth: AI is a collection of tools built by organizations that reflect the values, priorities, and power structures embedded in their design.[2]

Framing AI as a tool, not a being, provides a productive foundation for policy. It shifts the focus from fantasy to functionality.[3] It allows us to sidestep abstract philosophical puzzles and ask straightforward, practical questions: What does this tool do? What are its limitations and failure modes? Who controls it, and who benefits from its use? What safeguards exist to prevent misuse?

This "tool metaphor" is the necessary first step. It immediately clarifies where responsibility lies, namely not with the tool, but with its creators and deployers. This mental shift requires abandoning the fairy tale of enchanted algorithms. As author Ted Chiang observed, interacting with current AI systems resembles Tom Hanks's relationship with Wilson the Volleyball in *Cast Away*: the interaction provides psychological comfort through projection, but "there's no one else in there."[4]

The tools metaphor also puts the AGI discussions in perspective, as it is crucial that the philosophical puzzle about whether AI might eventually achieve genuine consciousness[5] does not paralyze present-day policy development.[6] The problem lies not in acknowledging uncertainty about the future, but in projecting sentience onto current systems that clearly lack it, thereby obscuring the human choices that determine their impact.

The Legal Framework: AI as a Product

Once we see AI as a tool, the legal path becomes clear. The most powerful principle for mature governance is to treat AI systems as products. By viewing AI not as a nascent consciousness but as a complex artifact sold in a marketplace, we can apply the robust, centuries-old body of product liability law.[7] This move has a powerful historical precedent in the automobile. In the early twentieth century, car accidents could have been framed as unavoidable consequences of a mysterious new technology. Instead, the common law, through landmark cases such as MacPherson v. Buick Motor Co.,[8] evolved to hold manufacturers accountable for the foreseeable harm caused by their defective products.[9] This principle did not stifle innovation; it channeled it, creating market incentives for seatbelts, airbags, and anti-lock brakes, which saved countless lives.

Product liability principles that apply to other industries should extend to systems that affect employment, healthcare, criminal justice, and other consequential domains.[10] Applying a similar framework to AI demystifies the technology and clarifies responsibility. It al-

lows us to ask a series of practical, legally established questions:

1. *Was there a design defect?* Was the system, even if built to specification, designed with foreseeable risks that a reasonable alternative could have avoided? This could be a crucial question for systems trained on biased data or given optimization goals that predictably lead to discriminatory or harmful outcomes.
2. *Was there a manufacturing defect?* Did a system deviate from its intended design in a way that made it unreasonably dangerous? This could, for instance, apply to a system corrupted during data processing or deployment, leading to harmful outputs.[11]
3. *Was there a failure to warn?* Did the manufacturer fail to provide adequate instructions about the product's limitations and foreseeable risks? For a powerful, general-purpose AI, this duty is immense. It requires transparency about the system's capabilities, error rates, and the contexts in which it is unsafe to use.

This framework is already being operationalized. The EU's AI Act, at its core, is a product safety regulation. It identifies "high-risk AI systems," many of which are embedded in other products such as medical devices or cars, and subjects them to the same rigorous conformity assessments and post-market monitoring that the public expects for any other potentially dangerous product.[12] This approach grounds the abstract anxieties of the AI discourse in the concrete world of consumer protection and corporate responsibility.

The Policy Toolkit

To move beyond reactive or oversimplified approaches to AI regulation, we need a policy toolkit grounded in democratic values and capable of handling complexity. That means rejecting overreach, paralysis, and false certainty in favor of adaptive governance and democratic stewardship. It means accepting there is a need to force fundamental changes in how companies design, deploy, and profit from their systems.[13]

1. Targeted Regulation, Not Blanket Bans

Democratic governance must focus on addressing specific, demonstrable harms—not banning the underlying tools. Technologies are tools, and any questions regarding their morality lie in their application and governance frameworks.[14] This means prohibiting harmful uses, not entire technologies. For instance, laws can be passed to ban the creation and distribution of non-consensual deepfakes, without affecting the underlying technology used in film, art, and education. Biometric surveillance can be restricted in public spaces to protect civil liberties, while facial recognition for phone unlocking is allowed.[15] An algorithmic hiring tool may be a functional tool for identifying certain skills, but it can be disastrously biased for others.[16] Effective governance is about making these context-aware distinctions.

Regulatory frameworks can establish conditional permissions that enable beneficial technological applications while preventing harmful ones. Public procurement offers a powerful tool for this kind of targeted intervention. By establishing high standards for gov-

ernment technology purchases, including requirements for auditing and fairness in AI systems, public agencies create strong market incentives for companies to develop more responsible products for everyone.[17]

2. Building Ethics into the Innovation Pipeline

Accountability cannot be an afterthought; it must be an architectural feature.[18] It starts with mandatory risk assessments that require companies and government agencies to evaluate the potential social impacts of an AI system before deployment.[19] Assessments should cover the effects across different communities, consider alternative approaches, and include plans for ongoing monitoring and adjustment based on empirical evidence rather than theoretical projections.[20]

It continues with transparency-by-design. When companies must document their design choices, training data sources, and intended applications, they create records that enable democratic evaluation while potentially influencing development decisions toward more responsible approaches.[21] Transparency must be meaningful rather than just technical, providing information in accessible formats that enable third party understanding and democratic evaluation.[22]

In the public space, it must also include participatory mechanisms, ensuring that communities affected by a technology have a voice in shaping how it works and what values it embodies, rather than having companies or government agencies design systems based on their own assumptions about user needs and preferences.[23]

Once a system is in the world, the guardrails must include ongoing oversight through independent auditing to monitor for discrimination, inaccuracy, and unintended consequences while enabling adjustments that improve performance over time.[24]

Finally, we need strong professional ethics for technologists, creating individual accountability that complements institutional rules and fosters a culture of social responsibility. Just as medical professionals follow ethical guidelines that prioritize patient welfare over commercial interests, technology workers could be held to standards that emphasize social responsibility and democratic values.[25]

3. Enforcing Corporate Responsibility

Accountability must have teeth. This means moving beyond voluntary commitments to concrete mechanisms, looking past the code and examining the business models that create incentives for harmful applications in the first place. When a platform's profits depend on maximizing engagement, it will inevitably develop systems that promote viral content even when it includes misinformation or hate speech. When surveillance technology companies profit from data collection, they have incentives to expand their monitoring capabilities, regardless of the privacy implications.[26]

Antitrust enforcement provides crucial tools for corporate accountability by preventing technology companies from accumulating market power that enables them to avoid responsibility for their social impact.[27] Internal checks on corporate power must be strength-

ened by protecting workers' rights. Whistleblowers and organized employees who can raise ethical concerns without fear of retaliation are one of the most effective lines of defense against irresponsible development and can help align corporate practices with public interests rather than just shareholder returns.[28]

Finally, accountability must be global. In our interconnected world, companies cannot be allowed to evade responsibility through regulatory arbitrage. When companies can move operations to countries with weaker oversight, they can escape accountability while continuing to operate in jurisdictions that demand higher standards. Coordinated international frameworks can establish minimum standards for corporate responsibility while allowing democratic variation in implementation approaches.[29]

4. Recognizing the Limits of the AI as a Tool Analogy
However, the tool metaphor does have limitations that must be acknowledged. Modern AI systems, particularly LLMs and multimodal systems, exhibit emergent behaviors that their creators didn't explicitly program and sometimes don't fully understand. These systems can generate novel outputs, combine concepts in unexpected ways, and exhibit capabilities that exceed their training objectives. While such emergent properties don't constitute consciousness or agency, they do represent a form of complexity that challenges simple tool metaphors.

Furthermore, AI systems increasingly operate in contexts where they shape human behavior and social

dynamics in ways that traditional tools do not. Recommendation algorithms shape the content billions of people consume, potentially influencing political opinions, consumer choices, and social relationships. Chatbots increasingly shape how people access information and understand the world. The systems function more like social infrastructure than individual tools, requiring governance approaches that account for their systemic effects.

Principle #2: Proportionality as Antidote to Madness

The second principle for mature governance is proportionality, which reframes policy intervention from a crisis response to tackling a public health challenge. This approach is the direct antidote to the emotional extremes of the Madness myth, which pushes policy between the fever of hype cycles and the chill of moral panics. A public-health approach, by contrast, is rooted in the sober, steady work of evidence-based risk assessment, targeted interventions, and a long-term focus on building societal resilience. Instead of the all-or-nothing logic of madness, it embraces proportionality, recognizing that not all AI systems pose the same level of risk, and that regulatory attention should be focused where the potential for harm is most significant.

The Mental Shift: Evidence Over Emotion

Moving beyond techno-solutionism means insisting on rigorous evidence before, during, and after deploying AI. It means treating technological solutions with the same skepticism applied to other high-stake poli-

cy interventions. If we are serious about abandoning mythological thinking, we need better tools when asking questions such as "Does this work?" "For whom?" "And at what cost?" Such an evidence-based approach helps us distinguish genuine breakthroughs from overhyped solutions.

The medical field provides a useful model. We do not deploy a new drug to millions based on a marketing pitch. New medical technologies must demonstrate safety and efficacy through controlled trials before receiving regulatory approval.[30] In the U.S., medical AI systems that receive approval from the Food and Drug Administration (FDA) must demonstrate effectiveness through controlled trials, leading to genuinely useful applications such as diabetic retinopathy screening and certain forms of medical imaging analysis.[31] However, many healthcare AI applications are deployed without rigorous evaluation, particularly in areas like clinical decision support and administrative automation.[32] Patients deserve better. For example, a Google AI system for retinal disease detection performed accurately during development but exhibited poor performance after deployment in Thai hospitals due to differences in image quality between the training and practice data.[33] Similar standards must apply to AI systems deployed in other critical domains.

IBM Watson Health's ultimate dissolution after billions in investment demonstrates how promises of salvation can consume enormous resources while delivering minimal clinical benefit.[34] The project failed not due to a lack of technology, but rather from unrealistic

promises that overlooked the complexities of medical decision-making. In contrast, DeepMind's AlphaFold was a genuine breakthrough that solved a 50-year-old challenge in protein folding, opening new frontiers in medicine.[35] It was technically sophisticated and practically valuable. This is a world away from most viral chatbots, whose technical sophistication is impressive but whose social value is often unclear. In fact, most AI chatbots solve problems that are already adequately addressed by existing technologies, such as search engines, customer service systems, and educational resources; however, they also create problems related to misinformation, privacy, and dependency.

Sometimes, the result can be mixed. When the COVID-19 pandemic upended livelihoods in Togo, the government launched the Novissi program. It set a specific goal to use satellite data and machine learning to improve the targeting of emergency cash assistance.[36] The system achieved a meaningful 17-point increase in accuracy, while also being honest about its limitations, including a 30% error rate in excluding truly poor individuals who did not match the algorithmic profile.[37] This transparency enabled productive discussions about how AI targeting could supplement, rather than replace, traditional poverty assessment methods.

As a rule of thumb, the goal of smart policy should not be to eliminate enthusiasm for technological innovation, but to channel it toward applications that demonstrably improve human welfare. It means fostering an environment where the next AlphaFold is celebrated as a greater achievement than the next viral chatbot.

The Legal Framework: Adaptive and Risk-Proportional Governance

If moral panic is a disease of governance, then a sense of proportion and timing is the cure. For any government faced with a new technology, the hardest question is "when to act?" Acting too soon on panic risks stifling innovation; acting too late risks allowing harm to take root.

The legislative rush to address deepfakes is a perfect lesson. Initial panic-driven laws in the U.S. focused on the hypothetical threat of video fakes in elections. When the real threat emerged years later as targeted audio manipulations, those early regulations were a poor fit.

The solution is not inaction, but "adaptive vigilance": designing laws that can evolve as our understanding of a technology grows. This requires iterative processes, for which we have decades of blueprints. We only need to be wise enough to borrow them. The introduction of automobiles, aviation,[38] pharmaceuticals,[39] and telecommunications[40] all involved periods of uncertainty and regulatory experimentation before the right balance was found. Each transition offers valuable lessons about striking a balance between innovation and public safety, mitigating unintended consequences, and adapting governance to technological advancements.

For example, the pharmaceutical industry's phased approach to drug development is a proven model, born from past tragedies and designed to prevent future ones. Phase I trials test basic safety, Phase II evaluates efficacy, and Phase III conducts large-scale comparisons, all before a drug is approved for public use.[41] In some in-

stances, AI deployment could easily follow similar stages: limited testing in controlled environments, evaluation of harms and benefits, and systematic comparison against existing methods before any wide release.

Automotive safety testing offers another relevant framework. Regulators do not just hope that cars are safe but require manufacturers to prove it through standardized tests. In the United States, the National Highway Traffic Safety Administration's Federal Motor Vehicle Safety Standards[42] require manufacturers to demonstrate compliance through standardized tests before vehicles can be sold. The standards are not static but evolve over time, based on real-world accident data and new technological capabilities, providing a model for creating adaptive safety requirements for AI.

Canada's Algorithmic Impact Assessment[43] requirement offers another model for regulatory approaches that resist hype.[44] It forces government agencies to evaluate a system's potential effects on accuracy, autonomy, procedural fairness, and human dignity before they launch, rather than simply measuring efficiency gains or cost savings.[45]

Regulatory sandboxes offer an alternative avenue for resisting salvation logic by enabling controlled experimentation under democratic oversight. The UK Financial Conduct Authority's sandbox allows fintech companies to test new products with real customers while maintaining consumer protections and, crucially, generating systematic evidence about their benefits and risks.[46] This model has been formally adopted by the EU AI Act,[47] whose Article 57 mandates the creation of

AI regulatory sandboxes in member states. These are defined as controlled environments where innovative AI systems can be developed, tested, and validated—sometimes in real-world conditions—for a limited time under regulatory supervision before being placed on the market. Educational AI systems could be piloted in selected schools with a comprehensive assessment of effects on student learning across different populations. Healthcare AI could be tested in controlled clinical settings with rigorous evaluation of diagnostic accuracy and patient outcomes. The key is to ensure the sandboxes serve democratic learning, not just corporate development, with public reporting of findings and a commitment to adjusting the rules based on evidence, not overoptimistic promises by vested interests.

What all these models share is a commitment to iteration and evidence. They feature reviews, pilot studies, and a willingness to adapt—a far cry from the reactive, all-or-nothing prohibitions born of panic.

A focus on impact also requires a critical shift in what we choose to measure. Much of the current regulatory discussion, particularly in high-level international forums, has been captivated by the idea of using computational power, or "compute," as a primary proxy for risk. The logic is simple: the more computing power a model requires, the more capable—and therefore potentially more dangerous—it must be. This is the approach taken by the Responsible AI Safety and Education Act (RAISE) in New York,[48] Senate Bill 53 in California,[49] and the EU's AI Act.

However, this is another seductive but flawed simplification. Risk is a product of context, not just capability. A small, computationally inexpensive model used to screen asylum applications or recommend medical treatments can cause immense harm, while a massive, high-compute model used to generate poetry is relatively benign. Focusing on compute as the key regulatory threshold is like judging a vehicle's danger solely on its horsepower, ignoring whether it's being driven at 10 mph in a school zone or 90 mph on a wet road by an inexperienced teenager.[50] Effective regulation understands that the context of deployment—the who, what, and where of its use—is a far more reliable indicator of risk than any technical specification. This approach is also more durable, as it is not rendered obsolete by inevitable advances in algorithmic efficiency that allow for more capable models to be run with less and less computational power.

The Policy Toolkit

Finally, the principle of proportionality is not just about *reacting* correctly; it is about building a society that is proactively resilient to the storms of hype and panic. In this context, resilience is the direct antidote to the "Madness" myth. It is a society's structural capacity to absorb the shock of technological change without breaking—that is, without succumbing to "governance whiplash" or "policy FOMO". A brittle system reacts with knee-jerk bans or reckless deregulation; a resilient system responds with a steady, evidence-based process.

This is achieved by intentionally creating "buffers," or institutional shock absorbers, at every level. We must

build resilience into our governing bodies, markets, international relations, and public sphere.

1. Institutional Resilience
Resilience requires independent, non-partisan bodies to provide clear-eyed analysis.

The United States had a model for this in its Congressional Office of Technology Assessment (OTA). From 1972 to 1995, it provided balanced analysis that resisted both industry hype and activist alarmism.[51] The OTA covered a variety of emerging technologies, from genetic engineering to computer security, helping legislators understand complex issues without relying solely on the input of interested parties.[52] While the office was ultimately defunded due to political pressures, its work demonstrated how independent technical expertise can support democratic decision-making about complex technologies.

The U.S. Congressional Budget Office (CBO) offers another model for independent analysis that provides factual foundations for policy debates.[53] CBO's economic projections and policy analysis, while not always accurate, create common ground for legislative discussions by establishing shared factual baselines. A modern equivalent, staffed with independent experts, could provide a trusted diagnosis of AI risks, free from the distortions of hype.

2. Market Resilience
Resilience implies aligning market incentives with public values. This means designing systems that reward companies for responsible behavior.

The FDA's "breakthrough device" designation provides a model for incentivizing transparency and responsible innovation.[54] Companies that demonstrate their devices address unmet medical needs and provide evidence of potential benefits receive expedited review and ongoing communication with FDA staff.

Environmental disclosure requirements offer another precedent for transparency incentives. The United States Environmental Protection Agency's Toxic Release Inventory requires companies to publicly report releases of toxic chemicals, creating market pressures for pollution reduction without mandating specific technologies.[55]

Financial services also demonstrate how transparency requirements can drive industry-wide improvements. The Basel III capital requirements require banks to publicly disclose their risk management practices, creating market incentives for better governance while enabling regulatory oversight.[56]

Similarly, for AI, transparency should be a competitive advantage, not a regulatory burden. Regulators can offer faster approvals for firms that open their models to rigorous testing and monitoring, while subjecting opaque players to heightened scrutiny. AI disclosure requirements could create competitive advantages for companies with better algorithmic governance practices while exposing those with inadequate practices.

3. Global Resilience

We must move beyond the "arms race" mentality (a key part of the Madness myth) and embrace international

cooperation. Challenges such as AI safety, ethics, and economic disruption require coordinated responses. The Montreal Protocol is an inspiring example of successful international technology cooperation.[57] When scientists identified the threat of ozone depletion from chlorofluorocarbons (CFCs), countries didn't compete to develop the most effective anti-ozone-destroying chemicals. Instead, they worked together to phase out CFCs and develop alternatives. The result was a combination of environmental protection and technological innovation that benefited all participants.

International cooperation is also essential in AI governance, especially in safety research, technical standards, and the management of economic disruption. The Global Partnership on AI represents early efforts to facilitate collaboration, though more structured approaches may be needed as AI technologies become increasingly pervasive.[58] Learning from successful international frameworks, such as aviation safety coordination and pharmaceutical regulatory harmonization, could inform more effective global approaches to AI governance and promote shared standards.

Yet even as the benefits of international cooperation become apparent, a counter-narrative has emerged that threatens to undermine collaborative approaches: the rhetoric of digital sovereignty. The EU has been particularly vocal in promoting this concept, framing technological independence as essential for political autonomy and economic security.[59] The move is a response to a legitimate fear of becoming a digital colony, dependent

on U.S. and Chinese platforms for critical infrastructure and data processing.[60]

The digital sovereignty agenda reflects legitimate concerns about technological dependence, but the rhetoric often slips into zero-sum thinking that mirrors the arms-race mentality it claims to reject.[61] When Ursula von der Leyen, President of the European Commission, pledged in 2014 that Europe must achieve "technological sovereignty" in critical areas such as AI,[62] she framed international cooperation as a luxury that Europe can ill afford rather than a necessity for addressing global challenges.

Such rhetoric creates its own competitive dynamics. If Europe seeks technological autonomy, other regions inevitably respond with their own sovereignty initiatives, fragmenting global technology ecosystems in ways that benefit no one. The result risks delivering not genuine independence but an expensive duplication of effort and a reduced collective capacity to address global challenges.

4. Public Resilience

Public resilience requires a commitment to critical thinking education, empowering citizens to evaluate information independently, understand how technological systems function, and participate in democratic processes that shape the governance of technology.[63]

The most effective long-term strategy implies a sustained, society-wide investment in robust digital and AI literacy, enabling society to navigate a complex world with wisdom and agency.[64] It means investing in insti-

tutional strength, ensuring schools, healthcare systems, and community groups are well-funded and accountable, so they can adapt to technological developments while maintaining their core purposes.[65]

It also depends on fostering diverse media ecosystems by supporting public media and local journalism. Such diversity contributes to resilience by ensuring that communities have access to multiple information sources rather than relying on a single platform or algorithm for news and communication.[66] More importantly, it provides an alternative to algorithmically optimized outrage and empowers citizens to distinguish between legitimate excitement and manipulative hype, building a collective immunity to madness.

Finally, resilience is sustained by developing the social capital and democratic participation infrastructure that allows communities to organize effectively around shared interests rather than remaining isolated and vulnerable to technological manipulation. It, in turn, enables communities to participate in finding adequate solutions to their problems, rather than waiting for a technological savior or fearing a digital demon.[67]

Table 1—The Guardrails Framework

1.	**CORE PRINCIPLE** Categorize Systems by Demonstrated Impact, Not Hypothetical Fantasy **RATIONALE & CORRECTIVE FUNCTION** **Illustration:** EU AI Act model that differentiates the obligations of, for example, high-risk medical devices vs. low-risk entertainment recommendations.[68] **Advantage:** Forces policymakers to address verifiable, real-world harms, not sci-fi scenarios. Focuses oversight where it is most needed.
2.	**CORE PRINCIPLE** Require Progressively Rigorous Testing Based on Potential Harm **RATIONALE & CORRECTIVE FUNCTION** **Illustration:** The FDA's approval process for medical devices, which requires progressively rigorous testing based on the severity of potential patient harm.[69] **Advantage:** Acts as a crucial brake on reckless hype cycles, providing a rational alternative to moral panics.
3.	**CORE PRINCIPLE** Mandate Systematic Hazard Identification & Human Accountability **RATIONALE & CORRECTIVE FUNCTION** **Illustration:** The FAA's Safety Management System in the aviation sector, which requires airlines to systematically identify hazards, assess risks, and implement controls proportional to an accident's potential severity and likelihood.[70] **Advantage:** Restores clear lines of responsibility, making it impossible to blame the algorithm for failures.

4. CORE PRINCIPLE
Incentivize Safety by Making Risk a Material Cost

RATIONALE & CORRECTIVE FUNCTION
Illustration: Financial regulation models, such as the Basel framework, require banks to hold capital proportional to their assessed risks.

Advantage: Replaces the need for blind faith in corporate benevolence with verifiable, economic incentives, forcing companies to internalize the costs of the risks their systems create.

Principle#3: Durability as Antidote to Heaven

The final principle for mature governance is durability. This approach conceives of AI regulation not as a series of reactive, technology-specific laws, but as a set of foundational principles. It is the direct antidote to the opposing temptations of the Heaven myth, which urges a dangerous deference to tech prophets, and the Sin myth, which demands reactionary prohibitions.

The Mental Shift: From Reactive Laws to Durable Principles

Instead of writing piecemeal legislation for each new technological marvel or abdicating governance entirely, this framework establishes a stable and future-proof set of high-level principles to guide all algorithmic systems.

Durability requires a shift in our long-term thinking. While the tool metaphor is useful for accountability, the infrastructure analogy is better for understanding AI's systemic influence. Like electricity or transportation networks, AI is becoming a foundational layer of society that requires robust governance frameworks,

clear technical standards, equitable access provisions, and comprehensive safeguards against abuse.[71] This framing forces us to consider public-interest principles such as universal access, reliability, and interoperability.

Additionally, the infrastructure approach alters our perspective on the concentration of the AI market. Rather than treating AI capabilities as commercial products subject only to market forces, infrastructure thinking acknowledges that access to AI tools is increasingly influencing opportunities for education, employment, and civic participation. Just as telecommunications and transportation infrastructure are subject to common carrier regulations and antitrust oversight, AI infrastructure might require similar governance mechanisms to prevent monopolization and ensure competitive markets.

Resilience and reliability become increasingly crucial as AI systems become more deeply integrated into critical social systems. Infrastructure frameworks emphasize the need for redundancy, fail-safe mechanisms, and disaster recovery planning. Such concepts translate naturally to AI governance through requirements for algorithmic auditing, backup systems, and transparency mechanisms that enable oversight and accountability.

Finally, a durable framework recognizes that technical fixes alone are insufficient. Many issues framed as "AI problems" are actually systemic, human problems amplified by new tools. A holistic approach requires embracing "complexity thinking": seeing social challenges not as bugs to be fixed but as outcomes of tan-

gled systems, histories, and power dynamics.[72] Poverty, for instance, is not a data distribution issue; it is a convergence of policy, labor, education, and structural injustice. Climate change is not just a technological problem; it requires coordinated changes in energy systems, transportation, and international cooperation. Technology can play a supporting role, but it cannot substitute for political will or public buy-in.

Complexity thinking looks beyond the algorithm to the broader social, economic, and political systems. Achieving this requires cross-sector collaboration, bringing together technologists, social scientists, educators, and community organizations to develop comprehensive strategies.[73]

- When we blame automation for unemployment, we must also look at economic policy, job training, and stronger safety nets.[74]
- When we blame platforms for loneliness, we must also support community development, public spaces, civic organizations, and programs that foster the face-to-face connections that no technology can replace.
- When we blame policy failures on inscrutable algorithms, we must invest in the cross-disciplinary research needed to build a comprehensive picture of their impacts, moving beyond the isolated perspectives of computer science or social science.

The Legal Framework: Principles-Based Regulation

The model for durability draws inspiration from the success of Principles-Based Regulation (PBR) in other complex, fast-moving domains. In fields such as finance

and environmental protection, experience reveals that attempting to write thousands of prescriptive rules is not a good idea, as they are likely to become obsolete quickly.[75] PBR, by contrast, sets broad, outcome-focused standards. For example, the UK FCA imposes a duty on firms to "act to deliver good outcomes for retail customers."[76] This approach endures in the face of innovation, forcing firms to engage with the spirit of the law and reducing the potential for creative compliance.

The Uniform Commercial Code (UCC) in the United States stands as a testament to this thinking. For over seventy years, the UCC has provided a stable and uniform legal foundation for commercial transactions, successfully adapting to the shift from an industrial economy to the era of e-commerce.[77] It achieves this durability not through detailed technological specifications but through clear, high-level principles governing contracts, sales, and secured transactions, successfully adapting from an industrial economy to the age of cryptocurrency.

An AI PBR would similarly establish a set of foundational, technology-neutral tenets:

- Human agency and oversight: A default requirement for meaningful human control over consequential decisions.
- Accountability and responsibility: A clear allocation of liability to the human actors who design, deploy, and profit from AI systems.

- Transparency and explainability: A right for individuals to understand the basis of algorithmic decisions that affect them.
- Fairness and non-discrimination: A prohibition on algorithmic systems that produce or perpetuate unlawful discrimination.
- Safety and reliability: A duty to ensure that AI systems operate safely and reliably in their intended contexts.

The Policy Toolkit: Reclaiming Democratic Control

After detailing the pervasive capture of our democratic institutions, is it not naive to propose democratic realism as the solution? This is not a contradiction; it is the core of the challenge. The fact that our institutions are under strain is not a reason to abandon them for technocratic prophets. On the contrary, it is the reason we must work to reclaim and strengthen them.

1. Tools for Democratic Participation

We already have powerful models for structured public engagement. As noted in Chapter 4, Ireland's Citizens' Assembly demonstrates that a representative group of everyday people, given time and access to independent expertise, can produce nuanced policy recommendations on deeply divisive issues such as abortion and climate change.[78] This same approach can be applied to AI governance, replacing the polarized debate of salvation versus damnation with evidence-based discussion. Imagine citizen assemblies systematically tackling the questions that the salvation narrative ignores. Should public schools use AI tutoring systems, and under what conditions? How should healthcare AI be designed to

preserve patient privacy? What role should algorithms play in the justice system, and what human oversight is essential? These are precisely the kinds of messy, value-laden debates that the salvation narrative seeks to override but that democracy is designed to navigate.

The Danish Board of Technology's consensus conferences have equally provided a model for public engagement on technological issues.[79] These conferences bring together diverse citizens to examine emerging technologies, hear from independent experts, and develop recommendations that have successfully influenced democratic policy.[80]

We even have examples of how digital tools can support this process, such as the vTaiwan platform mentioned in Chapter 3, which has enabled large-scale public consultation on complex issues such as ride-sharing and cryptocurrency regulation, using technology to enhance, not replace, democratic deliberation.[81]

2. Tools for Transparency and Agency

For democratic participation to be meaningful, it must be informed. It requires tools that empower citizens and embed human agency into the design of technology itself.

First, transparency must be designed for public understanding, not just for bureaucratic box-checking. Public registries of automated decision-making systems, such as those in Amsterdam[82] and Helsinki, are a crucial step.[83] They require government agencies to disclose how and when they use algorithmic systems in clear, accessible language, not technical jargon. The

Amsterdam registry details each system's purpose, decision-making logic, data sources, accuracy metrics, and human oversight procedures, allowing citizens to understand and evaluate the automated decisions that affect their lives and communities. The Helsinki registry goes a step further by requiring public consultation before deploying AI systems that could significantly affect citizen rights.[84] This consultation process creates opportunities for community input from the outset on whether algorithmic applications serve public purposes, while also enabling ongoing oversight of system performance and social impacts.

Second, transparency frameworks must also create a legal right to an explanation for citizens affected by automated decisions. France's algorithmic transparency law, for instance, allows citizens affected by automated government decisions about benefits, taxes, or other government services to demand an accessible explanation of how those decisions were reached, establishing a framework for accountability that goes beyond mere technical disclosure.[85] Corporate transparency requirements could follow similar principles by requiring companies to disclose the impacts of AI systems on workers, consumers, and communities in a meaningful manner. The EU's AI Act includes transparency requirements for certain AI applications, but these could be expanded to cover broader social and economic impacts that affect public welfare.[86]

Third, beyond transparency, we need community-controlled technology. Barcelona's Decidim offers an example of participatory democracy software designed

by and for civic organizations, rather than by a tech company optimizing for engagement.[87] The platform enables neighborhood assemblies, municipal councils, and civic organizations to organize democratic deliberation. Its governance is democratic, and its success in supporting participatory budgeting and neighborhood planning demonstrates how technology can serve public purposes. Cooperative ownership models, such as platform cooperatives, offer a similar path, enabling communities to collectively own and govern the systems that impact their lives. Platform cooperatives such as Stocksy United and Resonate demonstrate how worker and user ownership can align technological development with community interests, moving beyond a singular focus on profit maximization.[88] Public banking provides a related model for funding such initiatives, allowing community-controlled financial institutions to support local, democratic technology development, free from the commercial pressures that fuel the salvation myth.[89]

Fourth, we must build widespread technological literacy. This means rejecting the narrative of helplessness. Reclaiming agency begins with education that frames AI as a human creation that can be shaped through democratic engagement. Finland, for example, launched a program in 2020 to teach at least 1% of its citizens—and, by extension, civil servants—the basics of AI.[90] The campaign quickly exceeded its targets and offered a model for broader digital education.[91] The idea is simple: if lawmakers, judges, city council members, and voters understand what AI can and cannot

do, they are better equipped to ask tough questions.[92] Similarly, when workers can see the logic behind algorithmic management, they are better equipped to organize for fair treatment and oversight.[93] This literacy must extend to computer science education, which should include the analysis of algorithmic bias and ethical design, not just technical implementation.[94]

Finally, we must design systems that preserve human agency. This requires creating interfaces that allow users to understand and modify algorithmic decisions, providing options for human oversight, and maintaining human-centered workflows that utilize technology to augment, rather than replace, human judgment.[95]

3. Tools for Power-Balancing

More generally, a durable democracy must build structural resistance to myth-based capture. This means rebalancing power.

Media: Restructure the Pulpit

Technology journalism plays a crucial role in shaping public and policy understanding of AI, through editorial choices about coverage, sourcing, and framing that influence whether industry narratives are accepted uncritically or subjected to rigorous evaluation. It often amplifies industry perspectives rather than providing independent analysis. As the book *AI Snake Oil* notes, "If researchers and companies kindle the sparks of hype, the media fans the flames."[96]

This amplification isn't limited to industry hype; it extends to bizarre political theater. In October 2025, when Albania's Prime Minister announced his AI "min-

ister" was "pregnant" with 83 digital assistants,[97] he was framing a simple software rollout as a magical birth. Instead of ignoring this bizarre take, the media picked it up as a journalist at a European Commission briefing asked for a reaction to the happy news," and the Commission's spokesperson, rather than correcting the absurd anthropomorphism, played along, offering official "congratulations."[98] This is a perfect example of the media and policy institutions choosing to amplify a narrative rather than ignoring it or providing the critical analysis needed to hold power to account. Fostering a media ecosystem capable of more independent, critical analysis is therefore essential. Equally important in today's diverse media environment is investing in public media literacy, enabling citizens to critically assess the technological narratives presented to them. Fostering a resilient media ecosystem, such as by supporting public media or initiatives like the BBC's Verify service,[99] is crucial for countering misinformation and myths.

Research: Open the Black Box

AI research and development are often conducted in the laboratories of a few corporate giants, shielded from democratic oversight despite being fueled by public investment in the form of tax incentives, research grants, and infrastructure support. If democracy is to have any say in our technological future, it must have a seat at the table where that future is being designed.

Further, public investment must be used to create alternatives to corporate-controlled technology. When governments fund AI research, they can attach strings that

require public benefit, open development, and democratic accountability, creating a technology ecosystem that serves more than commercial advantage.[100]

Finally, academic institutions require structural reforms that reduce dependence on industry funding. They must be bastions of independent thought, not extensions of corporate R&D. Even though it is currently under pressure due to massive budget cuts from the Trump administration,[101] the National Science Foundation's historic support for fundamental research provides a model for public investment that generates knowledge serving broad social interests rather than narrow commercial applications.[102] The goal is not to eliminate industry collaboration but to ensure that academic research maintains the independence necessary for genuine evaluation of technological benefits and risks.

International cooperation, guided by frameworks such as the Montreal Declaration for Responsible AI,[103] can further align research priorities with shared democratic values, resisting the pull of a competitive race to the bottom in favor of a collaborative pursuit of the public good. Democratic countries could collaborate on research priorities that advance shared values of human rights, democratic governance, and social justice, instead of leaving AI governance to competition between national technology champions.

Market: Address Concentration

The extreme concentration of wealth and market power in the hands of a few tech companies warrants scrutiny. Robust antitrust enforcement is crucial, as it addresses the market concentration that enables these companies

to shape public discourse and eliminate competitive alternatives. When a few companies control the core infrastructure of AI development, their leaders naturally become like prophets, regardless of the merit of their visions. Addressing the potential negative impact of such concentration is a prerequisite for breaking the spell of the Heaven myth.

A Coherent System for Governance

The three frameworks of Accountability, Proportionality, and Durability are complementary layers of a coherent system for mature governance. The AI as Product framework provides the necessary sword, establishing clear liability for the platform's own actions and designs. The Public Health model creates the expert agencies needed to assess risk and inform both regulators and the public. Finally, the Principle-Based Regulation approach provides the overarching values that guide the system. It is a system of checks and balances for the algorithmic age, one that can simultaneously protect expression and hold powerful actors accountable for the harms their systems directly cause.

Ultimately, while the catalyst for this analysis has been the rise of AI, the principles of accountability, proportionality, and durability outlined are not unique to AI. They are the essential components of wise governance. The next transformative force may emerge from biotechnology, quantum computing, or a field yet to be imagined. If we learn the lessons of the AI era, we can build a resilient democratic framework ready for that

future: one that governs based on evidence and shared values, not on myth and fear.

Table 2—Flawed Mythologies vs. the Mature Governance Frameworks

MAGIC (Mystification)	**MADNESS** (Hype/Panic)
RESULTING POLICY FAILURE Policy Distraction & Techno-Solutionism	**RESULTING POLICY FAILURE** Emergency Governance & Policy Whiplash
MATURE GOVERNANCE PRINCIPLE Accountability	**MATURE GOVERNANCE PRINCIPLE** Proportionality
CORRESPONDING FRAMEWORK **AI as a Product:** Demystifies AI by applying product liability law, focusing on concrete harms and manufacturer responsibility.	**CORRESPONDING FRAMEWORK** **AI Governance as Public Health:** Replaces emotional extremes with evidence-based risk assessment and targeted interventions.
HEAVEN (Salvation)	**SIN** (Scapegoating)
RESULTING POLICY FAILURE Prophetic Deference & Democratic Abdication	**RESULTING POLICY FAILURE** Blame Deflection & Flawed Regulation
MATURE GOVERNANCE PRINCIPLE Durability	**MATURE GOVERNANCE PRINCIPLE** Shared Responsibility
CORRESPONDING FRAMEWORK **AI Regulation as PBR:** Establishes stable, high-level principles to prevent capture by utopian promises.	**CORRESPONDING FRAMEWORK** **Integrated System:** Holds human actors accountable (via liability) instead of blaming bots, while protecting platforms as forums.