A Pragmatic Blueprint for AI Regulation

An AI startup's proposal for fair, pro-growth, pro-AI, non-partisan, AI regulation

TRUSTIBLE™

Table of Contents

Executive Summary

AI is one of the most transformative technologies of the century, with the potential to accelerate scientific research, improve healthcare outcomes, and help small businesses compete with larger enterprises. The United States currently leads the world in AI development. Yet despite this leadership, a significant gap has emerged between AI's potential and its actual adoption. Many businesses remain on the sidelines, uncertain whether AI tools are reliable enough to deploy, unclear on their legal exposure, and unsure which vendors they can trust.

This adoption gap is the central challenge facing American AI policy today. It poses a direct risk to national competitiveness. China and other nations are investing heavily in AI deployment across their economies, and they will not wait for American businesses to build confidence. If the United States cannot translate its technological leadership into widespread adoption, that leadership will erode. There is also a domestic economic risk. Billions of dollars have flowed into AI companies on the expectation of transformative returns. If adoption stalls and revenue growth disappoints, a bubble correction could devastate the very industry the United States is counting on to maintain its edge.

Closing this gap requires trust. And trust requires a regulatory environment that establishes clear rules without stifling innovation. At Trustible, we define AI governance as the combination of processes, policies, and evaluations that manage and mitigate the risks of AI. Done well, governance does not slow adoption. It accelerates adoption by giving businesses the confidence to invest and deploy. Critically, trust cannot be mandated. Attempting to force AI on skeptical businesses, workers, or consumers will generate backlash. Sustainable adoption requires bringing stakeholders along willingly and building genuine confidence in the systems being deployed.

Right now, policymakers are not hitting the mark. The AI policy landscape is fragmented and uncertain. The European Union’s AI Act's rollout has been marked by repeated debates over timing and simplification. State laws in the United States face constant threat of federal preemption. High-profile lawsuits are working through courts with judges applying old frameworks to new problems. Meanwhile, the proposals on the table tend toward extremes: some are too heavy, imposing compliance burdens only the largest firms can absorb; others are too light, gesturing at concerns without creating real accountability.

The loudest voices in the debate have crowded out the reasonable middle. AI doomers treat the technology as an existential threat demanding precautionary restrictions. AI optimists dismiss concerns about harm as obstacles to progress. Neither camp addresses what most businesses actually need: a stable, predictable environment where they can adopt AI with confidence.

We call ourselves AI pragmatists. We believe AI will be genuinely transformative, but that transformation does not have to be catastrophic or ungoverned. We are not interested in hypothetical extinction scenarios, nor do we believe that market forces alone will solve every problem. Pragmatism means focusing on the actual barriers to adoption, the real harms that have materialized, and the practical compromises that can align incentives across the value chain.

At its core, good regulation allocates risk appropriately. It places accountability on those best positioned to manage it while protecting those who lack the information to protect themselves. No one wants to fly on an unregulated plane or receive care from an unlicensed professional. Thoughtful regulatory frameworks build trust in industries, and that trust allows markets to function and grow.

This paper offers policymakers a pragmatic framework built around five core positions: **a shared liability model** that distributes accountability across model providers, deployers, and end users; a **balanced approach to copyright** that protects creators while enabling beneficial AI development; **principles for protecting children** while building AI literacy; **content provenance systems** that help distinguish authentic from synthetic content; and **information-sharing mechanisms** that reduce uncertainty across the ecosystem. Each position reflects insights from our direct experience helping companies govern AI systems in practice, and each is designed to create conditions where responsible actors can thrive.

Our Point of View

Before digging into our specific policy positions, we at Trustible wanted to explain what has informed our perspective. This paper is written for policymakers and the broader policy community. Our goal is to share insights from the field, drawn from our daily work helping organizations implement AI governance, and to present a set of win-win policy options that can attract bipartisan support.

We understand the current AI legal, policy, and regulatory environment well. At Trustible, our team of experts reads every relevant piece of legislation, published standard, and major court decision about AI. We spend our days talking to legal teams at large regulated companies struggling to understand both the nature of AI and what compliance with laws should look like. We recognize which types of AI requirements are the hardest to implement, and which ones many businesses feel more comfortable deploying.

We understand AI. Our team has built and deployed AI systems into production before the large language model ("LLM") era, and we now integrate AI into both our daily work as well as into our product. We also contribute to research efforts on AI governance and risk management practices. We are able to read through the industry hype, build and understand our own AI agents and Model Context Protocol ("MCP") servers, and even develop new AI evaluation guidance.

We ourselves grapple with the challenges of adopting AI as a startup. We understand what kind of rules and regulations would actually impact us with heavy costs and burden, how to ensure cybersecurity and privacy compliance, and how laws and regulations translate into 'day to day' tasks. We understand the fast-changing AI ecosystem, and how effective specific AI use cases can be to help small businesses scale.

We are not "doomers." We agree there are safety concerns for AI but, for now, this should remain more an area of research rather than regulation.

We are not 'accelerationists.' We do not believe that AI alone will bring about a techno-utopia.

We are AI pragmatists. We think AI will be a transformative technology, but that transformation does not have to be widely harmful. AI can be a powerful tool that can bring personal, and economic benefits, but only if it's deployed in a trustworthy and responsible way.

Our Philosophy

Our positions are rooted in a few core assumptions and philosophical principles held by our team and informed by our experience. We believe:

AI Transformation is Inevitable

We are confident AI will be one of the most transformational technologies of our lifetime, and not one that can be paused or prevented. While its impacts have been modest so far, we think it is likely to accelerate in the next few years. There are now massive economic and national interests pushing AI forward, and will likely do so even if there is a short-term market correction in the AI space. Even if you assume AI systems like LLMs are 'stochastic parrots' (non-thinking statistical machines), it's been proven that stochastic parrots can be *useful*, and there will be consequences to the workforce. This doesn't have to be all bad news if policies anticipate and prepare for addressing the impact AI will have on the current balance, quality, and availability of work. With this in mind, we think policy makers need to think about widespread programs or policies for retraining the workforce around AI and the future of work.

Embrace the Genius of the 'AND', not the Tyranny of the 'OR'

Too many political debates devolve into entrenched positions of 'A' OR 'B'. These are great for social media, selling books, and TV talk shows, but they are not solutions for actual problems. We think we can have our cake and eat it too when it comes to AI. This means we want policies that assist both big tech companies and small tech startups. Policies that are fair to content creators and allow for innovative new uses of content. Policies that consider how to protect people from AI harms while also encouraging economic and personal benefits from AI.

We Need Market Incentives

Many regulations only ever set a 'floor' for what is acceptable. They primarily use the 'stick' approach. We also need carrots. Policies that encourage investment into trustworthy and responsible AI practices will be far more effective, innovation enhancing, and pro-growth than top down enforcement. There needs to be market pressures on AI model creators to disclose more information and conduct their own testing. Market incentives for deployers to mitigate foreseeable risks and implement protection layers as well as market incentives on users to choose how and when they use certain tools. We think there are good parallels in the cybersecurity domain, where many businesses 'invest' in their cybersecurity capabilities and receive market benefits, such as increased customer trust and sales, for doing so.

Go Beyond the Models

There has been a big regulatory focus on the 'frontier' models, and what protections should be put in place at that layer of the stack. However, every attempt to define what a 'frontier' model is has run into struggles, and the current regulatory approach focused on compute use or benchmarks has numerous flaws. In addition, many of the proposals for things like 'kill switches', or certain types of safety testing, have actually exceeded what is technologically possible at the movement, or didn't consider what an enforcement scheme would look like. Ultimately however, a model is simply billions of numbers sitting on a server, and is not an entity itself. Connecting a model into a system, and then deploying it in a specific context is what can create risks and harms.

There is No Such Thing as Unbiased AI

Regulators from varying ideological backgrounds will often all say they want 'unbiased AI' but they rarely mean the same thing. Models coming out of China that must reflect certain world views are 'unbiased' in China, but 'extremely biased' from more objective perspectives. There are deep philosophical and ideological disagreements about what constitutes 'fairness' that can be mutually exclusive, as we've seen with respect to, say, affirmative action in the United States. AI systems inevitably reflect society's biases, and even attempts to shift that can itself be a highly biased activity. Certain types of biases in AI systems need to be evaluated, and at least mitigated, but we cannot propose laws that require or assume that there is some end goal of a fully unbiased system.

Liability as the Primary Policy Lever

One final principle deserves emphasis because it runs through all our specific recommendations. We believe that the most effective way to improve AI outcomes is to allocate liability appropriately across the value chain. When model providers, deployers, and end users each bear responsibility proportionate to their control and capabilities, market incentives naturally drive harm-reducing behavior. This approach is more adaptable than prescriptive technical mandates, more enforceable than voluntary commitments, and more innovation-friendly than blanket prohibitions. It is the foundation on which our specific policy positions rest.

Our Policy Positions

These principles inform where we believe policy intervention is warranted, where market forces should lead, and where current proposals overreach, or are not technically feasible. The positions that follow attempt to put this philosophy into practice across five areas where we see the greatest need for pragmatic solutions.

Shared Liability Model

One of the biggest unsolved issues around AI is the question of liability across the AI value chain. Often, your position within the AI ecosystem will dictate your risk exposure preferences. Allocating liability fairly across the various actors in the AI chain is the primary way policymakers can spur market-driven incentives for harm-reducing mitigations. **We view the liability landscape through the lens of three main actors: model providers, AI deployers, and the end user**.

AI model creators, such as OpenAI, Anthropic, or Google, should be responsible for disclosing essential information about their systems, for conducting appropriate amounts of evaluation and testing, and for clearly transmitting information about limitations, vulnerabilities, and changes they make to the system. Disclosures do not need to reveal every possible trade secret about a model, but do need to provide relevant statistics, design justifications, and metrics that help deployers determine what risks they additionally need to mitigate.

This disclosure obligation should explicitly include information about known biases and limitations in model behavior. Model providers should communicate the fairness definitions and metrics they used during development, along with any documented performance disparities across different user populations or use cases. They should also acknowledge known limitations in model capabilities, edge cases where performance degrades, and categories of prompts or tasks where outputs may be unreliable. When model providers transparently document these characteristics, that transparency should provide meaningful protection against liability for harms that deployers were adequately warned about and chose to accept or failed to mitigate.

Independent or third-party evaluations (i.e. audits) are appropriate here as well, but should go beyond simple benchmarking. Model creators should also be responsible for collecting and analyzing information about model exploits and sharing information about them, and recommended guardrails or mitigations for deployers.

AI deployers should have a 'reasonable duty of care' standard (or similar non-US standard) for the systems they build and deploy. Negligence can be alleged against deployers for harms caused by their AI systems. However, the first step in these cases is proving that the deployer had some obligation to protect the user from harm. Laws should reflect that deployers owe a reasonable obligation to protect their users so that litigation can focus on whether the deployer breached that obligation and assess any damages.

Imposing a duty on deployers to their end users also creates incentives for more thoughtful, proportional diligence. Deployers should consider the foreseeable and reasonable risks that stem from choosing certain models and deploying certain use cases to develop effective mitigations.

Deployers will become more selective on which foundational AI systems they use, and optimize for ones that provide the fewest potential risks while meeting the intended goals. This can help create a healthier market for model selection that understands the model's risks, not just about cost or convenience. This also accounts for the evolving AI ecosystem, where foreseeable harms and appropriate mitigation steps will continually change, and organizations will be incentivized to keep implementing best practices. The reasonable standard is also flexible based on the resources and capabilities of the deployer, so that early startups with novel systems, and smaller usage have lower expectations and liability compared to large incumbents.

Finally, deployers should be responsible for providing downstream users with clear guidance on how to use their tools, any limitations, and any residual risks they need to accept. This information needs to be delivered in a format consumable by the intended end user (e.g. don't give your grandma a model card).

Users of AI systems should receive and acknowledge the relevant risks of using a system and should be responsible for harms if they use them outside of the clearly defined intended purpose. Detecting intentionally criminal actions should be a clear 'foreseeable risk' for certain types of systems, and deployers should have clear means of blocking or suspending malicious users, or reporting them to law enforcement.

There are strong analogies for this model already within the technology sector, particularly in cloud computing. Many software companies that leverage cloud resources operate under a shared responsibility model enforced through contractual terms. Cloud providers ensure appropriate safeguards in place for physical data centers and in-network communications, while developers building on them need to both do their own due diligence on the cloud and ensure their applications are themselves secure and compliant. This model has been successful in incentivizing a healthy cybersecurity and privacy assurance market, with cloud providers and software builders voluntarily undergoing different types of certifications and audits to prove trust to customers.

However, there are important limits to this analogy. The cloud shared responsibility model evolved over more than a decade, primarily in business-to-business contexts where sophisticated parties could negotiate contractual terms and assess technical risks. AI is moving faster and touching consumers more directly. Foundation models are being embedded into consumer applications, healthcare tools, educational products, and financial services within years of their development. Many end users have no meaningful ability to evaluate the AI systems they interact with, and no opportunity to negotiate terms. Contract-based liability allocation alone cannot adequately protect these users or create appropriate incentives for responsible deployment. This is why we believe statutory frameworks establishing baseline duties of care are necessary, rather than relying solely on market-driven contractual arrangements to allocate responsibility.

Copyright Issues

There are several open questions about copyrighted data related to AI systems. Even some of the world's heaviest AI regulations (like the EU AI Act) don't have a clear solution for them. While there is a rich set of common law precedents built up over decades on copyright, we believe certain aspects of AI call for fundamentally new practices and understanding to be built. In particular, the value of human content has actually increased because it can help create such powerful systems, while at the same time making recreation of it dangerously cheap. There needs to be a set of compromises established here that can benefit all the parties involved.

First is the question of using copyrighted data for model training. We support clearly allowing this to take place but also allowing a clear 'opt-out' scheme. If someone explicitly mentions content as being copyrighted, or puts technology in place to prevent that, intentional circumventing of that should be penalized. We encourage groups forming to voluntarily license out their generated content for AI, which would be similar to the approach used by ASCAP/BMI (royalty collection groups for the music industry).

Next is the question of whether AI generated content can itself receive copyright protections. We think a fair 'balance' to the system would be to disallow wholly generated content from receiving the same IP protections as human-created content. Any content that is significantly modified by a human can receive those protections. We support standards bodies creating clear guides for what constitutes a substantial modification. This is more or less a reflection of where the U.S. Copyright Office has landed in its guidance. We encourage lawmakers to codify this in relevant copyright laws and develop applicable guidance to help content creators understand the distinctions.

In addition, we think there should be a clear line between simply using content for model training versus obvious and intentional direct for-profit use of someone else's IP. In other words, there needs to be a legal recognition that simply training off of data is a different task than allowing someone to re-generate that content and commercially exploit it. This should also apply to systems that dynamically pull in content at run time and pass it off as their own (such as AI search engines). Deployer platforms should put standard guardrails in place to block commercial exploitation of this kind of generation and obey any takedown request made from the appropriate IP owners.

Protecting Kids & Education

There is a strong tension between trying to ensure that kids are protected from potential harms of AI while also figuring out how to create a highly AI literate workforce. Ironically, high AI literacy is one of the best ways to mitigate risks because users can understand its limitations and the various ways it can create misinformation. However, protecting kids has emerged as one of the biggest defenses used by state policymakers for passing their own legislation. This issue seemingly transcends normal partisan lines, with both left- and right-wing politicians supporting age restrictions. Strict age gating is difficult to implement and enforce, can create massive privacy issues, and doesn't really allow for appropriate education. Instead, we encourage dedicated standards and models tailored for educational or entertainment use appropriate for kids to use, and then reserve age gating practices for companies seeking to avoid the liability.

When it comes to education, AI has already created a massive challenge for educators, students, and parents. The entire world is essentially involved in a natural experiment with AI, with some students, schools, and educators trying out systems on their own and others opting out (or not even having access to AI). The results of these experiments may not be known for years or even decades. One of the biggest issues there is that the experiment is being fundamentally uncontrolled, and there are a lot of incentives to hide AI use by many of the parties involved. We need a heavily sponsored effort to ensure that data about use in AI is being collected appropriately, shared confidentially, and analyzed objectively.

At the moment few educational leaders are well equipped to measure the impacts or best practices of AI in their space. Only AI deployers themselves have that data and will inevitably analyze it for their own benefit, which may not align to positive student outcomes. To be clear, we think AI has the potential to be extremely beneficial for education, but discovering those uses and impact required structure, governance, and oversight.

Content Provenance

In an ideal world, all synthetically generated content would always be clearly watermarked or identified as such. That simply will not happen. Between open-source models, jailbreaking techniques, and watermark-removing technology, it will not be possible to detect and enforce this at scale. We still support standards for content provenance, and specific types of platforms should be compelled to disclose this information when available. But the law cannot expect to always know if something was generated.

It may be easier to take the opposite approach. Specifically, build technology and content provenance systems for content that is verified to be authentic. There are early versions of this for images that have gotten some adoption, and tools like Microsoft Word and Google Documents are working on similar capabilities for text content. This technology also helps AI companies avoid ingesting other generated content for training, which can contribute to model collapse. Blockchain technologies can be used to cryptographically create verifiable records of authentic content by storing a unique digital fingerprint on an immutable ledger so that any alteration can be detected and audited.

Beyond the technical benefits, a robust authenticity verification ecosystem creates meaningful market opportunities. Businesses whose value proposition depends on the quality of their content or the depth of their expertise can leverage verified provenance to distinguish themselves from competitors relying on generated material. Law firms, consultancies, research institutions, and creative agencies all have strong incentives to prove that what they are selling reflects genuine human judgment and skill rather than AI output dressed up as original work. On the demand side, buyers who want to patronize human-driven businesses or ensure they are paying for true expertise can rely on these verification systems for reassurance. This dynamic gives content creators a protected and marketable distinction that does not require government subsidies or mandates to sustain. The market itself will reward authenticity where authenticity matters, and many consumers and enterprises will willingly pay a premium for verified human content when that verification is reliable and accessible.

It is too early to target these schemes for hard regulation, but once the technology develops further, policymakers can play a role in encouraging adoption, supporting interoperability between competing standards, and ensuring that verification claims are not themselves fraudulent.

Information Sharing

A well-functioning AI ecosystem depends on the timely and reliable flow of information between model providers, deployers, and end users. Today, that flow is impeded by legal uncertainty and misaligned incentives, leaving businesses exposed to risks they cannot adequately assess or mitigate.

The current legal environment discourages transparency. Model providers that proactively disclose limitations or vulnerabilities risk having those disclosures used against them in litigation. As a result, even well-intentioned companies often withhold information that would help downstream deployers make better decisions. Meanwhile, foundation models are frequently updated without structured notification to deployers, changelogs, or versioning practices, making it difficult for businesses to maintain consistent, compliant AI systems.

This information asymmetry creates real costs. Deployers cannot conduct meaningful due diligence on systems whose characteristics are opaque. Enterprises struggle to explain AI-driven decisions to regulators or customers. And incidents that could be contained through early warning instead cascade across the ecosystem.

We support the development of a structured, government-facilitated information-sharing framework for AI, modeled on successful precedents in cybersecurity. Safe harbor protections for good-faith disclosures of AI limitations, vulnerabilities, and incidents would ensure that transparency does not create undue litigation exposure. Standardized notification requirements for material changes to foundation models, including versioning protocols and deprecation timelines, would give deployers adequate time to adapt their systems and maintain compliance. Confidential reporting channels would allow model providers and deployers to share incident data, emerging risks, and effective mitigations without public disclosure, similar to ISACs (Information Sharing and Analysis Centers) in critical infrastructure sectors. Finally, centralized incident analysis, potentially through a designated federal body, would help identify patterns across reported issues and disseminate anonymized guidance to the broader ecosystem.

These mechanisms would reduce uncertainty for businesses, enable more informed procurement and deployment decisions, and accelerate the development of effective risk mitigations across the industry.

Conclusion

These policy proposals are not meant to be comprehensive or final. There are issue areas we have not yet taken stances on, and new challenges arise regularly. We acknowledge that some of our positions will need further refinement. Our goal in publishing them is to show what a genuinely workable framework could look like, one that builds trust across the AI value chain and aligns incentives for model creators, deployers, and end users alike.

The adoption gap we described at the outset is not inevitable. It is a product of uncertainty, and uncertainty is something thoughtful policy can address. Businesses want to adopt AI. They recognize its potential to improve operations, serve customers better, and compete more effectively. What holds many of them back is not skepticism about the technology itself but doubt about the ecosystem surrounding it. Unclear liability rules, opaque model behavior, inconsistent requirements across jurisdictions, and insufficient information for sound decision-making all contribute to hesitation. These are solvable problems.

Too much of the current debate has been captured by doomers and optimists trading dire warnings past each other. One side prophesies catastrophe. The other promises utopia. Both generate attention and capture headlines, but neither addresses the practical barriers that are slowing AI's benefits from reaching businesses and consumers today. The pragmatist alternative is less dramatic but more useful. It asks how we structure an ecosystem where AI development is both innovative and trustworthy, and where benefits flow broadly rather than concentrating among a handful of dominant players.

We believe the framework outlined here represents what AI pragmatism looks like in practice. It is not a compromise where everyone loses something. It is an attempt to construct the conditions under which AI can deliver on its promise, for the companies building it, the businesses deploying it, and the people whose lives it will increasingly touch. Closing the adoption gap serves more than economic growth. It is how we ensure that AI's transformation benefits everyone.

NOTE: The policy positions expressed herein represent the views of Trustible and do not necessarily reflect the opinions or positions of the company's investors, board members, shareholders, or other affiliated parties or partners.

TRUSTIBLE™

Website	www.trustible.ai
E-mail	contact@trustible.ai
HQ address	1201 Wilson Blvd, Floor 25 Arlington, VA, USA 22209