THE AI TECH STACK: A PRIMER FOR TECH AND CYBER POLICY

Governance Layer

Table of Contents

The Governance Layer 5
Why Governance Matters 6
Threats and Vulnerabilities 7
Governance Across the AI Tech Stack 9
Governance Goals 10
Governing Autonomous AI Systems 13
Strategic Recommendations 15
Conclusion 20
Appendix A: Federal Governance Mapping by Stack Layer 21
Appendix B: State Governance Mapping by Stack Layer 22
References and Endnotes 23

ABOUT THIS REPORT

This report is intended as a primer for public policy and cybersecurity practitioners to better understand the dynamic, diffuse, and emerging landscape of Artificial Intelligence (AI) in which both audiences intersect—and must act. The audience should take away an understanding of the Governance Layer of the AI technology stack and be positioned to approach security solutions across and within each layer of the AI tech stack deliberately.

PALADIN
GLOBAL INSTITUTE

ABOUT THE AUTHORS: PALADIN GLOBAL INSTITUTE

The Paladin Global Institute's mission is to protect global critical infrastructure from cyber, AI, and deep tech threats and enhance the safety of people online by investing in research and advocacy, making informed policy recommendations, and bringing together the public and private sectors to share knowledge and create and invest in innovative technologies.

ACKNOWLEDGMENTS

Thank you to our reviewers including Trey Herr, Daniel Kroese, Connie LaRossa, Ciaran Martin, Florence Mottay, Katie Strand, and Wendi Whitmore.

Executive Summary

Artificial Intelligence (AI) is rewriting the rules of power, trust, and security faster than our ability to govern it. It is an operational reality shaping decisions in healthcare, finance, transportation, national defense, and the digital public square. Billions of online users leverage this technology—knowingly and not—every day. Within two months of making ChatGPT accessible, OpenAI gained 100 million unique users, arguably making AI the fastest adoption of any technology in history. Yet industry is racing ahead to deploy it without the understanding needed to ensure safety, security, accountability, and trust.

The AI Tech Stack framework introduced by the Paladin Global Institute in June 2025[1] was influenced by the Open Systems Interconnection (OSI) model[2] and the TCP/IP protocols as a first step to discovering and implementing layered security. When they were developed in the 1970s, the goal was to cure the fractured telecommunications infrastructure and encourage interoperable technology that underpins the internet; however, it overlooked the importance of governance. Today, our internet is technically interoperable, but the lack of governance integrated as part of the model has made it far too easy for nefarious actors to hold our digital infrastructure at risk. As we witness the accessibility and pervasiveness of AI systems, we have an opportunity *now* to learn from the gaps in previous frameworks and include governance, giving AI the best chance of delivering everything we expect from it.

Governance enables four key elements of AI adoption:

- **National Security**: AI systems increasingly underpin military operations, intelligence analysis, and critical infrastructure. Inadequate governance creates exploitable vulnerabilities. A compromised AI system integrated into our critical infrastructure is both a technical failure and a national security crisis.
- **Economic Competitiveness**: The United States is locked in a technological race with strategic competitors. Winning demands AI systems that global markets trust. Having invested in frontier models, the U.S. has a strategic advantage. But that advantage is ephemeral if we remain indifferent to governance. Fragmented, inconsistent, or absent governance undermines market confidence, creates compliance chaos, and cedes the U.S.'s strategic advantage to nations that establish interoperable international standards first.
- **Innovation**: AI is rapidly becoming foundational infrastructure for modern life, yet public confidence remains fragile. Without trust in AI systems and technology, adoption of AI stalls, and the transformative benefits of AI remain unrealized.
- **Trustworthiness:** Governance builds trustworthiness first, and only once a system is demonstrably trustworthy can it earn a user's trust. People are increasingly affected by AI-driven decisions, yet these systems are often not well understood. Without trust, AI adoption stagnates, limiting economic benefits, national security applications, and international adoption. The distinction between trusted AI and trustworthy AI is fundamental to AI governance frameworks and represents a critical gap that governance mechanisms must resolve.

Meeting these challenges requires a new governance paradigm. We must treat AI not as a monolith but as a layered technology stack, with distinct risks at each level. At the heart of this paradigm is the framework of protocols, principles, policy and law that wrap around the entire stack.

This paper makes the following policy recommendations:

- **Legislate for understanding first—then regulate with evidence.**
  Congress should prioritize AI system mapping and risk visibility before setting jurisdictional boundaries or preempting state law.
- **Move AI governance from static rules to dynamic, runtime oversight.**
  U.S. policy must reflect how AI systems operate across development, deployment, and continuous updates.
- **Codify and fund CAISI as the federal AI systems authority.**
  Establish CAISI at the U.S. Department of Commerce to map risks across the AI tech stack, resolve federal–state conflicts, and review incidents modeled on the NTSB.
- **Align governance to the AI tech stack and the model lifecycle.**
  Implement layered oversight—from data and models to infrastructure and applications—with pre-deployment testing, post-deployment monitoring, and continuous safeguards for high-risk systems.
- **Lead interoperable global AI standards while fixing U.S. fragmentation.**
  Set a national data security and privacy baseline and align with allies—especially the EU—to reduce regulatory patchwork and ensure U.S. values shape global AI norms.

Finally, this paper examines the Governance Layer as a concrete system establishing stewardship for data, models, infrastructure, and applications. Across the stack, certain threats and vulnerabilities are observable and foreseeable. The intent and capability of a threat to exploit vulnerabilities in the AI tech stack create identifiable risks. We identify potential threats and critical vulnerabilities in current approaches and propose strategic recommendations to reduce such risks and position the United States as the global leader in trustworthy AI development.

The Governance Layer

The Governance Layer of the AI Tech Stack[3] refers to the system of institutions, rules, and processes that set objectives, allocate responsibility, and verify performance for AI systems. It determines which values are embedded in AI, how risks are managed, and who is accountable for the outcomes. Governance encompasses the entire AI Tech stack, operating vertically, by imposing safeguards within each layer, and horizontally, through functions like standard-setting, risk assessment, certification, and enforcement carried out by regulators, industry, and civil society.

The Governance Layer operates through two complementary mechanisms: external and internal governance. *External governance* consists of laws, regulations, and standards developed by industry, governments, and international bodies. These include federal and state laws, implementation regulations, enforcement actions, judicial decisions, international agreements, and standards that set minimum requirements for AI systems and define prohibited behaviors.

Internal governance translates rules into organizational practice. It encompasses policies, processes, and practices that manage AI risks—such as model registries, documentation standards, change-control procedures, and independent ethics or safety functions. These measures place humans in the driver's seat, leveraging AI as it operates in a network. Internal governance also includes incident response protocols and escalation pathways for AI-related harms.

Effective governance depends on alignment between these two components: external governance establishes the baseline, operationalizing AI through decisions about data collection, model training, and deployment. They create a system that is both enforceable and adaptable to emerging challenges across the tech stack.

CORE COMPONENTS

The governance layer is built with four interlocking components that translate principles into practice.

Security protocols are foundational to ensure that AI systems operate within the defined values of an organization, meet established safety and security objectives, and perform as intended. These mechanisms include technical secure design principles and safeguards embedded into AI systems to prevent misuse, ensure resilience, and maintain performance under stress. These include automated risk detection, mitigation, failsafe mechanisms, and secure-by-design architectures.

Legal frameworks serve as the left and right boundaries of AI governance and setting a minimum baseline of what AI systems must not do (i.e., break the law) and establishing enforceable boundaries for behaviors to enable limitless beneficial innovation. These frameworks reduce systemic risk and must remain flexible and dynamic to mitigate emerging threats, vulnerabilities, and harms.

Ethics ensures that AI systems reflect enterprise values, protect individual rights, reduce harm, advance AI safety, and promote fairness. As an example, the 2025 America's AI Action Plan calls for systematic testing and mitigation of discriminatory outcomes in AI models and requires rigorous assessments of advanced AI systems to ensure alignment with constitutional rights and societal norms.[4] Innovators and investors involved in the development and deployment of AI systems should invest in pre-testing models and sandboxing AI systems prior to deployment in high-consequence environments.

Policy frameworks are built by stakeholders who innovate, invest, and use AI systems to enable a *common* baseline from which to build and deploy. Policies are the rules of engagement for risk management and continuous improvement. In the U.S. and indeed, globally, the leading edge of this work comes from the National Institute of Standards and Technology's (NIST) AI Risk Management Framework.[5] This framework provides a structured approach to identifying, assessing, and mitigating risks across the AI lifecycle and complements the NIST Cybersecurity Framework 2.0, which integrates AI-specific considerations into broader cybersecurity strategies. These frameworks aim to ensure that AI systems are secure by design and resilient in operation.[6] AI innovations stemming from these frameworks will catapult the U.S.'s AI leadership and pave the way toward interoperable, international AI standards.

These components are necessary elements of a governance architecture that is adaptable and capable of unleashing AI's benefits while safeguarding against misuse.

Why Governance Matters

Effective AI governance is a catalyst for innovation. As AI systems mature and the race for AI dominance intensifies, policy must evolve from patchwork compliance toward proactive system stewardship. The Governance Layer is where strategic enablement meets operational discipline. Without it, we risk building AI that is powerful but brittle, scalable yet unaccountable. But with deliberate, layered, and enforceable governance, we can architect an AI future that is not only competitive and innovative—but also reliable, secure, and aligned with universal values that respect human rights and fundamental freedoms.

As AI evolves into foundational infrastructure for economies, militaries, and societies, its safe deployment can no longer be left to fragmented oversight or sectoral silos. The Governance Layer becomes the connective tissue binding the AI Tech Stack. But without deliberate design, it also becomes the weakest link—either too rigid to adapt or too abstract to enforce. Securing AI systems through the Governance Layer is therefore a precondition for securing AI itself.

Yet today, governance is the least mature component of the AI stack. Governments are competing for AI dominance while key decisions are being made by corporate actors. In critical domains—from automated border control to algorithmic lending—organizations are deploying systems faster than frameworks can mature or be fully understood. This governance gap erodes public trust which is a foundational element for AI dominance. Today, while the technology is new and AI systems are maturing, is our best opportunity to close this gap.

To understand what is at stake in AI governance, then, it helps to see how different kinds of risk emerge at each layer of the AI technology stack—and how governance must adapt to each.

Threats and Vulnerabilities

Unlike technical vulnerabilities, which target specific system components, governance failures can create systemic risks lowering the cost for threat actors that seek to exploit technical vulnerabilities and increasing the costs for defenders. A lack of governance undermines the ability to prevent, detect, respond to, and recover from AI-related incidents. Governance failures can lead to uncontrollable, unmonitored, and potentially disastrous actions performed by AI tools. When oversight is weak or absent, isolated technical issues can escalate into cascading failures across critical systems.

National security implications dominate when AI governance failures enable adversarial exploitation of critical infrastructure, military systems, or intelligence capabilities. By proactively identifying and mitigating foreseeable issues where AI is deployed in high consequence environments, governance helps reduce the risk of catastrophic failures to critical infrastructure, ethical breaches, and erosion of public trust that can arise from unchecked or improperly managed AI systems.

Economic, trade, and cyber espionage concerns deepen these national security challenges. Just in 2026, Anthropic disclosed campaigns involving over 16 million exchanges across more than 24,000 fraudulent accounts;[7] OpenAI reported that DeepSeek employees circumvented platform controls to extract reasoning outputs, with the resulting model displaying structural patterns consistent with OpenAI's own systems;[8] and Google documented a single campaign targeting its Gemini system that involved more than 100,000 prompts aimed at extracting reasoning capabilities.[9] The U.S. House of Representatives' Select Committee on the Strategic Competition between the United States and the Chinese Communist Party characterized this conduct as a form of industrial espionage — one that existing export controls were not designed for.[10]

The governance gap here is precise: no legal framework currently treats adversarial distillation as a controlled technology transfer, no enforcement mechanism is calibrated to the speed or scale at which these campaigns operate, and no institution is tasked with detecting and attributing model-layer exploitation in real-time. Without governance structures that extend from the infrastructure layer to the model layer — and across layers, covering not just what hardware is shipped, but what capabilities can be extracted — America's frontier AI advantage functions as a subsidized training program for strategic competitors.

Finally, autonomous systems – like agentic AI – operating without adequate oversight can trigger unintended consequences that ripple across social, economic, and technological domains. Autonomous systems are ubiquitous throughout critical infrastructure and left unchecked, can have real life impact. For example, a misconfigured agent could have unintended but significant impacts such as exposing millions of records, including API tokens, private messages, and third-party credentials.[11] These risks underscore why investment in governance frameworks, institutions, and standards is essential to align AI development with national security and societal values.

Weak governance erodes accountability and transparency, making it harder to enforce security policies, audit system behavior, or ensure compliance with legal and ethical standards. Without thoughtful governance, we rely upon a patchwork of regulatory enforcement mechanisms to address this problem inconsistently on a case-by-case basis. Regulators are increasingly reactionary rather than building a reliable mechanism for mitigating identifiable risks.[12] It then also becomes far more difficult to enforce consistent security policies, audit system behavior, or ensure compliance with legal and ethical standards. This increases the likelihood of harmful outcomes going undetected and accelerates loss of public trust.

Trust is not a soft issue—it is a finite and strategic asset. A collapse in confidence in U.S.-developed AI technology stack will undermine adoption, stifle innovation, and weaken America's competitive position globally.[13] Recent dialogues with members of the European Union underscore the erosion of trust in the U.S. approach to AI deployments among allies.[14] A commitment to good governance, risk management, and compliance builds trust with investors, customers, and nations throughout the global AI ecosystem. Trust is key to winning the AI race.

Governance Across the AI Tech Stack

Each layer of the AI Tech Stack introduces distinct risks that require tailored governance approaches. At the foundation, the data layer's governance focuses on confidentiality, integrity, and availability. Governance structures that amplify confidentiality are rooted in principles of privacy. Provenance is a reliable tactic to ensure integrity. And structures that secure the supply of reliable data ensure availability.[15] It sets rules for what data can be collected, how it is stored and used, and ensures individuals retain rights over their information. Mechanisms include processes for identifying and normalizing data, data protection laws, privacy regulations, and data loss prevention tactics.

Moving up the stack, the model layer is governed by frameworks that oversee how models are designed, trained, and evaluated. At its essence, models are algorithmic mathematical functions, which enjoy the privileges of free speech under the U.S. Constitution. For this reason, this layer requires careful consideration for narrow and tailored governance mechanisms. Here, governance must answer critical questions: Which model capabilities warrant heightened oversight? What kinds of testing must be completed before models see broad deployment? What documentation is necessary to communicate model limitations, training data origins, and identified risks? Model governance should include testing for bias and robustness, documenting limitations and training data origins, and implementing Testing, Evaluation, Validation, and Verification (TEVV) processes. Policies that advance the use of tools like red-teaming, artifacts like model cards, and controllability standards are all central to this layer's secure deployment.

At the infrastructure layer, governance ensures the resilience and security of the physical and digital backbone—hardware, cloud services, and energy systems. This includes baseline cybersecurity standards, trusted supply chains, and secure access controls.[16] Because infrastructure often crosses physical borders and supports national critical functions, it carries significant national security implications. Cloud services, hardware, and energy are resources that are deployed for purposes beyond AI and cross jurisdictional boundaries. Streamlining AI governance structures that are optimized for resilience and security across jurisdictions to avoid duplicate, fractured, and overly prescriptive requirements.

At the application layer, governance addresses real-world scenarios such as lending, hiring, healthcare, law enforcement, and content moderation by integrating consumer protection and civil rights safeguards directly into AI systems. Effective governance at this level emphasizes transparency, ensuring that users are provided with clear information, notice, and opportunities to give informed consent. To uphold ethical standards and protect public trust, governance must establish accessible channels for recourse, allow individuals to challenge or appeal decisions made by AI, and strictly prohibit inappropriate or unlawful uses of AI.

The Governance Layer acts as connective tissue across these domains—aggregating risk insights, harmonizing standards, and ensuring lessons learned in one area strengthen safeguards throughout the system. Viewing governance through this structured lens, it becomes easier to identify where gaps and overlaps exist, and where regulatory reforms should be streamlined and prioritized. Governance frameworks also foster a shared vocabulary that enables policymakers, regulators, and practitioners to discuss the respective roles of federal and state actors in AI, paving the way for more coherent oversight across the technology landscape.

Governance Goals

AI governance will shape technological development and societal impact. Effective governance could simultaneously ignite innovation, advance national security interests, and build trust in AI systems. That's a tall order. Achieving these multiple objectives requires flexible frameworks that balance competing demands while remaining adaptive to technological change. This multifaceted challenge requires an approach that balances innovation with security and openness with control.

IGNITING INNOVATION

Innovation in AI rarely stalls because teams lack ideas; it stalls because they can't move trust and accountability across the stack. This means data moving safely into training pipelines, models moving reliably into infrastructure, and capabilities moving responsibly into applications. Governance ignites innovation when it reduces friction at those seams—by making it easier to access high-quality data lawfully, easier to prove provenance and permissions, and easier to deploy systems with confidence that security and compliance won't collapse on contact with reality.

Governance that reduces friction at the seams turns ambiguous risk into a predictable operating AI ecosystem where builders can share, reuse, and scale responsibly. Done well, it makes secure iteration cheaper than reckless speed: clearer rules for what data can be used and how, clearer expectations for lifecycle risk management, and clearer accountability when things go wrong.[17]

ADVANCING NATIONAL SECURITY

AI governance plays a vital role in hardening technological systems against adversarial threats. Robust adversarial testing and AI sandboxes, especially for critical infrastructure and high-consequence AI deployments, reduce the risk of malicious manipulation, data poisoning, or catastrophic system failure in these high consequence deployments. Governance mechanisms also enhance supply chain security by promoting traceability of AI model inputs, hardware origins, and training datasets, thereby mitigating the risks of foreign interference or espionage.

In imposing strict export controls on sensitive AI technologies, such as advanced semiconductors, frontier models, and high-performance compute resources, governance can limit the diffusion of potentially dangerous capabilities to adversarial states. These efforts, coupled with the integration of cybersecurity standards directly into AI development pipelines, ensure that AI becomes a defensive asset rather than a vulnerability.

BUILDING TRUST IN AI SYSTEMS AND TRUSTWORTHY AI

Equally critical is the need for AI governance to establish trust in trustworthy systems. The operative question for policymakers is not "do people trust this system?" but "has this system demonstrated the properties that warrant trust?" Those properties are not self-reported. They are shown through pre-deployment testing with disclosed results, adversarial red-teaming, documented performance across populations and contexts, independent audit, and clear accountability when failures occur—not mere adoption. A system that has undergone that scrutiny and published its findings has made a claim on trustworthiness. A system that has not, regardless of how confidently its developers speak, has not.

Trust in AI refers to AI systems that people trust and rely upon in practice. This is a subjective, empirical concept that captures current trust relationships between humans and AI systems. Trust, here, is a behavioral and psychological state—people use the system, depend on its outputs, and integrate it into decision-making processes of increasing significance.

Trustworthy AI refers to AI systems that possess the objective qualities, safeguards, and characteristics that warrant trust. This is a normative concept focused on the properties that make a system deserving of trust, which loosely follows the principles of privacy —reliability, transparency, accountability, fairness, privacy protection, and robust safety measures.[18] Governance frameworks that promote transparency, explainability, and open documentation help demystify AI operations and foster accountability.

Trustworthiness is about meeting established practices that justify confidence in the system's behavior and outcomes. The distinction matters because the remedies are different. When public confidence in an AI system erodes, the instinct is often to explain more, communicate better, or issue reassurances. But if the underlying system lacks demonstrable safeguards — if no one has tested it for bias, if incident reporting is absent, if it hallucinates, or if accountability is unclear — then better messaging is merely superfluous.

AI governance is the mechanism that makes this distinction enforceable. Without it, trustworthiness is a marketing assertion. With it, trustworthiness becomes a verifiable condition. It becomes a condition that regulators, deployers, and the public can interrogate, and one that earns trust precisely because it does not ask to be taken on faith. Interrogating AI systems for fairness and bias ensures AI systems do not perpetuate historical discrimination or generate new forms of algorithmic inequity. With appropriate measures in place, people can verify (not just assume) that AI is being used transparently, with documented evidence and independent review. Governance converts principles into practice and creates accountability loops—assigning responsibility to owners, requiring incident reporting and red-teaming for high-impact systems, and enabling independent certification or regulatory review. Done well, governance accelerates AI adoption by ensuring trustworthy systems and reducing uncertainty for users, businesses, and governments alike.

Governing Autonomous AI Systems

The governance goals outlined above—igniting innovation, advancing national security, and building trust—converge on a central technical challenge: AI systems increasingly operate with autonomy, making consequential decisions faster than humans can intervene and in domains where failures cascade rapidly. This autonomy spectrum, from narrow systems already deployed in critical infrastructure to potential future systems with transformative capabilities, represents the frontier where governance frameworks will either prove adequate or fail catastrophically.

Understanding this challenge requires distinguishing among fundamentally different types of near autonomous systems; each demanding tailored oversight mechanisms matched to observable risks. These challenges span three main categories:

(1) **narrow autonomous systems** currently operating in critical infrastructure,

(2) **agentic AI systems** with extended autonomy in bounded domains, and

(3) **advanced AI systems** with transformative potential.

AI governance must recognize, first, that these are systems, and governing systems can be complicated. And second, that governance must confront dynamic autonomy challenges rather than treating all AI systems under one static policy framework. For effective governance of **narrow autonomous systems**, enacted laws must be clear and not subject to wide interpretation.[19] Any law that authorizes regulation of AI systems must include some level of context; and therefore, must be commensurately prescriptive.[20] Effective governance therefore should enforce pre-deployment sector-specific certification, adversarial and edge-case testing, continuous monitoring, real-time incident reporting, standardized failsafe architectures, and explicit legal responsibility among developers, deployers, and operators. Designating empowered sectoral regulators (e.g., FDA for medical AI, NHTSA for autonomous vehicles, SEC and CFTC for finance, etc.) is crucial for enforcement.

A second set of challenges arises from **agentic AI systems** which are highly autonomous systems that respond autonomously to inputs and can perform most complex tasks nearly as well as humans.[21] Governance for these systems must prioritize transparency, accountability, and include clear audit trails, comprehensible explanations for decisions, defined rights and mechanisms for human interrogation, behavioral boundaries with automated detection of violations, and robust documentation of model limitations and performance. These are baseline standards necessary for maintaining accountability as AI's decision-making autonomy grows.

The third, most complex category involves **advanced AI systems** with potentially transformative and unpredictable capabilities. This is about systems whose behavior becomes hard to constrain or verify, particularly when they access critical information and information networks. The focus of governance here should be on preparatory measures with minimal costs, such as establishing capability thresholds that trigger heightened oversight, practicing red-teaming, restricted deployment until safety is proven, and structured access frameworks for advanced systems. International cooperation and information sharing become essential to prevent dangerous capability races and coordinate responses to unexpected AI behavior. As does investments in research, innovators, and technologies that keep humans in control of the machine.

Importantly, governance must avoid overreliance on technical controls like “kill switches”, which provide false confidence. Instead, the emphasis should be on development pathway oversight, staged deployment, and ensure that humans maintain meaningful control over decisions regarding the use and deployment of AI in consequential domains.

This three-tiered approach matches governance mechanisms to observable risks, ensuring humans maintain control over AI deployments.

Strategic Recommendations

The United States stands at a critical juncture where governance must evolve from a static point-in-time exercise into a dynamic, operational reality through runtime. The U.S. Courts have shaped the legal landscape within which U.S. law can influence AI governance. United States law governing AI must be narrowly tailored, clear, and contextual. First, algorithms enjoy the privilege of free speech granted by the First Amendment of the U.S. Constitution and laws that govern free speech must not restrict protected speech.[22] Second, the development and deployment of AI has significant political and economic consequence and therefore, Congress must clearly grant authority to federal agencies to regulate AI.[23] And, third, AI is a system of systems and to create laws to govern these systems, Congress must provide clear context.[24] As a matter of policy, federal agencies have advanced national frameworks—such as the White House's America's AI Action Plan, the NIST AI Risk Management Framework, and the TAKE IT DOWN Act.[25]

While Congress has yet to enact any law providing the contours of external governance, all fifty states in 2025 introduced or passed bills ranging from algorithmic accountability, data privacy, to sector-specific AI use.[26] AI developers, deployers, and users now face a complex and sometimes conflicting patchwork of requirements[27] Before lawmakers draw up new jurisdictional lines, it should demand a clear map of where AI risks sit. That does not mean Congress should wait on the sidelines as AI systems deploy and States ache for action. It means that Congress should legislate for understanding first and then do no harm.

Congress need not choose between understanding AI systems and acting on them, or between American leadership and global interoperability. It can act in a two-phase sequence (First for the U.S. Congress, then Policymakers generally) that makes the system more legible over time and positions the United States to drive—not just absorb—emerging AI standards. Congressional action sets the field, aligns basic security and safety protocols at home and abroad, and installs core transparency; From there, policymakers, including the Executive branch and state governments, must use that initial mapping to align federal and state roles without freezing innovators caught between a patchwork of AI policies and regulation.

RECOMMENDATIONS FOR CONGRESS: LEGISLATE FOR UNDERSTANDING

First, Congress should focus on building AI tech stack-literate capacity and guardrails. The priority is to create institutions, data sets and pipelines, and feedback loops that make smarter, agile lawmaking possible. That begins with authorizing and resourcing the Center for Artificial Intelligence Standards and Innovation (CAISI) to map risks and responsibilities across the data, model, infrastructure, and application layers. Similarly, Congress can authorize the U.S. Department of Justice to report annually on where federal–state law conflicts are emerging.

Next, Congress should require agencies to attach a short 'AI Tech Stack Impact' annex in major rulemakings, explaining which layers they are touching and how their rules interact with state law. Congress can direct GAO and CRS to produce committee-specific AI stack baselines (akin to a Congressional Budget Office score review), so authorizers and appropriators see the same picture. A Federal–State AI Governance Council, convened under CAISI, the Secretary of Commerce, the Office of the National Cyber Director, or the Office of Science and Technology Policy, can surface real conflicts and gaps instead of speculative ones.

This initial work is also the right moment to build interoperability with allied regimes, especially the European Union (EU) on shared terms. The EU AI Act and digital rules already reach U.S. firms that operate or sell into Europe but also take a broad-brush approach by limiting AI systems, rather than identifying specific components of the tech stack that pose different levels of risk.[28]

The Administration or Congress could direct the State Department, Department of Commerce, and the U.S. Trade Representative to work with EU counterparts to align basic taxonomies (e.g., what counts as "high-risk" AI deployments, define AI systems, etc.), testing and documentation practices, and harmonizing incident-reporting formats such that American priorities, innovation, national security, freedom of speech—values we share, like trust, safety, and security—are reflected in the global standards that are emerging around AI.. The goal is not to copy the EU model, but to ensure that U.S. companies are not forced to navigate completely incompatible rulebooks.

Only after that groundwork is laid does Phase Two begin. Refining jurisdiction with evidence. With CAISI's mapping, GAO's assessments, agency reports, and experience from interoperability efforts in hand, Congress can decide where to draw a federal floor—for example, for model and infrastructure safeguards and certain cross-cutting rights—and where to preserve space for state experimentation at the application layer. Targeted preemption, if needed, can be limited to specific conflicts documented by this infrastructure mapping, rather than imposed wholesale across all AI uses.

Build a Unified AI Governance Framework Aligned to the AI Tech Stack.

Policymakers should formalize a unified governance architecture that maps oversight responsibilities to each layer of the AI Tech Stack. This should include: (1) data integrity rules, (2) model transparency guidance, (3) infrastructure certification regimes, (4) application-level auditing, and (5) cross-cutting cybersecurity and accountability structures for shared risks. Such a framework should draw on the NIST AI RMF and related standards (e.g., OWASP Top Ten) but must be translated into enforceable policy. And policymakers should develop this framework with input from industry, states, civil society, innovators, investors, and affected communities, and require regular review as the technology changes.

Any framework must be adaptable to evolving technologies, providing clear guidance for emerging AI applications while maintaining robust safeguards against misuse. Regular reviews and updates should be institutionalized to reflect technological advancements and lessons learned from real-world incidents, ensuring the framework's continued relevance and efficacy.

Codify and Catalyze CAISI.

Establish, codify, and fund the CAISI at the U.S. Department of Commerce, through NIST or the Bureau of Industrial Security (BIS) authorization or a dedicated CAISI Act. The goal is to create an interagency, cross-sectoral entity empowered to set standards, coordinate incident response, conduct systemic risk reviews, and promote interoperable AI standards. CAISI should run rapid reviews after significant AI incidents, publish public findings, and recommend fixes—just as the National Transportation Safety Board (NTSB) does after major transportation accidents.[29]

CAISI's mandate should include the development of rapid response protocols for AI-related incidents, such as unexpected model failures or breaches of ethical guidelines. By fostering collaboration between government agencies, private sector organizations, and international partners, CAISI can help build a resilient ecosystem that identifies and mitigates systemic risks early. Additionally, CAISI should create public dashboards and transparency reports to keep stakeholders informed about AI safety trends and regulatory actions, reinforcing trust and accountability in the governance process.

Implementing this governance architecture will require recruiting and retaining technical personnel with expertise spanning AI systems, cybersecurity, and policy—an American workforce challenge that demands competitive compensation, fellowship programs, and rotation opportunities between government and industry with appropriate conflict-of-interest safeguards.

Harmonize Global AI Protocols Through Strategic Alliances.

The United States must lead in shaping interoperable global AI norms—on data provenance, model robustness, digital sovereignty, and trusted compute—through multilateral coalitions—just as it did with 5G in the last decade. Without this, domestic standards will be undercut by companies shopping for the weakest rules, foreign influence operations, and the pursuits of AI sovereignty.

To achieve meaningful regulatory harmonization, the U.S. should lead in international forums on AI standards. The State Department and Commerce (including CAISI), working with the U.S. Trade Representative (USTR) should lead an effort to create joint task forces and cross-border working groups that can ensure the interoperability of technical standards and protocols, enhancing collective security. By actively supporting

capacity-building efforts in developing countries and sharing best practices, the U.S. can help promote ethical AI adoption worldwide and safeguard against the proliferation of unsafe or unaccountable systems. While authoritarian regimes may achieve apparent efficiency through centralized control, democratic AI governance—built on transparency, accountability, and rule of law—creates trustworthy systems the world can trust and adopt.

RECOMMENDATIONS FOR POLICYMAKERS: DO NO HARM.

Institutionalize Model Lifecycle Governance.

Lifecycle governance should also include transparent documentation and continuous improvement protocols, requiring developers to update models in response to new threats and societal concerns. Requiring pre-deployment certification for high-risk models, including provenance tracing, adversarial robustness tests, and post-deployment behavioral monitoring. Sector-specific certification authorities (e.g., for health, finance, transportation) should be federally coordinated. Before deployment, high-risk models should face standardized tests for data quality, provenance, robustness, and misuse. After deployment, operators should be required to report serious incidents, monitor performance, and update safeguards as new threats emerge.

Default to Dynamic Protocols.

Failures in AI governance can have serious consequences, undermining innovation, security and trust while creating legal and financial liabilities for organizations. Preventing these failures requires designing AI frameworks that are technically feasible, strategically effective, and dynamic.

AI moves, matures, and develops fast. It is therefore necessary to build in, by default, a feedback loop between AI users, regulators, and developers that will enable the identification and resolution of unforeseen issues, preserve system reliability, and promote accountability. Equal measures should be given to sector-specific guidance across critical infrastructure to ensure that any new governance is practical, contextual, and appropriate. Sector-specific authorities should also maintain open channels for reporting adverse events, enabling timely interventions, and fostering a culture of safety and ethical responsibility throughout the AI lifecycle.

As AI systems grow in complexity and scale, the risk of regulatory oversight and ethical lapses increases. A risk-informed approach, in which AI systems are classified and governed according to their potential impact, and observed lapse, allows oversight to be both proportional and effective. Assigning clear liability to developers, deployers, and distributors of AI could close gaps that are otherwise easily exploitable. Such liability schemes failed to take root with the Internet but should be pursued in earnest in the age of AI.

AI has measurably compressed the attack lifecycle. According to Palo Alto Networks' 2026 Unit 42 Global Incident Response Report, the fastest quarter of intrusions in 2025 reached data exfiltration in just 72 minutes — down from nearly five hours the prior year — and attackers began scanning for newly disclosed vulnerabilities within 15 minutes of a CVE announcement.[30] Governance frameworks calibrated to annual reviews, or multi-year legislative cycles cannot respond at that tempo. Effective failure prevention therefore requires access to real-time threat telemetry: mandatory incident reporting, shared threat feeds, and machine-readable vulnerability data that allow oversight regimes to track actual, observed attack patterns rather than modeled projections without context. When governance is informed by what is happening on the network, rather than what analysts anticipate might happen, oversight can be proportional and timely — adjusting requirements as the threat landscape shifts rather than waiting for a crisis to force the update.

Organizations must also be incentivized—or required—to build internal governance mechanisms, such as AI policies, inventories, and impact assessment procedures to detect and mitigate risks before they cause harm. In other words, external governance (laws, regulations, and enforcement) only succeeds if it reliably triggers internal governance—concrete processes inside firms and agencies that make compliance, security, and alignment as the default, not the exception. When combined with the streamlining of U.S. governance standards and existing international frameworks such as the Organization for Economic Co-operation and Development (OECD) AI Principles, G7 Hiroshima AI Process Reporting Framework, General Data Protection Regulation (GDPR), or the European Union (EU) AI Act, this approach can serve to establish a coherent and interoperable global AI ecosystem.

Establish a National Data Security and Privacy Baseline.

Patchwork data security and privacy laws across state and international borders must be reduced. Congress should establish a comprehensive U.S. data security and privacy standard to streamline oversight burdens for innovators and promote international adoption of a U.S. standard. Without a federal baseline, U.S. AI companies will spend more time navigating conflicting rules than securing the data that feeds their models (or the systems in which their AI is deployed).

Such a comprehensive standard should prioritize individual rights and data minimization, mandating clear consent protocols and robust encryption requirements for sensitive information. A national standard must also enforce robust and adaptable data security requirements—spanning prevention to response—commensurate with today's dynamic cyber threat environment. The framework must be flexible enough to accommodate sector-specific needs, while ensuring baseline protections against unauthorized access and misuse. By streamlining compliance processes and providing clear guidelines, the standard can foster innovation without compromising security and position the U.S. as a global leader in responsible data stewardship.

Conclusion

AI systems are diffused across our global ecosystem yet rapidly becoming foundational architecture for our national critical infrastructure. They are embedding themselves in everything from credit decisions and logistics networks to intelligence analysis and critical infrastructure operations. That means governance can't be treated as an afterthought—or as a single law, agency, or compliance checkbox. Governance is how we translate broad values—trust, safety, and security—into concrete engineering requirements and enforceable obligations.

The question, now, is how—and by whom AI will be governed, so that AI does what humans intended AI to do. If we leave those choices to fragmented incentives and opaque decisions, we will end up with a world where trust is improvised, accountability is optional, and the most consequential systems fail in ways that are hard to detect and harder to contest. But if we build a governance layer that is rigorous, transparent, and adaptive—one that demands evidence before deployment, monitors systems after deployment, and assigns responsibility when harms occur—we can accelerate AI adoption, reduce systemic risks, and build trustworthy AI.

Appendix A: Federal Governance Mapping by Stack Layer

Goal: Streamline AI regulation while preserving legitimate state authority; use federal floors and targeted preemption only for documented conflicts.

TECH STACK LAYER	FEDERAL	CORE PRINCIPLES
Data	Set a national data security and privacy baseline to reduce patchwork burdens. Standardize minimum controls for sensitive data plus integrity/provenance expectations.	Federal floor, not ceiling: harmonize core definitions; allow stricter state rules only if interoperable. Preempt only where requirements directly conflict or make compliance impossible.
Model	Set a federal floor for model safeguards and transparency (model cards, red-teaming, incident reporting). Empower sector regulators require certification, adversarial testing, continuous monitoring, and clear liability in high-consequence domains.	Harmonize testing and documentation to avoid 50 different model compliance regimes. Preempt conflicting state-level model certification schemes; preserve state leverage through application rules. Work within technical limits of non-deterministic mathematical models. Consider open-source model weights provenance developed by adversarial nations.
Infrastructure	Establish trusted compute / infrastructure certification regimes, supply chain risk management best practices, export controls, baseline physical security, and cyber resilience standards for AI workloads and dependencies. Coordinate systemic risk reviews and national incident response playbooks.	Favor harmonization: infrastructure is cross-border by design. Narrow preemption policies where clear state- and federal authorities reside. Demand and procure secure-by-design hardware and software technologies.
Application	Sector regulators enforce use-case controls in federally regulated sectors (e.g., certification, monitoring, incident reporting). Support application-level auditing and baseline transparency/recourse where federal law applies.	Use safe harbors and targeted preemption only for clear conflicts with federal sector regimes or national security deployments. Consider application origin for apps developed by adversarial nations.
Governance	Require an “AI Tech Stack Impact” annex in legislation; convene a Federal-State council; set federal floors for AI systems. CAISI + GAO/CRS: Map responsibilities and publish standards; coordinate incidents and report where federal-state conflicts emerge. Develop interoperable standards with international partners, allies, and standards-making bodies.	Sequence: legislate for understanding first (“do no harm”), then refine roles. Targeted preemption only for documented conflicts; avoid wholesale wipeouts and moratoriums in favor of AI race-leading interoperable standards. Partner with industry, innovators, and investors to pursue internal and external governance best practices.

Appendix B: State Governance Mapping by Stack Layer

Goal: Streamline AI regulation while preserving legitimate state authority; use federal floors and targeted preemption only for documented conflicts.

TECH STACK LAYER	STATES	CORE PRINCIPLES
Data	Enact/enforce privacy and consumer protection; breach notification; data broker rules. Use state AG enforcement and procurement to drive compliance in sectors such as employment, housing, education, healthcare, and law enforcement.	Harmonize core definitions with federal and international standards (e.g., GDPR); allow stricter state rules only if interoperable.
Model	Regulate deployments in-state (impact assessments/auditability; prohibited practices) rather than trying to regulate every model directly. Attach documentation (e.g., model cards) and audit expectations (e.g. red-team reviews) to high-risk deployments.	Work within technical limits of non-deterministic mathematical models. Consider open-source model weights provenance and origin developed by adversarial nations.
Infrastructure	Critical infrastructure oversight and datacenter siting. Land permitting, utility and energy grid policies. Harmonize AI system definitions that align to national infrastructure policies.	Favor harmonization: infrastructure is cross-border by design; narrow policies where clear state- and federal authorities reside. Demand and procure secure-by-design hardware and software technologies.
Application	Primary locus for police powers: set guardrails for in-state uses (e.g., employment, housing, education, healthcare, law enforcement, etc.). Require notice/explanations/human review and appeal; set procurement rules for state agencies.	Use safe harbors only for clear conflicts with federal sector regimes or national security deployments. Consider application provenance and origin for apps developed by adversarial nations.
Governance	Provide enforcement signals, harms, and feasibility feedback through the Federal-State council. Pilot and iterate application rules; share best practices; align definitions where possible.	Sequence: legislate for understanding first ("do no harm"), then refine roles. Partner with industry, innovators, and investors to pursue internal and external governance best practices.

References and Endnotes

1 Walden, Kemba and Lynch, Devin. (June 6, 2025). The AI Tech Stack: A Primer for Tech and Cyber Policy. Paladin Global Institute. Retrieved from https://www.paladincapgroup.com/wp-content/uploads/2025/06/AI-Tech-Stack-Report.pdf.

2 The OSI model is a framework created by the International Organization for Standardization (ISO) that standardizes network communication that "provides a common basis for the coordination of standards development for the purpose of systems interconnection." It represents communications systems as seven distinct layers: Physical, Data Link, Network, Transport, Session, Presentation, and Application. ISO. "ISO/IEC 7498-1:1994." Accessed April 7, 2026. https://www.iso.org/standard/20269.html.

3 Walden, Kemba and Lynch, Devin. (June 6, 2025). The AI Tech Stack: A Primer for Tech and Cyber Policy. Paladin Global Institute. Retrieved from https://www.paladincapgroup.com/wp-content/uploads/2025/06/AI-Tech-Stack-Report.pdf.

4 The White House. (2025, July). America's AI Action Plan. Retrieved from https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf.

5 National Institute of Standards and Technology. (2023, January 26). Artificial Intelligence Risk Management Framework (AI RMF 1.0). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1.

6 National Institute of Standards and Technology. (n.d.). Framework for Improving Critical Infrastructure Cybersecurity. Retrieved September 24, 2025, from https://www.nist.gov/cyberframework.

7 Anthropic. (February 23, 2026). Detecting and preventing distillation attacks. Retrieved from https://www.anthropic.com/news/detecting-and-preventing-distillation-attacks.

8 OpenAI. (February 12, 2026). Memo to the U.S. House Select Committee on the Strategic Competition Between the United States and the Chinese Communist Party on "Updated Stakes for American-Led, Democratic AI." Retrieved from https://assets.bwbx.io/documents/users/iqjWHBFdfxIU/rRmql_jJcxb4/v0.

9 Google Threat Intelligence Group. (February 12, 2026). GTIG AI Threat Tracker: Distillation, Experimentation, and (Continued) Integration of AI for Adversarial Use. Retrieved from https://cloud.google.com/blog/topics/threat-intelligence/distillation-experimentation-integration-ai-adversarial-use.

10 U.S. House Select Committee on the Strategic Competition Between the United States and the Chinese Communist Party. (April 2026). Buy What It Can, Steal What It Must: China's Campaign to Acquire Frontier AI Capabilities. Retrieved from https://selectcommitteeontheccp.house.gov/sites/evo-subsites/selectcommitteeontheccp.house.gov/files/evo-media-document/buy-what-it-can-steal-what-it-must.pdf.

11 Trend Micro. "Viral AI, Invisible Risks: What OpenClaw Reveals About Agentic Assistants." February 6, 2026. https://www.trendmicro.com/en_us/research/26/b/what-openclaw-reveals-about-agentic-assistants.html; and Team, Microsoft Defender Security Research. "Detecting and Mitigating Common Agent Misconfigurations." Microsoft Security

Blog, February 12, 2026. https://www.microsoft.com/en-us/security/blog/2026/02/12/copilot-studio-agent-security-top-10-risks-detect-prevent/.

12 See e.g. Federal Trade Commission. "Rite Aid Banned from Using AI Facial Recognition After FTC Says Retailer Deployed Technology without Reasonable Safeguards." December 19, 2023. https://www.ftc.gov/news-events/news/press-releases/2023/12/rite-aid-banned-using-ai-facial-recognition-after-ftc-says-retailer-deployed-technology-without (Rite Aid was banned from using AI facial recognition technology for five years after the Federal Trade Commission found it deployed technology without reasonable safeguards.) and Monash Lens. "AI Has a Governance Problem." January 19, 2026. https://lens.monash.edu/responsible-ai-is-now-a-governance-risk-not-an-ethics-debate/ (xAI's Grok chatbot faced global scrutiny when it was used to create non-consensual sexual images and deepfakes, leading to immediate regulatory pushback, including enforcement of the Malaysian Online Safety Act of 2025).

13 Scott, Mark. "At Paris AI Summit, US, EU, Other Nations Lay Out Divergent Goals." Tech Policy Press, February 11, 2025. https://techpolicy.press/at-paris-ai-summit-us-eu-other-nations-lay-out-divergent-goals.

14 See, e.g., "Low Trust: Navigating Transatlantic Relations under Trump 2.0 | European Union Institute for Security Studies." October 14, 2025. https://www.iss.europa.eu/publications/chaillot-papers/low-trust; Scott, Mark. "At Paris AI Summit, US, EU, Other Nations Lay Out Divergent Goals." Tech Policy Press, February 11, 2025. https://techpolicy.press/at-paris-ai-summit-us-eu-other-nations-lay-out-divergent-goals; and Carnegie Endowment for International Peace. "Can Europe Trust the United States Again?" January 7, 2026. https://carnegieendowment.org/posts/2026/01/can-europe-ever-trust-the-united-states-again.

15 See Sherman, Justin "Securing data in the AI supply chain" Atlantic Council (September 5, 2025) available at Securing-AI-Data-Supply-Chain-Memo.pdf.

16 See Brackett, Sara Ann "Cloudbusting: Policy for Evaluating Trust in Compute Infrastructure" Atlantic Council (December 2025) available at cloudbusting-policy-for-evaluating-trust-in-compute-infrastructure.pdf.

17 See, e.g., the lifecycle risk management framing in the NIST AI RMF 1.0 https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf; the EU AI Act's requirements on risk management and data governance/training data quality https://artificialintelligenceact.eu/article/10/; and documentation norms for dataset provenance and intended use in "Datasheets for Datasets" https://arxiv.org/abs/1803.09010).

18 The Fair Information Practice Principles (FIPPs) are as follows: (1) access and amendment, (2) accountability, (3) authority, (4) minimization, (5) quality and integrity, (6) individual participation, (7) purpose specification and use limitation, (8) security, and (9) transparency. Information about the FIPPs is available at Fair Information Practice Principles (FIPPs) | FPC.gov

19 West Virginia v. EPA, 142 S. Ct. 2587 (2022) (concluding that Congress must clearly authorize regulations when regulatory action implicates significant economic and political consequences).

20 West Virginia v. EPA, 142 S. Ct. at __ (2022) (Interpreting a law that regulates emission systems because "of course almost anything could constitute such a 'system'; shorn of all context, the word is an empty vessel").

21 The World Bank (2024). "Navigating the AI Frontier: A primer on the Evolution and impact of AI Agents." Retrieved from: WEF_Navigating_the_AI_Frontier_2024.pdf; see also Kemba Walden, Chua Seah, and Dawn Song (2025) "How we enhance cybersecurity defences before the attackers in an AGI world." Retrieved from: https://www.weforum.org/stories/2025/10/how-we-enhance-cybersecurity-defences-before-the-attackers-in-an-agi-world/

22 Moody v. NetChoice LLC, 144 S. Ct. 2383 (2024); see also Mackenzie Austin and Max Levy, "Speech Certainty: Algorithmic Speech and the Limits of the First Amendment," 77 Stan.L.Rev.1 (2025). Retrieved from: https://review.law.stanford.edu/wp-content/uploads/sites/3/2025/01/Austin-Levy-77-Stan.-L.-Rev.-1.pdf.

23 West Virginia v. EPA, 142 S. Ct. 2587 (2022) (concluding that Congress must clearly authorize regulations when regulatory action implicates significant economic and political consequences).

24 West Virginia v. EPA, 142 S. Ct. at __ (2022) (Rejecting a law that regulates emission systems because "of course almost anything could constitute such a 'system'; shorn of all context, the word is an empty vessel")

25 The White House. (2025, July). America's AI Action Plan. Retrieved from https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf; National Institute of Standards and Technology. (2023, January 26). Artificial Intelligence Risk Management Framework (AI RMF 1.0). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1; Public Law No: 119-12.

26 Multistate.AI. (2026). "Artificial Intelligence (AI) Legislation Tracker 2026: All 50 States" https://www.multistate.ai/artificial-intelligence-ai-legislation.

27 Several AI law and regulatory trackers exist to advance understanding of the policy landscape, including: NCSL https://www.ncsl.org/technology-and-communication/artificial-intelligence-2025-legislation, White and Case https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-united-states; Morris Manning & Martin, LLP https://www.mmmlaw.com/news-resources/102kaxc-the-big-long-list-of-u-s-ai-laws/; and IAPP https://iapp.org/resources/article/us-state-ai-governance-legislation-tracker/.

28 The EU AI Act's Article 3 "Definitions," includes this term: https://artificialintelligenceact.eu/article/3/.

29 See, e.g., Text - S.3312 - 118th Congress (2023-2024): Artificial Intelligence Research, Innovation, and Accountability Act of 2024, and the America's AI Action Plan, America's AI Action Plan, each of which promote AI accountability in similar fashion.

30 Palo Alto Networks. "2026 Unit 42 Global Incident Response Report." Accessed April 7, 2026. https://www.paloaltonetworks.com/resources/research/unit-42-incident-response-report.