FAIR LABS

FAIR ARTIFICIAL INTELLIGENCE RESEARCH LABS

CONGRESSIONAL BRIEFING ◆ AI POLICY ◆ 2025

AI Policy Recommendations for Congress

Detailed, nonpartisan recommendations for federal AI regulation, governance, and oversight—built to endure across administrations and generations of technology.

PROGRAM	DOCUMENT TYPE	AUDIENCE	PUBLISHED
AI Policy	Congressional Briefing	Federal Legislators & Staff	2025

FAIR Labs · A 501(c)(3) nonprofit research institute · Washington, DC

CONTENTS

Table of Contents

EXECUTIVE SUMMARY

The bottom line for legislators

IN ONE PARAGRAPH

Congress does not need to resolve every open question about artificial intelligence to act responsibly now. It needs to establish durable, technology-neutral principles; concentrate binding obligations on the uses where a mistake can cost someone a job, a benefit, their liberty, or their life; require honest disclosure so that markets and courts can function; allocate liability clearly across a complex value chain; hold the government's own use of AI to the standard it would set for others; protect competition and open research from foreclosure; and fund the standards bodies and testing infrastructure without which any statute is unenforceable. This briefing translates each of these into specific legislative options, presented with their tradeoffs so that Members can choose with open eyes.

8
policy domains, each closing with numbered recommendations to Congress

6
drafting principles designed to keep legislation durable across technology generations

3
risk tiers that concentrate binding obligations where the stakes are highest

What we recommend, in brief

1. **Legislate principles and outcomes, delegate specifics.**
Write durable, technology-neutral duties into statute and delegate the fast-moving technical detail to an accountable agency operating under clear standards, so the law need not be reopened with every model release.

2. **Tier obligations to consequence.**
Reserve the heaviest requirements for high-consequence uses that affect rights, safety, and access to essential services; leave ordinary and low-stakes applications lightly governed so innovation is not smothered.

3 **Mandate transparency that is usable.**
Require disclosure calibrated to its audience—notice to affected individuals, documentation to regulators, and provenance signals to the public—rather than dense reports no one reads.

4 **Fund the referee before the game.**
Appropriate for the standards, evaluation infrastructure, and technical talent that make any AI law enforceable. A right without a capacity to verify it is a promise the government cannot keep.

The sections that follow develop each theme in depth, weigh the leading objections, and set out legislative options an oversight committee can act on immediately or stage over successive Congresses.

SECTION I

The Landscape & the Case for a Durable Federal Approach

Every generation of consequential technology has eventually drawn Congress into the work of setting national rules—railroads and the Interstate Commerce Act, securities and the New Deal disclosure regime, automobiles and federal safety standards, the internet and a patchwork that is still being litigated. Artificial intelligence has arrived faster than any of these, diffusing across the economy in a handful of years rather than decades. The question before Congress is not whether AI will be governed, but by whom, at what level, and with how much foresight. In the absence of a coherent federal approach, that governance is already being written—by state legislatures, by courts applying decades-old statutes to novel facts, by foreign governments whose standards American firms must meet to sell abroad, and by the internal policies of a few large developers.

This briefing takes as its premise that a durable federal framework is preferable to the alternatives—not because federal action is inherently wise, but because the alternatives impose costs that fall hardest on the public and on smaller American innovators. A framework worth enacting must be nonpartisan in construction: it should be defensible to a legislator who prizes economic dynamism and to one who prizes consumer protection, because a law that survives only one governing coalition is not durable at all.

Why the status quo is itself a policy choice

It is tempting to treat inaction as neutral—a prudent wait for the technology to settle. But inaction is not the absence of a rule; it is a decision to let the current rules apply by default, and those rules were written for a different world. Anti-discrimination law, consumer-protection statutes, tort doctrine, and sectoral regulation all reach AI, but they do so unevenly and with significant gaps. A hiring algorithm may fall under employment law but escape meaningful scrutiny because no one can see inside it; a generative system may defame or defraud with no clear allocation of responsibility; a medical decision-support tool may sit in a seam between device regulation and the practice of medicine. Congress's choice is not between regulating and not regulating. It is between shaping the rules deliberately and inheriting them by accident.

The costs of fragmentation

In the vacuum left by federal inaction, states have begun to legislate. This is a legitimate and often thoughtful exercise of state authority, and some state experiments will teach the nation valuable lessons. But a proliferation of divergent state regimes carries real costs. A developer building a single product must reconcile conflicting definitions of what counts as an "automated decision," differing disclosure triggers, and incompatible testing mandates. Compliance burdens of this kind fall disproportionately on smaller firms, which lack the legal departments to navigate fifty regimes; the

largest incumbents can absorb the cost, and fragmentation can thus entrench the very concentration many state laws were meant to check. Meanwhile the public receives uneven protection that depends on the accident of geography.

GOVERNANCE PATHWAYS AND THEIR CHARACTERISTIC TRADEOFFS

Pathway	Principal advantage	Principal risk
Federal inaction (status quo)	Preserves flexibility; avoids premature lock-in	Gaps in protection; de facto rules set by courts, states, and firms
Sectoral regulation only	Leverages expert agencies and existing law	Seams between sectors; inconsistent treatment of similar risks
State-by-state legislation	Laboratories of democracy; local responsiveness	Compliance fragmentation; burden falls on small firms; uneven protection
Comprehensive federal statute	National consistency; predictability for markets	Risk of rigidity; capture; one-size treatment of diverse uses
Federal framework + delegated rulemaking	Durable principles with adaptable detail	Requires sustained agency capacity and oversight

We present these pathways without prejudging which mix a given Congress will prefer; reasonable legislators weigh these tradeoffs differently. Our analysis in the sections that follow tends toward the final row—durable statutory principles paired with delegated, accountable rulemaking—because it is the option most capable of remaining relevant as the technology changes. But the design principles in Section II apply to any of these pathways, and a Congress that prefers a narrower, sectoral approach can still adopt them.

The federalism question, stated fairly

Any federal framework must confront preemption honestly. A national standard that displaces state law offers consistency but extinguishes state experimentation and may, if set too low, cap protection below what some states would choose. A framework that merely adds a federal floor while leaving states free to go further preserves experimentation but reintroduces some of the fragmentation it sought to cure. There is no costless answer. The most defensible course is usually a federal floor that guarantees baseline protection nationwide, paired with a deliberate choice—use by use—about where uniform national standards are important enough to justify preemption, such as in interstate markets for AI products where conflicting rules would be genuinely unworkable.

A NONPARTISAN FRAMING

The case for federal action does not rest on any view about how large government should be. It rests on a structural fact: AI markets are national and international, harms cross state lines, and the costs of a fragmented, accidental rule-set fall on consumers and small innovators alike. A Member skeptical of regulation and a Member eager for it can both prefer one clear national rule to fifty conflicting ones layered atop aging statutes.

What a durable framework must do

Before turning to specifics, it is worth naming the criteria by which any AI statute should be judged. A durable framework is *adaptable*—it does not hard-code today's technical assumptions. It is *proportionate*—it concentrates burden where consequence is greatest. It is *enforceable*—it comes with the institutional capacity to verify compliance. It is *interoperable*—it aligns where possible with allied jurisdictions and international standards so that American firms face one broad target rather than many. And it is *legitimate*—it is built through a process, and lands on a substance, that citizens across the political spectrum can accept as fair. The recommendations below are offered in service of those five criteria.

1. **Establish a federal baseline.**
   Enact a national floor of AI protections and obligations so that Americans receive consistent treatment regardless of state, and so that developers face a predictable national target rather than a growing patchwork.

2. **Decide preemption use-by-use.**
   Rather than a blanket preemption or none, direct that uniform national standards displace conflicting state law only where interstate markets or cross-border harms make a single rule genuinely necessary, preserving state experimentation elsewhere.

3. **Write to durable criteria.**
   Instruct that any AI statute and its implementing rules be tested against adaptability, proportionality, enforceability, interoperability, and legitimacy—and require periodic public reporting on how well the framework meets them.

SECTION II

Principles for Durable AI Legislation

A statute keyed to today's architectures will be obsolete before the ink dries. The task is to legislate the enduring questions—who is responsible, what must be disclosed, which uses demand scrutiny—while delegating the fast-moving technical answers to an accountable process.

The central drafting challenge in AI policy is the mismatch of clocks. Legislation moves on the timescale of a Congress; the technology moves on the timescale of a product cycle. A law that names specific model sizes, techniques, or capability thresholds risks being either overbroad or a dead letter within a year or two. The remedy is not to legislate vaguely, but to legislate at the right level of abstraction: fix the durable principles in statute, and delegate the perishable specifics to a body that can revise them without an act of Congress, under standards and oversight that Congress sets. Six principles, drawn from mainstream administrative-law practice and from the emerging consensus in the international standards community, should guide that drafting.

1. Technology neutrality

Regulate what a system does and the consequences it can cause, not the particular method by which it does it. A statute that governs "machine-learning models trained on more than N examples" will be gamed or bypassed by the next architecture; a statute that governs "automated systems used to make or materially inform decisions about a person's employment" will still bite whatever the underlying technique. Technology neutrality also serves fairness: it treats functionally identical risks identically, whether they arise from a neural network, a statistical model, or a hand-tuned rule engine.

2. Risk proportionality

Obligations should scale with the potential for harm. This principle, developed at length in Section III, is what keeps a durable framework from becoming either a rubber stamp or a straitjacket. A regime that treats a spam filter and a parole-recommendation system alike will over-burden the trivial and under-scrutinize the grave.

3. Outcome orientation

Where possible, require outcomes—non-discrimination, accuracy adequate to the use, meaningful human review—rather than prescribing the exact engineering steps to achieve them. Outcome-based duties survive technological change and let firms innovate on *how* to comply. Prescriptive process-based rules are sometimes necessary as a backstop, especially where outcomes are hard to observe, but they should be the exception, revisited on a schedule.

4. Delegation with guardrails

Congress should delegate technical specification to an agency, but delegation is only durable if it is disciplined. The statute should supply an intelligible principle to guide the agency, require notice-and-comment rulemaking, mandate periodic review, and preserve congressional oversight and the ordinary avenues of judicial review. Delegation done well multiplies congressional intent across a changing landscape; done poorly, it hands away the pen.

THE COLLINGRIDGE DILEMMA

Early in a technology's life it is easy to shape but hard to understand; later it is easy to understand but hard to shape. AI legislation must act under this dilemma, not pretend it away. The practical response is to legislate reversible, adaptable instruments now—principles, disclosure, delegated standards—and to build in the review cycles that let the framework tighten or loosen as understanding catches up with capability.

5. International interoperability

American firms sell into a global market, and American values are better served by shaping international standards than by ceding them. Where a federal framework can align with widely adopted reference points—the OECD AI Principles, the risk-management vocabulary of the NIST AI Risk Management Framework, the international management-system standards emerging from ISO/IEC—it lowers compliance costs and amplifies U.S. influence. Interoperability is not harmonization for its own sake; it is a recognition that a fractured global rule-set penalizes the firms most likely to be American.

6. Adaptability by design

Finally, a durable statute plans for its own revision. Sunset provisions on the most technical requirements, mandatory look-back reports, and a standing mechanism for updating definitions all build learning into the law. The goal is a framework that ages gracefully—one whose principles hold while its particulars evolve.

TWO DRAFTING STYLES, COMPARED

Dimension	Prescriptive / specific	Principles + delegated detail
Durability	Low—tied to current technology	High—principles outlast techniques
Predictability	High in the short run	Moderate; depends on rulemaking clarity
Speed of adaptation	Slow—needs new legislation	Fast—rules revised by agency
Risk of capture	Concentrated at drafting	Ongoing; mitigated by transparency and oversight
Democratic control	Direct but brittle	Indirect but sustained via oversight

> *Legislate the questions that do not change; delegate the answers that do.*
>
> — The organizing maxim of durable technology law

4 **Adopt technology-neutral definitions.**

Define regulated activity by function and consequence—"automated systems that make or materially inform consequential decisions"—rather than by technique, model size, or training method.

5 **Pair statutory principles with disciplined delegation.**

Fix the durable duties in statute and delegate technical specification to an accountable agency under an intelligible principle, with notice-and-comment, periodic review, and preserved oversight.

6 **Build in review and interoperability.**

Require look-back reports and, where appropriate, sunsets on the most technical provisions; direct agencies to align with recognized international standards where doing so protects the public without lowering the U.S. baseline.

SECTION III

Risk-Based Regulation

The single most important design decision in AI legislation is where to place the burden. Regulate everything at the same intensity and the framework will either paralyze ordinary commerce or, more likely, collapse into a formality that scrutinizes nothing. Regulate nothing and the gravest harms proceed unchecked. A risk-based approach—one that concentrates binding obligations on the uses where a mistake is most costly and hardest to reverse—resolves this by making rigorous oversight affordable exactly where it matters. This is the approach at the heart of the NIST AI Risk Management Framework, the OECD's work, and most serious legislative proposals in allied jurisdictions, and it commands unusually broad support across the political spectrum.

Tiering by consequence, not by capability

An important choice is what to tier *on*. Some proposals tier on the raw capability of the underlying model—its size, its general power. This has a role for the small number of frontier systems whose broad capabilities create diffuse risks, discussed below. But for the vast majority of the economy, the more useful axis is the *consequence of the use*: what happens to a real person if the system is wrong, biased, or manipulated? A modest model deciding who gets a mortgage warrants more scrutiny than a powerful model summarizing meeting notes. Tiering on consequence keeps the framework technology-neutral and focuses attention on the public's actual exposure to harm.

AN ILLUSTRATIVE THREE-TIER STRUCTURE FOR USE-BASED OBLIGATIONS

Tier	Illustrative uses	Core obligations
Tier 1 — Minimal	Spam filtering, content recommendation, productivity aids, internal drafting	General duties of care and honesty; no bespoke pre-market requirements
Tier 2 — Elevated	Systems that materially inform decisions about credit, housing, insurance, hiring, education	Documentation, testing for accuracy and bias, disclosure to affected persons, human review
Tier 3 — High-consequence	Systems central to decisions on liberty, essential benefits, safety-critical infrastructure, medical care	Independent evaluation, registration, ongoing monitoring, meaningful human oversight, audit access

The table is illustrative, not a proposed statute; the precise assignment of uses to tiers is exactly the kind of technical detail that should be developed through rulemaking and revised over time. What the statute should fix is the *structure*—that obligations scale with consequence—and the *principles* for assignment: severity of potential harm, the number of people exposed, the reversibility of a mistake, and the degree to which a human meaningfully intervenes before the outcome lands.

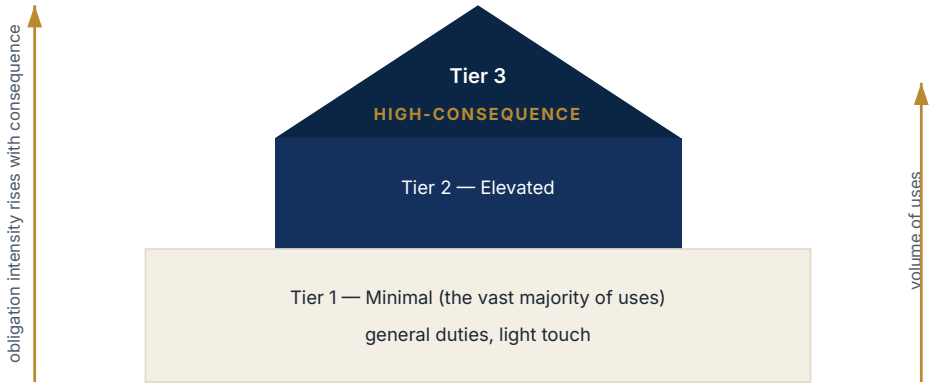

Figure 1. A risk-based framework inverts effort and volume: most AI uses sit in a lightly governed base, while binding obligations concentrate on the narrow apex of high-consequence applications. This is what makes rigorous oversight affordable where it counts.

What each tier should require

Tier 1 (minimal risk) should be governed by general, cross-cutting duties—honesty, a baseline duty of care, and existing consumer-protection and civil-rights law—without bespoke pre-market requirements. The overwhelming majority of AI uses belong here, and keeping this tier light is essential to preserving the innovation and productivity gains that make AI worth adopting.

Tier 2 (elevated risk) covers systems that materially inform consequential decisions about a person's economic life but leave a human meaningfully in control. Here the framework should require documentation of the system's design and data, testing for accuracy and for disparate performance across protected groups, disclosure to affected individuals, and a genuine avenue for human review. These obligations are demanding but not exotic; they resemble the diligence a prudent institution would perform in any event.

Tier 3 (high-consequence) is reserved for the narrow set of uses that determine or heavily steer decisions about liberty, essential benefits, safety-critical infrastructure, and health. Here the full apparatus is warranted: independent evaluation before deployment, registration with the relevant authority, continuous monitoring, meaningful and resourced human oversight, and audit access for regulators. The burden is significant, but so is the exposure of the public, who did not consent to be the test subjects of an unvetted system.

A DESIGN RULE FOR BORDERLINE CASES

When a use sits on the boundary between tiers, the framework should err toward the higher tier for the specific features that create the risk, not the whole system. Over-classifying wholesale breeds resentment and evasion; targeting the precise risky function preserves proportionality while still protecting the public. Rulemaking, not statute, is the right place to draw these lines, because they will need to move.

Frontier and general-purpose models

A small number of highly capable, general-purpose systems present a distinct challenge: their very generality means their risks cannot be neatly assigned to a single use. For this narrow category, a capability-informed overlay may be warranted—transparency about training and evaluation, testing for dangerous capabilities, and information-sharing with a designated authority—layered on top of the use-based tiers that govern downstream applications. Congress should treat this as a targeted addition for the frontier, not as the organizing principle for the whole economy, and should define the category by function and demonstrated capability rather than by a brittle numeric threshold that will not age well.

The objection, stated fairly

The strongest objection to risk-based regulation is that consequence is hard to assess in advance and that tiering invites both gaming (designing a system to fall just below a threshold) and regulatory drift (uses migrating upward in stakes without re-classification). These are real. The answer is not to abandon proportionality but to pair it with the disclosure and monitoring regimes of Sections IV and VII, and with an assignment process nimble enough to reclassify uses as evidence accumulates. A framework that tiers, and then watches whether its tiers are holding, is far better than one that either scrutinizes everything or nothing.

7 **Adopt a use-based, three-tier structure.**
Concentrate binding obligations on high-consequence uses affecting rights, safety, and access to essential services; keep the large base of ordinary uses lightly governed under general duties.

8 **Delegate tier assignment, fix the principles.**
Set in statute the criteria for tiering—severity, scale, reversibility, human control—and delegate the specific, revisable assignment of uses to rulemaking with public participation.

9 **Add a narrow frontier overlay.**

For the small class of highly capable general-purpose systems, layer targeted transparency, dangerous-capability testing, and information-sharing duties on top of the use-based tiers, defined by function rather than a fixed numeric threshold.

SECTION IV

Transparency & Disclosure Mandates

Markets, courts, and citizens cannot hold accountable what they cannot see. Transparency is the connective tissue of every other reform in this briefing—but only if it is designed for the people who must act on it.

Disclosure is the least intrusive regulatory instrument and often the most powerful. It does not ban or prescribe; it makes information available so that other mechanisms—competition, litigation, oversight, individual choice—can work. But transparency is frequently done badly, in ways that generate paperwork without generating understanding. The central design question is not *whether* to require disclosure but *to whom*, *of what*, and *in what form*. A single undifferentiated mandate produces documents that satisfy no one: too technical for the affected individual, too shallow for the regulator, too voluminous for the public. Effective transparency is layered, with each layer calibrated to its audience.

Three audiences, three kinds of disclosure

The first audience is the **affected individual**—the person whose loan, job, benefit, or claim an AI system has touched. What they need is notice that an automated system was materially involved, an explanation intelligible to a non-specialist, and a clear route to contest the outcome. The second audience is the **regulator or auditor**, who needs deeper documentation: how the system was built and tested, on what data, with what known limitations. The third audience is the **general public**, who benefit from broad signals—provenance markers on synthetic media, registries of high-consequence government systems—that support informed civic life without exposing trade secrets or security-sensitive detail.

A LAYERED DISCLOSURE REGIME, BY AUDIENCE

Audience	What they need	Form of disclosure	Chief risk if omitted
Affected individual	Notice, plain-language explanation, route to contest	Point-of-decision notice; accessible reasons; appeal path	Unappealable, opaque determinations
Regulator / auditor	Design, data, testing, limitations	Structured documentation; audit access under confidentiality	Unverifiable claims; unenforceable duties
General public	Provenance and existence signals	Content-provenance markers; public registries of high-stakes systems	Erosion of trust; undetectable manipulation

Documentation as infrastructure

For Tier 2 and Tier 3 systems, structured documentation should be a standing requirement, drawing on well-established practice in the research community. Instruments such as model cards and datasheets for datasets provide a mature template: a concise, comparable account of what a system is for, how it performs across conditions and groups, what data it learned from, and where it should not be used. Standardizing these artifacts lowers the cost of compliance and of oversight simultaneously, because a regulator reviewing a familiar format works faster and a developer producing a familiar format guesses less. Congress should direct the standards body of Section VIII to specify these formats rather than freezing them in statute.

THE USABILITY TEST

A disclosure requirement should be judged by a single question: does someone actually act differently because of it? Notice that an affected person cannot understand, documentation a regulator has no capacity to review, provenance signals no platform surfaces—all fail this test. Congress should resist the temptation to measure transparency by the weight of the paperwork it generates and instead require that disclosures be tested for comprehension and use.

Content provenance and synthetic media

A distinct transparency concern is the growing difficulty of distinguishing authentic from synthetic content. Provenance and watermarking standards—cryptographic signals that travel with content to indicate how it was made—offer a partial, imperfect remedy. Congress should encourage, and where appropriate require, the adoption of open provenance standards for consequential contexts, while being candid about the limits: watermarks can be stripped, and no technical measure substitutes for media literacy and robust institutions. Disclosure here should aim to raise the cost and lower the reach of deception, not to promise its elimination.

Guarding against transparency theater

Two failure modes deserve explicit attention. The first is *disclosure that buries*: dense, boilerplate documents that technically comply while defeating comprehension. The second is *disclosure that endangers*: mandates so detailed they expose security-sensitive information or legitimate trade secrets, chilling the very development the framework hopes to guide. The answer to both is calibration—tiered depth, audience-appropriate form, and confidentiality protections for genuinely sensitive material disclosed to regulators rather than to the world. Transparency is a means to accountability, not an end in itself, and a mandate that produces neither understanding nor verifiable compliance has failed regardless of how much it discloses.

10 **Require layered, audience-calibrated disclosure.**
Mandate plain-language notice and explanation to affected individuals, structured documentation to regulators under appropriate confidentiality, and public provenance signals—each designed for its audience rather than a single undifferentiated report.

11 **Standardize documentation formats.**
Direct the federal standards body to specify and maintain model- and data-documentation formats, building on established practice, so that compliance and oversight share a common, comparable language.

12 **Advance content provenance honestly.**
Support open provenance and watermarking standards for high-stakes contexts while acknowledging their limits and pairing them with investment in institutions and media literacy rather than treating them as a technical cure.

SECTION V

Accountability & Liability

When an AI system causes harm, the injured person confronts a distinctive problem: the chain from cause to consequence runs through many hands. A foundation model is built by one firm, fine-tuned by another, integrated into a product by a third, deployed by a fourth, and used by a fifth. Under existing tort and consumer-protection doctrine it is often unclear who is answerable, and that uncertainty serves no one well—not the victim who cannot find a defendant, and not the responsible actors who cannot price their exposure. Clarifying how responsibility is allocated across the value chain is among the most consequential things Congress can do, and among the most technically delicate.

Why the value chain complicates liability

Traditional product liability presumes a reasonably identifiable maker of a reasonably stable product. AI strains both presumptions. The "product" may change after sale through updates; its behavior emerges from data and training rather than from a specifiable design; and the harm often results from an interaction between the developer's model, the deployer's configuration, and the user's inputs. Doctrines built for defective toasters do not map cleanly onto systems that learn. The result, absent legislative clarity, is a slow accretion of case law that will be inconsistent across circuits and years in the making—during which both victims and firms bear the cost of uncertainty.

ALLOCATING RESPONSIBILITY ACROSS THE AI VALUE CHAIN

Actor	What they control	Duties best assigned to them
Foundation-model developer	Training data, base capabilities, documented limits	Disclose capabilities and known risks; test for dangerous behavior; honest documentation
Fine-tuner / integrator	Adaptation, guardrails, task-specific behavior	Test the adapted system for its intended use; document changes; pass through limits
Deployer	Context of use, configuration, human oversight	Ensure fitness for the actual use; provide notice and appeal; monitor in operation
End user	Inputs, adherence to instructions	Use within documented scope; report failures

The principle underlying the table is that **responsibility should track control**: each actor should bear duties commensurate with what they can actually observe and influence. A deployer cannot be expected to audit the internals of a foundation model it did not build; a foundation-model developer cannot foresee every downstream use. Assigning each party the duties it is best positioned to fulfill both protects the public and gives each actor a clear, insurable obligation. It also avoids the two

failure modes at the extremes: liability that lands entirely on the deepest pocket regardless of fault, and liability so diffuse that no one is answerable.

> *Responsibility should track control. Each actor should answer for what it can actually see and shape—no more, and no less.*
>
> — The allocation principle

Legislative options, with tradeoffs

Congress has several instruments, and they are not mutually exclusive. Each embodies a different judgment about how to balance victim protection, innovation, and administrability.

Duty-of-care standards

Congress can codify a duty of reasonable care tailored to each role in the value chain, giving courts a clear standard against which to judge conduct. This preserves the flexibility of the common law while curing its uncertainty about who owes what. The tradeoff is that "reasonableness" still requires case-by-case elaboration, which takes time to settle.

Rebuttable presumptions and burden-shifting

Because AI systems are opaque, a victim may be unable to prove exactly how a harm occurred. Congress can ease this by shifting the burden: where a claimant shows harm from a covered high-consequence system and a failure to meet documentation or testing duties, the burden shifts to the developer or deployer to show it acted with due care. This targets the information asymmetry directly. The tradeoff is added exposure for firms, which must be calibrated so it does not deter beneficial deployment.

Safe harbors tied to standards

To reward good practice and give firms a path to predictability, Congress can offer a safe harbor: an actor that follows recognized standards, maintains required documentation, and cooperates with post-incident investigation receives a rebuttable presumption of due care. Safe harbors align private incentives with public safety and reduce litigation for responsible actors. The tradeoff is that a poorly drawn safe harbor can become a shield for box-checking; it must be tied to substantive standards and lost in cases of bad faith or concealment.

A BALANCED PACKAGE

These instruments work best in combination: clear role-based duties of care, a narrow burden-shift for high-consequence systems to address opacity, and a safe harbor that rewards demonstrable adherence to standards. Together they protect victims where harm is gravest, give responsible firms a predictable path, and keep the deepest-pocket temptation in check. The precise calibration is a legitimate subject of debate; the goal of clarity is not.

Insurance and the pricing of risk

A functioning liability regime does more than compensate victims; it prices risk, and priced risk is managed risk. When each actor faces a clear, insurable duty, insurers and markets begin to reward safer practices and penalize reckless ones—an accountability engine that operates continuously, without a regulator in the room. Congress should view clear liability rules not merely as a remedy after the fact but as an ex-ante incentive for the whole ecosystem to internalize the cost of harm. Legal clarity is a precondition for this market to form; today's uncertainty is itself a barrier to the private discipline that good liability rules unlock.

13 **Codify role-based duties of care.**
Assign each actor in the value chain duties commensurate with what it controls, so that responsibility tracks control and every party has a clear, insurable obligation.

14 **Address opacity with a targeted burden-shift.**
For high-consequence systems, shift the burden to developers and deployers to show due care once a claimant demonstrates harm and a lapse in required testing or documentation, curing the information asymmetry that leaves victims unable to prove their case.

15 **Offer a standards-based safe harbor.**
Grant a rebuttable presumption of due care to actors who follow recognized standards, maintain required documentation, and cooperate after incidents—forfeited in cases of concealment or bad faith.

SECTION VI

Government's Own Use of AI

The federal government is among the largest purchasers and deployers of software on earth. How it buys, builds, and oversees its own AI sets a standard the private market will follow—and determines whether citizens can trust the systems that mediate their access to public services.

Two features make government's own use of AI a distinct policy priority. First, the state wields powers no private actor holds—to tax, to adjudicate benefits, to investigate, to detain—so that AI errors in government carry a gravity and a due-process dimension that market discipline alone cannot address. Second, federal procurement is a lever of extraordinary reach: the requirements the government writes into its contracts propagate through the vendor ecosystem and become de facto standards. Congress can therefore accomplish a great deal simply by governing the government well, and it can do so without the definitional battles that attend regulating the private sector, because the state is regulating itself.

Procurement as policy

The most powerful instrument here is the acquisition process. By requiring, as a condition of purchase, the documentation, testing, and contractual terms described earlier in this briefing, Congress can ensure that public-sector AI meets a high bar—and, in the bargain, create market demand for those same practices that benefits every buyer. Procurement rules should require vendors to supply structured documentation, to permit acceptance testing on the government's own data and population, to notify the government before materially changing a model, to cooperate with post-incident investigation, and to guarantee exit and data portability so the government is never locked into a system it needs to pause.

PROCUREMENT SAFEGUARDS FOR FEDERAL AI ACQUISITION

Safeguard	What it secures	Why it belongs in the contract
Documentation delivery	Design, data, testing, and limitations of the system	Enables independent review the government could not otherwise perform
Acceptance testing on agency data	Performance on the real population, not a vendor benchmark	Leverage is greatest before money changes hands
Change notification	Right to re-test before a model update takes effect	A silent update is an ungoverned redeployment
Incident cooperation	Logs and access needed to investigate failures	Cannot be secured after leverage evaporates
Auditability	Inspection rights to the depth the use requires	Trade secret is a legitimate interest, not a universal shield against public duty
Exit & portability	Ability to migrate and decommission without punitive cost	Lock-in raises the price of pausing a system that should be paused

Rights protection in public-sector deployments

Where the government uses AI in decisions that affect rights and benefits, additional safeguards are warranted beyond those any prudent buyer would demand. Citizens are entitled to know when an automated system is materially involved in a decision about them, to receive an explanation, and to contest the outcome before a human with authority to change it. High-consequence government systems should be inventoried in a public register, evaluated independently before deployment, and monitored in operation. These are not novel demands; they restate, in the AI context, the due-process commitments that have long constrained government action. The Government Accountability Office's accountability framework for federal AI and the National Institute of Standards and Technology's risk management work provide ready reference points for implementation.

LEAD BY EXAMPLE, NOT BY EXEMPTION

A recurring temptation is to exempt government's own systems—especially in law enforcement and national security—from the transparency and oversight the framework demands of others. Some genuinely sensitive uses require tailored, classified oversight rather than public disclosure. But a broad self-exemption would forfeit the government's standing to ask anything of the private sector and would remove scrutiny precisely where state power makes it most necessary. The right answer is adapted oversight—often through inspectors general and cleared congressional committees—not the absence of oversight.

National security and law enforcement, handled with care

Sensitive government uses of AI—in intelligence, defense, and policing—cannot always be governed by public disclosure without defeating their legitimate purpose or endangering operations. But sensitivity is a reason to tailor oversight, not to abandon it. Congress can require that such systems be subject to internal testing and documentation, to review by inspectors general, and to reporting to the appropriate cleared committees, preserving accountability through trusted channels. The history of surveillance authorities counsels that oversight designed at the outset is far more effective than oversight retrofitted after a controversy.

Building internal capacity to buy and build well

None of this works if agencies lack the expertise to evaluate what they are buying. A government that cannot tell a sound system from a slick demonstration will be captured by its vendors regardless of what its contracts say. Section VIII returns to this theme, but it bears stating here: the single best investment in responsible government AI is the technical talent inside agencies to exercise informed judgment. Procurement rules are only as good as the officials who apply them.

16 **Use procurement as a standard-setter.**
Require federal AI acquisitions to include documentation, acceptance testing on agency data, change notification, incident cooperation, audit rights, and exit provisions—propagating good practice through the vendor ecosystem.

17 **Protect due process in public deployments.**
Require notice, explanation, and a route to contest for consequential government AI decisions; inventory high-consequence systems in a public register; and mandate independent evaluation and ongoing monitoring.

18 **Tailor, do not exempt, sensitive uses.**
For national-security and law-enforcement systems, substitute trusted-channel oversight—inspectors general and cleared committees—for public disclosure rather than removing accountability altogether.

SECTION VII

Innovation, Competition, Open Models & Small-Developer Concerns

A framework that protects the public but hands the field to a handful of incumbents would be a poor bargain. American leadership in AI rests on a competitive, dynamic ecosystem—startups challenging incumbents, university labs pushing the frontier, open research diffusing capability widely. Poorly designed regulation can inadvertently entrench the largest firms by imposing fixed compliance costs that only they can bear. A durable framework must therefore be judged not only by the harms it prevents but by the competition and innovation it preserves. This is not a partisan concern; a vibrant, contestable market serves both the case for economic dynamism and the case for keeping any single firm from accumulating unaccountable power.

Regulation's uneven weight

Compliance is a fixed cost, and fixed costs are regressive across firm size. A documentation or testing requirement that a large developer absorbs with an existing compliance department can be prohibitive for a five-person startup. If a framework applies its heaviest obligations uniformly, it will—whatever its intent—advantage incumbents and raise barriers to entry. The risk-based structure of Section III is the primary safeguard here: by concentrating obligations on high-consequence uses rather than on all AI, it leaves most of the startup ecosystem in the lightly governed base. But Congress should go further and consider explicit accommodations calibrated to firm size and maturity.

DESIGN CHOICES AND THEIR EFFECT ON COMPETITION

Design choice	Effect on small developers	Mitigation
Uniform obligations for all AI	Regressive; entrenches incumbents	Adopt risk-based tiering; reserve heavy duties for high-consequence uses
Per-firm compliance staff mandates	Fixed cost falls hardest on small firms	Scale requirements to firm size and revenue; shared compliance resources
Restrictions on open model release	Chills academic and startup innovation	Target conduct and use, not publication as such; narrow, evidence-based limits
Complex, litigation-heavy liability	Deters entry; favors those who can absorb legal risk	Clear duties, safe harbors, proportionate exposure

The open-model question, presented even-handedly

Few issues in AI policy are as genuinely contested as the treatment of openly released model weights. The case *for* openness is substantial: open models drive research, enable independent safety scrutiny that closed systems resist, lower costs for startups and public-interest users, support reproducible science, and guard against the concentration of capability in a few firms. The case for *caution* is also serious: once weights are released they cannot be recalled, and some capabilities—if they mature—could be misused in ways that post-hoc controls cannot address. Both concerns are legitimate, and a responsible framework must weigh them rather than dismiss either.

A CALIBRATED POSTURE ON OPEN MODELS

The most defensible approach regulates *uses and conduct* rather than publication as such. Openness should be treated as a public good with a strong presumption in its favor, disturbed only by narrow, evidence-based restrictions where a specific, demonstrated capability creates serious risk that lighter measures cannot address. Congress should be wary of broad restrictions on releasing model weights: they are hard to enforce, they fall hardest on the academic and startup communities that most rely on openness, and they cede ground to firms and jurisdictions that face no such limits.

Positive measures, not just restraint

Protecting competition is partly a matter of avoiding harm, but Congress can also act affirmatively. Public investment in shared research infrastructure—compute, curated datasets, and evaluation resources made available to academics and small firms—can widen access to the inputs that only the largest players can otherwise afford. A national research resource of this kind, an idea with bipartisan pedigree, would help ensure that the next generation of American AI is not built solely inside a few corporate labs. Support for open standards, for interoperability, and for a healthy talent pipeline all serve the same end: a field in which many can compete, and in which no single actor becomes too central to challenge.

> *A framework should be judged not only by the harms it prevents, but by the competition and open inquiry it preserves.*
>
> — The innovation test

The balance, stated plainly

There is a real tension between safety obligations and the friction they impose on innovation, and it would be dishonest to pretend otherwise. But the tension is often overstated, because the largest gains from AI come from the very high-consequence uses—in health, finance, and public services—where public trust is the binding constraint on adoption. In those domains, credible safeguards are not a tax on innovation; they are what makes deployment possible at all. The task for Congress is to

place the friction precisely: heavy where consequence and public trust demand it, light everywhere else, and never so heavy as to be affordable only by the incumbents the framework should keep contestable.

19 **Scale obligations to firm size and risk.**

Pair risk-based tiering with accommodations for smaller developers—phased compliance, shared resources, proportionate requirements—so that regulation does not become a moat protecting incumbents.

20 **Preserve openness; regulate conduct.**

Maintain a strong presumption in favor of open research and model release, restricting publication only through narrow, evidence-based measures aimed at specific demonstrated risks rather than broad limits on openness itself.

21 **Invest in shared research infrastructure.**

Fund a national research resource providing compute, data, and evaluation tools to academics and small firms, broadening access to the inputs of frontier research and guarding against undue concentration.

SECTION VIII

Institutional Capacity: Standards, Testing & Coordination

Every reform in this briefing presumes a government able to write technical standards, test systems, and coordinate across agencies. Without that capacity, the finest statute is a paper promise. Building it is the least glamorous and most decisive recommendation we make.

Legislation confers rights and duties; institutions make them real. A law that requires high-consequence systems to be independently evaluated presupposes evaluators and methods; a law that requires standardized documentation presupposes a body to set the standard; a law that allocates liability presupposes courts and regulators able to understand the evidence. The recurring lesson of technology governance is that capacity, not statutory text, is the binding constraint. Congress can and should legislate ambitiously—but it must fund the machinery that turns ambition into enforcement, or it will have written expectations the government cannot meet.

Standards: the technical backbone

Much of the substance of AI governance is necessarily technical—how to measure accuracy across groups, what a robustness test should cover, what a documentation format must contain. These are not questions for statute; they change too fast and demand too much expertise. They are the proper work of a standards body operating through open, participatory processes. The National Institute of Standards and Technology has already produced a widely respected, voluntary AI Risk Management Framework that provides a common vocabulary and a structure for governing, mapping, measuring, and managing AI risk. Congress should build on this foundation, resourcing NIST and its associated efforts to develop the evaluation methods, benchmarks, and documentation standards that a risk-based framework requires.

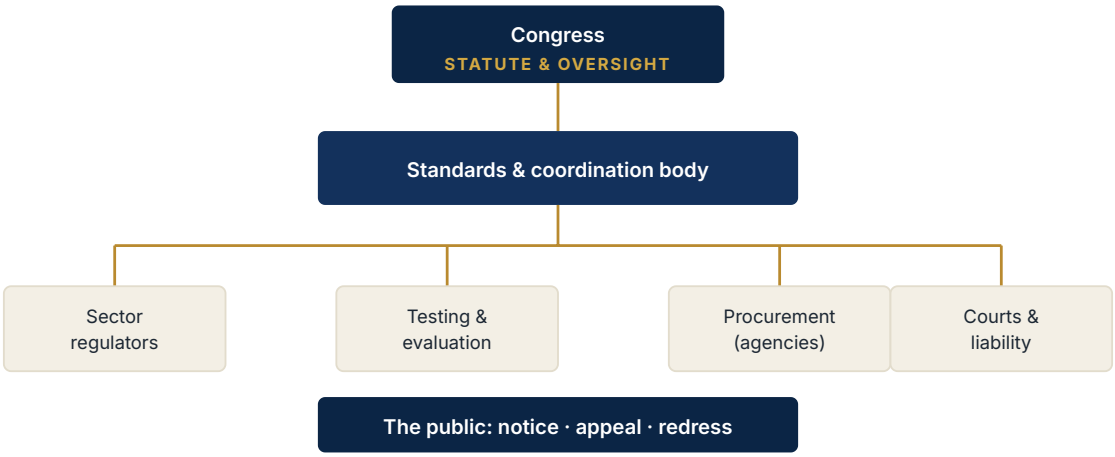

Figure 2. A workable governance architecture: Congress sets durable statute and conducts oversight; a standards and coordination body translates principles into technical specifications; sector regulators, evaluators, procurement offices, and courts apply them; and the public receives notice, appeal, and redress. Each layer depends on the capacity of the others.

Testing and evaluation infrastructure

A framework that requires high-consequence systems to be evaluated needs somewhere for that evaluation to happen and someone competent to perform it. Government cannot and need not test every system itself, but it must be able to set the methods, accredit qualified evaluators, and conduct or commission independent assessment where the stakes are highest. Public investment in shared testing infrastructure—evaluation suites, red-teaming capacity, secure environments for examining sensitive systems—benefits regulators, developers, and the public at once. An AI safety and evaluation capacity within government, coordinating with academic and independent testers, is the institutional counterpart to the risk-based obligations of Section III: without it, "independent evaluation" is a requirement with no one to fulfill it.

THE ENFORCEMENT GAP

The gravest risk to any AI statute is not that it will be too strict or too lax, but that it will be unenforceable—rights on paper that no agency has the talent or tools to verify. History offers many examples of ambitious technology laws hollowed out by under-resourced implementation. Congress should treat the appropriations that build standards, testing, and technical talent as inseparable from the substantive law, not as a discretionary afterthought.

The talent problem

The government competes for AI expertise with employers that can pay many times a federal salary. This gap is a structural threat to every reform here, because officials who cannot understand what they regulate or buy will be led by those who do. Congress should invest in technical talent through specialized hiring authorities, fellowship and rotation programs that bring experts into government for

a tour of service, and dedicated technical units that agencies can draw on. The return on this investment is disproportionate: a modest cadre of skilled evaluators inside government multiplies the effectiveness of every other dollar spent on AI oversight.

Coordination across a fragmented government

AI cuts across the jurisdiction of nearly every agency and congressional committee, which creates a coordination problem. Left unmanaged, it produces duplicative and conflicting requirements, gaps between agencies, and forum-shopping by regulated firms. The answer is not necessarily a single new super-regulator—an idea with real drawbacks, including the concentration of authority and the loss of sectoral expertise—but a coordinating function that harmonizes definitions, shares technical resources, and resolves overlaps, while sector regulators retain their domain expertise. The National AI Initiative established coordinating machinery across the federal enterprise that offers a foundation to build on. Congress should ensure that whatever body performs this function has the standing, resources, and technical depth to make coordination real rather than nominal.

A note on international coordination

Institutional capacity has an international dimension as well. The methods, benchmarks, and standards developed at home are more powerful when shared with allies, and American firms benefit when U.S. approaches shape the international reference points that define global market access. Investment in domestic standards and evaluation capacity is therefore also an investment in American influence abroad. A government that leads in the science of AI evaluation will help write the world's rules; one that does not will follow rules written by others.

22 **Fund standards and evaluation infrastructure.**
Resource NIST and associated bodies to develop evaluation methods, benchmarks, and documentation standards, and build shared testing capacity—including red-teaming and secure evaluation environments—so that statutory requirements have the machinery to be met.

23 **Close the talent gap.**
Establish specialized hiring authorities, fellowship and rotation programs, and dedicated technical units so that agencies possess the expertise to evaluate, buy, and oversee AI on equal footing with the firms they regulate.

24 **Coordinate without over-centralizing.**

Charge a coordinating function with harmonizing definitions, sharing technical resources, and resolving overlaps across agencies—preserving sectoral expertise—and align domestic standards with trusted international partners to extend American influence and lower compliance costs.

CLOSING NOTE TO MEMBERS

The recommendations in this briefing are offered as a menu, not a package to be swallowed whole. A Congress that adopts even a subset—durable principles, risk-based tiering, usable transparency, clear liability, exemplary government practice, protected competition, and the capacity to enforce any of it—will have given the country a far better foundation than the accidental rule-set that prevails today. None of it requires choosing between innovation and protection. All of it requires choosing deliberation over drift. That choice, at least, is squarely within Congress's power, and it belongs to no party alone.

APPENDIX

Notes & Further Reading

This briefing synthesizes mainstream expert understanding in AI governance, administrative law, and technology policy. The sources below are widely recognized entry points into the literature and practice that inform it; they are indicative rather than exhaustive, and their inclusion implies no endorsement of any specific pending legislation.

1. National Institute of Standards and Technology. *AI Risk Management Framework (AI RMF 1.0).* A voluntary framework organizing AI risk into the govern, map, measure, and manage functions.
2. Organisation for Economic Co-operation and Development. *OECD AI Principles.* Intergovernmental principles for trustworthy AI, including transparency, accountability, and robustness, adopted across dozens of countries.
3. U.S. Government Accountability Office. *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities.* On governance, data, performance, and monitoring for public-sector AI.
4. National Artificial Intelligence Initiative Act and the work of the National AI Initiative Office. On federal coordination of AI research, standards, and policy across agencies.
5. National Security Commission on Artificial Intelligence. *Final Report.* On U.S. competitiveness, talent, and institutional capacity in AI.
6. Executive Office of the President, Office of Science and Technology Policy. *Blueprint for an AI Bill of Rights.* On rights-protective principles for automated systems, including notice and human alternatives.
7. Mitchell, M., et al. *Model Cards for Model Reporting.* On structured documentation of model performance across conditions and groups.
8. Gebru, T., et al. *Datasheets for Datasets.* On documenting data provenance, composition, and intended use.
9. Selbst, A. D., et al. *Fairness and Abstraction in Sociotechnical Systems.* On why fairness and accountability cannot be resolved at the level of the model alone.
10. Raji, I. D., et al. *Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing.* On institutional audit processes across the development lifecycle.
11. Collingridge, D. *The Social Control of Technology.* On the dilemma of governing technologies whose effects are hardest to change once well understood.
12. ISO/IEC 42001, *Information technology — Artificial intelligence — Management system.* On international management-system standards supporting interoperable AI governance.