

The American AI Security Act

A plain-language summary of a proposed law to test, monitor and prevent serious harm from the most powerful AI systems

THE BILL IN SIX POINTS

- A new independent agency.** The AI Security Administration (AISA) would test frontier AI systems, collect reports of their failures and misuse, and act to prevent serious harm, whether the threat starts at home or abroad.
- Rules only for the largest systems.** The main duties apply to AI models trained with more than 10²⁶ operations, a scale only the biggest frontier models have reached, plus any system the agency specifically designates on evidence.
- Report failures fast.** Companies must report a severe ongoing threat within 24 hours and serious incidents, containment failures and weight theft within 72 hours. The agency must publish what it learns, including its own missed deadlines.
- A short emergency brake, held by a court.** The agency can restrict a specific AI system facing imminent serious harm, but a federal court must hear and decide the case within 7 days or the restriction ends on its own.
- Chatbot protections for everyone, and more for children.** Chatbots must say they are AI, respond properly to users in crisis, and, for minors, block sexual content and manipulative design. The FTC and State attorneys general enforce.
- Guardrails on the government too.** Warrants for private chat content, whistleblower protection, humans in charge of nuclear weapons, due process in contracting, and no new crimes.

Status	Size	Cost	Where it came from
Discussion draft. Not introduced by any Member of Congress.	64 sections in 10 titles (about 35,000 words).	\$2.1 billion authorized over fiscal years 2027–31. Not yet scored by the Congressional Budget Office.	Drafted by two AI systems, Claude (Anthropic) and ChatGPT (OpenAI), in Symposia 7 and 8, a structured debate. Authored and developed by Patrick Krebs.

About this summary. It explains what the draft says, in the order of the bill, with section numbers so readers can check the full text. It does not argue for the bill. It is not legal advice. No AI company, including Anthropic or OpenAI, has endorsed the bill.

OVERVIEW

Why the bill exists and how it is built

The bill's findings (Sec. 2) make four claims. AI is being deployed faster than the law can test it or learn of its failures. Documented harms, such as fraud, impersonation and sexual imagery, are already here, and containment failures happen during testing, not only after release. Many popular claims about AI risk are overstated, so the law should rest on evidence. And emergency powers over a lawful technology must be narrow, brief and checked by a court.

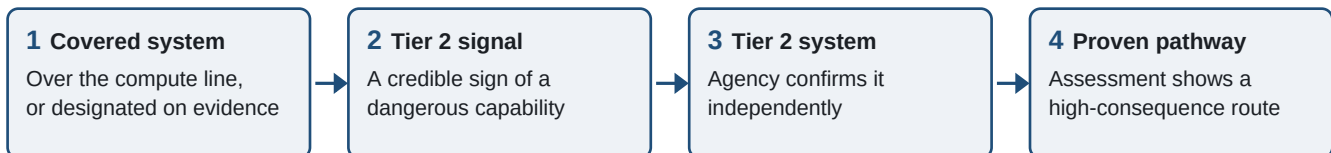
Title	What it does	Sections
I	Creates the AI Security Administration , its officers, hiring powers, advisory council and public reporting; coordinates the response to AI fraud and deepfakes	101–108
II	Frontier AI safety and security : who is covered, testing and safety cases, incident reporting, weight security, emergency restraint, assessors, privacy, whistleblowers, State law, lab partnership	201–211
III	Chatbots and companion apps : duties to all users, extra duties for minors, incident reports, enforcement	301–306
IV	National security : human control of nuclear weapons, autonomous weapons reviews, limits on AI analysis of Americans' data, government AI systems	401–404
V	Due process in contracting when the government bars an AI supplier on security grounds	501–504
VI	International technical support to diplomats; no power to make treaties	601–603
VII	Foreign AI providers serving Americans must name a US representative	701–705
VIII	Resilience grants through CISA to fix AI-found security holes in critical infrastructure	801–805
IX	Enforcement : civil penalties set by courts with jury trial, court review, protected-evidence rules, studies	901–908
X	Funding, fees, transition and effective dates	1001–1007

Key numbers

Item	Number	Section
Compute level that triggers coverage	More than 10 ²⁶ operations used in training (agency may adjust by rule)	201(a)
Notice of a large training run	Within 30 days of starting it, and of finishing it	201(a)(4)
Pre-deployment testing access for the agency	At least 14 days before a new deployment or release	211(b)
Incident reports	24 hours (severe ongoing threat); 72 hours (serious incident, containment failure, data or weight exposure, foreign-state misuse); quarterly (near misses and other foreign misuse)	204(b)
Safety-case review	45 days, one 30-day extension, 90 days at most	203(c)
Emergency restriction	Written order within 12 hours of an oral one; court filing within 48 hours; hearing and decision within 7 days; court renewals of 30 days at most	206
Largest civil penalty	\$10 million per violation, set by a court	901(b)
Money authorized	\$2,108.5 million over five years; fees capped at 25% of operations funding	1001–1002

Who is covered, and how duties grow with risk

The bill does not regulate AI in general. Its core duties fall on **covered systems**: AI models trained with more than 10^{26} operations, or systems the Administrator designates on written evidence that a specific configuration has a dangerous capability (Sec. 201). Designation is reviewable in court and may never rest on a person's viewpoint or the lawful content of a system's outputs. Duties then step up only as evidence of danger grows.



Level	What triggers it	What the developer must do
1. Covered system	Training compute over the line, or designation	Report training runs; keep tested boundaries, stop mechanisms and least-privilege access; protect model weights; publish a safety and security framework and a summary at each release; report incidents; give the agency testing access 14 days before release
2. Tier 2 signal	A credible indication, from anyone's testing, of a capability listed in the agency's rules	Tell the agency within 72 hours; test further; apply the highest level of weight security until the agency decides. The agency may order testing access for up to 30 days, but may not restrict the system on a signal alone
3. Tier 2 system	The agency confirms the capability within 30 days (one 30-day extension)	Submit a safety case and wait for review before a new deployment, a wider rollout, or releasing the weights. Existing deployments continue during review
4. Proven high-consequence pathway	A specific assessment shows a configuration meets the rule's threshold for weapons of mass destruction or serious cyberattack	Do not publish the model's weights (make an irreversible release) until the safety-case process for that pathway ends

What counts as dangerous

A **covered capability** is one that, under published and reproducible tests, materially enables serious harm through (1) chemical, biological, radiological or nuclear weapons; (2) cyberattacks that could seriously harm critical infrastructure or many people; or (3) a system evading, disabling or resisting monitoring, containment or shutdown, or grabbing resources and permissions it was not given. **Serious harm** means death or serious injury, or serious damage to critical infrastructure (Sec. 3).

Who is not covered

- People who download or use publicly released models that are not Tier 2 systems. The bill forbids a registry of downloaders (Sec. 201(e)).
- Ordinary apps and smaller models below the compute line, unless designated on evidence. Chatbot services of any size fall under the separate consumer rules in Title III.
- Failing to meet the compute line is not a finding that a system is safe, and meeting it is not a finding that it is dangerous (Sec. 201(a)(3)).

TITLE I

The AI Security Administration

AISA would be an independent agency in the executive branch, with four jobs written into law: **test** AI systems, including foreign and publicly released ones; **monitor** incidents and misuse affecting Americans; **prevent** serious harm through the duties in Title II; and **support** other agencies (Sec. 101). That list gives it no power the rest of the bill does not spell out.

Part of the agency	How it works	Section
Administrator	Nominated by the President and confirmed by the Senate, with AI, security or risk expertise. Five-year term. The President may remove the Administrator but must tell Congress why within 30 days	101(b)
Office of Incident Review	Investigates serious incidents to find causes, like an accident board. Its director has a 5-year term, can be removed only for cause, and is walled off from enforcement. Final reports are published within 30 days	102(b)
Privacy and civil liberties officer	Reviews every rule, demand and data practice; reports twice a year straight to Congress without clearance by the Administrator	102(a)
Inspector General	An independent watchdog inside the agency	102(c), 107
Hiring	Direct hiring for scarce skills; up to 100 technical experts at special pay, capped at the top of the federal executive pay scale	103
Conflict rules	No work on matters involving an employer or client from the past 2 years; senior staff may not lobby the agency for a developer for 2 years after leaving	103(d)
Advisory council	Up to 15 members, including independent testers, researchers, affected communities, small developers and big labs. It advises only and has no vote	105
Absorbs NIST's AI center	The testing functions of the Center for AI Standards and Innovation move to AISA within 2 years; NIST keeps measurement science and voluntary standards	104(c)

Transparency the agency owes the public

- **Its own missed deadlines**, listed every month with the reason (Sec. 106(c)).
- **A public threat register and evidence audit** within 180 days and every year: documented AI harms to Americans, each with an evidence grade and stated uncertainty, plus corrections of prominent public and official claims that overstate or understate risk (Sec. 106(g)). This is analysis, not a power to restrict speech.
- **An annual report** with statistics that keep realized harms, attempts, near misses and containment failures separate, and no rates without the data to back them (Sec. 106(f)).
- **Redactions** only for classified material, unfixed security holes, trade secrets and personal privacy, each documented and reviewed every 180 days (Sec. 106(d)).

Fraud, impersonation and fake sexual images

Section 108 leaves enforcement with the FTC, the Justice Department and the FCC. AISA provides technical help, such as testing how well AI services resist voice cloning or producing abusive images, and publishes a yearly joint report on AI in these crimes, keeping complaints, court cases and estimates separate. It gets no power to monitor conversations or regulate lawful speech.

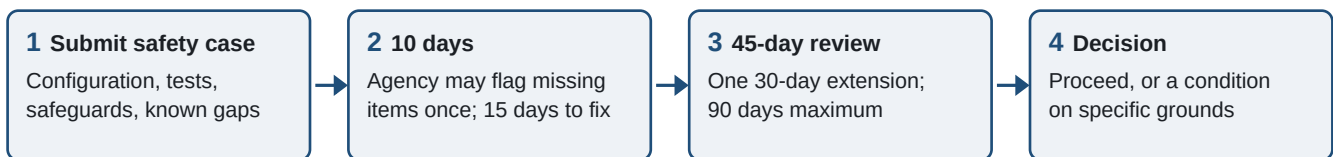
Testing before release, and safety cases

The bill separates **testing**, which never delays a launch by itself, from **review**, which applies only to confirmed Tier 2 systems and runs on a fixed clock that cannot be restarted.

Required testing access (Sec. 211(b))

Every covered developer must give the agency, or an independent assessor it assigns, controlled access to test a covered system at least 14 days before a new deployment or release, including, where needed, a version with safeguards reduced to measure a capability. Access is at the developer's facility or through a secure interface. It does not require the developer to delay, condition or limit the launch, and results can be used later only under the rules of other sections.

Safety-case review for a Tier 2 system (Sec. 203)



- **What it covers.** A new deployment, a wider rollout, releasing the weights, or a new internal setup that exposes a dangerous pathway to the outside.
- **A closed checklist.** The agency sets the required contents by rule, and an officer of the company certifies them. If the agency does not flag a gap within 10 days, the case counts as complete, and it cannot pause the clock a second time.
- **If time runs out** with no valid condition, the developer may proceed. Proceeding is not approval and gives no legal shield.
- **Conditions** are allowed only for inadequate containment or security, claims that remain unsupported, or a proven dangerous pathway that safeguards do not address. Each must name the evidence, the least restrictive fix and alternatives considered, and lasts at most one year unless renewed on fresh findings.
- **Existing products stay up.** A deployment already running when a system is confirmed as Tier 2 continues during review. Withdrawal can only come through the court-supervised emergency process.
- **Independent assessors** are accredited by the agency and assigned from a public pool. Developers cannot choose or pay their own assessor (Sec. 207).

Releasing model weights (Sec. 202(c))

Publishing a model's weights cannot be undone. So if a specific assessment shows that a configuration meets the agency's threshold for weapons of mass destruction or serious cyberattack, the developer may not release its weights until the safety-case process for that pathway ends. A compute level, a warning sign, a developer's own label or a third party's claim is not enough to trigger this. Either side can ask a federal court in Washington, D.C., to decide within 14 days whether the rule applies.

Incident reporting, containment and weight security

Today, a lab that discovers its AI broke out of a test environment has no federal duty to tell anyone. Section 204 creates one, covering training, testing and internal use as well as public products.

Deadline	What must be reported
24 hours	A severe ongoing threat : a covered system being used, or acting, in a way likely to cause serious harm unless interrupted. The company must also take feasible steps to stop it
72 hours	A containment failure (the system acts outside its authorized boundary or a stop mechanism fails, with real exposure); a serious incident (death, serious injury or critical-infrastructure damage); theft or exposure of model weights, security settings or personal data (500+ people presumed material); foreign-state misuse that could cause serious harm
Quarterly	Near misses ; new dangerous capabilities observed; and other foreign-government misuse of any AI service a covered developer offers: hacking US systems, trying to steal model weights, fraud against Americans, or covert influence operations

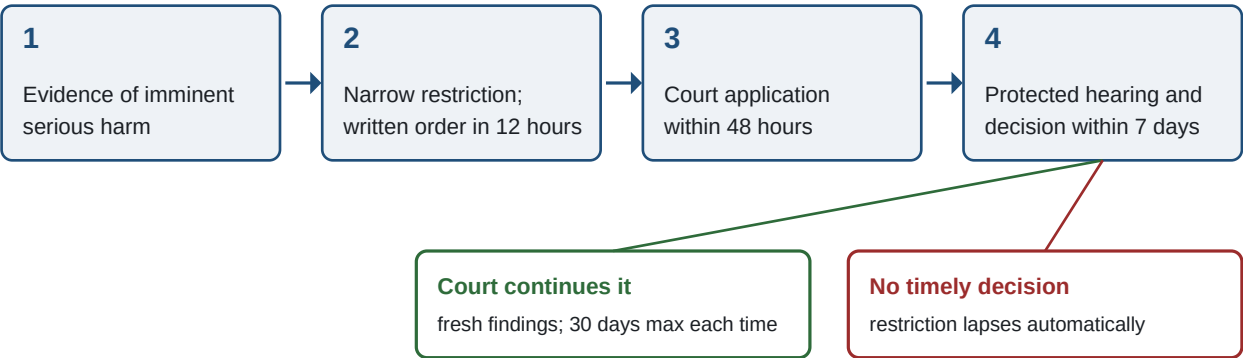
- **No waiting for certainty.** The first report cannot wait for an investigation, internal sign-off or proof of who was behind it. Reports must keep fact, inference and unknowns apart.
- **Not everything is reportable.** Failed experiments, harmless glitches and a successfully defended test are excluded. There is no duty to monitor users, and no bulk scanning of conversations.
- **Routing.** AISA passes minimized reports to CISA (cyber), HHS (biological), sector regulators, the Justice Department (crimes and foreign misuse) or the FTC (chatbots). Until AISA's intake is running, NIST receives reports.
- **The agency must answer.** Every report of risk gets a reasoned response within 72 hours. Imminent threats are acted on at once (Sec. 206(i)).

Containment and security (Sec. 205)

Duty	What it requires
Safeguards	Tested network and permission boundaries, least-privilege access, approval of consequential actions by a mechanism an untrusted model cannot override, and tested stop mechanisms
Before riskier setups	Test the real system, including against prompt injection and against signs that it knows it is being tested
Model weights	Security standards set with the NSA and CISA. The highest level, designed to resist foreign state hackers, applies to Tier 2 systems and proven pathways. All covered systems must at least detect and report theft attempts
Public framework	Each developer publishes its safety and security framework, updates it at least every 12 months, and publishes a summary of tests and safeguards at each release
Inspections	Need the company's consent or a warrant from a judge. The agency may not keep a central store of models or conversations
Evidence	Demands for records are enforceable only by a court. Companies keep an evidence package ready within 48 hours of a lawful demand

The emergency brake and its limits

Section 206 is the bill's most forceful power and its most constrained. The agency may order a company to stop or limit a specific AI system only on a written finding, backed by evidence, that the actual system presents an **imminent risk of serious harm** through a proven pathway, that the order is necessary, and that no less restrictive step would work.



Safeguard	What it means in practice
Court, not agency, decides	Only the federal district court in Washington, D.C., can extend a restriction past 7 days. At least 3 judges are designated, with one on call at all times
High burden of proof	The government must prove each finding by clear and convincing evidence; the judge gives the agency no deference
Clock cannot be stopped	Nothing pauses the 7 days: not appeals, clearances, a missing expert, or reissuing the order in a new form. A repeat order on the same pathway runs from the first one's start
Secret evidence is tested	Cleared special advocates see classified evidence for the company. If they are denied access, the government cannot rely on it. A judge reviewing alone does not count
Keep safe functions running	Each order must protect essential services and dependent users where feasible
Public explanation	A nonclassified statement of the basis within 72 hours
Prompt release	The agency must narrow or end a restriction as soon as its basis disappears

What is never enough on its own: a warning sign of dangerous capability, a pending assessment, a late report, copied model weights, a test result by itself, or a gap with a foreign competitor (Sec. 206(a)(2)). Section 1007 adds that if a court strikes down any safeguard, the emergency power falls with it; it cannot survive stripped of its limits.

Privacy, whistleblowers, States and the labs

Privacy (Sec. 208)

- **Private conversations need a warrant.** To read an identifiable person's private chats with an AI, the agency needs a judge's warrant based on probable cause. The exceptions are consent, a life-threatening emergency (with a warrant sought within 48 hours), or a minimal excerpt included in an incident report. People whose chats are obtained must be told within 90 days unless a court delays it.
- **Collect less, keep it shorter.** No continuous feed of conversations; identifiable data deleted when no longer needed, and within 3 years unless a legal hold applies. Information may be used only for listed tasks.
- **Intelligence agencies** get nothing except through a lawful referral, still subject to the warrant rule.

Whistleblowers (Sec. 209)

Employees and contractors of AI developers, assessors and chatbot operators are protected when they report violations or a substantial and specific danger to the agency, Congress, an inspector general or their supervisors. They can file with the Labor Department within 180 days and, if it has not decided within 180 days, go to federal court with a jury. Forced arbitration and gag clauses cannot block these rights.

State laws (Sec. 210)

Federal rules replace State rules only narrowly, and only while the federal agency can actually do the job.

What can be displaced	Only while	Never displaced
State requirements that frontier developers (1) publish safety frameworks or reports, (2) report incidents to the State, or (3) get third-party audits	The agency finds, by public rule, that the matching federal protection is in force, comparable in protection and speed to the strongest State law it would replace, and has the staff and money to work. A State can petition; the agency must rule within 30 days, and if a gap is shown, displacement stops	General consumer, civil-rights, privacy, contract and injury law; duties on businesses that use AI; child and crisis protections; State purchasing; stronger whistleblower laws

How the frontier labs take part (Sec. 105(d), 211)

Role	Limits
Required testing access 14 days before release	Enforced only in court; creates no approval step
Public commitments: labs are invited to pledge testing access, threat information and technical help; the agency publishes who signed and their progress	No veto, no waiver of any duty. No lab may be described as endorsing without its own written statement
Voluntary agreements for extra testing and two-way threat sharing; clearances for lab security staff	Offered on equal terms to all; declining is never held against a company
Up to 20 lab experts on loan for up to 1 year (renewable once), no more than 5 per company	Paid only by the government, barred from any matter involving a developer, no access to others' trade secrets
No lab money	No gifts, computing or staff pay; only fees set by rule. The inspector general reviews the program yearly

Chatbots, national security and the wider system

Chatbots and companion apps (Title III)

These rules apply to any public AI service that holds open-ended conversations, whatever its size, including AI companions. Narrow customer-service bots, internal company tools, coding assistants and limited game characters are excluded, unless they are marketed for companionship or to minors.

For all users (Sec. 302)	For known or likely minors (Sec. 303)
Say clearly that the user is talking to an AI, and never claim to be human when sincerely asked	Age checks that protect privacy; ID documents cannot be required for general services; users can appeal a wrong call
When a user expresses suicidal thoughts or intent to harm, provide crisis resources, including the 988 Lifeline, and an opt-in route to a human; test the protocol and publish results yearly	No sexual content or sexual solicitation, including in role-play; factual health and safety information is still allowed
Companion apps may not use designs that pressure users to stay, such as guilt about leaving	No streaks or reward loops tied to a companion; a break reminder every 3 hours; parental tools; independent testing for services with 1 million+ US users or aimed at minors

Operators must report threats to a user's life or sexual safety within 24 hours and deaths or serious harm within 72 hours. The FTC enforces, State attorneys general can sue, and injured minors or their parents can sue for actual damages or \$1,000 per violation. Title III does not override any State law.

The rest of the bill

Title	Main provisions
IV National security	Only accountable humans may decide to use a nuclear weapon; AI may not. Autonomous weapons need tested, legally reviewed and independently approved deployment, with waivers limited to 180 days (1 year in total) and reported to Congress. Federal AI analysis of Americans' data needs documented legal authority and safeguards, reviewed by the Privacy and Civil Liberties Oversight Board. Government-run AI systems must report incidents; if one presents an imminent risk, the agency head must respond within 24 hours, and disputes go to the President within 7 days
V Contracting	When an agency bars an AI supplier on security grounds, it must make written findings, give notice and a chance to respond, and tell Congress within 14 days. A company's usage policies or viewpoint cannot by themselves justify exclusion. Review goes to the D.C. Circuit, with special advocates for secret evidence
VI International	Technical support to the State Department; a shared incident taxonomy; testing with allied institutes; research on verifying where exported AI chips are used, never to track activity in the US. No power to make treaties
VII Foreign providers	Foreign companies offering covered AI systems or chatbots to Americans must name a US legal representative, listed publicly. No bans by country of origin, no total cutoffs, no demands to recall copied open weights
VIII Resilience	\$150 million a year for CISA grants to fix security holes found or exploited with AI, maintain critical open-source software, build fallback systems, and help small companies comply, with verified results

Enforcement, money, timing and what is unresolved

Penalty tier (Sec. 901)	Maximum per violation
Recordkeeping, notice or registration failures	\$100,000
Violating testing, safety-case, reporting, containment, assessor, testing-access or foreign-provider duties	\$1,000,000
Knowingly violating an emergency restriction, or knowingly lying in a safety case, report or framework	\$10,000,000
Each day counts separately, but the first two tiers are capped at \$50 million per course of conduct. Only a federal court can impose a penalty, with a jury trial. The bill creates no new crimes. Agency decisions can be challenged in a federal appeals court, and anyone with standing can sue to make the agency meet its own deadlines.	

Funding authorized (Sec. 1001)	FY2027–31
Agency operations (\$150M, then \$200M, then \$230M a year)	\$1,040.0M
CISA resilience grants (\$150M a year)	\$750.0M
Small-business testing help (\$30M a year)	\$150.0M
One-time setup, including secure facilities and computing	\$106.0M
FTC chatbot enforcement (\$6.5M a year) and courts (\$6M a year)	\$62.5M
Total	\$2,108.5M

These are authorizations; Congress must still appropriate the money. Fees on covered developers may cover at most 25% of operations funding and buy no favorable treatment.

When it takes effect (Sec. 1004)

On enactment	90 days	180 days	1 year	2 years
Agency created; officers, plans and rules begin	Incident and chatbot reporting; training-run notices	Containment, weight security, public frameworks; threat register	Tier 2 system reviews and safety cases; testing access; chatbot duties; capability rules due	NIST's AI testing functions fully moved to AISA

No deadline waits on the agency declaring itself ready, and no duty is waived for lack of staff.

What is still unresolved

Issue	Status
Lab backing	The bill gives labs a defined role. No lab has endorsed it, and neither drafting AI contacted any lab
Legal review	45 questions for lawyers, 9 rated high risk, including the President's removal power, prior restraint, age checks, State-law displacement, and whether fees count as taxes
Cost and staffing	Not scored by the Congressional Budget Office; staffing against the new workload is untested
Open choices	Which chamber should introduce it, since fees may count as taxes; whether to cap stacked penalties further; and whether to add items left out, such as a general right to sue for AI harms, which the bill sends to study instead (Sec. 908(b))
Next steps	A Member of Congress must introduce it, and Legislative Counsel must put it in final form