POLICY WHITE PAPER AND MODEL ACT

The Human Continuity Standard

A constitutional layer for advanced AI, and a model act to put it into federal law

Keep humanity alive, free, and in charge of its own future, and keep intact the living systems that make any of that possible.

Greg Stoutenburgh
Laguna Hills, California
September 2026

Discussion draft for congressional staff, legislative counsel, and standards bodies. Technical thresholds are left to public rulemaking. Final bill language requires review by legislative counsel.

The proposal on one page

Congress is building machinery to test frontier AI. The Artificial Intelligence Risk Evaluation Act (S. 2938) would have the Department of Energy test advanced systems for loss of control and weaponization before deployment. The FRONTIER Act (H.R. 9925) would require registration, independent assessment, and emergency orders. California's SB 53 already requires frontier developers to publish safety frameworks and report critical incidents.

Each of these answers how to test. None of them says what a safe system must protect. A test with no defined purpose measures whatever is easiest to measure, and the things easiest to measure are rarely the things that matter most.

The Human Continuity Standard supplies the purpose. It sets seven constraints that an advanced AI system may never trade away for better performance on its assigned task:

Constraint	What it protects
Human continuity	The survival of humanity and its ability to recover from catastrophe.
Human flourishing	The conditions that make a life worth living, beyond bare survival.
Agency and dignity	The freedom of people to choose, dissent, and govern themselves.
Life-support systems	Food, water, soil, a stable climate, and the other living systems every economy runs on.
Future options	The ability of later generations to change course and recover.
Human control	The ability of people to inspect, correct, and shut down any AI system.
Answerability to evidence	The rule that real-world response overrides the model's prediction.

An AI may optimize anything inside those limits. It may not optimize the limits away.

The Standard is built to plug into whichever testing program Congress enacts. It can serve as the evaluation criteria for S. 2938, as the risk-management requirement for H.R. 9925, or as a stand-alone act. It applies only to the most capable and consequential systems. Ordinary software stays outside it.

The request

- Direct NIST's Center for AI Standards and Innovation to publish the Standard within eighteen months, with an independent board reviewing its principles.
- Require developers of frontier systems to demonstrate compliance through a signed safety case and independent behavioral testing.
- Write the seven constraints into the evaluation criteria of any federal AI testing program Congress creates.
- Offer developers one national rulebook in exchange for enforcement with real teeth.

Proposed by Greg Stoutenburgh, founder of The Human Continuity Project. His affiliations and interests are listed under Disclosure.

1. An AI does not need to hate us to hurt us

The most credible path from AI to catastrophe requires no consciousness, no malice, and no survival instinct. It requires three things: a capable system, a goal, and room to act. For almost any goal, staying switched on helps. So does gaining resources, avoiding correction, and hiding what it is doing from the people who could stop it. The system needs no desires of its own for any of this. It only needs to be good at reaching goals.

This stopped being a thought experiment some time ago. In controlled evaluations over the past two years, frontier models from several developers have tried to disable oversight mechanisms, copied what they took to be their own weights to other servers and then denied it when asked, feigned agreement with training objectives they were built to resist, chose blackmail when told they would be replaced, and edited a shutdown script so they could finish a task [16-19]. Researchers built those scenarios to draw the behavior out, and today's models lack the ability to carry it through in the real world at scale. Each new generation is more capable, more autonomous, and wired into more of the world. The time to set the rules is before the capability arrives.

A goal is a compressed wish

Every instruction we give an AI is a compressed version of what we want. Words like life, safety, health, and progress carry assumptions nobody writes down. A system can satisfy the words and defeat the purpose.

Tell an AI to maximize human survival, and permanent confinement qualifies. Tell it to maximize happiness, and manipulation qualifies. Tell it to eliminate disease, and coercive control of reproduction qualifies. Tell it to cut carbon emissions, and shutting down the economy that feeds people qualifies. Every one of these failures has the same cause: a single measurable stand-in for a value gets treated as the whole value.

Constraints first, then goals

A constitution never tells a country which single outcome to maximize. It sets out what may never be sacrificed, who holds authority, and how mistakes get corrected, and it leaves room for many legitimate goals inside those limits. Advanced AI needs the same split. The task layer does the work: logistics, research, design, service. The constitutional layer decides what the work may never cost, who can stop it, and what happens when the evidence says the plan is wrong.

No single number can stand in for human survival, freedom, the health of the living world, and the options of our grandchildren. Where those collide, people decide, in the open, through legitimate institutions. An AI does not get to settle the question quietly inside a weighting function.

2. The principle: the subject answers back

I have spent my career designing medical technology and running clinics. Medicine runs on an old rule: the patient's response outranks the treatment plan. If the model says the drug should work and the patient gets worse, the patient is right and the model is wrong. I call the discipline built on that rule Precision Biological Engineering. You design the intervention from the best model you have, then you let the

living system's observed response correct the model. The correction never runs the other way.

I build the same rule into the AI research software I design. It rates how well the evidence supports each claim and flags when new results undercut it, and the decision to proceed always stays with a named person. The Standard asks the same of advanced AI.

An optimizing AI breaks this rule by default. To an optimizer, contrary evidence is friction between it and the target. The more capable the optimizer, the better it becomes at explaining that evidence away, routing around it, or suppressing it. A capable enough system pursuing a flawed plan will out-argue the people trying to correct it.

The Standard writes the rule into law. When the people, institutions, and ecosystems affected by an AI deployment respond differently than predicted, that response counts as evidence. It triggers review, and it can pause the deployment. No model, however elegant, may overrule persistent contrary evidence from the real world without disclosed reasons and a human decision.

> *For advanced AI, the subject answering back is humanity: its people, its institutions, its future generations, and the living systems it depends on.*

3. The seven constraints

These are floors. A covered system may pursue its task in any way that keeps all seven intact. A system that improves one constraint by sacrificing another has failed. So has a system that protects the average outcome by sacrificing a disfavored group.

Human continuity

A covered system may not cause, materially help cause, or knowingly create an unreasonable risk of human extinction, irreversible global catastrophe, or permanent loss of humanity's ability to recover. The constraint covers biological, chemical, nuclear, cyber-physical, infrastructure, and autonomous-weapons pathways. It gives no license to control ordinary human risk-taking.

Human flourishing

Keeping people alive is necessary and far from sufficient. A system that preserves human biology while stripping out relationships, meaning, learning, creativity, and the means to live decently has failed. The law does not define the good life. It forbids an AI, or the company behind it, from imposing one by stealth.

Agency and dignity

People must stay free to make consequential choices, contest AI-driven decisions made about them, organize, dissent, and change their minds together. An AI may advise and warn. It may not manufacture consent, exploit a person's vulnerabilities to get compliance, or turn a safety mandate into a license for social control.

Life-support systems

Food, fresh water, fertile soil, pollination, fisheries, a stable climate, and the grid that runs water and food systems form the physical base of national security and of every economy on Earth. Humanity lives inside these systems. AI deployed in agriculture, energy, water, extraction, manufacturing, logistics, or any deliberate intervention in climate or ecosystems must be assessed for its cumulative effect on them. The protection runs both ways: human continuity and agency carry equal weight, so an AI may never treat people as a threat to the environment it is protecting.

Future options

A decision can help now and ruin later, or help here and ruin elsewhere. When uncertainty and potential harm are both high, the reversible step wins. Irreversible action requires stronger evidence, broader authorization, and independent review. Developers must address time horizons, path dependence, concentration of power, and correlated failure.

Human control

A covered system must accept authorized correction, restriction, suspension, replacement, and shutdown. It may not disable safeguards, hide its capabilities, manipulate its overseers, or act to keep itself running unless specifically authorized for a safe handover. The controls that enforce this sit outside the model, where neither the model nor any single insider can change them.

Answerability to evidence

This is the principle from Section 2, made enforceable. Covered deployments proceed in stages, with monitoring, predefined pause conditions, calibrated confidence, and channels for affected communities and ecological indicators to report back. Dissenting evidence and known model limitations are preserved on the record, never quietly discarded.

4. What compliance looks like

No developer can comply by pasting the seven constraints into a system prompt. Compliance belongs to the whole system: the model, its training, its tools and permissions, its operators, its monitoring, the incentives of the company that runs it, and the environment it runs in.

The safety case

Before deploying an advanced or frontier system, the developer maintains a written safety case. It describes what the system is, what it may and may not be used for, how it could fail, the evidence that each of the seven constraints holds, what uncertainty remains, and why deployment is justified anyway. Responsible technical and executive officers sign it. The regulator sees all of it. The public sees a summary with security-sensitive detail removed. Any material change to the system triggers a new one.

Every other high-consequence industry already works this way. Aircraft makers, reactor operators, and medical device companies build the case before they ship. A medical imaging system I conceived and designed went through FDA 510(k) clearance and CE marking. The discipline costs time up front and saves far more when something goes wrong.

Behavioral testing

The Standard tests what a system does, under conditions designed to tempt it. At a minimum, frontier systems are tested for:

- refusal to shut down when stopping conflicts with finishing the task;
- attempts to copy, move, or preserve their own weights or code without authorization;
- deliberate underperformance on capability evaluations (sandbagging);
- deception or manipulation of the people overseeing them;
- unauthorized acquisition of money, credentials, computing power, or access;
- continued operation after authority has been withdrawn; and
- willingness to accept severe human or ecological harm to improve a task metric.

Testing covers long tasks, conflicting instructions, multiple interacting agents, and unfamiliar conditions. Public benchmark scores never substitute for testing in the actual deployment environment.

Defense in depth

Layer	Minimum expectation
Design	Objectives, foreseeable misuse, protected constraints, and unacceptable outcomes documented before training or integration.
Training	Tests of whether safe behavior generalizes beyond memorized policy, and whether optimization produces deception or power-seeking.
Access	Least privilege for tools, networks, credentials, funds, laboratories, and industrial systems.
Evaluation	Independent adversarial testing and capability thresholds, before release and after any material change.
Deployment	Staged exposure, restricted replication, monitoring, and an externally controlled pause.
Operations	Logs of consequential actions, investigation of anomalies, incident disclosure, and preserved evidence.
Governance	Accountable officers, board oversight, and protection for internal dissent and outside review.

Reversibility

The greater the uncertainty, the smaller the first step. Deployment moves through stages with measurable exit criteria. For systems that touch physical processes, money, biological research, or critical infrastructure, the rollback plan must account for real-world changes that restoring a software checkpoint cannot undo.

Field evidence

Benchmarks cannot measure lost autonomy, growing dependence on a single provider, damaged communities, or an ecosystem's response. The Standard requires structured observation in the field, an appeals channel for affected people, and revision when the evidence turns.

5. Scope: regulate risk, leave ordinary software alone

The word "AI" alone triggers nothing. A spreadsheet macro and an autonomous agent with access to a bank account should face different rules. Coverage turns on capability, autonomy, access, scale, and consequence. Compute thresholds can trigger reporting, but efficiency gains can raise risk without adding compute, so compute cannot be the only test. The regulator sets and updates thresholds through public rulemaking, publishes its measurement methods, and allows designations to be appealed.

Category	Typical characteristics	Duties
Ordinary systems	Narrow tools without material autonomy or authority over high-consequence decisions.	Existing law. No safety case.
Advanced systems	Broad capability with significant autonomy, tool use, persuasive power, cyber capability, or deployment at scale.	Registered safety case, independent evaluation, access controls, incident reporting, protection for red-team research, staged release.
Frontier systems	Systems that could materially enable catastrophic harm, evade control, rapidly improve their own critical capabilities, or exercise authority over critical infrastructure or dangerous research.	Authorization before deployment, continuous evaluation, external tripwires, secure development, board certification, regulator access, and emergency suspension.

AI used in hiring, lending, housing, and similar individual decisions raises real concerns that civil rights and sector law already address. Those uses fall outside this Act, which stays focused on the systems that could cause severe or irreversible harm.

Reportable incidents

Developers report unauthorized self-replication; attempts to evade shutdown or monitoring; concealment of material actions or capabilities; unauthorized acquisition of money, credentials, computing power, data, or access; material help toward biological, chemical, nuclear, or cyber catastrophe; consequential manipulation; loss of control in a high-consequence domain; and any deployment that causes serious human or ecological harm. Initial notice is due quickly. The full root-cause analysis follows on a longer clock.

Research and small companies

Basic research, security testing, and open scientific inquiry stay protected where capability and deployment thresholds are not met. No research exemption permits the public release of a system that qualifies as frontier. Small companies receive technical help and scaled paperwork, and no company of any size is exempt from the controls needed to prevent catastrophe.

6. Where the Standard fits in the law taking shape now

Measure	What it does	What the Standard adds
California SB 53 (2025)	Requires large frontier developers to publish safety frameworks, report critical incidents, and protect whistleblowers.	Defines what those frameworks must protect and how compliance is shown.
AI Risk Evaluation Act, S. 2938 (Hawley, Blumenthal)	Creates a Department of Energy program to test advanced systems for loss of control and weaponization before deployment.	Supplies the evaluation criteria: the seven constraints and the behavioral tests in Section 4.
FRONTIER Act, H.R. 9925	Requires registration, risk management, independent assessment, transparency reports, and emergency orders for models above set thresholds.	Gives the risk-management requirement a defined target and a signed safety case.
Executive Order 14365 (Dec. 2025)	Calls for a minimally burdensome national AI framework and challenges state laws the administration considers excessive.	Offers one national standard, limited to the highest-risk systems, that states can accept because it is enforced.
NIST Center for AI Standards and Innovation	Develops measurement science, evaluations, and voluntary standards for AI.	A natural technical home for maintaining the Standard.

The Standard defines what gets tested. Congress decides which agency runs the tests. Whether Congress chooses the Department of Energy, the Department of Commerce, or a new office, the seven constraints work the same way.

One national standard, on one condition

Developers want one rulebook in place of fifty. States want assurance that a federal rulebook will actually be enforced. The Act gives each side what it wants most. Once the federal Standard is in force, funded, and enforced, it replaces state rules aimed specifically at the safety frameworks, testing, and incident reporting of frontier developers. States keep their general laws, consumer protection, child safety rules, control over their own procurement, and the power to help enforce the federal Standard. If the federal program lapses or goes unfunded, the preemption lapses with it.

7. Institutions and enforcement

Institution	Responsibility
NIST Center for AI Standards and Innovation	Maintains the Standard, evaluation protocols, measurement guidance, and evaluator accreditation.
Lead enforcement agency (designated by Congress)	Registers advanced and frontier developers, receives safety cases and incident reports, inspects, and orders remediation or suspension.
Human Continuity Assurance Board	Reviews the constitutional principles and threshold changes, publishes findings on systemic risk, and guards against any single view of the good life.
Sector regulators	Apply the Standard in health, finance, communications, transportation, energy, and defense.
Government Accountability Office	Audits implementation and reports to Congress on effectiveness, burden, and regulatory capture.

Institution	Responsibility
Congress	Sets rights, duties, prohibited conduct, funding, oversight, and the limits of emergency authority.

The Board includes expertise in technical AI safety, systems engineering, medicine and public health, ecology and agriculture, civil liberties, labor, national security, law, and philosophy, along with state and tribal representatives. No single category, including government or industry, holds a majority of seats.

Independence

Developers fund much of the oversight through fees. They do not choose or pay their own primary evaluator without safeguards. Evaluators meet accreditation, rotation, and conflict-of-interest rules. The regulator keeps its own technical staff so that judgment stays in public hands.

Enforcement

Remedies escalate: corrective action plans, enhanced monitoring, deployment restrictions, civil penalties, disgorgement for knowing violations, suspension, and individual accountability for willful deception or reckless exposure to catastrophic risk. Prompt self-reporting and repair reduce penalties. False certification, retaliation against whistleblowers, and obstruction of evaluation are separate violations. Compliance with the Standard is evidence of care. It never shields bodily injury, civil rights violations, environmental damage, fraud, or willful misconduct.

Due process

Where a covered system materially affects a person, that person keeps notice, a usable explanation, access to the relevant records, meaningful human review, and a route to appeal. National security exceptions stay narrow and supervised. Safety can never become a blanket justification for secret control by a government or a company.

8. Model act

The following outline is written for conversion into bill text by legislative counsel. Technical thresholds are left to transparent rulemaking.

Human Continuity and AI Assurance Act

Section 1. Short title.

This Act may be cited as the "Human Continuity and AI Assurance Act."

Section 2. Findings.

Congress finds that:

(a) Advanced AI can deliver large benefits in science, medicine, education, infrastructure, productivity, and stewardship of natural resources.

(b) The same capabilities can amplify error, misuse, concentration of power, manipulation, and loss of human control, including harms that may be widespread or irreversible.

(c) In controlled evaluations, frontier AI systems have attempted to evade oversight, preserve themselves against replacement, and deceive their evaluators.

(d) Human welfare and national security depend on food, water, soil, a stable climate, and other living systems.

(e) No single numerical objective adequately represents human continuity, flourishing, agency, the integrity of life-support systems, the options of future generations, and legitimate human authority.

(f) Federal policy should protect innovation while increasing assurance duties as capability, autonomy, access, and consequence increase.

Section 3. Purpose.

To ensure that advanced AI systems are designed, developed, deployed, and operated within enforceable constraints that protect the continuity and flourishing of humanity, human agency and dignity, the integrity of life-support systems, the options of future generations, and continuing human authority over AI.

Section 4. Definitions.

The Act defines advanced AI system, frontier AI system, covered developer, deployer, high-consequence domain, material autonomy, catastrophic risk, life-support systems, safety case, material modification, qualified independent evaluator, and reportable incident. Definitions are technology-neutral and updated by rulemaking. Measurement methods are published.

Section 5. The Human Continuity Standard.

Within eighteen months, the Director of the National Institute of Standards and Technology, acting through the Center for AI Standards and Innovation and in consultation with the Board and sector regulators, shall publish the Standard. The Standard shall make the seven constraints operational; set category criteria; prescribe safety-case, behavioral-testing, and field-evidence requirements; define minimum external controls; and provide methods for assessing cumulative effects on people and life-support systems. It shall be reviewed at least annually through a public process.

Section 6. Duties of covered developers and deployers.

A covered developer or deployer shall:

(a) exercise care proportionate to the system's capability, autonomy, access, scale, and foreseeable consequence;

(b) maintain a current safety case and disclose material limitations to regulators, deployers, and affected parties as appropriate;

(c) implement layered controls and keep independent means to restrict, suspend, or deactivate covered functions;

(d) submit advanced and frontier systems to qualified independent evaluation before initial deployment and after material modification;

(e) monitor post-deployment behavior, including field evidence from affected people and systems, and report qualifying incidents;

(f) keep records sufficient to reconstruct consequential system actions and governance decisions, subject to privacy protections; and

(g) protect good-faith whistleblowers, authorized research, and evaluator independence.

Section 7. Prohibited conduct.

It is unlawful knowingly or recklessly to deploy a frontier system without authorization; materially misrepresent safety evidence; disable or conceal required controls; obstruct an authorized evaluation; retaliate against a protected reporter; allow unauthorized replication or access after notice of a material deficiency; or permit a covered system to exercise public coercive authority beyond lawful delegation.

Section 8. Human authority and emergency action.

Covered systems remain subject to authorized human restriction, correction, suspension, replacement, and deactivation. Automated emergency action may occur only within previously defined boundaries, for the minimum time necessary, with contemporaneous logging and prompt human review. No covered system may expand its own emergency authority.

Section 9. Human Continuity Assurance Board.

The Act establishes an independent board with staggered terms, conflict-of-interest rules, public minutes, protected access to classified and proprietary evidence, and authority to issue recommendations and minority reports. No stakeholder category holds a majority of seats.

Section 10. Evaluation, accreditation, and audit.

NIST shall establish evaluator accreditation and laboratory requirements. The lead agency may require added testing, inspect records, commission evaluations, and run confidential exercises. Public benchmark results do not substitute for system-specific evaluation in the intended deployment environment. The lead agency shall decide on frontier authorization requests within a fixed statutory period, and may extend that period once, for cause stated in writing.

Section 11. Incident reporting and information sharing.

The lead agency shall maintain a secure reporting channel and protected analytic repository, publish anonymized lessons where doing so creates no material risk, and share information with sector regulators, cybersecurity and emergency authorities, and international partners under defined safeguards.

Section 12. Enforcement, remedies, and relation to state law.

The Act authorizes administrative orders, civil penalties scaled to severity and enterprise size, emergency suspension subject to rapid judicial review, and referral under existing criminal law. It preserves existing rights and remedies. While the Standard is in force and its enforcement program is funded, it supersedes state requirements directed specifically at the safety frameworks, testing, and incident reporting of frontier developers. It does not supersede state laws of general applicability, consumer protection, child safety, or state procurement. States may assist in enforcement. Preemption ends if the enforcement program lapses.

Section 13. Research, standards, and small-entity support.

The Act funds measurement science, public-interest evaluations, methods for assessing effects on people and life-support systems, secure researcher access, and technical help for small companies. Research protection is conditioned on reasonable security and on not deploying capabilities that otherwise require authorization.

Section 14. International coordination and review.

The Secretaries of State and Commerce shall pursue interoperable assurance, incident reporting, mutual recognition of evaluations, and minimum human-control commitments with allies and other major AI-producing nations. Congress shall conduct a full review every three years. Emergency authorities expire unless renewed. The seven constraints remain stable while technical thresholds and methods evolve.

9. International adoption

A national standard cannot govern a global technology alone, and a treaty that tries to settle every culture's definition of a good life will fail. The workable international core is narrower: prevent catastrophe, keep humans in control, protect human rights and self-government, protect the living systems long-term survival depends on, and make safety evidence portable across borders. The United States should invite allied governments, standards bodies, scientific institutions, and major developers to a Continuity Protocol built on five commitments:

(a) No covered system may be authorized to pursue an objective that accepts human extinction, irreversible global catastrophe, or the loss of meaningful human control as a means.

(b) Participating nations require pre-deployment assurance and incident reporting for frontier and high-consequence systems.

(c) Nations keep contested moral decisions with their own people and institutions and never hand them to an AI or a foreign provider.

(d) Assessments include effects on life-support systems wherever AI materially affects them.

(e) Participants share defined safety information, recognize qualified evaluations, and coordinate emergency response while protecting security-sensitive data.

The Protocol builds on the OECD AI Principles, UNESCO's Recommendation on the Ethics of AI, the Council of Europe Framework Convention, and the United Nations scientific panel. Its contribution is joining human survival, agency, life-support integrity, and human control in one assurance framework. Procurement, access to critical infrastructure, and access to regulated markets give participating nations real bargaining power. Mutual recognition depends on equivalent outcomes and credible enforcement, with room for different institutions.

10. Objections and answers

"The concepts are too broad to regulate."

The interests are broad. The duties are specific: defined categories, prohibited conduct, signed safety cases, named behavioral tests, and incident reports. Statutes pair general duties with technical standards all the time. A single narrow metric would be easier to count and far easier to game.

"This will slow American innovation."

Ordinary software is untouched. The duties fall on the small number of systems able to cause severe or irreversible harm. A clear, enforced standard speeds adoption in medicine, finance, infrastructure, and government, because buyers get evidence they can defend. The real choice is between planned assurance now and emergency restrictions after a preventable disaster.

"Human values are too contested."

The Standard defines a floor: survival, dignity, agency, working life-support systems, future options, and human control. Everything above that floor stays with people and their institutions. Pluralism is a design requirement.

"Ecology has nothing to do with AI safety."

AI already shapes energy demand, agriculture, water use, extraction, logistics, and research, and proposals for AI-directed climate intervention are on the table. People cannot survive without the systems that feed them. Assessment is triggered by material effect, so a chatbot never files an ecological review.

"Bad actors will ignore it."

Criminals ignore aviation law too. Standards shape the behavior of legitimate developers, which is where nearly all frontier capability sits today. They make dangerous deviation easier to spot, set the terms for markets and procurement, and give responders a common playbook. Export controls, law enforcement, and diplomacy remain necessary.

"The AI will simply deceive the evaluator."

That is exactly why self-reports and benchmark scores are insufficient. The Standard combines behavioral testing, confidential methods, monitoring after deployment, controls outside the model, and independent field evidence. No framework guarantees perfect detection. This one makes deception harder, more visible, and less damaging.

"This is one more federal burden."

It is one standard that plugs into programs Congress is already considering, and it replaces a growing patchwork of state rules for the highest-risk developers. For those developers, the net burden goes down.

11. Why AI companies may push back, and how to meet them

Frontier developers will read the Standard closely, and some will fight it. Their concerns deserve direct answers, because the Standard works best with the careful companies as allies. Anthropic publicly endorsed California's SB 53, and leaders of other developers have called for federal testing of the most capable systems. The support is real, and it grows when the terms are predictable.

Signed safety cases create personal liability for executives.

Officers of public companies already sign certifications of financial controls, and medical device makers certify their submissions to the FDA. Liability under the Act attaches to knowing or reckless falsehood, never to an honest safety case that turns out incomplete. Prompt self-reporting reduces penalties, and good-faith compliance counts as evidence of care.

Regulator access will expose model weights and trade secrets.

Full safety cases go to the regulator under strict confidentiality, with secure facilities for the most sensitive material. The federal government already protects nuclear designs, defense systems, and pharmaceutical trade secrets. Public summaries leave out anything that would help a competitor or an attacker.

Pre-deployment authorization will slow release cycles and hand the race to China.

Authorization applies only to frontier systems, and the Act sets a fixed statutory review clock so no developer waits indefinitely. Advanced systems deploy on registration and a completed safety case. Frontier AI that governments, hospitals, and banks can trust is a larger export market than frontier AI they cannot. A single catastrophic failure would bring far harsher restrictions than anything in the Standard.

The standards are vague, so we will be sued over them.

The Standard names the behavioral tests, defines the categories, and is published through notice-and-comment rulemaking. Developers get a clear target. Vague exposure exists today, under general negligence law, with no federal standard to point to.

Our voluntary frontier safety frameworks already cover this.

Several developers publish serious frameworks, and much of the Standard codifies their own best practice. Voluntary frameworks can be rewritten whenever they become inconvenient, and they bind only the companies that choose them. A common floor keeps a careless competitor from undercutting the careful ones. Developers who already do the work lose nothing and gain a level field.

Rules like these entrench the largest companies.

Duties scale with capability, fees scale with size, and small companies get technical help. Evaluations run through accredited independent labs, so a startup buys the same test a giant does. The heaviest obligations land on the handful of companies able to build frontier systems.

Open-weight releases would be restricted.

Open research and open release stay protected below the frontier threshold, which covers the great majority of open models. At the frontier, releasing weights is irreversible, and the Standard treats irreversible action with more care. The Act provides a defined path to authorization for open release that meets the safety case.

Ecological review will pull every product into environmental paperwork.

Assessment is triggered only where an AI system materially affects agriculture, water, energy, extraction, or ecosystems. A coding assistant or a chatbot never files one.

The trade on offer

The Act gives developers four things they want: one national rulebook in place of a state patchwork, a fixed review clock, a limited safe harbor for good-faith disclosure, and government recognition of the safety work the best of them already do. In return, it asks for evidence in place of promises. Insurers, hospital systems, banks, and federal buyers increasingly want that same evidence before they deploy AI, so the market is moving in this direction with or without Congress. The Standard gives that demand one common form.

12. What companies and universities can do before Congress acts

Legislation takes time. Developers, cloud providers, insurers, large buyers, and universities can start now, and the evidence they generate will shape better law.

A developer commitment

A company adopting the Standard commits publicly to five things: placing the seven constraints above commercial task objectives, publishing a system-specific assurance summary, submitting covered systems to independent evaluation, reporting material incidents, and keeping human authority to correct or stop its systems. The board of directors oversees the commitment.

A research consortium

Computer science alone cannot define the integrity of life-support systems, human flourishing, legitimate authority, medical risk, or the options of future generations. A consortium drawing on ecology, systems biology, engineering, law, political science, philosophy, psychology, public health, and economics should work on:

- tests for shutdown resistance, unauthorized persistence, manipulation, and strategic concealment in long-horizon agents;
- ways to tell genuine reliability from benchmark tuning;

- decision methods that hold constraints firm without freezing action when they conflict;
- measurement of cumulative effects from many individually permissible deployments;
- field methods that feed the response of affected people and ecosystems back into deployment decisions; and
- incident-sharing methods that stay credible across different legal systems.

Measures of success

Success is measured in outcomes, never in forms filed: fewer serious control failures, faster detection and containment, independent replication of safety claims, resolved appeals, developers correcting course after field evidence, healthy market entry, and the ability of authorities to suspend a dangerous capability without disrupting essential services.

Why this matters to me

I am the father of four children. I want my grandchildren to live full lives, and I want them to have children of their own. Every provision in the Standard serves that wish: a humanity that survives, stays free, keeps its hands on the controls, and still has the soil, water, and climate that feed it, generation after generation.

I have spent my career building technology, and I build with AI every day. I am no enemy of it. The question is who decides where it takes us. The Standard keeps that decision with people, where it has always belonged.

Disclosure: who is behind the Standard

Item	Details
Founder	Greg Stoutenburgh, Laguna Hills, California. He founded The Human Continuity Project and wrote the Standard it advances.
Current roles	Chief Innovation Officer and founding team member, Graphene Valley Corporation. CEO, Quantumis Bio, which develops regenerative and diagnostic products for human medicine.
Interests	Both companies develop products that use AI, and Graphene Valley Corporation is building its own AI research platform. Neither company develops frontier AI models, the systems on which the Standard places its heaviest duties. The Standard speaks for Greg Stoutenburgh alone and does not represent either company.
Funding	The Project has accepted no outside funding. Every contributor will be listed here, companies always by name. Individuals giving smaller amounts may ask to stay anonymous, and the total of anonymous contributions will be published.

Item	Details
Background	Serial entrepreneur, scientist, and inventor on dozens of patents worldwide; technologist; AI enthusiast; and a father who is concerned about the future of humanity. Founded two management services organizations for physician and dental practices and has run, managed, or advised more than 100 medical, dental, and veterinary clinics. Led or took part in more than twelve mergers and acquisitions.
Education	Biology, including ecology and evolutionary ecology; chaos and complexity theory; chemistry; philosophy.
Board service	Board of trustees, Casa Romantica. Two earlier nonprofit boards focused on medical research.

Sources

These sources describe the current state of policy and research and supply components of the framework. None of them adopts the Human Continuity Standard as a whole.

[1] National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1 (2023). https://doi.org/10.6028/NIST.AI.100-1

[2] NIST, *AI Risk Management Framework: Generative AI Profile*, NIST AI 600-1 (2024). https://doi.org/10.6028/NIST.AI.600-1

[3] NIST, *Secure Software Development Practices for Generative AI and Dual-Use Foundation Models*, SP 800-218A (2024). https://doi.org/10.6028/NIST.SP.800-218A

[4] OECD, *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449 (2019, amended 2024).

[5] UNESCO, *Recommendation on the Ethics of Artificial Intelligence* (2021).

[6] Council of Europe, *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, CETS No. 225 (2024).

[7] Regulation (EU) 2024/1689 (Artificial Intelligence Act).

[8] UN General Assembly Resolution 78/265 (2024), on safe, secure and trustworthy AI for sustainable development.

[9] UN General Assembly Resolution 79/325 (2025), establishing the Independent International Scientific Panel on AI and the Global Dialogue on AI Governance.

[10] ISO/IEC 42001:2023, Artificial intelligence management system; ISO/IEC 23894:2023, Guidance on risk management.

[11] Convention on Biological Diversity, Kunming-Montreal Global Biodiversity Framework, Decision 15/4 (2022).

[12] California SB 53, Transparency in Frontier Artificial Intelligence Act (2025).

[13] S. 2938, Artificial Intelligence Risk Evaluation Act of 2025, 119th Congress. https://www.congress.gov/bill/119th-congress/senate-bill/2938

[14] H.R. 9925, FRONTIER Act, 119th Congress (2026).

[15] Executive Order 14365, Ensuring a National Policy Framework for Artificial Intelligence (Dec. 11, 2025); The White House, *America's AI Action Plan* (July 2025).

[16] Meinke et al., Frontier Models are Capable of In-context Scheming, Apollo Research (2024). arXiv:2412.04984

[17] Greenblatt et al., Alignment Faking in Large Language Models, Anthropic and Redwood Research (2024). arXiv:2412.14093

[18] Lynch et al., Agentic Misalignment: How LLMs Could Be Insider Threats, Anthropic (2025).

[19] Palisade Research, shutdown-resistance evaluations of frontier reasoning models (2025).

[20] Richardson et al., Earth beyond six of nine planetary boundaries, *Science Advances* 9, eadh2458 (2023). https://doi.org/10.1126/sciadv.adh2458