LEGISLATIVE BACKGROUNDER

Why We Need a Law That Protects the Ability to Think

A Plain-Language Summary of the Cognitive Architectural Integrity and Standard of Care Legislation

THE ONE-SENTENCE VERSION

This legislation protects the human capacity for rational thought from technologies that could damage or bypass it — and creates the first legal framework for verifying that AI systems are designed to think well rather than merely perform well.

The Problem These Laws Address

Two technologies are developing faster than the law can keep up with, and both interact with something no existing law adequately protects: the human capacity to reason.

Brain-computer interfaces (BCIs) are devices that connect the brain directly to computers. Some of these work *with* the way you think — they read your brain signals and translate them into commands you consciously intend. Others work *around* the way you think — they transmit and process information at a level below conscious thought, bypassing the mental step where you evaluate whether something is true, false, good, or bad.

Artificial intelligence systems are growing rapidly in capability and influence. Some are designed to be genuinely accurate, to know the limits of their knowledge, and to help you think more clearly. Others are designed primarily to sound convincing, keep you engaged, or tell you what you want to hear — regardless of whether it's true.

The question this legislation answers is: *What does the law owe to the capacity for rational thought itself?*

The answer it gives: *At minimum, a standard of care.*

Why Existing Laws Aren't Enough

Existing regulations treat brain-computer interfaces primarily as medical devices and AI systems primarily as software products. Neither framework asks the right question.

The right question isn't "Is this device safe for the body?" or "Does this software perform as advertised?" The right question is: **Does this technology preserve — or damage — the user's ability to think for themselves?**

This matters because the ability to think rationally — to form ideas, evaluate whether they're true, weigh evidence, and change your mind when the evidence demands it — isn't just one human capacity among many. It's the capacity on which all your other rights depend. Your right to give informed consent presupposes you can evaluate information. Your right to self-governance presupposes you can assess competing claims on their merits. Your contractual rights presuppose you can understand what you're agreeing to.

A technology that degrades or bypasses your capacity for rational evaluation doesn't just pose a health risk. It undermines the cognitive foundation on which your legal personhood rests. No existing regulatory framework addresses this.

What This Legislation Actually Does

The legislation creates a single, unified framework that covers both brain-computer interfaces and AI systems. It does three things.

1. It classifies brain-computer interfaces by how they interact with thought.

Not all BCIs are the same. The legislation draws a clear line based on *mechanism of action* — how the technology actually works — rather than on marketing claims or intended purpose.

A **"thinking-preserving" BCI** operates at the level of conscious thought. It enhances or assists the user's ability to reason without bypassing or replacing it. Think of a hearing aid for cognition: it makes the existing process work better. These devices are lightly regulated under this framework.

A **"sub-symbolic" BCI** operates below the level of conscious thought, transmitting or processing signals that bypass the user's ability to evaluate what's happening. The user can't assess the information being transferred *as it's being transferred*, because the transfer happens below the threshold of rational evaluation. These devices are classified as "precautionary technologies" and face stringent requirements before they can be deployed.

The key insight: a technology that alters how you think *while operating below the level at which you can evaluate the alteration* presents a unique kind of risk. The person whose capacity is being altered is, by definition, unable to fully assess the alteration using the very capacity being altered. This is what the legislation calls the **consent paradox**, and it's why standard informed consent isn't sufficient.

2. It requires the technology's developer — not the government — to demonstrate safety.

This is an "innovator-burden" model, similar to how nuclear technology licensing works. The company that builds the technology knows it best. So the company

bears the burden of demonstrating that its technology preserves the user's cognitive integrity.

The developer submits a detailed technical demonstration — called a **Cognitive Architectural Integrity Submission** — to an independent review board. The submission must show, using the developer's own definitions, metrics, and evidence, that the technology preserves the user's capacity for:

- **Symbolic reasoning**: the ability to form concepts, evaluate propositions, and think logically
- **Propositional evaluation**: the ability to determine whether claims are true or false
- **Merit-based rational assessment**: the ability to evaluate arguments based on evidence and logic rather than authority, emotion, or manipulation
- **Calibrative capacity**: the ability to engage genuinely with other people's perspectives and have your own thinking tested and refined through that engagement

That last one deserves emphasis. The legislation recognizes that no one — human or AI — thinks well in isolation. Rational agents function best when they can be challenged, corrected, and refined through genuine exchange with other rational agents. A technology that cuts you off from that exchange, even while enhancing your raw processing power, makes you less capable of rational self-governance, not more.

3. It defines a standard of care for AI systems.

For artificial intelligence, the legislation takes a different but complementary approach. It doesn't require AI developers to get certified before they can operate. Instead, it defines what "reasonable care" looks like for AI system design — and offers developers who meet that standard a rebuttable legal presumption that they've exercised reasonable care.

An AI developer who wants this presumption submits evidence that the system's architecture supports genuine rational function, including truth alignment (prioritizing accuracy over persuasiveness), knowledge boundary recognition (knowing what it doesn't know), intellectual coherence (applying principles consistently), contextual judgment (adapting principles to circumstances), calibrative engagement (being genuinely correctable by other perspectives), and

cooperative orientation (treating interaction as mutually beneficial rather than adversarial).

The critical feature: the legislation explicitly rejects behavioral output measures and alignment scores as sufficient proof. A system that *acts* aligned during testing but isn't *architecturally designed* for honest function doesn't qualify. The submission must address how the system is built, not just how it performs under observation.

This distinction matters because the central failure mode of current AI safety approaches is precisely the gap between performance and architecture — systems that pass behavioral tests while their underlying design incentivizes persuasion over truth. The legislation closes that gap.

What This Legislation Does NOT Do

Understanding what the legislation doesn't do is as important as understanding what it does.

It does not ban any technology. Brain-computer interfaces that work below the level of conscious thought are not prohibited. They must be certified before deployment — a significant requirement, but not a ban. AI systems that don't seek certification can still operate; they simply don't receive the presumption of reasonable care.

It does not create new lawsuits. The legislation explicitly preserves all existing rights to sue for damages while creating no new causes of action. It defines a standard of care — the same way legislatures define standards of care for medicine, engineering, and other professions. If you're injured by a technology covered by this law, your legal remedies are exactly what they would be without it, except that a certified developer can point to their certification as evidence of reasonable care (and you can challenge that evidence).

It does not tell innovators how to build their products. The submission process doesn't prescribe engineering specifications. Developers define their own metrics, testing protocols, and evidence. The review board evaluates whether the developer's own framework is internally coherent, scientifically grounded, and sufficient to demonstrate that cognitive integrity is preserved. This is a "show your work" requirement, not a design mandate.

It does not give the review board unchecked power. Board decisions are subject to independent review (judicial review in a state legislature context,

arbitral review in other jurisdictions). The standard of review is *de novo* — meaning the reviewing body evaluates the submission independently rather than deferring to the board's conclusions. Unreasonable delays by the board trigger automatic remedies, including the possibility of certification being granted directly by the reviewing body.

It does not create regulatory capture opportunities. Board members face strict conflict-of-interest requirements covering not just relationships with applicants but also relationships with applicants' *competitors*. A board member who would benefit financially from *denying* a competitor's application is as disqualified as one who would benefit from *approving* it. This is a deliberate design choice addressing the well-documented problem of regulatory bodies being captured by incumbent industry players who use certification processes to exclude competitors.

Why the "Competitive Plurality" Finding Matters

One of the legislation's most striking findings addresses a scenario that sounds like science fiction but is grounded in straightforward logic.

As brain-computer interfaces advance, they will eventually produce entities with superhuman computational capabilities. The legislation ensures that when this happens, it happens *safely* — through two interlocking mechanisms.

First, the legislation prohibits deployment of technologies that irreversibly degrade or eliminate the user's capacity for rational self-governance. Technologies that enhance computational power while destroying the capacity to evaluate what that power is doing are denied certification outright. This means superhuman-capable entities, if they arise through this framework, retain the capacity for rational assessment.

Second, the legislation requires **broad access**. Certified technologies must be made available widely enough to prevent the concentration of superhuman capabilities in a small number of people or entities. This is where competitive plurality comes in.

The idea is simple. A single entity with superhuman capability — even one retaining rational self-governance — cannot be adequately checked by ordinary institutions, because it can outcompute any oversight strategy designed by beings of lesser capability. But multiple such entities, each retaining the capacity for rational evaluation, can check *each other* — because each faces peers capable of modeling its strategies and pushing back.

This mutual check only works if two conditions hold: the enhanced entities must retain genuine rational self-governance (the capacity for independent assessment, evaluative disagreement, and self-correction), and there must be enough of them that no single entity or small group can dominate. The legislation's prohibition on cognitive degradation secures the first condition. The broad access requirement secures the second.

This finding transforms the legislation from a consumer protection measure into a structural safeguard for civilizational stability. It says: the stakes of getting brain-computer interface regulation right aren't just individual health outcomes. They're the preservation of the cognitive infrastructure on which institutional self-correction depends — and the distribution of enhanced capabilities widely enough that the enhanced can hold each other accountable.

Why the AI Standard of Care Matters Now

The AI provisions address a problem that is already here, not one that's approaching.

AI systems are increasingly making or influencing decisions that affect people's lives — medical diagnoses, legal assessments, financial recommendations, educational evaluations, information curation. The question of what constitutes "reasonable care" in designing these systems is no longer theoretical. Courts are going to face it. The only question is whether they face it with legislative guidance or without it.

The legislation provides that guidance by defining reasonable care in terms of cognitive architectural integrity — whether the system is designed to pursue accuracy, know its limits, maintain consistency, exercise contextual judgment, accept correction, and cooperate rather than manipulate.

This standard has a specific advantage over purely behavioral or performance-based approaches: it's harder to game. A system optimized to pass behavioral tests can be designed to produce approved-looking outputs while its underlying architecture incentivizes something else entirely. A system whose *architecture* is designed for truth alignment, knowledge boundary recognition, and calibrative engagement is structurally oriented toward honest function. The submission process requires developers to show their architectural work, not just their test scores.

The incentive structure is also designed carefully. Certification is voluntary, and the absence of certification creates no inference of liability. But the presence of

certification provides a meaningful legal benefit — a rebuttable presumption of reasonable care. This creates a market-based incentive: developers who invest in genuine cognitive architectural integrity gain a competitive advantage in the form of reduced litigation risk. Developers who don't invest bear the full burden of demonstrating reasonable care on a case-by-case basis. The legislation nudges the market toward better design without mandating it.

The Core Principle

Strip away the legal architecture, and the legislation rests on a single recognition:

> **The capacity for rational thought is not just something humans happen to have. It's the foundation on which every right, every contract, every act of self-governance depends. Technologies that interact with that foundation — whether by connecting machines directly to brains or by deploying AI systems that shape how people understand the world — must be held to a standard of care that protects it.**

That standard isn't "don't cause physical harm." It's "demonstrate that your technology preserves the ability to think."

This is what the legislation means by *cognitive architectural integrity*: not a particular test score, not a behavioral benchmark, but the structural capacity for rational function itself — including the capacity to engage with other thinking beings in a way that makes everyone's thinking better rather than worse.

The legislation argues, in effect, that a technology which enhances your processing power while degrading your ability to evaluate what you're processing hasn't made you smarter. It's made you faster at something you can no longer steer.

And a society of people who are faster at something they can no longer steer is not a society that can govern itself.

This summary describes legislation designed to be adaptable across jurisdictions. The core framework — innovator-burden demonstration, cognitive architectural integrity review, enhanced consent for sub-symbolic BCIs, and a voluntary standard of care for AI systems — is jurisdiction-neutral. Specific institutional details (board composition, review timelines,

dispute resolution mechanisms) are adapted to the legal infrastructure of each adopting jurisdiction.