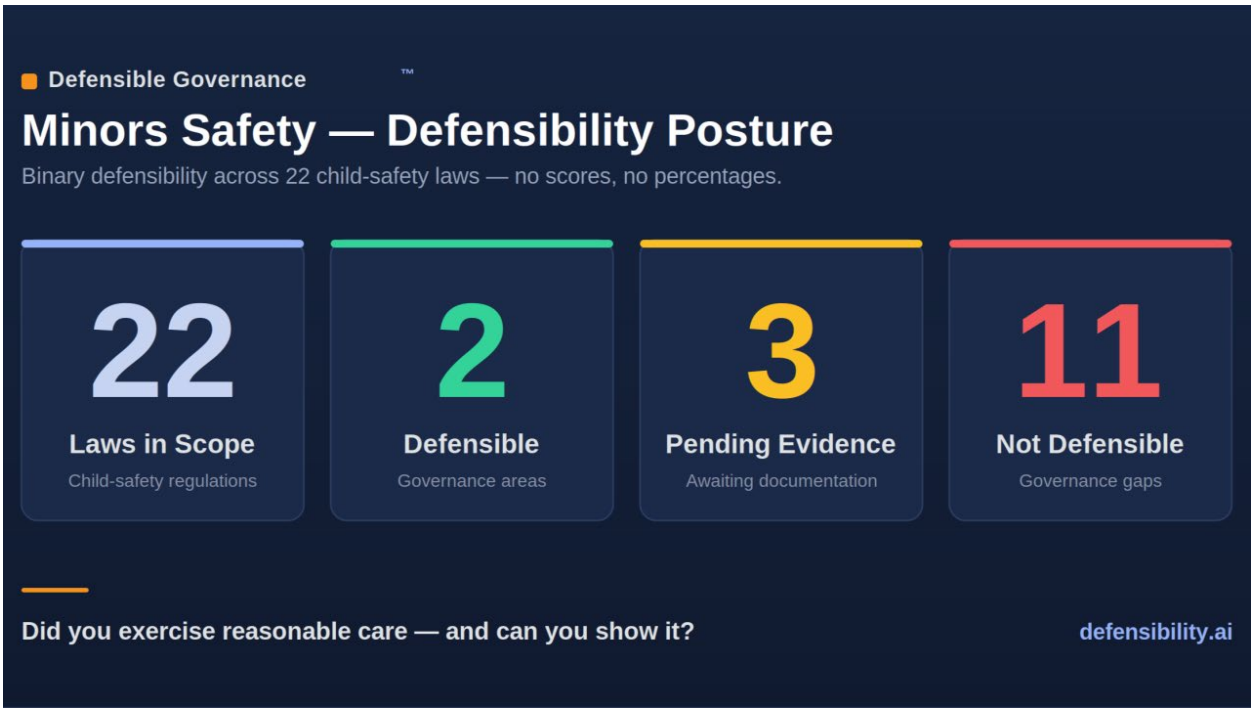

Gaming, Social Media, and EdTech companies now have a mandate to focus on Psychological welfare

A gaming, social media, or EdTech company run from the US already answers to a federal baseline, fifty state regimes, and the EU, UK, and places like Australia at once. Psychological welfare is where the principles go quiet — and the binding stack gets loudest.

*By John Johnson, Founder & CEO, **Defensibility.ai***

The G7 digital and tech ministers this week announced a Common Set of Principles for protecting minors online. International alignment on this question is overdue, and the participating governments deserve credit.

But **if you run** a **gaming**, **social media**, or **EdTech company** out of the **US** — with users who don't stop at the border — the **G7 principles are the least of what you answer to**. Long before any voluntary principle reaches you, you are already operating inside a dense, overlapping stack of binding law: a federal baseline, a fast-moving patchwork of state statutes, and the EU, UK, and Australian regimes that apply the moment a minor in those jurisdictions opens your app. Much of it **carries personal liability** for **named executives. The principles are a floor**. The stack is the reality.

And the place the principles go quietest is exactly where that binding stack is loudest. The seven **principles address *risk management*, *age verification*, *privacy defaults*, *CSAM prevention*, *parental tools*, *digital literacy*,** and ***research access*** — every one of them about access and exposure: keeping certain things away from minors and giving adults the tools to supervise. ***Not one addresses what happens to a minor's psychological and behavioral wellbeing*** once they are inside the service.

That omission would matter less if the law had also overlooked it. It hasn't. **Across** the same **jurisdictions**, l**egislators have already moved past safety-by-design** and into prescriptive **regulation** of the **design features** that **affect minors' mental health, sleep, attention, and behavior**. Those **laws are in force**. Enforcement is underway. And the **dollar figures are** already **substantial**.

The principles stop where psychological harm begins

The G7 principles call for "recommendation systems designed to minimise excessive online engagement." That is an aspiration. It is not a defined obligation, it carries no defaults, and it specifies no penalty.

The binding statutes that have actually passed across the same jurisdictions are far more specific. They **legislate** the **design features that produce psychological harm** — not **the harm** itself, but the **algorithmic** and **product mechanics** that **produce it.** Feed defaults. Notification windows. Curfew hours. Dark patterns. Algorithmic amplification of harmful content. Behavioral profiling. Manipulative AI directed at vulnerable groups.

This is the **layer the principles do not reach**. And it is the **layer where executive accountability** has already **attached**.

The stack a US-based global platform is already inside — and where the sharp edge is

The binding **obligations do not sit** in **one statute**. They stack — a **federal** baseline, a fast-moving **layer of state law**, and the **international** regimes that set the global standard. A platform that minors use, are likely to use, or simply cannot exclude is operating inside all three layers at once.

The federal baseline: COPPA.

The Children's Online Privacy Protection Act still governs the collection of personal data from children under 13, and it still requires verifiable parental consent. It remains mandatory — and it is no longer sufficient. COPPA does not reach teenagers aged 13 to 17, and it says nothing about design harm, compulsive use, or addictive engagement. It was written for a web of forms and data fields, not feeds and recommendation engines.

The state battleground — where the real shift is happening.

This is where legislators have moved from "did you collect data lawfully?" to "did you design this product responsibly for a child?" Three patterns have emerged.

Design-code and duty-of-care laws now legislate the product itself.

- **California Age-Appropriate Design Code Act (CAADCA).** Mandates the highest privacy settings by default for known minors and prohibits dark patterns — design choices that nudge minors toward decisions against their interests. Specific protective defaults are defined by statute. Subject to ongoing litigation, but on the books and shaping policy well beyond California.
- **Vermont Age-Appropriate Design Code Act.** The clearest psychological-harm-and-design-liability law yet written in the United States. Enacted in June 2025 and effective January 1, 2027, it imposes a minimum duty of care: a covered business's data practices and design choices must not cause reasonably foreseeable emotional distress or reasonably foreseeable compulsive use by a minor — with "compulsive use" defined as repetitive use that materially disrupts sleeping, eating, learning, or concentrating. It carries a private right of action and applies once a defined share of a service's users are minors. This is the duty-of-care standard the G7 principles only gesture at, written into binding law.

Platform-regulation laws legislate the mechanics of engagement.

- **New York SAFE for Kids Act.** Chronological feed by default for minors. Algorithmic amplification disabled unless parents affirmatively consent. Night-hour notifications disabled by default. The statute names the psychological mechanism — engagement-maximizing recommendation systems — and legislates against it as a default.
- **Utah Social Media Regulation Act.** Mandatory curfew defaults for minor accounts, 10:30 PM to 6:30 AM, with no organizational discretion over the window. Sleep disruption is treated as a regulated harm, met with a default the platform cannot override.
- **Florida HB 3.** Restricts social-media accounts for minors under 14, requires parental consent for ages 14 and 15, and targets the addictive design features the legislature concluded were causing psychological harm.
- **Texas SCOPE Act.** Restricts the data collection that fuels behavioral targeting of minors, prohibits targeted advertising to known minors, and gives parents controls over algorithmic amplification.

- **Tennessee Protecting Kids from Social Media Act and Mississippi's Walker Montgomery Protecting Children Online Act.** Age verification, parental consent, and harm-mitigation duties — extending the same expectations into additional states and confirming this is a national pattern, not a coastal one.

Enforcement-driven laws are already producing lawsuits.

- **Florida Digital Bill of Rights.** Restricts the handling of "known child" data and bans re-identification — and is the subject of active enforcement by the state Attorney General, including the action against Roku. This is the layer where statutes become litigation.
- Beyond the named statutes, minor-facing platforms remain exposed under state unfair-and-deceptive-practices (UDAP) laws, the FTC's Section 5 authority over dark patterns and unfair design, California's CPRA (which sets higher penalties for misuse of minors' data), and CSAM statutes in all fifty states.

Europe and the UK — the global standard.

The strictest and most influential frameworks sit outside the United States, and they apply to any platform serving European or British minors.

- **EU AI Act, Article 5.** An absolute prohibition on AI practices that exploit the vulnerabilities of specific groups, including minors — no proportionality test, no cure period, no defense based on consent. It carries the Act's highest penalty tier: up to €35M or 7% of global annual turnover. The Article 5 prohibitions became applicable in February 2025.
- **Digital Services Act, with the Guidelines on the Protection of Minors.** Binding obligations on very large online platforms to assess and mitigate systemic risks to minors, including risks to physical and mental wellbeing, and a ban on targeted advertising to minors.
- **General Data Protection Regulation.** Governs the lawful basis for processing minors' data, requires data protection impact assessments, and sets the threshold for parental consent — between 13 and 16 depending on the member state.
- **UK Online Safety Act.** Requires in-scope services to complete a documented Children's Risk Assessment before launch and on material change, covering the risk of psychological harm — including content that normalizes self-harm and eating disorders — and, in defined circumstances, attaches personal criminal liability to senior managers.

Australia.

- **Australia Online Safety Act.** Children's-safety obligations enforced by the eSafety Commissioner, including provisions newly aimed at preventing under-16s from creating accounts on certain platforms. Both platform liability and individual officer liability attach.

Count the **psychological-welfare layer alone** — the laws that target emotional distress, compulsive use, sleep disruption, addictive design, and manipulative AI rather than data and access — and there are **ten** binding regimes: **California**, **Vermont**, **New York**, **Utah**, **Florida** HB 3, **Texas**, the **EU AI Act**, the **DSA**, the **UK Online Safety Act**, and the **Australian Online Safety Act**. A **platform operating across the US, EU, UK, and Australia** answers to **every one of them** at once.

*"The G7 principles name the right problem — and it's the same one we built Defensibility.ai to help with. **Principles tell** a platform **what good looks like**; they **don't give leaders** a **way to show**, later, that **they weighed** the **risk to minors before a feature shipped**. We give leaders a disciplined way to **see what is legally expected** of them, and to **document the child-safety decisions** they made, **before a product ships** rather than reconstructed under examination."*

— Maverick James, Co-Founder and Product Counsel, Defensibility.ai

What's coming, and what we've built into the product

Two more laws are advancing and worth tracking by every executive at a minor-facing platform.

Kids Online Safety Act (KOSA). Passed the U.S. Senate 91–3 in July 2024 and reintroduced in 2025. In March 2026, the House Energy and Commerce Committee approved a rewritten version within the broader Kids Internet and Digital Safety (KIDS) Act — but stripped the duty-of-care provision that was KOSA's sharpest legal edge. House–Senate reconciliation will determine the final shape.

COPPA 2.0 (Children and Teens' Online Privacy Protection Act). Extends COPPA's protections from under-13 to under-17 and restricts behavioral advertising to minors. Passed the Senate unanimously in March 2026. House action pending.

We have built both into the Defensibility.ai Minors Safety & Child Welfare application as best-practice safeguards. Platforms operating defensibly against these standards today are positioned to flip them to compliance mode the day either becomes law, without rework.

What every one of these laws is converging on

Strip away the jurisdictional differences and the same five obligations surface in nearly every regime:

1. **Privacy by default.** The most protective settings, on automatically, for any user known or estimated to be a minor.
2. **Age assurance.** Know — or reasonably estimate — who is a minor, and treat them differently.
3. **Design accountability.** No dark patterns, no manipulative UX, no exploitative engagement loops.
4. **Psychological harm and compulsive use.** Assess the mental-health impact of the product and address its addictive features — the obligation now most explicit in Vermont, the UK OSA, and the EU DSA.
5. **Duty of care and evidence.** Identify foreseeable harm, evaluate alternatives, document the decision, and be able to prove leadership acted reasonably.

The first four describe what a platform must build. The **fifth describes what it must be able to prove**.

The shift that actually matters: from "did you comply?" to "did you exercise reasonable care?"

This is the change underneath all of it. The question regulators and courts are now asking has moved from ***did you comply****?* to ***did you exercise reasonable*** *or* ***adequate care****?* It is no longer enough to point to what you built. You have to be able to show *why* you built it that way — and that the decision was reasonable in light of the foreseeable risk to a child.

That is **product-liability logic layered on top of privacy law**, and it falls hardest on social media and gaming, because those platforms are built on the exact mechanics regulators are now scrutinizing: engagement loops, reward systems, recommendation engines, social reinforcement, and retention mechanics — **viewed through** the lens of **compulsive use**, **behavioral influence**, and **psychological impact**.

A platform used by minors, likely to be accessed by minors, or unable to exclude them is now governed by all three at once: privacy law, product-liability logic, and duty-of-care governance.

More than $1B in fines — and every one tied to harm to minors

Enforcement track records reveal what regulators have actually concluded — not what they say in principles documents.

- **Epic Games (Fortnite) — $275M FTC penalty, plus $245M in consumer refunds.** December 2022. COPPA violations and dark patterns directed at children. The largest COPPA enforcement action ever.
- **Instagram (Meta) — €405M.** September 2022. Ireland's Data Protection Commission, under GDPR. Public-by-default settings on minor accounts and exposure of children's contact information via business accounts. The first major EU-wide decision focused specifically on children's data.
- **TikTok — €345M.** September 2023. Ireland DPC, under GDPR. Public-by-default settings for child users, deficient age verification, and failures in supervisory features. Under appeal.
- **TikTok — £12.7M (~$15.7M).** April 2023. UK Information Commissioner's Office. Allowing roughly 1.4 million children under 13 onto the platform without parental consent.
- **YouTube — $170M.** September 2019. FTC and the New York Attorney General, under COPPA. Tracking the viewing history of minors to support targeted advertising — exactly the behavioral-profiling pattern the newer state laws now legislate against directly.
- **TikTok (then Musical.ly) — $5.7M.** February 2019. FTC, under COPPA. Failure to obtain parental consent before collecting children's information.

More than **$1.2 billion** in **publicly disclosed fines**, against six of the **most recognizable platforms** in the **world**, ***every one tied specifically to the handling of minors***. These are not careless companies. They are among the most sophisticated operators on the internet, with substantial legal and compliance functions. The fines happened anyway.

The next wave is not about privacy. It is about harm.

The fines above were privacy cases — unlawful data collection, deficient consent, deceptive settings. The **enforcement** now **building against gaming platforms** is a **different animal**. The theory has moved from did you handle children's data lawfully? to did you design a product that **exposed children** to **foreseeable psychological**, **behavioral**, and **physical harm** — and **did** your **leadership act reasonably** once it knew? The Epic Games settlement above — $520 million in all — was the privacy era's high-water mark. What has come since is a different kind of case.

Loot boxes and addictive monetization.

- Genshin Impact (HoYoverse / Cognosphere) — $20M FTC settlement, January 2025. The FTC alleged COPPA violations alongside unfair and deceptive loot-box mechanics: a confusing multi-tier virtual-currency system and misrepresented odds that, the agency said, pushed children and teens to spend on prizes they had little chance of winning. The settlement bars selling loot boxes to under-16s without parental consent and requires real-money purchase options and odds disclosure. It is one of the first federal actions to connect minors, monetization mechanics, and gambling-like behavioral design in a single case. HoYoverse settled while disputing many of the allegations.

Grooming, exploitation, and platform design — the Roblox litigation.

Roblox has become the test case for the next generation of child-safety litigation, and the claims have moved well past privacy. Everything below is an allegation in pending matters; Roblox disputes the claims and no liability has been established.

- State attorneys general. Louisiana (alleging the platform lets predators "thrive, unite, hunt and victimize kids"), Kentucky (alleging Roblox created "a hunting ground for child predators"), and Texas (alleging the company marketed the platform as safe while failing to prevent predators from contacting minors), with additional states including Florida, Iowa, Nebraska, and Tennessee filing suits or opening investigations through 2025 and into 2026. A Texas court has allowed the state's core claims about misleading safety representations to proceed past dismissal.
- Local government. Los Angeles County sued in 2026, alleging the platform's design lets adults masquerade as children and that Roblox failed to implement readily available safeguards — age verification, default communication restrictions, meaningful parental controls — to reduce foreseeable harm.
- Federal MDL. Dozens of family lawsuits alleging child sexual exploitation and grooming have been centralized in a multidistrict litigation in the Northern District of California (In re: Roblox Corporation Child Sexual Exploitation and Assault Litigation, MDL No. 3166). The docket grew from roughly 85 cases at the start of 2026 to about 146 by April. Some of the most serious filings tie grooming and sextortion that began on the platform to the later deaths of teenagers.
- Arbitration and class actions. In December 2025, a California court held that Roblox could not force these sexual-assault claims into private arbitration, citing the federal Ending Forced Arbitration Act; separate class actions challenge the platform's data collection and marketing.

The through-line in every one of these is the same sentence, and it is the one that should concern any executive at a minor-facing platform: *you knew the risk existed*, and you *failed to mitigate foreseeable harm*.

That is **not a privacy standard**. It is a **duty-of-care standard —** the same **standard Vermont,** the **UK Online Safety Act**, and the **EU's frameworks** are now **writing into statute**, and the same question a court asks when it assigns liability after the fact. The statutes and the lawsuits are converging on one demand: show that leadership weighed the foreseeable harm before the product reached the child.

What this means for executives in gaming, social media, and EdTech

The job of a **Chief Privacy Officer**, **General Counsel**, **CISO**, or **CTO** at a minor-facing company has changed.

Producing a defensible record now requires:

- Mapping obligations across the binding psychological-welfare regimes — not just the privacy and access ones.
- Documenting which named executive role is accountable for each obligation, in each jurisdiction.
- Setting and sealing — before product release — a **contemporaneous record** of the **reasonable decisions** you made **around risk of psychological and behavioral harm**.
- Holding that record in a form that survives examination two years later, when a regulator asks how a feature was approved.

No spreadsheet, annual **policy review**, or **risk-register** exercise **produces** the **artifact regulators** are now **asking for**. The shape of the obligation has changed faster than the governance tooling.

What we built

The **Defensibility.ai** **Minors Safety & Child Welfare application** covers 20+ child-welfare laws and standards out of the box — including the psychological-welfare provisions discussed above — and adding a new one is a configuration process, not a feature request. KOSA, COPPA 2.0, and additional regimes are built in as best-practice frameworks that activate to compliance mode when enacted. The application produces the artifacts regulators typically ask for: named-executive accountability, threshold decisions sealed before product release, and a cryptographically sealed evidence record.

The G7 principles are a meaningful floor. The binding **psychological-welfare statutes** are the **ceiling** many **platforms have not yet absorbed**. The distance between the two — the engagement-maximizing recommendation systems, the addictive design features, the algorithmic amplification of harmful content, the manipulative AI practices — is where companies, and the executives who run them, are now operationally exposed.

If you operate in gaming, social media, or EdTech, the binding **obligations are real**, the **enforcement** record is in the **public domain**, and the gap most platforms operate inside is documentable harm to minors. The question is no longer whether the regulatory landscape has moved. It has. The **question** is whether you **can show**, when asked, that your **leadership weighed** the **risk** of **psychological harm** to **minors before** that harm reached them.

That is the work. **We built** the **tool that does it.**

Want to know where you actually stand? We're offering a no-charge Functional Defensibility Assessment (normally $12,500) to a small cohort of gaming, social media, and EdTech platforms. It maps your obligations across these laws, flags which ones require a sealed pre-release assessment, and hands you a board-ready exposure report named to the executives who carry the risk. Message me, or start at **defensibility.ai**.

Defensibility.ai builds the legal-defensibility layer for executive leadership and boards. Its two commercially available applications — Defensible Governance and Minors Safety & Child Welfare — are the subjects of pending U.S. provisional patent applications. We do not guarantee a specific legal outcome and are not a replacement for legal advice from a licensed attorney.