Accountability Social & AI Age Verification

A Risk Architecture and Governance Framework for AI–Social Platform Interaction

Jason D. McCoy — Pre-Failure Research Group LLC

What This Is

A structured analytical framework examining accountability, risk propagation, and age verification challenges where AI systems and social media platforms intersect. It's built to be methodologically explicit and implementation-aware — moving past general concern toward structured risk classification, governance mapping, and intervention models suitable for legislative review.

Nothing in this framework constitutes legal advice or accusation. All legal references reflect publicly documented actions, settlements, or regulatory proceedings, with clear distinction maintained between adjudicated violations, settlements, ongoing litigation, and regulatory investigations.

This framework does not track user behavior, collect personal data by default, perform content moderation, enforce ideology, or operate as a surveillance layer. It's an age *verification* architecture — not a behavioral governance engine. It is also structurally and philosophically distinct from OmniSniffer: related in audit discipline, but independent, and does not inherit any of OmniSniffer's enterprise trust-enforcement mechanisms.

The Four-Layer Risk Model

Layer	Function	Primary Risk
L1 — Exposure	Passive consumption of platform content	Unfiltered content delivered without age screening
L2 — Data Behavior	Platform capture of behavioral telemetry	Profiling minors without verified consent

Layer	Function	Primary Risk
L3 — Governance Gap	Mismatch between policy and enforcement	Weak age controls, opaque moderation logic
L4 — Amplification	Algorithmic acceleration of content reach	Viral propagation to minor audiences

Platforms scoring high on both amplification and youth exposure (L4 combined with L1 risk) represent the highest-priority regulatory targets. Cross-platform sharing multiplies L4 exposure without adding L3 enforcement.

The framework applies this model against ten major social platforms and eleven AI systems, scoring each on five dimensions — Amplification Velocity, Youth Exposure Probability, Verification Robustness, Audit Transparency, and Legal Exposure Intensity — cross-referenced against publicly documented regulatory and legal history (FTC consent decrees, COPPA actions, GDPR investigations, and more).

Tiered Age Verification — The Core Proposal

The design principle: **only age eligibility is verified, never identity.** The system proves "over threshold," not "who you are."

Tier	Access Level	Requirement
Tier 0	Anonymous / minimal	None — public browsing only, no posting or AI generation
Tier 1	Self-declared minor	Self-declared age; restricted content, flagged for monitoring
Tier 2	Verified minor	Photo ID or guardian verification; limited engagement
Tier 3	Verified adult	Photo ID verification; full platform access
Tier 4	Guardian-linked minor	Guardian ID + minor confirmation; supervised full access

Privacy mechanics: Verification tokens are cryptographically signed and time-limited, proving

only an age threshold — never name, address, or other identity data — and stored completely separately from content and behavioral logs. No long-term biometric storage unless legally mandated.

Governing Invariants

A short list of design constraints the framework treats as immutable — none can be modified without a formal governance review:

- **Minimum Necessary Proof** — only age eligibility is proven, never identity, unless legally required
- **Zero Default Retention** — no long-term storage of verification artifacts
- **Separation of Verification and Platform** — the platform receives a token, never the underlying evidence
- **Ephemeral Credentialing** — signed tokens with automatic expiration
- **No Behavioral Expansion** — verification can never drift into risk scoring, content tracking, or trust ranking
- **Fail-Closed** — any ambiguity or technical failure defaults to the most restrictive tier
- **Anti-Database Formation** — the system cannot aggregate verification data into identity profiles

across sessions

Explicit non-goals: the system is designed to never track user behavior, collect personal data by default, moderate content, enforce ideology, or function as a surveillance layer.

Toward Legislative Action

The framework translates into five legislative modules built on existing federal authority — COPPA, FTC Act Section 5, and Commerce Clause precedent for interstate digital commerce:

1. **Systemic Risk Statement** — documenting how algorithmic amplification, behavioral profiling, and cross-platform chains create measurable youth exposure risk
2. **Empirical & Legal Record** — the platform-by-platform regulatory history underlying the framework
3. **Quantitative Risk Classification** — a reproducible scoring rubric for prioritizing enforcement targets
4. **Legislative Proposal Framework** — a federal age-tier verification mandate, audit logging requirements, and third-party certification for verification providers

5. **Implementation Roadmap** — a phased rollout from FTC rulemaking through platform compliance to ongoing independent audit

The framework also builds in failure-mode safeguards for legislative drafters — addressing over-blocking, privacy risk from over-verification, small-business compliance burden, and jurisdictional conflicts — so the statutory language is designed to withstand judicial scrutiny rather than just political pressure.

Pre-Failure Research Group LLC | Jason D. McCoy

Full technical compendium, including platform-by-platform risk scoring and the complete legislative appendix, available on request.