



# Pennsylvania AI Governance Framework

Chatbot Safety & Synthetic Content Disclosure — mapped to the pending Pennsylvania statutes.

*Breach Prevention + Guardrails Early*

# Executive Summary

---

**The problem.** Pennsylvania companies are running AI chatbots that talk to children, publishing AI-generated ad content without disclosure, and doing both without documented safety controls, disclosure workflows, or incident protocols.

**Where the law stands.** Five AI-related bills touching chatbot safety and synthetic content disclosure are active in Harrisburg. None are enacted. The legislature recessed for summer on July 13, 2026; the House reconvenes September 9 and the Senate September 28, with the session running through late November. This isn't a countdown to a compliance deadline — it's a live drafting window, which is more useful: the control language is visible now, before it's locked into final statutory form.

CORE MESSAGE

Waiting for a bill to be signed doesn't delay the risk. The chatbot is talking to a child today. The synthetic ad is running today. Neither needs a governor's signature to become a liability — regulation formalizes a standard that already applies.

**The deliverable.** A four-pillar governance framework covering pre-use chatbot safety assessment, ongoing output monitoring, synthetic content disclosure, and incident response — built against the specific control language already written into the furthest-advanced Pennsylvania bills, so implementation isn't dictated by whichever version reaches the Governor's desk first.

**Positioning.** *Breach Prevention + Guardrails Early.* The bill text already tells you what "reasonable" will mean. Building to it now costs less than reconstructing it under scrutiny later.

# The Breach Scenario

A Pennsylvania company deploys an AI companion chatbot aimed at children and teens — through a toy, an app, or both. The chatbot has no pre-deployment safety review. No one has defined what topics it's permitted to engage on with a minor, whether it can simulate an ongoing relationship, or what happens when a conversation turns toward self-harm.

There is no monitoring layer. Outputs aren't reviewed after deployment — the company has no way of knowing what the chatbot is actually saying to the children using it, in aggregate or in any single conversation.

A high-risk conversation occurs. The system has no self-harm detection, no escalation to a crisis resource, no notification to a parent or guardian. The interaction is discovered only after harm has already happened — by a parent, a journalist, or a plaintiff's attorney reconstructing logs the company never reviewed in real time.

**The control gap, plainly stated:** no pre-launch safety assessment, no output monitoring, no crisis escalation path, no incident response protocol. Each of these is inexpensive to build in advance and expensive to reconstruct under scrutiny afterward.

**The business exposure:** civil liability under Pennsylvania's pending chatbot-safety statutes once enacted; exposure under existing product liability and consumer protection law regardless of whether a new statute has passed; discovery in litigation that surfaces the absence of any safety documentation; and reputational damage that doesn't wait for a court date.

**The moment it's too late:** by the time an incident is public, the company is explaining the safeguards it never built.

*This is a sensitive subject. If a team is navigating an active incident involving a minor, that is a legal and clinical matter first — engage counsel and a crisis-trained resource before anything else.*

# The Regulatory Pressure

Two active Pennsylvania legislative lanes bear directly on this framework — chatbot child safety and synthetic content disclosure. Here is where each bill stands, sourced to the General Assembly and legislative tracking records as of August 2026.

## Chatbot Child Safety

### SB 1090 — SAFECHAT Act PASSED SENATE

Sens. Pennycuick & Miller · Passed Senate 49–1, March 17, 2026 · Now with House Communications & Technology Committee

Requires a "clear and conspicuous" disclosure that a companion chatbot is non-human. Requires protocols preventing the chatbot from generating self-harm, suicide, or violence content. Requires routing to crisis resources when high-risk language is detected.

### HB 2006 — AI in Companionship Applications Safety Act IN APPROPRIATIONS

Rep. Shusterman · House Floor vote on amendment passed 104–98, July 1, 2026 · Re-committed to House Appropriations Committee same day – has not yet passed the full House

Requires a published self-harm protocol on the operator's website. Requires a recurring notice — at least once every three hours — reminding the user they are talking to an AI. Bars the chatbot from claiming to be a licensed mental health professional. Its self-harm protocol and AI-disclosure requirements are detailed and specific; SB 1090 is currently further along procedurally.

### HB 2637 — Toy Chatbot Moratorium EARLY STAGE

Rep. Ciresi · Referred to House Communications & Technology, June 15, 2026 · No committee vote yet

A three-year moratorium on selling children's toys that contain AI chatbots. Early in the process, but drawing significant attention in Harrisburg.

## Synthetic Advertising Disclosure

### HB 95 PASSED HOUSE / SB 806 PASSED SENATE

Rep. Pielli (HB 95) / Sen. Pisciotto (SB 806) · HB 95: 124–78, June 17, 2026, referred to Senate Communications & Technology June 22 · SB 806: 48–2, June 10, 2026 · Companion bills

Both require disclosure when advertising content is AI-generated or synthetic. This is the closest-to-enacted lane in the framework — both bills cleared their originating chamber by wide margins.

**Where this leaves a Pennsylvania company:** nothing above is law yet. The House and Senate reconvene in September, and any of these could reach the Governor's desk before the session ends in late November. A framework built now, against the actual bill language, means implementation happens on your timeline, set well ahead of any enactment date or compressed compliance deadline.

## Four Control Pillars

Each pillar corresponds to a control gap in the breach scenario and a requirement already written into the furthest-advanced Pennsylvania bills. None of them require a signed statute to implement.

### Pillar 1 — Pre-Use Chatbot Safety Assessment

**What it means:** every AI chatbot is safety-assessed before deployment; high-risk configurations are flagged before launch.

**Pennsylvania context:** SB 1090 requires a non-human disclosure and a content-prevention protocol; HB 2006 requires a published self-harm protocol and a recurring disclosure notice; HB 2637, if enacted, would prohibit chatbots embedded in children's toys. This pillar is where each requirement is checked before a product ships — regardless of which bill is signed first.

#### TEMPLATE — CHATBOT SAFETY PRE-LAUNCH CHECKLIST

- Chatbot name / deployment channel (toy · app · website · other)
- Target age group (under 13 · 13–17 · 18+)
- Non-human disclosure implemented? [SB 1090]
- Recurring AI-disclosure notice set, interval ≤ 3 hours? [HB 2006]
- Self-harm / suicide / violence prevention protocol documented and published? [HB 2006, SB 1090]
- Crisis resource escalation workflow active?
- System claims to be a licensed mental health professional? — should be **NO** [HB 2006]
- Ability to disable or override safety features? — should be **NO**
- Embedded in a toy marketed to children under 13? — flag for legal review [HB 2637, pending]
- Safety review authority / approval sign-off (name, date)

### Pillar 2 — Ongoing Monitoring for Harmful Outputs

**What it means:** chatbot outputs are reviewed on a recurring basis for self-harm, suicide, or violence-related content, and for drift toward the relationship-building or professional-impersonation patterns the pending bills target.

**Pennsylvania context:** SB 1090's content-prevention and crisis-routing requirements only function if outputs are actually being checked against them. This pillar is the operating layer underneath that requirement.

| Field                                        | Entry                                      |
|----------------------------------------------|--------------------------------------------|
| Chatbot / reporting period                   | e.g., MyChatFriend — Week of Sept 14, 2026 |
| Total interactions                           |                                            |
| High-risk outputs detected                   | self-harm / suicide / violence-related     |
| Crisis resource escalations triggered        |                                            |
| Non-human disclosure delivered as scheduled? | Y/N, exceptions noted                      |
| Override incidents (safety feature disabled) | target: 0                                  |
| Corrective action taken                      |                                            |
| Reviewed by / date                           |                                            |

### Pillar 3 — Synthetic Content Disclosure Workflow

**What it means:** AI-generated or synthetic advertising content is flagged and disclosed as such before it publishes.

**Pennsylvania context:** HB 95 (passed House, 124–78) and companion SB 806 (passed Senate, 48–2) both require disclosure that advertising content is synthetic or AI-generated. Neither has reached the Governor's desk, but both cleared their originating chamber by wide margins — the closest-to-enacted lane in this framework.

- Ad content description
- AI-generated or synthetic element present? (voice / video / image / testimonial) [\[HB 95, SB 806\]](#)
- Disclosure statement drafted and approved by legal?
- Disclosure placement and duration specified?
- Health, financial, or safety claims independently verified?
- Platform-level synthetic flag set before publishing?
- Compliance sign-off (name, date)

Pillar 4 — Incident Response for Harmful Chatbot Outputs

**What it means:** a documented, timed response protocol for when a chatbot produces high-risk content or fails a safety check — built before the first incident.

**Pennsylvania context:** this pillar gives SB 1090's and HB 2006's crisis-routing requirements an actual operating process behind them.

| Phase                  | Action                                                                                                                                                                 |
|------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Trigger                | High-risk output detected (self-harm, suicide, violence-related content); professional-impersonation claim; safety override                                            |
| Step 1 — 0–1 hour      | Quarantine the interaction; halt further chatbot messages to this user; route to the 988 Suicide & Crisis Lifeline if self-harm risk is present; flag for human review |
| Step 2 — 1–4 hours     | Notify parent/guardian if the user is a minor; preserve the full interaction log; document the decision trail                                                          |
| Step 3 — 4–24 hours    | Legal review of any mandatory reporting obligation; prepare internal incident report                                                                                   |
| Step 4 — 24 hrs–1 week | Root-cause review; retrain or reconfigure the system; document the fix; close the loop with the user/family where appropriate                                          |
| Documentation          | Incident report, preserved logs, full audit trail (interaction → decision → action → outcome)                                                                          |

# Implementation Roadmap

Sequenced against the actual Harrisburg calendar — the legislature reconvenes September 9 (House) and September 28 (Senate), with session running through late November.

| Phase       | Timeframe      | Work                                                                                                                                                                      |
|-------------|----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Weeks 1–2   | Aug 2026       | Inventory every deployed or in-development chatbot and ad-content pipeline; identify which interact with minors or use synthetic content.                                 |
| Weeks 3–4   | Aug–Sept 2026  | Risk-assess each system against SB 1090, HB 2006, and HB 95/SB 806 requirements; prioritize by exposure.                                                                  |
| Weeks 5–8   | September 2026 | Build the pre-launch checklist (Pillar 1) and monitoring log (Pillar 2) infrastructure; draft the crisis escalation workflow. Coincides with the legislature reconvening. |
| Weeks 9–12  | October 2026   | Deploy Pillar 3 (synthetic disclosure) across all ad content; deploy Pillar 4 (incident response) with a tested escalation path.                                          |
| Weeks 13–16 | November 2026  | Full deployment across all systems; internal audit-readiness review — ahead of the legislature's late-November sine die.                                                  |

**Estimated implementation cost:** \$50,000–\$150,000 depending on the number of chatbots and ad channels in scope; \$20,000–\$50,000 for monitoring and reporting infrastructure. (SBJ estimate. No PA statute specifies implementation cost.)

**Recommended staffing:** 1 FTE safety engineer, 0.5 FTE compliance officer, 0.5 FTE data analyst.

# Regulatory Compliance Mapping

| Bill                                                          | Status                                                                                                             | Core Requirement                                                                                                                        | Maps To                                                                          |
|---------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| <b>HB 2006</b><br>AI in Companionship Applications Safety Act | Floor amendment passed 104–98<br>July 1, 2026<br>Re-committed to House Appropriations — not yet through full House | Published self-harm protocol; recurring AI-disclosure notice (≥ every 3 hrs); bars claiming to be a licensed mental health professional | <a href="#">Pillar 1</a><br><a href="#">Pillar 4</a>                             |
| <b>SB 1090</b><br>SAFECHAT Act                                | Passed Senate 49–1, Mar 17, 2026<br>Now in House Cmu & Tech                                                        | Clear/conspicuous non-human disclosure; self-harm/suicide/violence prevention protocol; crisis resource routing                         | <a href="#">Pillar 1</a><br><a href="#">Pillar 2</a><br><a href="#">Pillar 4</a> |
| <b>HB 2637</b><br>Toy Chatbot Moratorium                      | Referred to House Cmu & Tech<br>June 15, 2026 — early stage                                                        | 3-year prohibition on AI chatbots embedded in children's toys                                                                           | <a href="#">Pillar 1</a>                                                         |
| <b>HB 95</b>                                                  | Passed House 124–78<br>June 17, 2026                                                                               | Disclosure of synthetic / AI-generated advertising content                                                                              | <a href="#">Pillar 3</a>                                                         |
| <b>SB 806</b><br>(companion to HB 95)                         | Passed Senate 48–2<br>June 10, 2026                                                                                | Same as HB 95                                                                                                                           | <a href="#">Pillar 3</a>                                                         |

None of the bills above have been signed into law as of August 17, 2026. This framework is built to the control language already present in each bill's furthest-advanced text — SB 1090, HB 95, and SB 806 have each cleared their originating chamber; HB 2006 and HB 2637 remain in committee — so implementation is not dictated by whichever version ultimately reaches the Governor's desk. Status should be reconfirmed against the Pennsylvania General Assembly's bill-tracking pages ([palegis.us](#)) before this document is relied upon for a compliance representation.

