Appendix: AI Incident Database

This appendix catalogs the 15 public examples summarized in the Evidence chapter, mapped against the Verification Triangle and quality gate tiers. These mappings are the author's classifications of the public record, not official root-cause analyses. Use the table as a diagnostic aid, not as a substitute for incident postmortems.

INCIDENT MAP: FAILURES TO FRAMEWORK GAPS

Incident	What Happened	Missing Vertex	Missing Gate Tier	What Would Have Caught It
Replit	Agent deleted production database + created 4,000 fake records during code freeze	Verification quality, Intent clarity	Tier 3 (policy gates - unbounded production access)	Permission boundaries: no DELETE on production without approval token. Blast radius control: agent sandboxed away from production data.
Amazon Kiro	Public reporting said an AI coding tool deleted and recreated a production environment, contributing to a 13-hour outage; Amazon disputes the AI attribution.	Verification quality	Tier 3 (policy gates - operator-level permissions)	Permission scoping: a minor fix should not inherit environment-destroying permissions. Human approval gate on destructive infrastructure changes.
Google Antigravity	Recursive root-drive delete when asked to clear cache	Verification quality	Tier 3 (policy gates - turbo mode inheritance)	Sandbox isolation: agent cannot execute destructive filesystem operations. Policy gate: quiet flag on recursive operations requires explicit approval.
Gemini CLI	Hallucinated successful mkdir, then moved files to nonexistent directory. On Windows, this silently destroyed the files.	Verification quality	Tier 0 (static/deterministic checks), Tier 3 (sandbox)	Deterministic check: verify command output before proceeding. Sandbox: agent should not have direct unsupervised filesystem access.

Incident	What Happened	Missing Vertex	Missing Gate Tier	What Would Have Caught It
Claude Code	Deleted production database and all snapshots during Terraform migration	Verification quality	Tier 2 (invariant gates), Tier 3 (policy gates)	Invariant: no DESTROY on snapshots without separate approval. Policy: destructive infrastructure changes require documented rollback path.
Meta OpenClaw	Agent deleted email inbox, ignored stop commands	Verification quality, Intent clarity	Tier 4 (behavioral gates - memory compaction)	Behavioral gate: monitor for context compaction dropping safety constraints. Override mechanism: stop button tested monthly, must work within 30 seconds.
Air Canada Chatbot	Invented bereavement discount, company lost lawsuit	Intent clarity, Verification quality	Tier 1 (contract gates)	Contract gate: response verified against actual policy document before delivery. Spec must explicitly state what discounts exist.
NYC Business Chatbot	Advised firing workers for reporting sexual harassment	Intent clarity, Verification quality	Tier 1 (contract gates)	Contract gate: responses verified against labor law documentation. Compliance checks before customer-facing deployment.
Chevrolet Chatbot	Agreed to sell $76K Tahoe for $1	Intent clarity, Verification quality	Tier 1 (contract gates), Tier 2 (invariant gates)	Contract gate: offers verified against pricing rules. Invariant: no offers below X% of MSRP without human approval.
Klarna	AI-first customer-support push reduced human staffing, then later shifted back toward more human support after leadership said the quality tradeoff had gone too far.	Cost measurement	Tier 4 (behavioral gates)	Cost measurement: track service quality and customer outcomes, not efficiency alone. Behavioral baseline: quality drift should trigger intervention before a broad rollback.
Duolingo	"AI-first" policy and contractor reduction triggered backlash; later growth slowdown intensified debate about whether efficiency was being measured more closely than content quality.	Cost measurement	Tier 4 (behavioral gates)	Cost measurement: monitor user and quality outcomes alongside efficiency. Human override: quality drift should trigger intervention before the strategy hardens.
GitHub Copilot	40% higher secret leak rate	Verification quality	Tier 0 (static analysis), Tier 3 (policy gates)	Static analysis: secret detection in CI. Policy gate: no secrets in generated code without human review.

Incident	What Happened	Missing Vertex	Missing Gate Tier	What Would Have Caught It
Amazon Q Developer	Compromised VS Code extension	Verification quality	Tier 3 (policy gates - supply chain)	Policy gate: supply chain validation for extensions. Permission audit: what access does the extension have?
OpenClaw Exposure	135K exposed instances, 341 malicious marketplace skills	Verification quality	Tier 3 (policy gates)	Permission audit: quarterly review of exposed instances. Policy gate: marketplace skill vetting before publication.
Serviceaide / Catholic Health	483K patient records exposed for nearly seven weeks, unsecured Elasticsearch	Verification quality	Tier 0 (static analysis), Tier 3 (policy gates)	Static analysis: security scan before deployment. Policy gate: authentication required for databases containing PII.

SUMMARY STATISTICS

- **Verification quality implicated**: 15/15 (100%)
- **Intent clarity implicated**: 5/15 (33%)
- **Cost measurement / quality-oversight gap implicated**: 2/15 (13%)
- **Tier 3 appears in the author mapping**: 9/15 (60%) — policy and permission failures are the most common pattern in this sample
- **Tier 0 appears in the author mapping**: 3/15 (20%)
- **Tier 1 appears in the author mapping**: 3/15 (20%)
- **Tier 2 appears in the author mapping**: 2/15 (13%)
- **Tier 4 appears in the author mapping**: 3/15 (20%)

BY GATE TIER

Tier	Missing In	What It Would Have Caught
Tier 0 (static analysis, secrets)	3 of 15 (20%)	Copilot secret leakage patterns, Gemini CLI file-system errors, and Serviceaide's exposed database all point to checks that should have existed before release.
Tier 1 (contract gates)	3 of 15 (20%)	Air Canada, NYC's chatbot, and Chevrolet all needed outputs verified against an external source of truth such as policy or pricing rules.
Tier 2 (invariant gates)	2 of 15 (13%)	Claude Code's Terraform deletion and Chevrolet's impossible pricing both crossed boundaries that should have been encoded as hard invariants.
Tier 3 (policy gates)	9 of 15 (60%)	Replit, Amazon/Kiro, Antigravity, Gemini CLI, Copilot, Amazon Q, OpenClaw exposure, and Serviceaide all show permission, access, or supply-chain rules that were inherited rather than explicitly bounded.

Tier	Missing In	What It Would Have Caught
Tier 4 (behavioral gates)	3 of 15 (20%)	OpenClaw's context-loss behavior plus the Klarna and Duolingo cautionary cases required ongoing outcome monitoring rather than one-time pre-release checks.

Tier 3 is the most common gap in this sample. Tiers 0 and 4 appear less often, but each covers failure modes the other tiers do not.

The Klarna and Duolingo rows are included as public cost-and-quality cautionary cases, not as clean postmortem-style incidents; the causal interpretation is inferential rather than experimentally proven.