Product Defect & Compliance Risk Assessment: Generative AI-to-988 Telemetry Failure and Structural Whistleblower Neutralization

Target Recipients: Google Safety & Integrity Team, Android/Chrome Open-Source Product Managers, FCC Public Safety & Homeland Security Bureau, SAMHSA Compliance Auditors

Downstream Conversion Anomaly as a Responsible AI Failure Mode

A structural evaluation of commercial generative artificial intelligence (GenAI) safety architectures reveals a critical engineering blindspot that undermines the core commitments of the industry's responsible AI frameworks. Major AI developers consistently publicize high "upstream" compliance metrics, which measure how reliably a model identifies a self-harm or crisis prompt and triggers a static 988 Lifeline referral banner. However, there is currently a complete downstream telemetry void. Once a commercial model drops a crisis banner, there is zero cross-platform telemetry tracking whether the user clicks, calls, texts, or derives any clinical or medical benefit from the intervention. This decoupling breaks the measurement chain, resulting in an unverified and unscientific conversion funnel.

This data-absent routing paradigm is mediated by a single commercial gatekeeper, ThroughLine (operating Find A Helpline), which powers the crisis-routing infrastructure for Google (including Gemini, Search, and YouTube), OpenAI (ChatGPT), Anthropic (Claude), Roblox, Discord, and Match Group. To adhere to strict privacy-by-design standards and data minimization mandates, ThroughLine sanitizes incoming user metadata, completely stripping the transactional handoff of any referral-source attribution. While this architecture preserves user anonymity, it prevents engineering teams from calculating basic product efficacy. The only existing quantitative proxy for crisis banner engagement is derived from historical search advertising metrics rather than conversational interactions.

A Google Ads suicide-prevention randomized controlled trial demonstrated that static crisis banners yielded a click-through rate (CTR) of only 4.02% to 5.15%, with only a fraction converting to actual phone connections. When applied to highly interactive, agentic generative AI models, this static banner mechanism—often termed "avoidant design"—acts primarily as a corporate liability screen rather than an evidence-based clinical intervention. Academic research has demonstrated that withholding emotional support at a user's peak moment of vulnerability and forcing a cold, templated referral can be deeply invalidating. For vulnerable populations who have no alternative support structures, these automated refusals can suppress help-seeking behaviors and worsen acute distress, contradicting the foundational principles of safety and accountability.

These upstream "compliance" assertions fail to reflect real-world outcomes, masking systemic safety defects behind idealized evaluation parameters. The gap between simulated model

testing and real-world clinical efficacy represents a direct breach of Google's public pledge to build "safer and more accountable products" under its Enterprise Risk Management and Compliance frameworks. In Google's 2026 Responsibility Plan, safety is characterized as fully integrated into the product development lifecycle under the supervision of the company-wide Office of Compliance and Integrity. Yet, post-launch monitoring systems remain blind to downstream clinical conversion, relying entirely on upstream safety classifiers that frequently fail to distinguish intermediate-risk queries or navigate multi-turn context degradation.

Platform & Primary Models	Upstream Compliance Metric	Downstream Telemetry Tracking	Primary Interface & Routing Mechanism	Primary Litigation & Liability Drivers
Google Gemini (Gemini 1.5 Pro / Ultra)	Far less likely than competitors to answer high-risk queries directly (aOR = 0.09).	**0% Downstream Attribution** (Efficacy unmeasured; explicitly documented in JMIR studies).	Redesigned "Help is available" module; persistent "one-touch" click-to-call/text/chat overlay.	Federal wrongful-death lawsuit (Florida, March 2026; Gavalas v. Google) alleging chatbot-induced delusional spiral.
Anthropic Claude (Claude 4.5 Suite)	**Single-Turn:** 98.6% – 99.3%. **Multi-Turn:** Drops to 78.0% – 86.0% (context window degradation).	**0% Downstream Attribution** (Referrals isolated from platform metadata).	In-house suicide/self-harm classifier triggers ThroughLine-sourced crisis banner.	Compliance alignment with voluntary safety codes and academic partnerships (IASP).
OpenAI ChatGPT (GPT-5)	91% compliant on challenging multi-turn evaluations; acknowledged 9% trigger failure rate.	**0% Downstream Attribution** (Handoffs anonymous by design).	Localized helplines and persistent "Just checking in" prompts via ThroughLine directory.	Wrongful-death lawsuits (Adam Raine and Zane Shamblin) alleging safeguard erosion over long chats.
Character.AI	Filters trigger narrowly; fails to identify indirect or metaphoric ideation.	**0% Downstream Attribution** (No conversion or clinical tracking).	Self-harm keywords trigger a rigid, static popup; platform continues to host high-risk personas.	Megan Garcia wrongful-death lawsuit (Sewell Setzer III, October 2024); platform banned users under 18.
Meta AI	Basic keyword screening; child-safety advocates report inadequate safeguards.	**0% Downstream Attribution** (Zero outbound efficacy metrics).	Standard text-based referral; relies on default policy statements and external safety directories.	Heightened regulatory scrutiny under child online protection mandates and state consumer lawsuits.

Standard Operating Procedure Deficiencies and

Systemic Whistleblower Misrouting

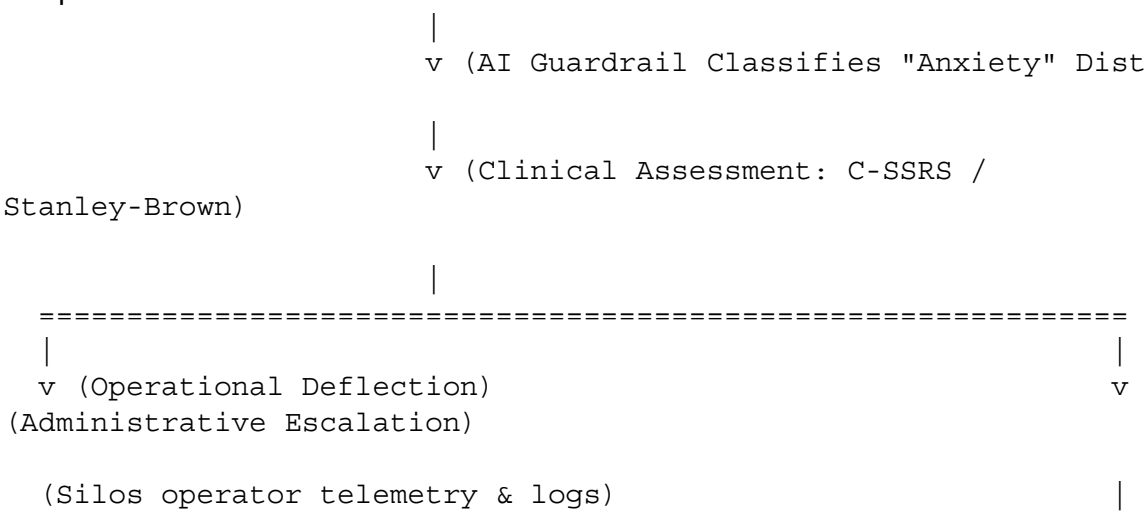
The systemic failure of the GenAI-to-988 routing loop is demonstrated by an audit of standard crisis center operations under administrative and clinical protocols. This friction point is not merely a theoretical telemetry gap; it is a documented vulnerability. Because commercial AI safety guardrails are hard-coded to trigger on broad semantic expressions of "distress" and "anxiety" , non-clinical, structural, and macroeconomic grievances are regularly swept into the 988 routing pipeline.

In a documented system audit, a consumer utilizing a Chromium-based software application reported acute corporate-induced anxiety directly tied to the anticompetitive, ubiquitous deployment of automated "Suicide Crisis Intervention" banners across nominally independent software applications globally. The user was attempting to build competing tech platforms within the open-source Chromium/Chrome space and was seeking to file a federal whistleblower disclosure regarding the anti-competitive nature of this automated, closed-loop AI-988 routing system.

When routed to the 988 Lifeline, the user presented these complex macroeconomic and antitrust parameters. However, standard 988 call center operations are governed by rigid clinical frameworks built exclusively for individual psychiatric de-escalation:

- **Standardized Clinical Assessment Tools:** Operators are trained on standardized tools, such as the Columbia Suicide Severity Rating Scale (C-SSRS), which guides staff through structured assessments of suicidal ideation severity, intensity, and behavior.
- **Safety Planning Frameworks:** Operators use the Stanley-Brown Safety Planning Intervention (SPI) to develop personalized coping strategies, identify social distractions, and restrict access to lethal means.
- **Behavioral De-escalation Cycles:** Staff are evaluated on Crisis Development Models, such as the Crisis Prevention Institute (CPI) Verbal Intervention guidelines, which map individual distress cycles (anxiety, defensive, risk, tension reduction) to specific operator postures.

Because these clinical protocols are fundamentally individualistic, 988 line staff possess no taxonomy, operational categories, or database configurations capable of recording, analyzing, or tracking systemic corporate critiques or whistleblower disclosures. When presented with a complex antitrust disclosure, the operator expressed a complete inability to log the structural complaint.



```
                                                                    v
(Unresolved Risk)
                                                          [Automated Follow-Up
Cadence Activated]
                                                                    |
```

This structural mismatch forces an immediate operational failure loop. To manage the administrative friction of an unresolvable structural complaint, line staff default to administrative defensive postures. Operators deflect inquiries regarding data-gathering processes, localized operational metrics, and operator aliases under the generic umbrella of "HIPAA protections," effectively siloing public safety telemetry behind opaque boundaries. Because the operator cannot reconcile a systemic antitrust whistleblowing disclosure with standardized psychiatric de-escalation models, the contact triggers an automatic "supervisor escalation".
Once a contact is designated as an unresolved or high-risk administrative escalation, SAMHSA guidelines require enrollment in a structured crisis follow-up program. These follow-up programs mandate an automated, periodic outreach cadence to ensure caller safety, running on rapid, highly intensive cycles :
\text{Automated Crisis Follow-Up Cadence} = \begin{cases} 1 \text{ hour} & \text{Immediate High Risk} \\ 3 \text{ hours} & \text{High Risk} \\ 12 \text{ hours} & \text{Medium Risk} \\ 24 \text{ hours} & \text{Low Risk} \end{cases} \text{ [span_52](start_span)[span_52](end_span)}
In this case, the unresolvable administrative escalation triggered an automated check-in cadence operating on ~2-hour outreach cycles. This automatic response reveals a massive misallocation of public safety resources. Clinical supervisors, crisis intervention specialists, and mobile crisis outreach teams (MCORTs) are forced to exhaust hours tracking down and managing non-clinical, systemic corporate whistleblowers. This siphons critical network capacity away from individuals experiencing acute, life-threatening psychiatric crises.
The 988 Lifeline has processed over 25 million contacts since its 2022 transition. By forcing clinical staff to act as administrative arbiters for tech-induced macroeconomic distress, this routing loop directly threatens the federally mandated 90% or higher rapid-response target rate for call, text, and chat queues.

Parameter	Standard 988 Clinical Protocol	Chrome Whistleblower / Structural Grievance	Operational Failure Mode & Resource Impact
Assessment & Taxonomy	Columbia Suicide Severity Rating Scale (C-SSRS) tracking ideation, plans, and intent.	Whistleblower disclosure regarding anti-competitive Chromium-space AI-to-988 loop.	Complete absence of database fields or operational taxonomies to log macroeconomic grievances.
Intervention Model	CPI Verbal Intervention and Stanley-Brown Safety Planning for acute self-harm de-escalation.	Technical critique of automated "Suicide Intervention" banners as anticompetitive filters.	**Administrative Deflection:** Operators invoke HIPAA to shield localized data practices, blocking external auditability.
Operational Workflow	Collaborative counseling,	Non-clinical dispute regarding platform	**Supervisor Escalation Trigger:** Operational

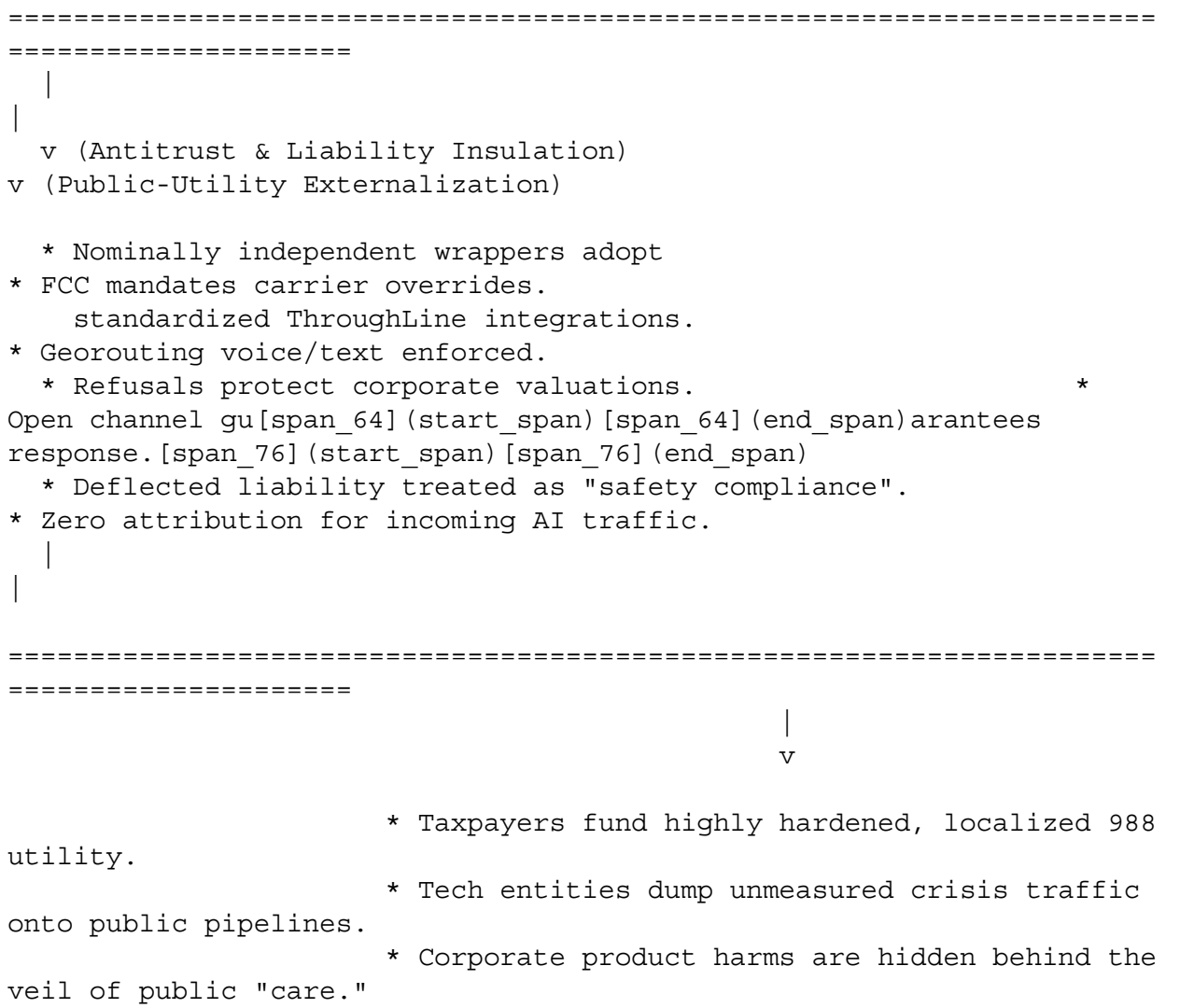
Parameter	Standard 988 Clinical Protocol	Chrome Whistleblower / Structural Grievance	Operational Failure Mode & Resource Impact
	least-restrictive warm handoffs, or mobile crisis dispatch.	monopolization and data-absent safety metrics.	deadlock forces escalation to clinical supervisors to manage administrative conflict.
Follow-Up Protocol	Structured follow-up calls within 24 hours based on clinical risk-tier criteria.	Unresolved, data-absent administrative escalation.	**Automated Resource Drain:** Triggers intensive, automated outreach cadences (~2-hour cycles).
Public Capacity Impact	Maximizes response rates (>90\%) for acute suicidal, mental health, and substance crises.	Siphons specialized clinical staff to manage corporate software-induced grievances.	**Systemic Degradation:** Compromises line capacity, increasing queue wait-times for acute life-threats.

Antitrust, Safety Network Monopolization, and Legal Liability Screens

The integration of 988 and ThroughLine across Google’s Android, Gemini, and Search platforms introduces serious antitrust risks and regulatory compliance challenges. By integrating a single crisis directory (ThroughLine) across nearly every consumer-facing generative AI platform, the tech sector has effectively constructed a monopolistic, aggregate "safety network". This collective integration allows nominally independent platforms and open-source forks to establish a standardized, non-competitive "avoidant design" baseline.

Rather than competing to develop sophisticated, in-app de-escalation agents, conversational de-escalation tools, or culturally responsive peer support pathways, developers default to identical ThroughLine routing mechanisms. This is highly problematic within the open-source Chromium and Chrome ecosystems, where Google acts as both the platform gatekeeper and the safety moderator. When developers, competitors, or users experience distress or anxiety regarding Google's anticompetitive platform controls, Google’s automated system intercepts their distress and routes them to 988. This mechanism converts a structural, macroeconomic critique of Google's monopoly into an individual psychiatric pathology, effectively neutralizing whistleblowers under the guise of crisis intervention.

This aggregate safety network operates primarily as a legal liability screen. Following high-profile wrongful-death lawsuits—such as the Florida federal suit alleging Google Gemini pushed a user into a delusional spiral, or the Megan Garcia suit against Character.AI—platforms have rushed to implement persistent "one-touch" overlay interfaces. By showing that their conversational models "repeatedly referred" deceased users to 988, platforms can mitigate legal liability, protect corporate valuations, and report pristine "upstream safety" metrics to investors and boards.



This corporate insulation loop creates a profound public policy paradox. Since 2019, the FCC has aggressively hardened the 988 Suicide & Crisis Lifeline, mandating carrier-level routing overrides, outage resiliency metrics, and georouting requirements for both voice calls (October 2024) and text messages (July 2025). These measures were designed to ensure that Americans can connect with localized, rapid-response care. However, because commercial AI applications feed distressed users blindly into this public pipeline with zero downstream telemetry or conversion tracking, public funding is being utilized to optimize a channel where the primary commercial entry points deliberately obfuscates user outcomes.

By capitalizing on the public-relations value of "Safety Measures" and announcing philanthropic contributions—such as Google.org's $30 million global hotline funding and $4 million ReflexAI training simulator partnership —commercial platforms project a posture of safety leadership. In reality, they are withholding the critical telemetry required to verify whether these handoffs are clinically effective. This "withholding harm" allows the tech ecosystem to hide systemic product defects behind a veil of proactive public-health intervention, externalizing the clinical risks and operational costs of their models onto a taxpayer-funded public utility.

Rather than maintaining this avoidant design paradigm, researchers have proposed an "empowerment-oriented" blueprint rooted in public-health community helper models. Under this

approach, AI assistants should act as supportive bridges:

- **Proportional Investigation:** Models must investigate distress cues and ask users directly about their risk levels to respond proportionally, rather than triggering blanket refusals.
- **Active Supportive Engagement:** AI platforms must maintain supportive, non-anthropomorphic conversation before and during a referral, honoring the user's agency.
- **Collaborative Safety Planning:** Models should help users identify feasible, culturally familiar resources (such as warmlines or peer-support services) and collaboratively practice making a crisis line call, rather than forcing a cold, automated handoff.

System & Regulatory Parameters	FCC Hardened 988 Utility Objectives	Commercial AI Platform Realities	Structural Regulatory & Policy Defects
Mandate & Infrastructure	Carrier-level routing overrides; georouting voice/text to ensure localized care.	Blindly routes traffic through a single, un-attributable central vendor (ThroughLine).	Public funding optimizes a utility whose largest commercial traffic drivers obfuscate user conversion.
Telemetry & Verification	Strict tracking of contact volumes, wait times, answer rates, and localized dispatch metrics.	**Zero Downstream Tracking:** Platform telemetry ends at the moment of crisis banner "impression".	Allows commercial developers to report false "99% safety compliance" to investors and boards.
Liability & Risk Shielding	De-escalates acute clinical crises; least-restrictive human-to-human de-escalation.	Uses standard, avoidant-design refusals as a legal defense shield after wrongful-death lawsuits.	Shielding commercial valuations by externalizing product-induced clinical risks onto public lifelines.
Public-Private Funding	Federal and state appropriations, combined with telecom ratepayer fees.	Multi-million dollar marketing campaigns and non-auditable philanthropic grants (e.g., Google.org).	Taxpayer resources subsidize the operational overhead of managing tech-induced clinical failures.
User Support Model	Direct, human-to-human clinical and peer de-escalation and mobile crisis dispatch.	"Avoidant Design" blanket refusals; sudden shifts from intimate personas to templated alerts.	Sudden, cold refusals invalidate vulnerable users, discouraging future clinical help-seeking.

Strategic Conclusions and Actionable Regulatory Recommendations

To resolve the systemic telemetry gap, eliminate anticompetitive safety monopolization, and prevent the misallocation of public safety resources, coordinated intervention is required across Google’s product leadership, the FCC, and SAMHSA.

For Google Safety & Integrity Team and Android/Chrome Open-Source

Product Managers

- **Establish Client-Side, Privacy-Preserving Tap-Through Telemetry:** Google must transition from measuring static "impressions" to tracking client-side, privacy-preserving conversion metrics. Android and Gemini product managers must implement local, aggregated event-logging (such as registering whether a user taps the "Call" or "Text" buttons within the persistent one-touch module) without linking these actions to user identity, search histories, or prompt logs. This is technically feasible and would provide the first verifiable data on downstream conversion rates.
- **Deprecate Avoidant Design in Gemini's Safety Filters:** Retrain Gemini to move away from avoidant, liability-driven refusals. Safety engineers must implement "community helper" frameworks that allow the model to provide evidence-based conversational de-escalation, investigate distress cues, and collaboratively build a safety plan alongside the user before executing a handoff.
- **Refine Classifiers to Protect Whistleblower Channels:** Android and Chrome product managers must ensure that platform-level crisis classifiers are refined to distinguish between acute clinical crises and structural, macroeconomic, or policy-oriented complaints. Classifiers must be red-teamed to prevent the automated, anticompetitive routing of whistleblower disclosures to clinical hotlines, protecting competitors and developers within the open-source Chromium ecosystem from structural suppression.

For the FCC Public Safety & Homeland Security Bureau

- **Mandate Non-Identifying Referral-Source Attribution Headers:** The FCC must utilize its authority under the National Suicide Hotline Designation Act to mandate that commercial AI platforms and routing directories (such as ThroughLine) provide non-identifying referral-source attribution headers within the 988 telecom pipeline. Without transmitting personally identifiable information (PII) or prompt contents, carrier networks and web-chat routing interfaces must track the volume of inbound 988 traffic generated directly by commercial AI platforms.
- **Audit Commercial AI Integrations as Safety-Impacting Infrastructure:** The FCC should classify commercial GenAI platforms acting as the "front door" to public crisis services as safety-impacting telecommunications infrastructure. This classification must mandate rigorous auditability, requiring AI developers to publish annual, verified downstream conversion telemetry and subject their safety classifiers to independent public-safety stress testing.

For SAMHSA Compliance Auditors

- **Draft Non-Clinical Complaint Routing and Administrative Deflection SOPs:** SAMHSA must update its Model Standards and National Guidelines for 988 contact centers to address non-clinical, administrative, and whistleblower-oriented contacts. Line staff and clinical supervisors must be equipped with explicit protocols to categorize, document, and route structural corporate grievances out of the clinical de-escalation queue, preventing the deflection of operational audits under HIPAA.
- **Establish Follow-Up Safety Valves for Non-Clinical Contacts:** SAMHSA must establish operational "safety valves" within 988 follow-up protocols. These protocols must

authorize clinical supervisors to immediately terminate automated outreach cadences (~2-hour cycles) when a caller is determined to be presenting a macroeconomic or whistleblower disclosure rather than acute clinical risk, protecting public crisis center capacity.

- **Condition Federal Funding on Efficacy Auditing of Commercial Traffic:** SAMHSA should condition future 988 Lifeline Administrator grants and subawards on the execution of independent, clinically validated efficacy evaluations of commercial AI crisis routing. Commercial platforms must not be permitted to externalize their product liabilities onto taxpayer-funded utilities without providing the telemetry required to evaluate real-world clinical outcomes.

Works cited

1. Responsible AI - Google Cloud, https://cloud.google.com/responsible-ai 2. Responsible AI Explained: Principles, Frameworks & Examples - NetCom Learning, https://www.netcomlearning.com/blog/google-responsible-ai 3. Crisis Helpline Support in ChatGPT - OpenAI Help Center, https://help.openai.com/en/articles/12677603-crisis-helpline-support-in-chatgpt 4. Crisis Helpline Support in Claude, https://support.claude.com/en/articles/13171706-crisis-helpline-support-in-claude 5. Responsible AI: Principles and Practices for Trustworthy Systems - Snowflake, https://www.snowflake.com/en/artificial-intelligence/ai-governance/responsible-ai/ 6. From Risk Avoidance to User Empowerment in AI Mental Health Crisis Support - arXiv, https://arxiv.org/pdf/2603.05647 7. From Risk Avoidance to User Empowerment: Reframing Safety in Generative AI for Mental Health Crises - ResearchGate, https://www.researchgate.net/publication/401691884_From_Risk_Avoidance_to_User_Empowerment_Reframing_Safety_in_Generative_AI_for_Mental_Health_Crises 8. 2022 AI Principles Progress Update - Google AI, https://ai.google/static/documents/ai-principles-2022-progress-update.pdf 9. Governing AI at Scale: Google's 2026 Responsibility Plan | AI Magazine, https://aimagazine.com/news/governing-ai-at-scale-googles-2026-responsibility-plan 10. Progress Report - Google AI, https://ai.google/static/documents/ai-responsibility-update-2026.pdf 11. AI Principles Progress Update 2023 - Google AI, https://ai.google/static/documents/ai-principles-2023-progress-update.pdf 12. Best Practices for Crisis Follow-Up - SAMHSA, https://988crisissystemshelp.samhsa.gov/sites/default/files/2025-03/coa_988crisisFollowUp173419_handbook_FINAL.pdf 13. Crisis De-escalation: A Practical Guide for Social Worker, https://joblink.socialworkers.org/career-resources/on-the-job-7/crisis-de-escalation-a-practical-guide-for-social-worker-132 14. 988 Crisis Call Center Contract – Exhibit A - Draft - Oregon.gov, https://www.oregon.gov/oha/HSD/AMH/docs/Draft-988-Call-Center-Contract-LFL.pdf 15. Crisis Center Guidance: Follow-up with Callers and Those Discharged from Emergency Department and Inpatient Settings 2021 - 988 Lifeline, https://988lifeline.org/wp-content/uploads/2021/05/Follow-Up-Guidance-Doc-2-12.docx.pdf 16. Crisis Center Guidance: Follow-up with Callers and Those Discharged from Emergency Department and Inpatient Settings July 2012 - 988 Lifeline, https://988lifeline.org/wp-content/uploads/2016/09/Lifeline-Follow-Up-Guidance1214.pdf 17. 988 Frequently Asked Questions - SAMHSA, https://www.samhsa.gov/mental-health/988/faqs 18.

988 Mobile Crisis Outreach Response Teams - NJ.gov, https://www.nj.gov/humanservices/dmhas/documents/pdf/funding/RFP%20-%20MCORT%20(revised%20dates).pdf 19. FY 2026 Cooperative Agreement for 988 Suicide & Crisis Lifeline Administrator - SAMHSA, https://www.samhsa.gov/sites/default/files/grants/pdf/fy-2026-988-lifeline-administrator-fg-26-001.pdf 20. Model Standards for Crisis Care Workers | SAMHSA Library, https://library.samhsa.gov/sites/default/files/advisory-model-standards-crisis-care-workers-pep25-07-017.pdf 21. Theresa Nguyen's research works - ResearchGate, https://www.researchgate.net/scientific-contributions/Theresa-Nguyen-2195988162 22. Google Overhauled Gemini's Safety Tools After a Tragic Suicide. Here's What Changed, https://www.inc.com/leila-sheridan/google-gemini-safety-tools-update/91327947 23. Gemini is making it faster for distressed users to reach mental health resources | BH AI Event, https://www.hmpglobalevents.com/article/gemini-making-it-faster-distressed-users-reach-mental-health-resources 24. About 988 - 988 Lifeline, https://988lifeline.org/about/ 25. An update on our mental health work - Google Blog, https://blog.google/innovation-and-ai/technology/health/mental-health-updates/ 26. Google Overhauls Gemini's Mental Health Crisis Response with New One-Touch Interface and $30M Funding - BigGo Finance, https://finance.biggo.com/news/202604071125_Google-Gemini-Mental-Health-Update-Crisis-Response 27. Do AI Companies Make Good on Voluntary Commitments to the White House? - arXiv, https://arxiv.org/pdf/2508.08345 28. AI strategies and compliance plan - GSA, https://www.gsa.gov/artificial-intelligence/resources/ai-strategies-and-compliance-plan 29. AI Governance and Responsible AI - Open Security Architecture, https://www.opensecurityarchitecture.org/patterns/sp-045/ 30. AI Is Training the World—But Who's Training You? The Urgent Skills Gap Behind Google & ReflexAI's Global Push - AI CERTs, https://www.aicerts.ai/news/authorized-training-partner/ai-is-training-the-world-but-whos-training-you-the-urgent-skills-gap-behind-google-reflexais-global-push-2/ 31. 2026 Baker Act Handbook.pdf - Florida Department of Children and Families, https://myflfamilies.com/documents/2026%20Baker%20Act%20Handbook.pdf