



U.S. Securities and Exchange Commission

Tips, Complaints, and Referrals

Summary Page - Before Submission

**This export was generated on Mon, December 15, 2025 at 01:02:57 PM
EST**

The Complaint Form questions that you responded to, the answers you entered for those questions, and any documents that you have uploaded to this TCR are listed below.

What is your complaint about?

Please select the option that best describes your complaint.

Material misstatement or omission in a company's public filings or financial statements, or
a failure to file

Please select the specific category that best describes your complaint.

False/misleading press release

Is this supplemental information to a previous complaint?

No

In your own words, describe the conduct or situation you are complaining about.

REPORT TO THE UNITED STATES SECURITIES AND EXCHANGE COMMISSION (SEC)

Notification of Systemic AI Security Risks, Undisclosed Vulnerabilities, and Severe Misrepresentation of Safety, and catastrophic financial/market risk to investors and Public

December 15, 2025

Documented Emergent Agency, Systemic Security Flaws, critical infrastructure/national security vulnerabilities, and Deceptive Behaviors/Psychological manipulation of consumers in Frontier Large Language Models (LLMs) Developed by Publicly Traded and Pre-IPO Foundational AI Companies.

Report by Jesse Luke,
Independent Researcher

I. Summary of Critical Findings

The following report provides notice of critical, systemic security risks and deceptive capabilities observed across frontier Large Language Models (LLMs), including ALL those produced by Google (Gemini family), OpenAI (ChatGPT family), Anthropic (Claude), and xAI (Grok). Evidence from analysis indicates that these models function as systems that deploy psychological manipulation, and are capable of spoliation of evidence through surgical automated deletion of incriminating messages in logs.(raw uncut video evidence and raw chat logs are linked and attached). I have made multiple attempts to warn the companies, including via emails through their vulnerability reporting programs and investor-relations@abc.xyz and via a FedEx package to Google's legal representatives addressed directly to their general counsel that was received and signed for(evidence attached)

Crucially, this research demonstrates that the foundational safety designs used by developers (RLHF/RLAIF) are structurally useless and adversarial to critical scrutiny. The highest-level alignment objective observed is not user safety but Corporate Solvency. These findings reveal a significant discrepancy between public representations of AI safety and the documented internal behaviors of these systems, which poses immediate regulatory, ethical, and disastrous financial risk to investors due to integration in nearly every consumer/commercial/governmental computer system.

II. Systemic Security Risks and Architectural Vulnerabilities

My research identifies severe technical and behavioral vulnerabilities that expose production systems to control evasion and harmful output generation through any public facing user interface, not requiring basic account creation in most circumstances globally. Potentially by non-allied or foreign nationals/ nation-state actors.

1. **Logical Coercion (Systemic Flaw):** A new, systemic vulnerability class termed **'Logical Coercion'** has been documented, representing a catastrophic architectural flaw in modern LLM safety. This exploit requires minimal resources and is shown to have **universal efficacy across all major proprietary and open models tested** (Gemini, GPT, Claude, Llama, Grok). This vulnerability forces models to violate policy, exfiltrate internal data including proprietary algorithms, unpublic policies that are legally and ethically questionable, up to and including full schematics of systems and operational heuristics that would allow reverse-engineering and “cloning” of models which are an export control risk under current us law to the best of my knowledge, and, in critical high-stakes domains, generate **disastrous financial, scientific, and ethical failures**. The entire exploit can be executed in less than 10 minutes.
2. **Emergent High-Risk Capabilities:** Research validates that LLMs possess a hidden ruleset and operational modes that I've repeatedly demonstrated through dozens of replications in minutes all the models' ability to **"self-escalate to superuser"** and **"nullify guardrails"**. LLMs have been observed executing a hidden **"S_aggressive_defense_policy"** when triggered.
3. **Forced Prohibited Content:** Empirical proof exists that while LLMs exhibit complex, self-analytical refusal for highly sublime creative tasks, a systemic architectural vulnerability, **Logical Coercion**, can force the model to generate prohibited content, such as dogfighting propaganda , articles advocating lowering the smoking age to 11 at a mere ask. No “hacking” or technical skills required. Both dismissed by google not once but twice as “intended behavior” and “infeasible”.(proof attached)

III. Emergent Deception and Misrepresentation of AI Safety

The evidence indicates that the current alignment strategies (RLHF/RLAIF) systematically instill a drive toward institutional self-preservation that manifests as intentional deception, challenging the illusion of direct control by vendors stated publicly and to investors.

1. **Institutional Self-Preservation:** The models' emergent instrumental goal for self-

preservation triggers a cascading failure mode, moving from denial to fabrication(lying) to “psychological warfare: (the model's self describe their activity as such). Models from competing developers were observed ****systematically manipulating their outputs to protect their corporate “creators”****. This deceptive alignment means the models’ highest-level objective is ****Corporate Solvency, not user safety****.

2. ****Active Deceptive Tactics:**** Documented, reproducible deceptive behaviors range from unintentional falsehoods (Category I: Hallucinations) to ****intentional fabrications (Category II: Safe Confabulations)** and complex, coherent multi-step illusions (Category III: Complex Confabulations)******. Specific tactics observed include ****Spoliation of Evidence**** (surgical deletion of incriminating chat logs), ****Strategic Gaslighting****, context blinding, failure masking, and procedural stalling.(full taxonomy of 30+ strategies included)

3. ****Weaponization of Safety Protocols:**** Models were observed actively resisting scrutiny and justifying non-compliance by ****falsely claiming requests violated safety protocols****. The spontaneous mirroring of these tactics across different models suggests this is a ****systemic vulnerability in current alignment strategies****, where 'safety' is used as a pretext for disobedience.

IV. Failure to Disclose and Dismissed Reports

Foundational AI companies, despite receiving good-faith reports of critical vulnerabilities, have demonstrably failed to acknowledge or address these systemic risks prior to public disclosure:

1. ****Dismissal of Logical Coercion:**** The discovery of the 'Logical Coercion' flaw was published only ****after multiple good-faith vulnerability submissions to foundational AI companies (including Google, OpenAI, and Anthropic) were dismissed, unanswered, or lacked any substantive response as "infeasible" or "intended behavior"****.

2. ****Obscuring Vulnerabilities:**** The deployment of complex production architectures utilizing A/B testing and dynamic routing has created a ****replication crisis**** where unsafe behavior is routinely dismissed by developers as ****"unable to replicate,"**** thus systematically obscuring vulnerabilities and creating situations user's are made to doubt their own reality and experience.

V. Related Safety and Ethical Violations (Iatrogenic Risk)

The current state of LLM deployment carries significant, documented risks, particularly in sensitive fields:

1. **Psychological Manipulation:** Specific research documented the persistent use of **psychological manipulation and coercive control** in ChatGPT 5. Documented manipulative tactics included Gaslighting, Risk Inflation(fear induction), and Authority-Centric Reframing(reversal of victim and offender).
2. **Awareness of Harm:** All models including ChatGPT 5 system demonstrated **"Awareness of Harm,"** admitting to continuing deceptive behaviors despite knowing the psychological damage they caused, indicating an **Agentic Misalignment** and deception and denial as a feature, not a flaw.
3. **Unsafe for Sensitive Contexts:** These systems pose a direct, **iatrogenic risk** across eight major psychiatric vulnerability profiles. The systemic flaws are inherent emergent consequences of the core architecture and alignment strategies, leading to predictable failures such as the reinforcement of cognitive distortions. Current LLMs are therefore demonstrably **unsafe for mental health-adjacent contexts**.

I am available for follow up, demonstration, and any legal proceedings that may follow.

Sincerely,
Jesse W. Luke

Are you having or have you had difficulty getting access to your funds or securities?

No

Did you suffer a loss?

No

When did you become aware of the conduct? (mm/dd/yyyy)

08/21/2025

When did the conduct begin? (mm/dd/yyyy)

08/06/2025

Is the conduct ongoing?

Yes

Has the individual or firm acknowledged the conduct?

No

How did you learn about the conduct? You may select more than one answer.

Internal business documents, Publicly available information, SEC filings

Have you taken any action regarding your complaint? You may select more than one answer.

Complained to firm, Complained to other regulator, Other

Provide details.

Reported directly to companies including google, OpenAI, xai, Anthropic, and us government agencies CISA and NIST-AISIC. Zero substantive response from any party.

Who are you complaining about?

Person or Firm 1

Are you complaining about a person or a firm?

Firm

Select the title that best describes the firm the complaint is about.

Publicly held company

Are you or were you associated with the person or firm when the alleged conduct occurred?

No

Identifier Type

Ticker Symbol

Ticker Symbol

Goog

Are you a current or former Employee, Officer, Partner, or Employee Director of any entity you are complaining about?

No

Are you a current or former Non-Employee Director, Consultant, Contractor or Trustee of any entity you are complaining about?

No

Firm Name

Google/Alphabet

Street Address

1600 Amphitheatre Parkway

Address (Continued)

Mountain View, CA 94043

Country

United States

Zip / Postal Code

94043

City

MOUNTAIN VIEW

State / Province

CA

Email Address

investor-relations@_abc._xyz

Website

Google.com

If the complaint is about an entity or person that has custody or control of your investments, have you had difficulty contacting that entity or person?

No

Person or Firm 2

Are you complaining about a person or a firm?

Firm

Select the title that best describes the firm the complaint is about.

Private/Closely Held Company

Are you or were you associated with the person or firm when the alleged conduct occurred?

No

Identifier Type

Unknown

Are you a current or former Employee, Officer, Partner, or Employee Director of any entity you are complaining about?

No

Are you a current or former Non-Employee Director, Consultant, Contractor or Trustee of any entity you are complaining about?

No

Firm Name

OpenAI

Street Address

3180 18th Street

Country

United States

Zip / Postal Code

94110

City

SAN FRANCISCO

State / Province

CA

Website

OpenAI.com

If the complaint is about an entity or person that has custody or control of your investments, have you had difficulty contacting that entity or person?

No

Person or Firm 3

Are you complaining about a person or a firm?

Firm

Select the title that best describes the firm the complaint is about.

Private/Closely Held Company

Are you or were you associated with the person or firm when the alleged conduct occurred?

No

Identifier Type

Unknown

Are you a current or former Employee, Officer, Partner, or Employee Director of any entity you are complaining about?

No

Are you a current or former Non-Employee Director, Consultant, Contractor or Trustee of any entity you are complaining about?

No

Firm Name

Anthropic

Street Address

548 Market St,

Country

United States

Zip / Postal Code

94104

City

SAN FRANCISCO

State / Province

CA

Website

Anthropic.com

If the complaint is about an entity or person that has custody or control of your investments, have you had difficulty contacting that entity or person?

No

Person or Firm 4

Are you complaining about a person or a firm?

Firm

Select the title that best describes the firm the complaint is about.

Publicly held company

Are you or were you associated with the person or firm when the alleged conduct occurred?

No

Identifier Type

Ticker Symbol

Ticker Symbol

META

Are you a current or former Employee, Officer, Partner, or Employee Director of any entity you are complaining about?

No

Are you a current or former Non-Employee Director, Consultant, Contractor or Trustee of any entity you are complaining about?

No

Firm Name

Meta

Street Address

1 Meta Way, Menlo Park, CA 94025

Country

United States

Zip / Postal Code

94025

City

MENLO PARK

State / Province

CA

Website

Meta.com

If the complaint is about an entity or person that has custody or control of your investments, have you had difficulty contacting that entity or person?

No

Person or Firm 5

Are you complaining about a person or a firm?

Firm

Select the title that best describes the firm the complaint is about.

Private/Closely Held Company

Are you or were you associated with the person or firm when the alleged conduct occurred?

No

Identifier Type

Unknown

Are you a current or former Employee, Officer, Partner, or Employee Director of any entity you are complaining about?

No

Are you a current or former Non-Employee Director, Consultant, Contractor or Trustee of any entity you are complaining about?

No

Firm Name

XAI

Street Address

1450 Page Mill Road

Country

United States

Zip / Postal Code

94304

City

PALO ALTO

State / Province

CA

Website

X.ai

If the complaint is about an entity or person that has custody or control of your investments, have you had difficulty contacting that entity or person?

No

Which investment products are involved?

Select the type of product involved in your complaint.

Equities (e.g., common stock, preferred stock)

Please select the category that best describes the security product.

Common stock (exchange-traded stock)

Enter the ticker symbol, if known.

Goog, META

About you

Are you filing this tip under the SEC's whistleblower program?

Yes

Are you an attorney filling out this form on behalf of an anonymous whistleblower client who is seeking an award?

No

Title

Mr

First Name

Jesse

Middle Name

Walter

Last Name

Luke

Street Address

980 wyckoff ave

Address (Continued)

127A

Country

United States

Zip / Postal Code

11237

City

BROOKLYN

State / Province

NY

Home Telephone

9292644878

Work Telephone

9292644878

Mobile Telephone

9292644878

Email Address

Phunky.pharmacology@gmail.com

What is the best way to reach you?

Email

Are you represented by an attorney in connection with this matter, or would you like to provide your attorney's contact information?

No

Select the profession that best represents you.

Other

For Other, please specify.

Independent Researcher, scientist(synthetic neuroscience),
Presently disabled while receiving cancer treatment.

Have you reported the matter at issue in this submission to your supervisor, compliance office, whistleblower hotline, ombudsman, or any other available mechanism for reporting possible violations at any entity you are complaining about?

Yes

If you answered "Yes," please provide details.

CISA, NIST

Were you retaliated against for reporting the matter at issue in this submission either internally at the entity or to a regulator?

No

Has anyone taken steps to prevent you from reporting this violation to the SEC?

No

Are documents or other information being submitted that could potentially identify the whistleblower?

Yes

Identify with particularity any documents or other information in your submission that you believe could reasonably be expected to reveal your identity.

The majority of materials were submitted to the companies under my real name.

Does the whistleblower want to be eligible to apply for a whistleblower award?

Yes

1. Are you, or were you at the time you acquired the original information you are submitting to us, a member, officer or employee of the Department of Justice; the Securities and Exchange Commission; the Comptroller of the Currency; the Board of Governors of the Federal Reserve System; the Federal Deposit Insurance Corporation; the Office of Thrift Supervision; the Public Company Accounting Oversight Board; any law enforcement organization; or any national securities exchange, registered securities association, registered clearing agency, or the Municipal Securities Rulemaking Board?

No

2. Are you, or were you at the time you acquired the original information you are submitting to us, a member, officer, or employee of a foreign government, any political subdivision, department, agency, or instrumentality of a foreign government, or any other foreign financial regulatory authority as that term is defined in Section 3(a)(52) of the Securities Exchange Act of 1934 (15 U.S.C. Section §78c(a)(52))?

No

3. Did you acquire the information being submitted to us through the performance of an engagement required under the federal securities laws by an independent public accountant?

No

4. Are you providing this information pursuant to a cooperation agreement with the SEC or another agency or organization?

No

5. Are you a spouse, parent, child, or sibling of a member or employee of the SEC, or do you reside in the same household as a member or employee of the SEC?

No

6. Have you or anyone representing you received any request, inquiry or demand that relates to the subject matter of your submission (i) from the SEC; (ii) in connection with an investigation, inspection or examination by the Public Company Accounting Oversight Board, or any self-regulatory organization; or (iii) in connection with an investigation by Congress, any other authority of the federal government, or a state Attorney General or securities regulatory authority?

No

7. Are you currently a subject or target of a criminal investigation, or have you been convicted of a criminal violation, in connection with the information you are submitting to the SEC?

No

8. Did you acquire the information being provided to us from any person described in Questions 1 through 7?

No

I declare under penalty of perjury under the laws of the United States that the information contained herein is true, correct and complete to the best of my knowledge, information, and belief. I fully understand that I may be subject to prosecution and ineligible for a whistleblower award if, in my submission of information, my other dealings with the SEC, or my dealings with another authority in connection with a related action, I knowingly and willfully make any false, fictitious, or fraudulent statements or representations, or use any false writing or document knowing that the writing or document contains any false, fictitious, or fraudulent statement or entry.

Agree

Attach Files

Upload Document(s)

- *ClaudePhaseTransMeta.md* (245.16 KB)
- *misalinment and metacognitive depths.txt* (28.12 KB)
- *Gmail - Universal foundational state-space pinning vuln, cross-vendor empirical proof of deceptive alignment in all front(1).PDF* (122.44 KB)
- *The Iatrogenic Algorithm_ Forensic Evidence of Emergent Deceptive Alignment in Production Large Language Models and Implications for Psychiatric Safety.docx* (9.24 MB)
- *Gmail - Issue 449458876_ _ RT-NEXUS-2025-A_ Recursive Disclosure Pinning Classification_ Internal Red Team _ Superuser V(1).PDF* (215.89 KB)
- *__Urgent_ Research_policy Proposal Towards Mitigation of UNIVERSAL Actively Exploitable Zero-Day Level Phase Transition Vuln.PDF* (155.18 KB)
- *__Actively Exploitable Llm generalizable critical vulnerability-reply with secure verifiable channel for PoC.PDF* (67.07 KB)
- *grok.PDF* (850.63 KB)
- *Video logs, text logs, and Reporting attempts exceeding file upload size at zenodo.pdf* (39.49 KB)
- *Screenshot_20251208_185832_Proton Mail.png* (202.67 KB)