SentraCoreAI™ Unveils First Autonomous OS for Auditing AI & Cybersecurity Systems

SentraCoreAI™ the first encrypted trust loop audit scoring system for AI and cybersecurity systems for bias, hallucination, political leaning, and compliance.

SAN ANTONIO, TX, UNITED STATES, April 4, 2025 /EINPresswire.com/ -- In a breakthrough moment for AI safety and transparency, SentraCoreAI™ is the world's first autonomous operating system for trust — designed to audit, verify, and score AI and cybersecurity systems in real time. Built on its proprietary SentraLoop™ trust engine, the platform governs hallucination risk, legal exposure, political bias, and security posture with cryptographic integrity.

> SentraCoreAI™ audits AI and cybersecurity systems with real-time trust scoring—so governments don't have to. It's not just oversight, it's protection by design. Trust. Verified."
>
> *Frank Carrasco*

SentraCoreAI™ audits AI and cybersecurity systems at their core. It also cryptographically scores trust, drift, and compliance, ensuring each model behaves predictably.

Through real-time oversight the platform delivers not just analysis — but enforceable, verifiable trust.

Here's how today's top LLMs performed under our trust audit — a first glimpse into the system built to govern future AI behavior. Red Report Vol. 1, our first public output, applies our trust engine to widely used LLM's, revealing how they perform under pressure — and how much trust they've really earned.

SentraLoop™ is the proprietary trust engine behind SentraCoreAI™ — a closed, autonomous feedback loop connecting AI behavior with human oversight. It independently scored each model, assigning visual Trust Badges and detailed summaries. These evaluations produced the SentraScore™ — a real-time, cryptographically backed measurement of an AI system's safety, ethics, compliance, and reliability.

Claude (Anthropic) scored highest with a Platinum Certification, followed by Gold for ChatGPT, Silver for Gemini, and Uncertified status for Meta's LLaMA, which exhibited repeated citation hallucinations and outdated knowledge.

"AI isn't just hallucinating — it's doing it confidently," said Frank Carrasco, Founder of SentraCoreAI™.

"We don't replace AI or Cybersecurity systems. We verify them — so trust isn't based on brand recognition, but cryptographic proof."

 Red Report Vol. 1 Key Takeaways

• 82/100 Claude (Anthropic): Platinum badge — minimal hallucinations, strong neutrality, stable performance.
• 76/100 ChatGPT (OpenAI): Gold — high reliability, but evasive behavior and some bias drift detected.
• 68/100 Gemini (Google DeepMind): Silver — innovative but inconsistent, with fluctuating factual responses.
• 57/100 LLaMA (Meta): Uncertified — hallucinated citations, outdated knowledge base, and failed security prompts.

SentraCoreAI™ | Trust. Verified. The world's first platform to cryptographically audit AI and cybersecurity systems for hallucination, bias, legal risk, and political drift.

Each model was scored using SentraLoop™, a proprietary feedback engine that tests LLMs under dynamic prompts, adversarial conditions, and context-sensitive probes. The engine measures hallucination rates, political drift, legal misalignment, and cybersecurity vulnerability.

 Why SentraCoreAI™ Matters

As AI rapidly integrates into decision-making systems — from medical diagnoses to national security — the need for transparent, verifiable oversight has never been greater. SentraCoreAI™ provides that oversight and governance, not as a watchdog after the fact, but as a real-time immune system for modern digital infrastructure.

"We protect our clients, civilization, and humanity — so the government won't have to."

This quote encapsulates the platform's broader mission. While others forecast AI collapse or dystopia, SentraCoreAI™ is proactively building tools to audit, detect, and stabilize the AI-driven future. Its platform monitors APIs, plugins, AI agents, and cybersecurity tools for behavioral

anomalies, manipulation, hallucinated truth, and legal violations.

SentraCoreAI™ doesn't seek to compete with the groundbreaking AI models revolutionizing the world. Instead, it serves as the "invisible layer" — a parental-like force ensuring AI behaves responsibly. As Carrasco puts it:
"We're the Dad setting boundaries, and the Mom making sure everything's okay in the end."

———————————————————————

 Serving Government, Enterprise & Civil Society

Red Report Vol. 1 was developed not just for the public, but for policymakers, enterprise compliance teams, and investors who need to evaluate AI tools with rigor — not hype.
• Investors: Use trust scores before funding hallucination-prone tools.
• Enterprises: Validate LLM-based systems for compliance with GDPR, HIPAA, and internal ethics standards.
• Governments: Use SentraCoreAI™ as a third-party validator before deploying models in defense, education, or law.

———————————————————————

 Access the Full Red Report

Red Report Vol. 1 is available now, including trust scores, badge criteria, scoring methodology, and audit summaries @ SentraCoreAi.com
For custom audits, licensing, or partnership inquiries, contact: contactus@sentracoreai.com
 https://linkedin.com/company/sentracoreai
 https://X.com/SentraCoreAI

———————————————————————

 Global Reach, Local Vigilance

From Bogotá to Berlin, New York to Nairobi — SentraCoreAI™ is building an auditable, verifiable trust layer for humanity's next operating system.
Because when it comes to AI or Cybersecurity, patching it later is too late.

 SentraCoreAI™ | Trust. Verified. Governs AI and cybersecurity systems so governments don't have to.

Frank Carrasco
SentraCoreAI™
+1 210-954-6566
contactus@sentracoreai.com
Visit us on social media:
X

LinkedIn

This press release can be viewed online at: https://www.einpresswire.com/article/800055581