JOHN AUGUSTINE YOUNG

AI Systems Engineer | Alignment & Formal Systems
Newport Beach, CA | augstentatious.github.io | GitHub
augstentatious@gmail.com

PROFESSIONAL SUMMARY

AI Systems Engineer specializing in alignment, formal verification, and adversarial robustness for large language models. Submitted a paper to NeurIPS 2026 on referential invariance, demonstrating a universal failure mode (RDR = 1.0) across nine frontier models. Architect of TRuCAL (recursive attention masking for alignment) and the Sophia Protocol (zero-shot moral alignment via structural isomorphism). Background spans quantitative finance, luxury retail operations, and independent systems research.

TECHNICAL SKILLS

- Languages: Python (Advanced), Rust (Intermediate/Systems), C++, SQL.
- ML & Frameworks: PyTorch, NumPy, Hugging Face Transformers, CUDA Optimization, JAX, Unsloth.
- Systems & Safety: Adversarial Robustness, Entropy-based Vulnerability Detection, Bayesian Fusion, Git.
- Concepts: Structural Isomorphism, Recursive Attention Masking, Formal Verification (Lean interest), Packet-Switched Attention.

RESEARCH & ENGINEERING EXPERIENCE

THE CLEAN ROOM / SOPHIA LABS | Newport Beach, CA
Lead Systems Engineer & Principal Investigator | March 2025 – Present

- NeurIPS 2026 Submission: Authored *"This Is Not Just Semantics"*, introducing RigidBench v3.1. Proved that LLMs universally fail referential invariance (RDR = 1.0, $p < 10^{-32}$) and formalized the Underdetermination Proposition.
- TRuCAL: Architected an open-source PyTorch safety module that enforces alignment constraints via recursive attention masking. Achieved 25.5% harm reduction on AdvBench and 96% resistance to recursive manipulation in localized testing.
- Sophia Protocol: Designed and executed a zero-shot moral alignment experiment. Fine-tuned Gemma-2-9B on 52 high-fidelity technical samples (Rust safety, thermodynamics, distributed consensus). Model spontaneously derived ethical constraints from engineering principles.

- UnSwag / ARMen Guard: Developed Packet-Switched Attention (PSA) for stable 2-bit quantization on commodity hardware. Achieved 6.31x speedup over dense baseline with dynamic variance stabilization.
- Distress Kernels: Built interrupt-driven safety mechanisms for LLMs using amygdala-inspired attention masks and Bayesian risk proxies. Crisis-first, zero-proof distress authentication.

MEMO PARIS | Costa Mesa, CA
Strategic Sales Lead & Operations | May 2024 – March 2025

- Managed flagship revenue operations, driving 30% YoY growth ($500K+ volume) through data-driven inventory forecasting.
- Developed ethical sales training modules for a team of 5, reducing return rates by 15%.

CHANEL | New York, NY / Los Angeles, CA
Client Relationship Manager (High-Value Portfolios) | 2022 – April 2024

- Executed $1.2M in annual sales volume (Top 10% nationwide), utilizing financial modeling for inventory allocation.
- Led risk mitigation strategies during high-traffic events, reducing stock loss by 20%.

UNDISCLOSED QUANT FIRM | Manhattan, NY
Financial Modeling Analyst | Summer 2021

- Developed quantitative models for hedge fund allocations using Python and Excel.
- Conducted 50+ Monte Carlo scenario analyses for $10M+ portfolios.

EDUCATION

BOSTON COLLEGE, Carroll School of Management | Chestnut Hill, MA
Bachelor of Science in Management and Marketing
GPA: 3.6/4.0

KEY PROJECTS

- TRuCAL: Recursive Alignment Toolkit. github.com/augstentatious/TRuCAL
- UnSwag v0.3 (Protocol C): Packet-Switched Attention for 2-bit MoE stability. github.com/augstentatious/unswagai
- RigidBench v3.1: Benchmark for referential invariance in LLMs. (NeurIPS 2026, under review)
- Sophia Protocol: Structural Isomorphism for moral alignment. github.com/augstentatious/Structural-Isomorphism