

JAKE KANTOR

jake.t.kantor@gmail.com | [linkedin.com/in/jakekantor](https://www.linkedin.com/in/jakekantor) | jakekantor.io | New York, NY (relocating)

AI security practitioner specializing in adversarial ML and AI red teaming, with governance depth — tests model robustness, evaluates AI outputs, and translates findings into NIST AI RMF and EU AI Act controls. USPTO provisional patent holder in adversarial ML; pursuing an M.S. in Cybersecurity (AI Security) at Old Dominion University, expected Aug 2027. Security+ and AWS Certified Cloud Practitioner.

EXPERIENCE

AI Model Evaluation & Assurance — Mercor (Aug 2026 – Present)

Frontier model evaluation / RLHF

- Evaluate frontier AI model outputs across documents, slides, and spreadsheets against defined accuracy, quality, and completeness criteria, providing human-in-the-loop signal for RLHF training.
- Assess reasoning and output integrity through an adversarial-ML lens, flagging behavioral and quality regressions against baseline.
- Author structured assessments and comparative rankings that translate model failures into actionable evaluation feedback.

Product Architect — Modern Launch (Jan 2024 – Jan 2026)

web3 / smart-contract development

- Architected a decentralized game with non-custodial wallets on Base (EVM) and Telegram integration; led design through implementation.
- Built Solidity contracts with access controls and reentrancy safeguards; integrated Chainlink VRF for verifiable randomness and developed React/TypeScript features.

Software Engineer — World Space (Jan 2022 – Jan 2023)

blockchain game engineering

- Engineered blockchain-based games (Unity, C#, Solidity) with integrated wallet and NFT systems.
- Built front-end and UX features in JavaScript and Python.

SELECTED PROJECTS

AI Governance & Technical Auditing Framework — [GitHub](#)

- Red-teamed a facial-recognition model with a custom FGSM attack — an imperceptible perturbation collapsed accuracy to 14.5% — then evaluated adversarial training as a defense; also assessed demographic fairness (gender, five racial groups) and privacy–utility tradeoffs (differential privacy vs. federated learning) under the NIST AI RMF.
- Produced governance recommendations mapped to EU AI Act Articles 9/10/13/15 and NYC Local Law 144, plus reusable SOPs, risk-assessment and policy templates, and a vendor AI-governance questionnaire. (*3-person team; led the adversarial-security assessment.*)

AI Agent — Generative SOPs — [demo](#)

- Claude-powered tool generating audit-ready Standard Operating Procedures aligned to NIST AI RMF, EU AI Act, and ISO/IEC 42001; deployed on Cloudflare Workers.

Provisional Patent — Adversarial ML for Media Protection (USPTO, 2026)

- Filed a provisional patent covering adversarial machine learning techniques to protect creative media from unauthorized AI ingestion — at the intersection of AI security, IP law, and digital forensics.

ETHDenver Hackathon — Winner

- Built a blockchain gaming MVP with a team at ETHDenver.

EDUCATION

M.S. Cybersecurity, AI Security concentration — Old Dominion University (Expected Aug 2027)

A.S. Cybersecurity — Tidewater Community College (2025)

B.A. Musical Theater — James Madison University (2018)

CERTIFICATIONS

CompTIA Security+ | AWS Certified Cloud Practitioner | IAPP AIGP (in progress) | HTB AI Red Teamer (in progress)

SKILLS

AI Security & Assurance: Adversarial ML (FGSM), model evaluation & red-teaming, LIME explainability, fairness/bias analysis, differential privacy, NIST AI RMF, EU AI Act

ML & Data: Python, scikit-learn, TensorFlow/Keras, PyTorch, pandas, NumPy

Security & Cloud: AWS, cloud security, cryptography, network security

Engineering: Solidity/EVM, TypeScript, React, Git