

Lindsay Gross

lindsaygross32@gmail.com | [LinkedIn](#) | [Website](#) | [GitHub](#)

AI Product Manager | LLM Agents · Evals · Safety | Trust & Safety Background | Duke MEng in AI

EXPERIENCE

Associate Product Manager (Rotational APM Program), Visa, San Francisco, CA Sept 2026 – Present

- Own go-to-market for a B2B payments product: the pricing model, partner integrations, and delivery across engineering, legal, and compliance, working with VP/SVP leadership and external partners.
- Drove AI adoption across the ~400-person Core Products team: shipped 7 LLM agent workflows with 10+ automated quality checks each, packaged them as an installable kit, and ran a 20-person training.

Managing Editor, Editorial Standards, The AI Collective, Remote May 2026 – Present

- Launched The Stress Test, a weekly AI accountability column that became the top-converting format in a 187K-subscriber newsletter (1.87% click-to-open vs. 1.27% average).
- Packaged the house style guide into 2 Claude skills the editorial team uses for drafting and QA.

Research Manager, Deep Tech at Duke, Durham, NC June 2026 – Aug 2026

- Developed a full-semester technology policy course in 9 weeks: a 27-session syllabus and 19 course documents, including 8 policy-failure case studies (SB 1047, Colorado AI Act, NYC Local Law 144) and a 16-mode failure taxonomy.

Research Fellow, Deep Tech at Duke, Durham, NC Feb 2025 – May 2026

- Designed the measurement method for a study of AI-generated music on Spotify: a Python screening pipeline over Spotify API data and a weighted confidence rubric with inter-rater tracking.
- Led a risk assessment of AI agent deployments (rogue-agent failure modes, security vulnerabilities, accountability gaps); findings adopted into Paladin's published risk framework.

Trust & Safety Analyst, Tremau, Remote May 2024 – Sept 2024

- Conducted EU Digital Services Act risk assessments for a major social media platform, identifying 15+ risk pathways; built mitigation roadmaps and escalation workflows adopted by leadership and the client's Trust & Safety Council.

AI Ethics, Policy & Research Fellow, Metaphysic.ai, Remote Feb 2024 – July 2024

- Co-developed a safety review framework for hyper-realistic generative AI where no industry benchmarks existed; onboarded 8 clients in 3 months with 100% adherence.

Tech & Public Policy Fellow, Beyond the Screen, Remote May 2023 – Aug 2023

- Partnered with Facebook whistleblower Frances Haugen to build a database of social media harms; authored 5 literature reviews on moderation, misinformation, and polarization.

SELECTED PROJECTS

Impunity Index: NLP Evaluation Pipeline | [Website](#) | [GitHub](#)

- Shipped an end-to-end NLP pipeline over 1.4M DOJ Epstein-case documents (spaCy, Sentence Transformers, LinearSVC, Flask, GCP) scoring 1,264 named individuals; 77–98% faced no consequences.

SAFE-T: AI Transportation Equity Audit | [Website](#) | [Demo](#)

- Placed 2nd at Duke's Society-Centered AI Hackathon with an equity audit of AI-driven transportation planning in Durham, NC: 4 bias tests surfaced ~25% undercounting in low-income areas and a 29.5% disparate impact ratio.

LUCID: Short-Form Video Manipulation Analyzer | [Website](#) | [GitHub](#)

- Engineered a full-stack app scoring psychological manipulation in TikTok videos with a fine-tuned DistilBERT classifier, Claude API LLM-as-judge labeling, and Whisper transcription; deployed on Next.js + FastAPI.

EDUCATION

MEng, Artificial Intelligence, Duke University, Durham, NC May 2025 – May 2026

GPA: 3.87 | Coursework: AI Product Management, Deep Learning, Reinforcement Learning, Intelligent Agents, Explainable AI

BA, Public Policy (Digital Intelligence Certificate), Duke University, Durham, NC Aug 2021 – May 2025

GPA: 3.94 | Cum Laude | Published at KU Leuven AI & Democracy Symposium | All Tech Is Human Board

SKILLS

AI Product & Safety: Eval Design, LLM-as-Judge Labeling, Agent Workflows, Guardrail Design, AI Risk Assessment, Pricing Strategy, Launch Readiness

Technical: Python, SQL, TypeScript, Node.js, React, FastAPI, PyTorch, Hugging Face, spaCy, Claude Code, Anthropic API, Playwright, Docker, GCP

Policy & Languages: Trust & Safety, AI Governance, EU Digital Services Act | English (Native), Spanish (Professional)