

Lindsay Gross

lindsaygross32@gmail.com | (973) 634-3182 | [LinkedIn](#) | [Website](#) | [GitHub](#)

AI engineer and product builder at the intersection of safety, governance, and user experience, with hands-on work across trust and safety operations, evaluation systems, and cross-functional product development.

WORK EXPERIENCE

- Research Fellow, Deep Tech at Duke, Durham, NC** Feb 2025 – Present
- Building an AI-powered detection system to classify AI-generated content on Spotify, designing scoring rubrics and validation methodology against human review; informing ongoing platform policy decisions around AI-generated media.
 - Led a governance case study for Paladin’s “AI Tech Stack” initiative, evaluating security risks, rogue-agent failure modes, and accountability gaps in production AI systems; findings adopted into Paladin’s published risk framework.
- Managing Editor, Editorial Standards, AI Collective, Remote** May 2026 – Present
- Serving as senior editorial voice for a daily AI newsletter reaching 225,000+ readers across 40+ countries; writing two weekly columns, running final editorial review on every send, and reviewing sponsor creative for voice and quality.
- Trust & Safety Analyst, Tremau, Remote** May 2024 – Sept 2024
- Conducted risk assessments for a major social media platform under the EU Digital Services Act, identifying 15+ compliance gaps across online safety, AI accountability, and content moderation; built remediation roadmaps adopted by leadership.
 - Built enforcement review processes and escalation workflows from scratch with the client’s Trust & Safety Council, translating regulatory requirements into operational runbooks and documentation now used across the moderation team.
- AI Ethics, Policy and Research Fellow, Metaphysic.ai, Remote** Feb 2024 – July 2024
- Co-developed a legal and ethical review framework for hyper-realistic generative AI where no internal precedent existed, defining review criteria, escalation workflows, and compliance documentation; onboarded 8 clients in 3 months with 100% adherence.
 - Authored a 20+ page policy paper analyzing US copyright law implications for AI training data, translating legal findings into product recommendations that shaped the company’s data sourcing practices.

SELECTED PROJECTS

- Impunity Index: NLP Evaluation Pipeline** | [Website](#) | [GitHub](#)
- Shipped an end-to-end NLP pipeline processing 1.4M documents using spaCy, sentence embeddings, and LinearSVC for multi-label classification; surfaced a 77–98% no-consequence rate across 1,264 scored individuals. Stack: Python, VADER, Sentence Transformers, D3.js, Flask, GCP.
- SAFE-T: AI Transportation Equity Tool** | [Website](#) | [GitHub](#)
- Launched a full-stack equity audit tool (Python, GeoPandas, Vite, Leaflet.js) analyzing bias across 65 Durham census tracts; uncovered ~25% undercounting in low-income areas and a 29.5% disparate impact ratio. Now operating as a nonprofit; presented findings to Durham Transportation Dept.
- LUCID: Short-Form Video Manipulation Analyzer** | [Website](#) | [GitHub](#)
- Built a full-stack app analyzing psychological manipulation in TikTok videos using a fine-tuned DistilBERT multi-label classifier, Claude API for LLM-as-judge labeling, and Whisper for transcription. Deployed via Next.js + FastAPI on Vercel/Railway; model hosted on HuggingFace Hub.

EDUCATION

- MEng, Artificial Intelligence, Duke University, Durham, NC** May 2025 – May 2026
GPA: 3.87 | Coursework: Deep Learning, Reinforcement Learning, Intelligent Agents, AI Product Management, Explainable AI
- BA, Public Policy | Digital Intelligence Certificate, Duke University, Durham, NC** Aug 2021 – May 2025
GPA: 3.94 | Cum Laude | Dean’s List | VP & Co-founder, Duke Tech for Change | Published at KU Leuven AI & Democracy Symposium | All Tech Is Human Board

SKILLS

Technical: Python, SQL, PostgreSQL, JavaScript, TypeScript, Node.js, React, Flask, FastAPI, Streamlit, Anthropic API, HuggingFace, scikit-learn, spaCy, Pandas, NumPy, BeautifulSoup, D3.js, Docker, Git, GCP, AWS

Safety & Governance: Trust & Safety Operations, Risk Assessment, AI Governance, Policy Enforcement, Evaluation Frameworks, Compliance Auditing, Cross-functional Coordination, Process Design, Stakeholder Documentation

Languages: English (Native), Spanish (Professional Proficiency)