

# Samantha Pease (She/Her)

(330) 940-9424 | Sam@Walking-Stick.com | SamPease.github.io | github.com/SamPease | linkedin.com/in/sam-pease

## SUMMARY

PhD-trained ML engineer building production-grade LLM + retrieval systems. Built RAG over 20K webpages (300K+ chunks) shrinking a 1.8GB embedding store to 28MB (IVFPQ, 384-d) and trimming per-answer context 25K→3K tokens (~88%) via hierarchical summarization + manual hallucination review. Deployed FastAPI + LangChain/LangSmith service (token/latency/error tracing) and compared Anthropic vs OpenAI for cost/quality. Experience: graph ML (GNN link prediction AUC 0.935), Playwright scraping, 3D SfM + Gaussian Splatting, classical + deep ML, strong math foundations. 9 years teaching → clear specs & rapid iteration; excited to leave academia and apply these skills to passions like fostering autonomy and travel.

## SKILLS

**LLM / Retrieval / Eng:** RAG (FAISS IVFPQ, MMR, hierarchical summarization), FastAPI (async), caching, latency/cost & prompt optimization, LangChain/LangSmith tracing, Anthropic & OpenAI eval  
**ML / Data / Graph:** Python, PyTorch + PyG, classical ML, NetworkX, SQL/SQLite, Playwright, Git

## EXPERIENCE

- Machine Learning Engineer Intern** Summer 2024  
Covar Durham, NC
- Built two production-candidate prototypes: (a) video→3D Gaussian Splatting pipeline (SfM + segmentation) over 2K–10K frame inputs; (b) object identification visualization module using Segmentation (SAM).
  - Adapted & stitched open-source SAM and Gaussian Splatting codebases for rapid comparative experiments.
  - Distilled 20+ papers into architecture tradeoffs; demoed to stakeholders & external client informing roadmap; produced clear specs under ambiguity.

- Math Instructor** 2017–Present  
Duke University & Rutgers University–Newark Durham, NC & Newark, NJ
- Taught Applied Calculus & Calculus I and supported large (100+ student) Precalculus, College Algebra, Applied Calculus lectures; created syllabi, assessments, structured problem sessions.
  - Translated abstract math for 1000+ students into clear, stepwise explanations; praised for effective teaching.

## PROJECTS

- Trans Advice Agent** Summer 2025
- Built large-scale RAG over 20K webpages (300K+ chunks) using FAISS IVFPQ (shrunk raw store 1.8GB → 28MB) + MMR + query expansion + hierarchical summarization (~25K→3K tokens).
  - Reduced answer latency & context cost via async batch summarization and caching; manual hallucination review + prompt tuning; evaluated Anthropic vs OpenAI for cost/quality.
  - Deployed FastAPI (POST /ask) service with LangSmith step-wise token/latency/error tracing & provenance UI (paginated sources) on Render.

- Instagram Network Analysis** Summer 2025
- Scraped Instagram mutuals with Playwright (design limiting scope to mutuals) building directed graph (~700 nodes / 23K edges); visualized communities (Louvain) via PyVis.
  - Trained PyTorch Geometric GNNs for link prediction (AUC 0.9352, AP 0.9349) and community classification; engineered structural features & optimized scraping throughput.

### Additional Projects

- Built an image classification neural net in pure NumPy; analyzed architectural tradeoffs (ML course project)
- Applied persistent homology to LiDAR forest canopy data to differentiate forests (TDA research project)

## EDUCATION

- Rutgers University–Newark - Ph.D. Mathematics** May 2026
- Duke University - B.S. Mathematics & Computer Science, with Distinction** May 2020