YASH LOTHE

US Citizen | ylothe@andrew.cmu.edu | (412) 287-0757 | https://www.linkedin.com/in/yashlothe/

EDUCATION

Carnegie Mellon University **May 2027**
Master of Science in Artificial Intelligence and Innovation (MSAII) *GPA 3.8/4.0*
Selected Coursework: LLM Systems, Deep Reinforcement Learning, Advanced Natural Language Processing, Large Language Model Applications, Search Engines, AI Engineering

Birla Institute of Technology and Science (BITS Pilani – Goa) **May 2025**
Bachelor of Engineering in Computer Science *GPA 7.69/10.0*
Relevant Coursework: Artificial Intelligence, Deep Learning, Brain-Inspired Deep Learning, Human-Computer Interaction, Data Structures & Algorithms, Design & Analysis of Algorithms, Probability & Statistics

SKILLS

Programming: Python, Java, C++, SQL, JavaScript
Machine Learning: Transformers, Retrieval-Augmented Generation (RAG), Recommender Systems, Ranking & Retrieval, Reinforcement Learning, Deep Learning, LLM Evaluation, Bias & Robustness Analysis, A/B Testing
LLM & AI Systems: LangChain, LangGraph, LiteLLM, Prompt Engineering, Model Evaluation Pipelines
Frameworks & Libraries: PyTorch, TensorFlow, Hugging Face, Scikit-learn, Pandas, NumPy, SciPy
Data & Infrastructure: FAISS, BigQuery, Oracle, Redis, FastAPI, Docker, PostgreSQL, AWS, Flask, Streamlit

EXPERIENCE

Texas Instruments **Jun 2026 - Aug 2026**
AI Solutions Engineer Intern *Dallas, TX*

- Built per-account personalization for TI's production recommender using a **Thompson sampling contextual bandit** with 3-tier fallback (per-company / sector / global), optimizing for new product category discovery. Added per-impression reward attribution, hierarchical Bayesian priors, and warm-start on 6 months of historical impressions. Serves ~2,048 tier-1 companies with nightly mapping updates.
- Designed the **RAG subsystem for an on-prem LLM agent** that auto-generates SPICE testbenches for analog circuits: curated a 25K-chunk corpus, implemented hybrid search with cross-encoder reranking and LLM query rewriting, reaching Recall@10 = 0.78 across 12 circuit classes.
- Shipped a **new email recommendation channel** for the recommender to production, serving ~400 daily campaigns. Built the recommendation logic, FastAPI serving endpoint, drift-gated automated testing, and MLflow deployment across two production repos.

Harvard University **Aug 2024 - Jul 2025**
Research Assistant, Banaji Implicit Social Cognition Lab *Cambridge, MA*

- Led a Responsible AI evaluation of large multimodal models, auditing **face-to-character inference biases** in GPT-4o, GPT-5, Claude Sonnet 4.5, and Gemini 3 Flash.
- Built preprocessing, bias-quantification, and multi-model evaluation pipelines (8,160 trials) identifying systematic trait generalization across 14 attributes and 6 judgments.
- Co-author on "Like humans, GPT-4o demonstrates face-to-character biases" (**accepted, PNAS Nexus**).

Ericsson **Jun 2024 - Aug 2024**
AI/ML Intern *Bangalore, India*

- Designed and implemented **synthetic data generation pipelines** (100,000+ samples) using GANs, VAEs, and Kernel Density Estimation in Python, improving model robustness and enabling production-scale ML training.
- Evaluated 15+ open-source LLMs on synthetic enterprise benchmarks (ROUGE, BERTScore, COMET) to guide model selection for internal deployment.

ACADEMIC PROJECTS

Large-Scale Recommender System using Two-Tower Retrieval & Deep Ranking Models **Oct 2025 - Dec 2025**
Carnegie Mellon University *Pittsburgh, PA*

- Built an end-to-end recommendation pipeline combining a **Two-Tower retrieval model with a deep neural re-ranking network**, learning dense user/item embeddings from interaction data.
- Implemented scalable retrieval → candidate generation → ranking workflows using FAISS vector similarity search and a PyTorch MLP ranker.
- Achieved strong improvements over matrix factorization and GBDT baselines: +18% Recall@50, +12% NDCG@10, +9% MRR.

Agentic LLM Red-Teaming System using LangGraph **Jan 2026 – Mar 2026**
Carnegie Mellon University *Pittsburgh, PA*

- Built an **agentic LLM safety evaluation framework using LangGraph** that autonomously generates adversarial prompts and multi-turn attack conversations to test deployed LLM applications.
- Implemented a multi-agent workflow (planner, generator, executor, judge, reporter) with shared state to orchestrate automated red-teaming, structured evaluation, and coverage-driven attack planning.
- Developed reproducible evaluation pipelines using HarmBench, JailbreakBench, and XSTest, measuring jailbreak success, over-refusal, prompt injection, and tool misuse across configurable LLM endpoints.