Saiakhil Chilaka

(984) 260-2526 | akhilchilakawork@gmail.com | linkedin.com/in/akhilchilaka | github.com/akhilc08 | akhilchilaka.com/

Education

Cornell University, College of Arts & Sciences — Ithaca, NY
B.A. in Computer Science — *Expected May 2028*

- **Relevant Coursework:** Distributed Systems for ML, Machine Learning, Computer Systems, Data Structures

Experience

Generative AI at Cornell — August 2026 – Present
Applied AI Researcher — *Ithaca, NY*

- Built a compaction method for coding agents with 93% instruction retention vs. 85% for full context at 1/3 the tokens
- Beat 27 common compaction methods on LongMemEval at 98.7% accuracy & matched the best method on τ-bench at 8%

MotherDuck — May 2026 – August 2026
Software Engineering Intern — *New York, NY*

- Built the internal eval harness benchmarking AI agents on 40 user workflows via the MotherDuck CLI and MCP
- Examined failed agent trajectories to write CLI guides for 3 features, taking task category accuracies from ∼16% to 100%
- Reduced agent token cost on the MotherDuck CLI up to 57% through A/B testing 5 CLI output optimizations

Duke University — August 2024 – December 2024
Research Intern — *Durham, NC*

- Published first-author IEEE paper on CNN ensemble for Multiple Sclerosis diagnosis; 97.5% accuracy & 97.8% AUC-ROC
- Fine-tuned VGG19/ResNet50 layers via depth-wise separable convolutions, reducing model size 14x for edge deployment
- Injected Gaussian white noise alongside rotation, shear, and zoom augmentation to improve generalization by up to 7%

North Carolina State University — February 2024 – September 2024
Research Intern — *Raleigh, NC*

- Built a skin cancer classifier for edge devices with knowledge distillation, shrinking it 18x and lifting accuracy 19%
- Trained a generative adversarial network to synthesize training data, improving overall performance score by 13%

Projects

ForeSight: Calibrated Decision Models as Opponent Forecasters | *Submitted to AAAI '27*

- Built a Pokemon game agent using a system one opponent forecaster, matching published LLM agents win-rate (56%)
- Fine-tuned a 421M System One model (Laya) to forecast opponent moves, lifting prediction accuracy from 43% to 80%
- Showed 2-turn lookahead adds 13–20 pts to win rate over one-step play, using 4% of per-turn budget for lookahead search

Attributing Tool Failures in Bulk Agent Operations | *Submitted to NAACL '27*

- Benchmarked LLM agents on AppWorld, finding they misreport which tool call failed in long runs 67% of the time
- Built a drop-in MCP proxy that echoes call arguments into results, restoring 94–100% accuracy across Claude & GPT

Agentic Polymarket Predictor | *Python, Neo4j, Qdrant*

- Built a 5-stage agentic predictor for Polymarket, achieving 17% ROI during forward testing & 30.5% ROI for backtesting
- Architected a hybrid RAG pipeline over Neo4j & Qdrant via Reciprocal Rank Fusion with a BGE reranker
- Beat or matched the market on Brier scores on the 53% of markets the model chose, validated across multiple evaluations

Technical Skills

Languages: Python, TypeScript, SQL
Frameworks & Libraries: PyTorch, TensorFlow, scikit-learn, LightGBM, verl, FastAPI, HuggingFace Transformers
Databases & Infrastructure: PostgreSQL, Neo4j, Qdrant, Modal, AWS, Docker, Git
ML/AI: MCP, RAG, Agent Evaluation, LLM Fine-tuning, LoRA, Reinforcement Learning, Knowledge Distillation

Publications

Chilaka et al. "CNN Ensemble for Multiple Sclerosis Diagnosis on Embedded Systems" | IEEE MCSoC '24