

Michael Liu

[Email](#) | [LinkedIn](#) | [GitHub](#) | [Blog](#)

Skills

- Programming: Python, Prompt Engineering, Java, Go, JavaScript, Ruby
- Cloud Services: AWS (EC2, S3, DynamoDB, CloudFormation, SQS, SNS, IAM, Lambda)
- Data Processing: Large-scale distributed systems, high-throughput pipelines, CI/CD pipelines
- AI/ML: Transformer, RL, Evaluation, RAG/Agents, Multimodal, Pytorch, Einops, Tensorflow, WandB

Professional Experience

Amazon Web Services (AWS) - Software Development Engineer II (promoted Q3 2024)

Aug 2022 - October 2024

- Oversaw global AWS data transfer metering—processing 140 PB of data across 12 million batches daily and generating \$11 million in daily revenue
- Ensured 24/7 system reliability through on-call rotations that resolved critical outages, unblocked CI/CD pipelines, and managed customer onboardings and inquiries
- Led a team of 3 engineers in implementing zero-touch automation for new AWS region builds—coordinating tooling adoption and cross-team efforts to accelerate data center expansion
- Designed and executed an S3 storage optimization project that cut costs by 90% on the team's largest S3 bucket, saving ~600PB/month and boosting operational efficiency

Personal Projects

Nov 2024 - Present

ML's Blog ([Visit Blog](#))

- Built a personal technical blog using Hugo, documenting explorations in AI concepts.

Transformer Implementations and Mechanistic Interpretability Experiments ([View in GitHub](#))

- Trained GPT-2 transformer from scratch with various sampling techniques integrated with KV cache
- Used TransformerLens to reverse engineer model behaviors (attention patterns, induction circuits)

Reinforcement Learning Implementations ([View in GitHub](#))

- Implemented and compared classical and deep reinforcement learning algorithms (SARSA, Q-learning, DQN, RLHF) in PyTorch, analyzing their performance in various gym environments.

LLM Alignment Evaluation and Agent Toolkit ([View in GitHub](#))

- Designed threat models and generated evaluations using LLMs to test for model alignment
- Built a LLM Agent to play Wikipedia Racing Game using elicitation methods (ReAct, Reflexion, etc)

Diffusion and Multimodal Models Implementation ([View in GitHub](#))

- Trained diffusion models from scratch with various sampling methods (DDPM, DDIM) and implemented multimodal architectures (CLIP).

Certifications

- RAG++/Evaluations/Structured Outputs - Weights & Biases (Jan '25)
- Machine Learning Engineering for Production (MLOps) Specialization - DeepLearning.AI (Feb '24)
- LLM101/102x: Foundation Models/Application Through Production - Databricks (Dec '23)
- TensorFlow Developer Certificate - Google (Oct '22)
- Deep Learning/Natural Language Processing Specialization - DeepLearning.AI (Jun '22, Jun '23)

Education

University of Maryland, College Park

B.S. Computer Science (Machine Learning Specialization) Spring 2022