Allen Thomas

217-298-6572 | allenthomasdev@gmail.com | LinkedIn - /AllenThomasDev | Github - /AllenThomasDev

Education

University of Illinois at Urbana-Champaign August 2024 - December 2025
Master of Computer Science *Illinois*
Distributed Systems, Systems for Gen AI, Topics in LLM Agents, Deep Learning

Experience

Data Engineer June 2022 – June 2024
Helpshift *Pune, India*

- Led migration of Helpshift's Issues Analytics API and PowerBi API from HBase to Amazon Redshift, handling a 10x traffic increase and 50x surge in queries with zero downtime
- Drove petabyte-scale pipeline migration to AWS, resulting in monthly savings exceeding $16,000
- Led the successful migration of customer-facing APIs, ensuring zero downtime and improving response times by 40%, enhancing user experience and reducing support requests
- Migrated legacy stream processing pipelines from Apache Storm and HBase to Apache Flink and YugaByte
- Played a key role in decommissioning legacy infrastructure, and supporting cost-saving initiatives
- Led a data democratization initiative, reducing support tickets for the engineering team by 40%, by training support staff to use Metabase, enabling them to generate reports and resolve data queries without engineering assistance.

Projects

Control Vector-Based LLM Steering and Analysis | *LLMs, Hugging Face, OpenAI* March 2025

- Developed a framework to train control vectors that modulate LLM-generated text across personality traits such as optimism vs. pessimism and introversion vs. extroversion.
- Engineered a pipeline to generate, collect, and analyze LLM responses at various control strengths, producing enriched datasets with automated personality scoring.
- Integrated evaluation compatibility with `lm-evaluation-harness`, enabling systematic benchmarking

Stream Processing Framework | *Golang, Distributed Systems, Multithreading* November 2024

- Architected and implemented a high-performance stream processing system achieving exactly-once processing semantics with advanced fault tolerance mechanisms
- Engineered optimized task scheduling algorithms for enhanced real-time data processing throughput
- Integrated custom-built HyDFS storage layer ensuring reliable data persistence and recovery

Hybrid Distributed File System (HyDFS) | *Golang, Distributed Systems, Fault Tolerance* October 2024

- Implemented a scalable distributed file system supporting data operations across 10 nodes with failure tolerance
- Implemented consistent hashing for optimal data distribution, improving file replication and routing performance
- Synthesized HDFS and Cassandra architectural principles to create a robust, scalable storage solution

Research

Independent Research on LLM Inference Scaling | *RunPod, Qwen, DeepSeek, OpenAI* 2025

- Investigated inference-time scaling laws and reasoning token manipulation to enhance LLM performance.
- Benchmarked and fine-tuned *s1* and *s1.1* models, analyzing "Wait"-based budget forcing for reasoning extension.
- Conducted large-scale inference experiments using distributed GPU clusters, optimized for cost efficiency.

Agentic Research Code Reproducibility System | *Docker, Google Gemini, OpenAI* October 2024

- Engineered an autonomous agentic AI system that analyzes research papers and reconstructs their environments
- Developed self-directed exploration capabilities enabling automated reproducibility testing through intelligent command execution in containerized environments

Multi-Agent Reinforcement Learning-Implementation of Hide and Seek | IEEE 2021

- Researched and implemented multiple reinforcement learning algorithms for multi-agent systems, developing a novel hide-and-seek simulation environment inspired by OpenAI research on Multi-agent Autocurricula

Technical Skills

Languages: Golang, Java, Python, Clojure, SQL, R, Scala, JavaScript
Frameworks & Platforms: Apache Spark, Airflow, Hadoop, Kafka, Flink, Hive, HBase, AWS Redshift, S3, Athena, EMR, dbt, PyTorch, Docker, pandas, NumPy, scikit-learn, Hugging Face Transformers
Developer Tools: Neovim, Git, Gerrit, Jira, AWS, Terraform
Databases: PostgreSQL, MySQL, MongoDB, Cassandra, Elasticsearch