

SAMARTH SOMANI

somani.samarth@gmail.com • (248)-346-6767

[linkedin.com/in/samarth-somani/](https://www.linkedin.com/in/samarth-somani/)

WORK EXPERIENCE

Ally Financial – Charlotte, NC

May 2021 – Present

Senior Specialist of Data Science - Generative AI (July 2025 - Present)

- Created reusable evaluation code to assess vision models (Gemma3, GPT-4oV) for contract processing, recommending the optimal model
- Designed experiments with LLMs, document models, chunking & retrieval methods, benchmarking & optimizing the RAG pipeline

Specialist of Data Science - Generative AI (July 2023 - July 2025)

- Built reusable codebases & playbooks for finetuning, GenAI evaluation, LLM best selection, & data synthesis, driving adoption org-wide
- Presented on LLMs, Vector Search Engines, summarization methods, OPM, and AI bias at events to promote AI literacy across Ally
- Integrated retrieval augmented generation & semantic search across multiple use cases, collaborating with various cross-functional teams
- Partnered with mortgage leaders to develop AI system for identifying high-risk appraisals, improving review processes & mitigating risk
- Created NLP tool using sentiment analysis, topic modeling, & summarization to track customer email feedback for credit card campaigns
- Engineered for-purpose AI model for earnings call summarization, using advanced regular expressions & various chunking strategies
- Designed & tested semantic & keyword similarity tool to evaluate AI-generated content & address copyright concerns in Generative AI
- Collaborated with ML Ops, Architecture, Model Risk Management, & UI teams to productionize models for end-user adoption

Associate Specialist of Data Science - Conversational AI (July 2022 – July 2023)

- Created Streamlit-based Python tools to enhance chatbot monitoring, focusing on escalation triggers and rich media content
- Developed topic modeling solution using clustering and text mining to identify new intents, increasing chatbot containment rate
- Analyzed auto-chatbot performance using SQL and Python, identifying growth opportunities and enhancing reporting

Data Science Intern (May 2021 – August 2021)

- Leveraged Power BI, Python, and SQL to analyze chatbot customer interactions, optimizing metric reporting for product owners
- Spearheaded initiative to identify and remove high false-positive intents with the creation of interactive Power BI dashboards
- Enhanced chatbot efficiency with Conversational AI team by analyzing intent performance, improving user experience and reducing costs

Ignis Consulting – Remote

June 2020 – June 2021

Co-Founder & VP Client Development

- Founded pro-bono consulting group promoting social change; recruited 6 board members, secured 10+ clients, & created 20 student roles
- Developed project plans and timelines for Just Speak Inc. and Detroit Cycling Studio, guiding Project Managers in execution

PROJECTS

Patent-Pending: GenAI Hybrid Summarization Model (Ally)

- Designed & filed patent for an all-purpose summarization model using topic modeling, extractive & abstractive summarization techniques
- Addressed positional bias, token costs, context window constraints, & difficulty generating larger summaries in long-form summarization
- Integrated model in Ally.ai platform & Ally Model Hub for users to generate free-form summaries for larger PDF documents

Auto QA Generation Tool (Ally)

- Built Auto QA tool that improved RAG accuracy by 16% via curated context inputs; also used to create test sets for RAG evaluation
- Crafted end-to-end UI using React for Q&A generation, incorporating a human-in-the-loop workflow for real-time feedback and control
- Designed a Flask backend supporting user-defined parameters and iterative editing, saving, and regeneration of Q&A pairs

Company-Wide Generative AI Prompt Tracking and Anomaly Detection (Ally)

- Built ETL pipeline & streamlit app tracking prompt usage via keyword extraction, embeddings, weighted scoring & anomaly detection
- Reduced AI risk by identifying over-reliance on topics & ensuring monthly usage aligns with expected patterns across lines of business

Simplifying AI Research for Everyone (Masters Capstone Project)

- Built a tool to simplify research papers with summarization, term highlighting, visual insights, and a RAG chatbot
- Developed a summarization & RAG pipeline using Qdrant, PostgreSQL, & Airflow ETL for parsing, chunking, embedding, and retrieval

EDUCATION

University of North Carolina - Charlotte, School of Data Science, Charlotte, NC

August 2025 - Present

PhD of Data Science

- Current PhD Level Coursework: Fundamentals of Artificial Intelligence, Statistics for Data Science, Research Design for Data Science

University of Michigan, School of Information, Ann Arbor, MI

September 2022 – May 2025

Master of Applied Data Science, GPA 3.92/4.00

- Key coursework includes NLP, Unsupervised & Supervised Learning, Deep Learning, Data Mining, Information Visualization, etc

Indiana University, Kelley School of Business, Bloomington, IN

August 2018 - May 2022

Bachelor of Science in Information Systems & Business Analytics, GPA 3.73/4.00

- Winner, PwC Case Competition (company relocation); Finalist, K-204 & K-303 Deloitte Case Competitions (Top 10 of 60+ teams)

TECHNICAL SKILLS

Artificial Intelligence

LLM | RAG | Evaluation | OCR | Vision | Finetuning | Prompting | Embeddings | Vector DBs | Chunking | Doc Ingestion | Pytorch | TensorFlow

Data & Tools

Python | R | AWS | SQL | Topic Modeling | Snowflake | Airflow | DAG | Git | Postman | Streamlit | React | Flask | Power BI | Excel | Streamlit