

MICHAEL L. GEIS, PHD.

michael.l.geis.phd@gmail.com ♦ [LinkedIn](#) ♦ [Github](#) ♦ [Personal Website](#) ♦ Chicago, IL

SUMMARY

Data scientist with experience defining and solving concrete business problems with mathematical methods. As a mathematics PhD, I have taken my experience solving novel research problems and translated that into an ability to define and solve problems with real world impact using data-driven approaches. Seeking data scientist, machine learning engineer, and data engineer roles.

SKILLS

General Tools Python, SQL, Bash, Git, Docker, Azure Cloud, DBMS, Pandas, Plotly, Flask, Sklearn, Pytorch, Transformers, OpenAI
Machine Learning ETL, Data Cleaning, EDA, Feature Engineering, Hyperparameter Tuning, Regularization, Model Calibration, Deployment
Soft Skills Project Management, Client Communication, Technical Communication, Presentation Skills, Independent Research

WORK EXPERIENCE

RedMane Technology, R&D Team - Data Scientist

11/2023-Present

- **Triage/Prioritization Algorithm Development For Child Welfare Reports**
 - Led the development of a binary classification model serving as the core of a larger prioritization algorithm.
 - Designed and implemented a robust data ingestion pipeline, sourcing data from PDF files and an Azure synapse data warehouse; wrote custom code to securely download, parse PDFs, clean data, and join sources into views for EDA, model development, and evaluation.
 - Conducted EDA, identifying structured data limitations and focusing on unstructured text, performing word frequency, and demographic analysis to assess feasibility of modeling task.
 - Led model experimentation assessing feasibility of KNN, transformer-based classifiers, XGBoost, GLMs, SVMs on featurized text.
 - Trained final model with CV in Azure ML Studio; tuned hyperparameters; reduced overfitting with early stopping and L2 weight decay.
 - Deployed dockerized Flask API for inference; ensured fairness of model predictions across demographic groups using Fairlearn.
- **Scalable Document Processing Pipeline for OCR, Indexing, and Search**
 - Engineered end-to-end document processing pipeline to digitize and index over 800,000 historical documents for child welfare cases.
 - Designed a modular ASP.NET Core web API that orchestrates NLP tasks including OCR (Azure Form Recognizer), summarization (GPT-4o), and custom NER (spaCy) to extract domain-specific entities.
 - Implemented hybrid search using an Azure AI Search built from Azure Data Lake Storage. Search strategy combines traditional keyword search, vector-based semantic search, and meta-data filtering, improving retrieval accuracy for complex queries.
 - Wrote a Python library to achieve high-throughput bulk ingestion, processing up to 200 docs/hour into Azure Data Lake Storage.

Northwestern University, Mathematics Department - Graduate Researcher/Teaching Assistant

09/2015-09/2022

- High-Frequency Eigenfunction Asymptotics Research Program
 - Led a research program in which we calculated new high-frequency asymptotic limits of ladder sequences on CROSS spaces.
 - Presented research findings and complex ideas to both technical and non-technical audiences. Attended interdisciplinary conferences.
 - Developed a routine of continuous, independent learning in order to adapt to new, state-of-the-art methods in a rapidly evolving field.
- Teaching Experience
 - Taught 18 math courses to students with diverse backgrounds including engineering, social sciences, biology, physics, and mathematics; ranked in the **top 20%** of all graduate TAs according to average student feedback scores from 2018-2022.
 - Wrote and published a set of expository notes on the [geometry of Lie groups](#) that has been downloaded **over 5,500** times.

SELECTED INDEPENDENT PROJECTS

arXiv Mathematics Subject Classifier

07/2023-08/2023

- Built an end-to-end solution to a multi-label text classification problem, classifying mathematics paper abstracts by subject.
- Trained and deployed a huggingface transformer model with a front-end for inference written in Gradio.

Scientific Paper Recommender System

05/2023-06/2023

- Pitched project concept, organized and led a group to create a CBF recommender system for arXiv articles based on their abstracts.
- Sourced and cleaned data; built a KNN-based system using transformer embeddings; added topic analysis with BERTopic.

LEADERSHIP & AWARDS

Mathematics Department, Northwestern University

- Organized Seminar on the Symbol Calculus of Fourier Integral Operators
- Awarded *NSF RTG Fellowship* in Geometric Analysis

09/2019-06/2020

09/2015

Rutgers University

- Graduated summa cum laude
- Mathematics Honors Track

05/2015

09/2013-05/2015

EDUCATION

Data Science Bootcamp, Erdős Institute

05/2023-06/2023

Mathematics Ph.D., Northwestern University

09/2015-09/2022

Thesis: [Empirical Measures for Integrable Eigenfunctions Restricted to Invariant Curves](#)

Advised by Steven M. Zelditch

B.S. Material Science, Rutgers University

09/2010-09/2015

Double major in Mathematics