

# Steven Haworth

## SKILLS

- **Languages** - Python, R, SQL, Scala, CQL, Powershell, MATLAB, Julia, C++
- **Frameworks** - MNE, NumPy, Pandas, SK-Learn, TensorFlow, PyTorch, Spark, Kafka, AWS, FastAPI, spaCy, Docker, Tableau, Matplotlib, Seaborn, Snowflake, Git, GitHub, GitLab, MLFlow, Postgres
- **Applications** - EEG Signal Processing, Machine Learning, Deep Learning, Computer Vision, Databases, Large Language Models, Retrieval-Augmented Generation (RAG)

## EXPERIENCE

### UWHealth Hospital & Clinics

Madison, WI

*ML Research Assistant*

May 2024 - July 2025

- Applied Bayesian Network modeling and graph-based feature extraction to EEG data for mental state classification; achieved 0.88 ROC-AUC using supervised learning on network-derived embeddings
- Forecasted 72-hour seizure recurrence using EEG features and Cox regression, visualizing survival probabilities and identifying interpretable risk factors; results accepted for presentation at international conference
- Revealed meaningful manifold geometry in a previously inseparable feature space using manifold learning and hyperparameter optimization, leading to sleep stage differentiation (Precision: 82%, F1: 76%)

### WashU - St. Louis, Massachusetts General Hospital, Harvard Medical

Partially Remote

*Graduate Researcher*

Sep 2024 – Present

- Collaborated with multiple institutions to forecast seizure risk using Convolution Neural Networks, LLMs, and fundamental machine learning methods (tree-based classifiers, linear/kernel methods)
- Queried and joined up to 30 TB of EEG and metadata from AWS S3 using Apache Spark; sampled a confounding variable controlled subject population to ensure fair and accurate seizure risk modeling
- Tuned a CNN-based embedding pipeline in PyTorch to convert long-form EEG signals into compact feature vectors, supporting temporal modeling of seizure recurrence

## PROJECTS

### Transformer-Based EEG Topic Miner [GitHub Link](#)

- Built a Retrieval-Augmented Generation (RAG) system by combining FAISS-based semantic retrieval with transformer-based classification of EEG-relevant literature across 40,000+ biomedical papers
- Containerized the full NLP pipeline with Docker and integrated CI/CD hooks to detect model drift to retrain on new incoming literature
- Fine-tuned SciBERT using HuggingFace Transformers with cross-entropy loss and the AdamW optimizer on weakly labeled PubMed abstracts; achieving 97% accuracy on a hold-out set

### Optimal EEG Montage Builder [GitHub Link](#)

- Built three progressive mixed-integer programming models in GAMSPY to optimize electrode selection for dream state classification, reducing electrode count by 70% while exceeding baseline ROC AUC by 0.2
- Engineered multiple cost-aware objective functions, penalizing artifact-prone electrode placements and adding soft symmetry constraints for bilateral montage balance
- Conducted Pareto analysis across nine constraint levels and validated montages via 5-fold stratified cross-validation, identifying inverted-U performance curve with optimal 56-electrode configuration

### Synthetic Time Series with Diffusion Models [GitHub Link](#)

- Developed a denoising diffusion model (DDPM) to generate realistic synthetic S&P 500 price series; achieved distributional similarity in kurtosis, volatility, and autocorrelation
- Tuned a deep Bayesian network with attention-based architecture to stabilize training pass and match empirical statistics; improved standard deviation similarity from 1.12 to 0.58
- Benchmarked model performance against GAN and ARIMA baselines; synthetic time series preserved structural trends and volatility in holdout sets

## EDUCATION

### University of Wisconsin-Madison

May 2025

*Bachelors in Data Science, Dean's List Student*

### University of Wisconsin-Madison

Expected May 2027

*Master of Science in Data Science, funded through the Highly Qualified Candidate Scholarship*