# Steven Haworth

262-949-8102
0218steven@gmail.com
GitHub
LinkedIn
Personal Website

## 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; manuscript published in Journal in Neurology
- 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 RAG system to accelerate EEG literature discovery for research teams by combining FAISS-based semantic retrieval with transformer-based classification 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 EEG 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 0.12
- 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

**ARGUS v1.3, Artifact Rejection & Seizure Detection System** *GitHub Link*

- Developer of the spatiotemporal deep learning system "Argus", made to classify seizures from raw EEG; achieved 94.43% sensitivity and 93.43% specificity on neurologist-annotated recordings
- Designed a CNN-based architecture with temporal attention to model seizure dynamics across 10-second windows; trained on clinically derived 22-channel bipolar EEG montages
- Built and evaluated a two-stage artifact filtering and seizure detection pipeline; accepted for presentation at the American Clinical Neurophysiology Society (ACNS) 2026 and validated on holdout EEG data

## 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*