Joshua Mason

Saint Paul, MN | +1 (651) 315-5708 | joshuam1008@gmail.com | www.linkedin.com/in/joshuam1008/

EDUCATION

North Carolina State University, Raleigh, NC
Master of Computer Science (MCS) Jan 2024 - Dec 2024
Cumulative GPA: **3.93** | **Park Scholar** (highly selective merit-based scholarship)
Relevant Coursework: Artificial Intelligence, Data Mining, Convex Optimization, Databases, Software Engineering, Cloud Architecture

Bachelor of Science in Computer Science; Minor in Economics Aug 2020 - Dec 2023
Cumulative GPA: **4.00** | **Summa Cum Laude** | Park Scholar | **Critical Language Scholar** | President of Cybersecurity Club

WORK EXPERIENCE

Thomson Reuters Twin Cities (Eagan), MN
Applied Science (NLP/ML) Intern May 2024 – Aug 2024
- Collaborated with tax and finance subject matter experts (SMEs) to iteratively develop, label, and refine a high-quality dataset for a complex natural language processing (**NLP**) **classification task** tailored to a **Retrieval-Augmented Generation (RAG) pipeline**.
- Designed a novel zero-shot classifier in an AWS **Sagemaker Studio** environment (Python + Jupyter) leveraging **LangChain**, AWS **Bedrock**, and third-party **LLMs**, incorporating **chain-of-thought prompt engineering** and a tunable feature for achieving either **100% recall** or **100% precision**.
- Created custom Python scripts to dynamically inject context for **stress-testing a production RAG pipeline**.
- Conducted thorough performance evaluations of the RAG pipeline using the classifier, identifying **three potential failure modes** and recommending actionable debugging experiments to improve reliability and robustness.

BAE Systems, Inc. Research Triangle Park (Durham), NC
Data Science (R&D) Intern May 2023 – Aug 2023
- Led independent research and experimentation within the Geospatial R&D team to develop **cloud segmentation** techniques for real-time, low-altitude, multi-band, multi-resolution live stream feeds.
- Curated a custom dataset and developed three models: a traditional **computer vision** (CV) model, a classical **machine learning** (ML) model, and a **U-net deep learning** (DL) model, using tools such as **OpenCV**, **scikit-learn**, and **PyTorch**.
- Achieved over **75% accuracy** and **15 FPS processing** on visual and IR band spectrums using a traditional computer vision model with advanced feature detection and filtering, implemented in **MATLAB** and **Python**.

Laboratory for Analytic Sciences (research lab) Raleigh, NC
Data Science Intern May 2022 - Jun 2022
- Built two web scrapers using Python and **Beautiful Soup** to collect and construct a Mandarin-language newspaper dataset for **NLP event extraction** research. Applied Mandarin language skills to manually annotate the dataset for downstream NLP tasks.
- Performed **model inference** on event extraction and event coreference resolution models, diagnosing and resolving key failure modes.

Punch Cyber Analytics Group (cybersecurity research and consulting firm) Reston, Virginia
Cybersecurity Intern May 2021 - Aug 2021
- Engineered a **weak-labeling pipeline** for security tags using a local **Apache Spark cluster**, processing **over 1TB** of unlabeled corporate **Zeek** (Bro) **network logs** from a simulated red team/blue team cyberwarfare exercise.
- Leveraged **OSINT techniques** and tools like **Maltego** to investigate potential advanced persistent threats (APTs), applying insights gained from four intensive workshops.

People for PSEO (grassroots nonprofit focused on state-level education issues) Minneapolis, MN
Policy Research Analyst; Chair of Student Committee Nov 2018 - Jan 2021
- Researched education statistics, creating **Matplotlib** and **Seaborn** visualizations to support state-level education policy advocacy.

ACADEMIC & PERSONAL PROJECTS

Lunar Lander AI Agent Jan 2024 - May 2024
- Developed a PyGame-based simulation replicating the physics of the Apollo Lunar Lander in a **partially observable, continuous, real-time,** and **unknown environment**. Created both a **policy-based expert system** and a **Q-learning agent** (PyTorch + Numpy) that optimized landing efficiency by minimizing time and fuel consumption through careful **reward engineering**.

NBA All-Star Prediction Aug 2023 - Dec 2023
- Evaluated **22 single and ensemble models** for predicting NBA All-Star selections. Trained a custom **hybrid ANN** using PyTorch, incorporating the most diverse and performant models, **achieving a superior F1 score** over all surveyed approaches.

Data Gandalf - A Dataset Recommender Aug 2023 - Dec 2023
- Proposed and developed a web application addressing a real-world data discovery challenge sponsored by **SAS**. Built a multi-step **recommendation system** using **TF-IDF** and **BM25**, enabling precise dataset recommendations tailored to user selections.

SKILLS

Programming: Python (Jupyter), Java, C, C++, JavaScript, HTML/CSS, MATLAB, SQL (PostgreSQL, MySQL, SQLite)
Data Mining/ML/NLP: Pandas, NumPy, OpenCV, scikit-learn, NumPy, PyTorch, NLTK, spaCy, gensim, LangChain
Cloud: AWS (CLI, S3, EC2, SageMaker Studio, EKS), Azure (Boards, Repos)
Tools & Frameworks: git/GitHub, REST APIs (FastAPI), GitHub Actions, VS Solutions, Docker, Linux, OpenSearch
Languages: Mandarin Chinese (ACTFL OPI Advanced-Mid Certification; proficient in spoken and written)