Sagi Shaier

Boulder, CO | Sagi.Shaier@colorado.edu | LinkedIn | GitHub | Portfolio

EDUCATION

University of Colorado Boulder — (expected) Spring/Summer 2025
PhD Computer Science
Dissertation: "Factual Knowledge-enhanced Question Answering in Dynamic Environments"

University of Colorado Boulder — May 2023
MS Computer Science

Kennesaw State University — August 2014-December 2018
BS Computational and Applied Mathematics, (summa cum laude)
Concentration in Epidemiology, Minor in Statistics, Pre-Med.

SKILLS

Programming Languages: Python, Bash, SQL, Java, MATLAB
Frameworks: PyTorch, TensorFlow, HuggingFace, OpenCV, Keras, NumPy, Pandas, sklearn, scikit-learn, Seaborn, AWS
Tools: Git, Docker, TensorBoard, WandB, Linux/Unix, LaTeX, OpenAI
Project Management: strategic planning, budgeting, goal posting, delegation and supervision
Communication: scientific and analytical writing, public speaking and presenting, teaching and training

PROFESSIONAL EXPERIENCE

University of Colorado Boulder — August 2020-present
Doctoral Researcher

- **Minimized inference requirements by over 90% in "foundational" models** using biologically-inspired algorithms, accelerating progress towards artificial general intelligence (AGI) and highly-scalable generative AI.
- Developed a framework for **building more accurate and trustworthy language models** that prioritize factuality and mitigate hallucinations by grounding their generation with citations.
- **Increased accuracy by up to 42%** in large language models (LLMs) knowledge assessment through the development of a novel evaluation method, which also **reduces language generation redundancy by up to 40%**.
- Managed 10+ groups of graduate students on a variety of natural language processing projects, such as information retrieval, multihop question answering (QA), knowledge graphs, multilingual AI, and dialogue systems.
- **Improved accuracy by up to 36% in retrieval augmented generation (RAG)** models using attention-based strategies.
- **Awarded the prestigious Social Impact Award** out of 363 international candidates for identifying fairness concerns in biomedical QA systems, paving the way for more reliable and fair AI in high-stakes healthcare applications.

Oracle — January 2024-April 2024
Research Intern

- Designed and implemented scalable multi-GPU machine learning systems for large scale training and inference, which **supports 70B parameter models** in both supervised and unsupervised settings.
- **Developed 5 novel datasets** to assess language models' ability to answer complex, ambiguous questions with citations, spanning multiple domains and requiring challenging multihop reasoning.
- **Increased LLMs' question answering accuracy by 19.4%** and **citation generation accuracy by 86.7%** through innovative prompt engineering and fine-tuning techniques, yielding more accurate, trustworthy, and reliable AI.

National Institute of Health (NIH) — May 2023-August 2023
Machine Learning Research Intern

- Collaborated with several researchers teams in **developing multimodal NLP applications**, such as knowledge representation and QA systems in the biomedical domain using both structured and unstructured data.
- Produced intuitive visualizations that **simplify complex algorithms**, making them accessible to a broader audience.
- **Designed a biologically-inspired continual learning algorithm** that leverages a sparse mixture of experts, allowing computer vision models to retain knowledge, improve transfer learning, and adapt to dynamic environments.

Pacific Northwest National Laboratory (PNNL) — May 2021-October 2021
National Security Research Intern

- **Revealed novel patterns in high-dimensional text data and word embeddings** using various techniques intersecting topology and natural language processing, uncovering intrinsic logic and advancing natural language understanding.
- **Improved factual knowledge representation** with non-Cartesian normalization methods.

Quantum Metric December 2019-May 2020

Data Scientist (Research Team)

- Optimized query performance and analytics on massive datasets by integrating Google's BigQuery, AutoML, and Google cloud computing, and enabling efficient processing of massive datasets.
- **Maximized business outcomes** through predictive behavioral analytics, leveraging deep learning techniques like XGBoost and K-Means to drive revenue growth and customer engagement in diverse industries.
- **Led research initiatives** that applied advanced ML methods to detect real-time app crashes and service disruptions, saving millions annually by minimizing downtime and improving customer satisfaction.

Welocalize May 2019-August 2019

Machine Learning Intern

- **Grew project management capabilities** by developing predictive regression models using advanced data analysis techniques, complex SQL queries, Pandas, Numpy, and data visualization with Plotly.
- Delivered actionable insights through data analysis, utilizing clustering, Seaborn correlations, and dimensionality reduction to identify organizational bottlenecks, and optimizing the time tracking system for improved performance.

VOLUNTEER EXPERIENCE

National Institute of Mental Health (NIMH) August 2023-present

Machine Learning Researcher

- **Accelerate neural systems learning speed and accuracy** by developing biologically-inspired continual learning algorithms.

Hack Oregon February 2019-September 2019

Machine Learning Researcher

- Created a spatial distribution map of potential casualties of a Cascadia Earthquake to **inform medical response strategies**.

HONORS & AWARDS

University of Colorado Boulder, Outstanding Research Papers Award	($500)	May 2024
University of Colorado Boulder, Outstanding Service Award	($500)	May 2024
University of Colorado Boulder, Publication Recognition Award x4	($2,000)	August 2023, Spring 2024
AACL, Social Impact Award	($0)	August 2023
Nelson A. Prager Family Fund, Nelson Prager and James H Martin Graduate Award	($2,000)	August 2023
University of Colorado Boulder, Outstanding Student Award	($500)	August 2023

PUBLICATIONS

S. Shaier, F. Pereira, K. von der Wense, L. Hunter, and M. Jones. More Experts Than Galaxies: Conditionally-overlapping Experts With Biologically-inspired Fixed Routing (**Under Review**) 2024.

S. Shaier, A. Kobren, and P. Ogren. Adaptive Question Answering: Enhancing Language Model Proficiency for Addressing Knowledge Conflicts with Source Citations. Empirical Methods in Natural Language Processing (**EMNLP**) 2024.

S. Shaier, L. Hunter, and K. von der Wense. It Is Not About What You Say, It Is About How You Say It: A Surprisingly Simple Approach For Improving Reading Comprehension. Association for Computational Linguistics (**ACL Findings**) 2024.

S. Shaier, L. Hunter, and K. von der Wense. Desiderata For The Context Use Of Question Answering Systems. European Chapter of the Association for Computational Linguistics (**EACL**) 2024.

S. Shaier, K. Bennett, L. Hunter, and K. von der Wense. Comparing Template-based And Template-free Language Model Probing". European Chapter of the Association for Computational Linguistics (**EACL**) 2024.

S. Shaier, L. Hunter, and K. von der Wense. Who Are All The Stochastic Parrots Imitating? They Should Tell Us!. Asia-Pacific Chapter of the Association for Computational Linguistics (**AACL**) 2023.

S. Shaier, K. Bennett, L. Hunter, and K. von der Wense. Emerging Challenges In Personalized Medicine: Assessing Demographic Effects On Biomedical Question Answering Systems. Asia-Pacific Chapter of the Association for Computational Linguistics (**AACL**) 2023.

S. Shaier, M. Raissi, and P. Seshaiyer. Data-driven approaches for predicting spread of infectious diseases through DINNs: Disease Informed Neural Networks (**Letters in Biomathematics)** 2022.

RELEVANT COURSEWORK

Artificial Intelligence, Biologically Inspired Multi-Agent Systems, Statistical Methods, Computer Vision, Conversational AI, Deep learning, Dialogue Systems, Graph Theory, Information Retrieval, Machine Learning, Natural Language Processing, Neural Networks, Reinforcement Learning, Linear Algebra, Mathematical Biology, Calculus, Dynamic Models in Biology