Sagi **Shaier**

📄 Google Scholar   ✉ Sagi.Shaier@colorado.edu   ⬡ GitHub   🌐 Personal Website   in LinkedIn   ↗ Towards Data Science Blog

Research Interests_____

Language models, dialogue & QA systems, biomedicine, factuality, continual learning, knowledge graphs, information retrieval

Education_____

University of Colorado Boulder   *Boulder, Colorado*

PH.D. IN COMPUTER SCIENCE   *Aug. 2020 - Aug. 2025 (estimate)*
Co-advised by Katharina Kann and Lawrence Hunter

Kennesaw State University   *Kennesaw, Georgia*

BACHELOR'S OF SCIENCE   *Aug. 2014 - Dec. 2018*
Major in Computational and Applied Mathematics, Statistics minor, Pre-Med track. 3.96 GPA

Publications (Accepted)_____

2023   **Shaier**, Hunter, Kann, **"Who Are All The Stochastic Parrots Imitating? They Should Tell Us!"** The 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (**IJCNLP-AACL**)

2023   **Shaier**, Bennett, Hunter, Kann, **"Emerging Challenges in Personalized Medicine: Assessing Demographic Effects on Biomedical Question Answering Systems"** The 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (**IJCNLP-AACL**). **Won Social Impact Award.**

2022   **Shaier**, Hunter, Kann, "Mind the Knowledge Gap: A Survey of Knowledge-enhanced Dialogue Systems" (**arXiv** submission)

2022   **Shaier**, S., Raissi, M. and Seshaiyer, P., "Data-Driven Approaches for Predicting Spread of Infectious Diseases Through DINNs: Disease Informed Neural Networks", **Letters in Biomathematics**., Vol. 8, Iss. 1, pp. 71-105. Impact factor: 2.78.

2019   **Shaier**, S, Burke, M. "A Mathematical Model for the Effect of Domestic Animals on Human African Trypanosomiasis (Sleeping Sickness)," **The Kennesaw Journal of Undergraduate Research**: Vol. 6 : Iss. 1, Article 1.

Publications (Under Review)_____

2023   **Shaier**, Hunter, Kann, **"Desiderata For The Context Use Of Question Answering Systems"**.

2023   **Shaier**, Bennett, Hunter, Kann, **"Factual Probing Techniques: Comparing Template and Template-free Approaches"**.

Professional Experience_____

The National Institutes of Health

Machine Learning Research Special Volunteer   *May. 2023 - Present*
• Developing biologically-inspired continual learning algorithms

University of Colorado Boulder

Graduate Research Assistant   *Jan. 2022 - Present*
• Validating and improving our internal biomedical knowledge graph and knowledge graph creation methods

Graduate Teaching Assistant   *Aug. 2020 - Jan. 2022*
• Algorithms: graded proof-based assignments and taught weekly hour-long recitations on the fundamentals of algorithms and various algorithmic strategies
• Human Computer Interaction: directed weekly hour-long recitations on techniques for user-centered design and the development of interactive technologies

Pacific Northwest National Laboratory

National Security PhD Intern   *May. 2021 - Oct. 2021*
• Applied topological structures to high dimensional word embeddings
• Used topological data analysis methods to identify logic in NLP

- Experimented with novel word representation approaches

Quantum Metric
Data Scientist - Research Team   *Dec. 2019 - May. 2020*
- Spearheaded strategic projects such as predictive behavioral analytics, predictive churn risk score, loyalty score, & more
- Researched, constructed, and implemented various algorithms to predict customers' psychological nature across diverse industries such as airlines, retail, & banking
- Integrated Google's BigQuery, AutoML, & servers to optimize queries and analyses on big data while utilizing numerous supervised & unsupervised algorithms such as XGBoost, K-Means, Regressions, DBSCAN, PCA, Random Forest, & Neural Networks

Kepler Computing
Machine Learning Research Engineer (Contract)   *July. 2019 - Oct. 2019*
- Identified key computation characteristics of AI workloads in areas such as self-driving, vision, speech, and AR/VR
- Worked with the chief computer architect and HPC engineers to co-design groundbreaking AI ASIC
- Studied cutting edge AI workloads and tracked innovations happening in AI

Welocalize
Machine Learning Intern   *May. 2019 - Aug. 2019*
- Built regression models to predict project duration times and to detect outliers
- Gathered, engineered, and visualized various data types
- Identified bottlenecks within the organization using clustering analysis, Seaborn correlations, dimensionality reduction, and H2O.AI RF

Hack Oregon
Data Scientist on the Disaster team (Volunteer)   *Feb. 2019 - Oct. 2019*
- Mapped the distribution of likely casualties in the 9.0 Cascadia Earthquake in order to better understand medical response needs

Awards_____

Fall 2023 **IJCNLP-AACL 2023 Social Impact Award** AACL

Fall 2023 **Publication Recognition Award** University of Colorado Boulder

Fall 2023 **Endowed Founder's Fellowship** University of Colorado Boulder

Fall 2023 **Nelson Prager and James H Martin Endowed Graduate Fellowship** University of Colorado Boulder

Fall 2023 **Conference Support Fellowship** University of Colorado Boulder

Summer 2021 **Dean's Summer Research Fellowship** University of Colorado Boulder

Fall 2020 **Early Career Professional Development Fellowships** University of Colorado Boulder

Fall 2020 **Departmental Fellowship** University of Colorado Boulder

Fall 2017, Spring 2018, Fall 2018 **National Science Foundation Scholarship** Kennesaw State University

Summer 2018 **Society for Mathematical Biology Award** Kennesaw State University

Summer 2018 **Birla Carbon Scholarship** Kennesaw State University

Spring 2018 **Undergraduate Research Award** Kennesaw State University

Talks_____

Nov. 2023	Emerging Challenges in Personalized Medicine: Assessing Demographic Effects on Biomedical Question Answering Systems	Asia-Pacific Chapter of the Association for Computational Linguistics (IJCNLP-AACL) in Bali, Indonesia
Dec. 2022	On Language Models, Medicine, and Factuality	International Society for Computational Biology in Aspen, Colorado

Posters_____

Nov. 2023	Who Are All The Stochastic Parrots Imitating? They Should Tell Us!	Asia-Pacific Chapter of the Association for Computational Linguistics (IJCNLP-AACL) in Bali, Indonesia

Jan. 2023	Extreme Multi-hop Question Answering on a Massive Biomedical Knowledge Graph	Pacific Symposium on Biocomputing in The island of Hawaii, Hawaii
Dec. 2022	On Language Models, Medicine, and Factuality	International Society for Computational Biology in Aspen, Colorado
Jul. 2018	A Mathematical Model for the Effect of Domestic Animals on the Basic Reproduction Number of Human African Trypanosomiasis (Sleeping Sickness)	Annual Meeting of the Society for Mathematical Biology in Sydney, Australia
Apr. 2018	A Mathematical Model for the Effect of Domestic Animals on the Basic Reproduction Number of Human African Trypanosomiasis (Sleeping Sickness)	KSU Symposium of Student Scholars in Kennesaw, Georgia
Jul. 2017	A Mathematical Model for the Effect of Domestic Animals on Human African Trypanosomiasis (Sleeping Sickness)	Annual Meeting of the Society for Mathematical Biology in Utah
Apr. 2017	A Mathematical Model for the Effect of Domestic Animals on Human African Trypanosomiasis (Sleeping Sickness)	KSU Symposium of Student Scholars in Kennesaw, Georgia
Mar. 2017	A Mathematical Model for the Effect of Domestic Animals on Human African Trypanosomiasis (Sleeping Sickness)	Southern Regional Honors Conference in Asheville, North Carolina
Apr. 2016	A Mathematical Model of African Sleeping Sickness	KSU Symposium of Student Scholars in Kennesaw, Georgia

Mentoring Master Students

Fall 2023	Proposing a new task and baselines on evaluating when text requires a citation	Arvind Sreenivas
Fall 2023	Multihop question answering for language models with extreme context size	George Baker
Fall 2023	Multihop question answering for language models with extreme context size	Ankush Raut
Fall 2023	Developing a method to automatically evaluate dialogue systems	Mario Choy
Fall 2023	Developing a method to automatically evaluate dialogue systems	Chiranthan Sridhar
Spring 2022	Evaluating when knowledge graphs are better than free-form text	Satviki Sanjay Pathak
Spring 2023	Using language models to generate text with citations to evaluate its factuality	Chiranthan Sridhar
Spring 2023	Evaluating evaluation methods that measure the factuality of text	Arvind Sreenivas
Spring 2023	Developing approaches that use existing medical textbooks to improve the performance on the United States Medical License Exams dataset (MedQA-USMLE)	George Baker
Fall 2022	Developing a method to automatically evaluate dialogue systems	Satviki Sanjay Pathak

Military

Israel Defense Forces

Israel
Aug 2010 - Aug. 2013

- Prevented further escalations in terrorism throughout the Middle East