Ghazal Khalighinejad

gk126@duke.edu | Google Scholar | GitHub | ghazalkhalighinejad.github.io

Education

Duke University September 2021 - Present
Ph.D. Student in Computer Science; GPA: 3.92/4
Advisors: Bhuwan Dhingra, Sam Wiseman

Sharif University of Technology September 2017 - May 2021
Bachelor of Science in Computer Science

Research Interest

I'm interested in advancing transformer-based models beyond natural language, with applications in sciences. My research explores how neural architectures affect model performance, methods for embedding multimodal data, and for handling challenges associated with long-context tasks.

Research Experience

Google Research Student Researcher May 2025 - Present
Training a retrieval model to detect suboptimal code patterns using code embeddings.

The Simons Foundation & Polymathic AI Guest Researcher March 2025 - May 2025
Developed a multimodal foundation model for retrieval over astronomical observations.

Adobe Research Research Intern May 2024 - August 2024
Developed a consistent and fault-tolerant text-to-video retrieval system.

Publications

1. MatViX: Multimodal Information Extraction from Visually Rich Articles. Ghazal Khalighinejad, Sharon Scott, Ollie Liu, Kelly L. Anderson, Rickard Stureborg, Aman Tyagi, Bhuwan Dhingra. ***NAACL* 2025,** Oral Presentation.

2. Training Neural Networks as Recognizers of Formal Languages. Alexandra Butoi, Ghazal Khalighinejad, Anej Svete, Josef Valvoda, Ryan Cotterell, Brian DuSell. ***ICLR* 2025.**

3. IsoBench: Benchmarking Multimodal Foundation Models on Isomorphic Representations. Deqing Fu*, Ruohao Guo*, Ghazal Khalighinejad*, Ollie Liu*, Bhuwan Dhingra, Dani Yogatama, Robin Jia, Willie Neiswanger (*Equal contribution). ***COLM* 2024.**

4. Extracting Polymer Nanocomposite Samples from Full-Length Documents. Ghazal Khalighinejad, Defne Circi, L.C. Brinson, Bhuwan Dhingra. ***Findings of ACL* 2024**.

5. How Well Do Large Language Models Understand Tables in Materials Science? Defne Circi, Ghazal Khalighinejad, Anlan Chen, Bhuwan Dhingra, L.C. Brinson. *Integrating Materials and Manufacturing Innovation (**IMMI 2024**).*

6. Approximating CKY with Transformers. Ghazal Khalighinejad, Ollie Liu, Sam Wiseman. ***Findings of EMNLP* 2023**.

7. Exploring the Effect of Frequency Resolution in FNet. Gregory Szumel, Ghazal Khalighinejad, Rickard Stureborg and Sam Wiseman. ***SustaiNLP @ ACL* 2023**.

8. Retrieval of Synthesis Parameters of Polymer Nanocomposites using LLMs. Defne Circi, Ghazal Khalighinejad, Shruti Badhwar, Bhuwan Dhingra, L. Brinson. ***AI4MAT @ NeurIPS* 2023**.

9. Galloping in Fast-Growth Natural Merge Sorts. Elahe Ghasemi, Vincent Jugé, Ghazal Khalighinejad. ***ICALP* 2022**.

Preprints

1. Open-Domain Document Retrieval with Vision Language Models. Ghazal Khalighinejad, Raghuveer Thirukovalluru, Bhuwan Dhingra. *In submission*, 2025.

2. It's LIT! LLMs with Interpretable Tool Calling. Ruixin Zhang, Jon Donnelly, Zhicheng Guo, Ghazal Khalighinejad, Haiyang Huang, Alina Jade Barnett, Cynthia Rudin. *In submission*, 2025.

Research Experience

Multimodal Large Language Models for Sciences Current

- Created a large dataset and evaluation method for extracting structured data from visually rich scientific articles using vision-language models.
- Proposed novel evaluation metrics for assessing model performance on tasks involving multimodal data extraction, with a focus on curve similarity and hierarchical structure alignment.

Algorithmic Reasoning in Transformers 2022-2023

- Trained transformers to approximate CKY parsing, replacing CKY in modern constituency parsers without accuracy loss and improving runtime from cubic to quadratic dependence on sentence length.

Sorting Algorithms 2021-2022

- Proved that several merge sorting algorithms are as efficient as TimSort when employing its galloping sub-routine.

Awards & Achievements

aiM National Science Foundation Fellow: Awarded a full-tuition scholarship and funding for research in AI + Materials.

ACM-W Computer Science Research Conference Scholarship: Awarded a scholarship to attend NeurIPS 2022.

CRA-WP Scholarship: Awarded a scholarship to attend CRA-WP Grad Cohort for Women Workshop.

TEACHING EXPERIENCE

Graduate Teaching Assistant:

Introduction to Natural Language Processing Fall 2022
Instructor: Bhuwan Dhingra

Design and Analysis of Algorithms Spring 2021
Instructor: Rong Ge

Relevant Coursework

Machine Learning: Neurosymbolic Machine Learning, Advanced Natural Language Processing, Deep Learning, Causality and Interpretability

Algorithms and Theory: Distributed Algorithms, Algorithmic Game Theory, Information Theory, Probability and Statistics, Algorithms, Operations Research, Mathematical Analysis, Linear Algebra

Skills

Programming: Python, Java, C/C++, Matlab

Libraries: PyTorch, JAX, TensorFlow

Others: Git, Bash Shell Scripting, LaTeX