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 the capabilities of transformer-based models beyond natural language, with a focus on addressing problems in physics, chemistry, and materials science. My research explores how neural architectures shape model performance, methods for embedding multimodal data, and strategies for handling challenges associated with long-context tasks.

EMPLOYMENTS

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

PREPRINTS

1. MatViX: Multimodal Information Extraction from Visually Rich Articles. Ghazal Khalighinejad, Sharon Scott, Ollie Liu, Kelly L. Anderson, Rickard Stureborg, Aman Tyagi, Bhuwan Dhingra. *In submission*, 2024. Available at https://arxiv.org/abs/2410.20494

2. Training Neural Networks as Recognizers of Formal Languages. Alexandra Butoi, Ghazal Khalighinejad, Anej Svete, Josef Valvoda, Ryan Cotterell, Brian DuSell. *In submission*, 2024. Available at https://arxiv.org/abs/2411.07107

3. 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*, 2024.

4. Reflections from the 2024 Large Language Model (LLM) Hackathon for Applications in Materials Science and Chemistry. Yoel Zimmermann, Adib Bazgir, Zartashia Afzal, ... Ghazal Khalighinejad, ... Ian Foster, Ben Blaiszik, et al. *In submission*, 2024.

PUBLICATIONS

1. 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.**

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

3. 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**)*.

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

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

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

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

Research Experience

Multimodal Large Language Models for Sciences <div align="right">Current</div>
- 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 <div align="right">2022-2023</div>
- 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 <div align="right">2021-2022</div>
- 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 <div align="right">Fall 2022</div>
Instructor: Bhuwan Dhingra

Design and Analysis of Algorithms <div align="right">Spring 2021</div>
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