Amirhossein Kazemnejad

6650 Rue Saint-Urbain, Montreal, QC H2S 3G9

amirhossein.kazemnejad@mila.quebec • Website • kazemnejad • Google Scholar • a_kazemnejad

Research Interests

Compositional Generalization, Length Extrapolation, Computational Power of Transformers, Positional Encoding, Representation Learning for Sequential Data, Natural Language Processing (NLP)

Education

McGill University / Mila *Jan 2021 - Nov 2023*
MSc (Thesis-based) in Computer Science *GPA: 4/4*
Supervisor: Prof. Siva Reddy
Thesis: *The Cure or the Curse: Investigating the Role and Challenges of Positional Encoding in Transformers*

Iran University of Science and Technology (IUST) *Sep 2015 - May 2020*
BSc in Computer Engineering *GPA: 17.90/20 (Rank 2/35)*
Supervisor: Prof. Mohammad Taher Pilehvar
Thesis: *Analyzing The Effects of Semantic Violations on Computational Language Models*

Publications

The Impact of Positional Encoding on Length Generalization in Transformers [link] 2023
Neural Information Processing Systems **(NeurIPS)**
Amirhossein Kazemnejad, Inkit Padhi, Karthikeyan Natesan, Payel Das, Siva Reddy

Measuring the Knowledge Acquisition-Utilization Gap in Pretrained Language Models [link] 2023
Findings of Empirical Methods in Natural Language Processing **(EMNLP)**
Amirhossein Kazemnejad, Mehdi Rezagholizadeh, Prasanna Parthasarathi, Sarath Chandar

The Curious Case of Absolute Position Embeddings [link] 2022
Findings of Empirical Methods in Natural Language Processing **(EMNLP)**
Koustuv Sinha*, Amirhossein Kazemnejad*, Siva Reddy, Joelle Pineau, Dieuwke Hupkes, Adina Williams

Paraphrase Generation by Learning How to Edit from Samples [link] 2020
Association for Computational Linguistics **(ACL)**
Amirhossein Kazemnejad, Mohammadreza Salehi, Mahdieh Soleymani

**equal contribution*

Research Experience

Present – Jan 2021 **Mila - Québec AI Institute**
Graduate Research Assistant | *Supervisor: Prof. Siva Reddy*
- Working on improving human-like and compositional generalization and adaptive computation in Transformers. Collaborated with teams from AI2, Meta AI, and IBM Research.
- **Focus**: Out-of-Distribution Generalization, Transformer Architecture, Positional Encoding

Aug 2023 – Jul 2022 **Noah's Ark Lab - Montréal Branch**
Part-time Research Intern @ NLP Team, employed via Quantum.ca | *Supervisor: Dr. Mehdi Rezagholizadeh*
- Investigating the gap between world knowledge acquisition and utilization in pre-trained language models.
- **Focus**: Parametric Knowledge, Downstream Knowledge Transfer, Scaling Behaviour of LLMs

Jun 2020 – Aug 2018 **Machine Learning Lab - Sharif University**
Undergraduate Research Assistant | *Supervisor: Prof. Mahdieh Soleymani*
- Worked on equipping sequence-to-sequence Transformers for natural language with a retrieval-based editor module to improve conditional text-generation.
- **Focus**: Paraphrase Generation, Conditional Text Generation, Retrieval-Augmented Models

Selected Research Projects

Length Generalization in Transformers *Sep 2022 - Sep 2023*
Advisor: Prof. Siva Reddy
- Evaluating the impact of positional encoding on length generalization and its interaction with scratchpad (chain-of-thought).
- Demonstrated the inefficacy of currently used positional encoding and the surprising performance of Transformers with no positional encoding.

- Analyzed the differences in algorithms implemented by various positional encodings both theoretically and empirically. Work done in collaboration with IBM Research and published in **NeurIPS 2023**

Usable Parametric Knowledge in Pre-Trained Language Models *Jul 2022 - Aug 2023*

Advisor: Dr. Prasanna Parthasarathi

- Designed a novel framework to measure how much of parametric knowledge can be actually utilized in downstream tasks.
- Showed the assumption that all parametric knowledge is readily usable does not hold true in practice for a variety of models.
- Demonstrated the gap between usable and acquired knowledge still exists even when scaling model sizes by finetuning OPT models from 125M to 13B parameters. Submitted a conference paper to **EMNLP 2023**

The Shift Invariance Property in Absolute Position Embeddings *Mar 2022 - Jul 2022*

Advisor: Prof. Siva Reddy

- Studied absolute position embeddings (APEs) in Transformers and their ability to capture relative position information.
- Demonstrated over-reliance on positional shortcuts and poor performance on non-zero positions of models trained with APEs.
- Provided extensive experimental results over 8 finetuning and prompting tasks. Work done in collaboration with Meta AI and published at **EMNLP'22** Findings. Also, presented at **BlackboxNLP'22**

Contrastive Learning For Structured Prediction *Jan 2021 - Dec 2021*

Advisor: Prof. Siva Reddy

- Proposed a contrastive learning framework to enforce consistency of local structures in the output space.
- Applied the framework to the representations of the Transformer's decoder to improve compositional generalization.
- Improved out-of-domain performance in SCAN and a few semantic parsing tasks with no modifications to the model. Work done in collaboration with AI2.

Improving Transformer Architecture for Retrieval-Augmented Text Generation *Aug 2018 - Jun 2020*

Advisor: Prof. Mahdieh Soleymani

- Incorporated the retrieve-and-edit framework into Transformers with a novel module for a fine-grained editing process.
- Utilized the framework in paraphrase generation for improved quality and data augmentation for improved diversity.
- Demonstrated the generation improvements in terms of automatic metrics and human evaluation. Work published at **ACL'20**.

Teaching Experience

Teaching Assistant, McGill, COMP 330 Theory of Computation (Prof. Prakash Panangaden) *Fall 2021*
Teaching Assistant, McGill, COMP 204 Computer Programming (Prof. Yue Li) *Winter 2021*
Teaching Assistant, Sharif Univerisity, CE719 Deep Learning (Prof. Mahdieh Soleymani) *Winter 2020*
Teaching Assistant, IUST, Deep Learning (Prof. Mohammad Taher Pilehvar) *Winter 2019*
Teaching Assistant, IUST, Intro. to NLP (Prof. Sauleh Eetemadi) *Winter 2019*
Teaching Assistant, IUST, Intro. to AI and Expert Systems (Prof. Mohammad Taher Pilehvar) *Fall 2018*

Professional Activities

Organiser	GenBench Workshop 2023 (website), EACL 2021 (website), Virtual Conference of EMNLP 2020 (website)
Program Committee	EMNLP 2023, ACL 2023, EMNLP 2022, ACL Rolling Review Dec. 2021

Services and Volunteer work

Open-Source Contribution

- **TensorFlow 2.0** (GitHub links: #375, #503, #511, #534, #546, #535, #673, #603, #335)
  Contributed to the Seq2Seq module's features, bug fixes, and documentation.
- **GenBench Collaborative Benchmarking Task** (GitHub Repo)
  A Python framework for collaborative task submission and an automated pipeline for task evaluation and verification.
- **Jekyll Academic Template** (GitHub Repo)
  Feature-rich Jekyll template for academic courses with 44 starts and 82 forks on GitHub.
- **PLM Research Codebase** (GitHub Repo)
  PyTorch/Huggingface codebase, Seq2Seq tasks, decoder-only architectures, support for dependency injection, and SLURM.

Blog Posts

- **Transformer Architecture: The Positional Encoding** (link)
  Top Google result for *Positional Encoding* with nearly 5K monthly views. Referenced by Stanford, CMU, and MIT courses.
- **How to do Deep Learning research with absolutely no GPUs** (link to parts #1 and #2)
- **TensorFlow 2.0 Tutorial** (link)

Honors and Awards

(2023) **Best Reviewer** Award at EMNLP 2023.

(2021) Awarded **Graduate Excellence Fellowship** by McGill University.

(2020) Ranked **2nd** among Computer Engineering students of the 2015 batch.

(2017-2019) Recipient of **Outstanding Student** Award by Iran University of Science and Technology.

(2015) Top **99.2nd** percentile in national university entrance exam among nearly 200,000 participants.

Skills

Programming Languages	Python, Java, C++, Bash
Frameworks & Libraries	PyTorch, HuggingFace, DeepSpeed, TensorFlow, Keras, WandB, Matplotlib, Plot9
Other Tools	Git, Docker, Singularity, SLURM, GitHub Actions
Languages	Persian (Native), English (Fluent), Arabic (Limited)