Mohamed Elgaar

https://mohdelgaar.github.io

Email : melgaar@cs.uml.edu
Mobile : 978-221-1549

Research Interests

- **Curriculum Learning:**: Exploring methods to enhance training speed, robustness, and performance of machine learning models through curriculum learning techniques. Additionally, investigating the use of linguistic features to analyze datasets and models, including identifying and addressing biases towards specific linguistic features, to inform the development of robust models. Lastly, researching the extension to data augmentation strategies for improving model generalization and addressing identified weaknesses through linguistic analysis.
- **Controlled Generation and Data Augmentation:**: Researching controlled generation techniques in the scope of linguistic features.

Education

University of Massachusetts Lowell — Lowell, MA

- *Ph.D. in Computer Science* — *2021 ∼ currently*
  PI: Hadi Amiri

Korea Advanced Institute of Science and Technology (KAIST) — Daejeon, South Korea

- *B.S. in Computer Science* — *2016 ∼ 2020*
  Interdisciplinary Minor: Industrial, Electrical Engineering and Math

Awards

- KAIST College of Engineering Innovator Award (2020)

Publications

Preprints

- Mohamed Elgaar, Hadi Amiri, **Multi-Attribute Linguistic Tuning for Controlled Paraphrase Generation**, Under Review
- Hadi Amiri, Nidhi Vakil, Mohamed Elgaar, Jiali Cheng, Mitra Mohtarami, Adrian Wong, Mehrnaz Sadrolashrafi, Leo Anthony G Celi, **Analysis of Race, Sex, and Language Proficiency Disparities in Documented Medical Decisions**, medRxiv 2024.07.11.24310289

Peer-Reviewed

- Mohamed Elgaar, Jiali Cheng, Nidhi Vakil, Hadi Amiri, Leo Anthony Celi, **MedDec: A Dataset for Extracting Medical Decisions from Clinical Narratives**, Findings of the Association for Computational Linguistics: ACL 2024
- Jiali Cheng, Mohamed Elgaar, Nidhi Vakil, Hadi Amiri **CogniVoice: Multimodal and Multilingual Fusion Networks for Mild Cognitive Impairment Assessment from Spontaneous Speech**, Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH 2024)
- Mohamed Elgaar, Hadi Amiri, **Ling-CL: Multiview Curriculum Learning using Linguistic Complexity**, Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)
- Mohamed Elgaar, Hadi Amiri, **HuCurl: Human-induced Curriculum Discovery**, Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL 2023)
- Mohamed Elgaar, Jungbae Park, Sang Wan Lee, **Multi-Speaker and Multi-Domain Emotional Voice Conversion Using Factorized Hierarchical Variational Autoencoder**, 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2020)

Patents

- Mohamed Elgaar, Jungbae Park, **Methods for Emotional Voice Conversion**, Korean patent (10-2277205)

Experience

- **NAVER Clova** — Seongnam, Korea
  ML Research Intern — *Sep 2020 - Jan 2021*
  - Investigated novel methods for feature representation and fusion to enhance model performance across multiple speech datasets.
  - Research on advancing Spoken Language Understanding (SLU) through the development of a multi-task learning model. Achieved notable improvements in accuracy and efficiency.

  Reference: Ryan Seongjin Shin, Tech Lead. sungjin.712@navercorp.com, +82-10-2907-1833

- **Humelo** — Seoul, Korea
  ML Research Intern — *June 2019 - May 2020*
  - Lead the research project on emotion conversion.
  - Achieved multi-speaker, many-to-many emotion conversion, while maintaining speaker identity.

  Reference: Jungbae Park, Research Team Leader. jbpark0614@gmail.com, +82-10-5190-1589

- **Crazing Lab** — Seoul, Korea
  ML Engineering Intern — *June 2018 - Aug 2018*
  - Created a pipeline for data collection and annotation, invloving the integration of Realsense depth sensors, OpenCV, and image registration.
  - Created a model for real-time human and body parts tracking using RGB&Depth sensors.
  - Extended the YOLOv3 object detection framework to utilize 16-bit depth maps, involving CUDA programming.

  Reference: JaeYoung Lee, Senior Robotics Engineer. jy@wom.ai, +82-10-4305-5579, +1-800-890-6895

Invited Talks

- **CS Reading Group, Qazvin Islamic Azad University.**: Understanding NLP Models through Linguistic Curricula. July 9, 2024.

Academic Service

- **Review**:
  - **Association for the Advancement of Artificial Intelligence (AAAI)**: 2021, 2022
  - **International Conference on Machine Learning (ICML)**: 2021, 2022
  - **Conference on Neural Information Processing Systems (NeurIPS)**: 2022, 2023
  - **Association for Computational Linguistics (ACL)**: 2023
  - **Empirical Methods in Natural Language Processing (EMNLP)**: 2023
  - **ACL Rolling Review (ARR)**: Oct'23*, Dec'23

  $\star$: Outstanding review.

- **Mentoring Students**:
  - **Aryan Nagpal**: Fall 2023
  - **Dipika Boro**: Spring 2022
  - **Manu Hegde**: Fall 2021

Undergraduate Research

- **Machine Learning & Intelligence Lab** KAIST, Daejeon, Korea
  Feb 2019 - June 2019 *PI: Eunho Yang*
  - Implemented and reproduced paper results, studied recent advances in machine learning, learned to identify new research problems.

  Reference: Professor Eunho Yang, MLI Lab. http://mli.kaist.ac.kr, eunhoy@kaist.ac.kr, +82-042-350-3572

- **Users & Information Lab** KAIST, Daejeon, Korea
  Jan 2018 - June 2018 *PI: Alice Oh*
  - Reproduced paper on source code summarization, implemented a machine translation model, investigated few-shot learning, developed a sequence to sequence model.

  Reference: JinYeong Bak, PhD candidate in U&I Lab. http://uilab.kaist.ac.kr, +82.10.3120.9382, jy.bak@kaist.ac.kr

Other Information

- **Languages**:
  - Arabic: Native
  - English: Fluent
  - Korean: Intermediate
- **Extra-curricular Activities**:
  - Egypt country representative at KAIST (3 years)
  - International students mentor (2 years)