

## RESEARCH INTERESTS

---

- **Curriculum Learning:** Investigating data selection and ordering strategies to advance how we train large language models, aiming to enable next-generation models at a fraction of current computational costs.
- **Controlled Generation:** Developing techniques that enable precise steering of language model outputs.
- **Model Interpretability:** Analyzing model behavior to understand training dynamics and generation steering.

## EDUCATION

---

- **University of Massachusetts Lowell** Lowell, MA  
• *Ph.D. in Computer Science* 2021 ~ currently  
*Advisor: Hadi Amiri*
- **Korea Advanced Institute of Science and Technology (KAIST)** Daejeon, South Korea  
• *B.S. in Computer Science* 2016 ~ 2020  
*Interdisciplinary Minor: Industrial Eng., Electrical Eng., Math*

## PUBLICATIONS

---

### Preprint

- [Mohamed Elgaar](#), Hadi Amiri, **P-Masking: Power Law Masking Improves Multi-attribute Controlled Generation**, arXiv:2410.24201 [PDF]
- [Mohamed Elgaar](#), Hadi Amiri, **Multi-Attribute Linguistic Tuning for Controlled Paraphrase Generation**, arXiv:2410.24199 [PDF]
- 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 [PDF]

### 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 ACL 2024 [PDF] [Code]
- Jiali Cheng, [Mohamed Elgaar](#), Nidhi Vakil, Hadi Amiri **CogniVoice: Multimodal and Multilingual Fusion Networks for Mild Cognitive Impairment Assessment from Spontaneous Speech**, INTERSPEECH 2024 [PDF] [Code]
- [Mohamed Elgaar](#), Hadi Amiri, **Ling-CL: Multiview Curriculum Learning using Linguistic Complexity**, EMNLP 2023 [PDF] [Code]
- [Mohamed Elgaar](#), Hadi Amiri, **HuCurl: Human-induced Curriculum Discovery**, ACL 2023 [PDF] [Code]
- [Mohamed Elgaar](#), Jungbae Park, Sang Wan Lee, **Multi-Speaker and Multi-Domain Emotional Voice Conversion Using Factorized Hierarchical Variational Autoencoder**, ICASSP 2020 [PDF] [Demo]

### Software & Tools

- [Mohamed Elgaar](#), Hadi Amiri, Mitra Mohtarami, Leo Anthony Celi, **MedDecXtract: A clinician-support system for extracting and visualizing medical decisions in clinical notes** [Demo] [Video]
- [Mohamed Elgaar](#), Hadi Amiri, **LingConv: An interactive toolkit for controlled paraphrase generation with linguistic feature control** [Demo] [Video]

## PATENTS

---

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

## AWARDS

---

- KAIST College of Engineering Innovator Award (2020)

## EXPERIENCE

---

- **NAVER Clova** Seongnam, Korea  
*ML Research Intern* *Sep 2020 - Jan 2021*
  - Developed and analyzed a multi-task learning framework for speech understanding that jointly optimizes intent classification, phoneme recognition, and emotion detection tasks.
  - *Reference:* Ryan Seongjin Shin, Tech Lead. sungjin.712@navercorp.com
- **Humelo** Seoul, Korea  
*ML Research Intern* *June 2019 - May 2020*
  - Led research on multi-speaker emotional voice conversion, achieving many-to-many emotion conversion while preserving speaker identity, resulting in a patent and ICASSP publication.
  - *Reference:* Jungbae Park, Research Team Leader. jbpark0614@gmail.com
- **Machine Learning & Intelligence Lab (PI: Eunho Yang)** Daejeon, Korea  
*Undergraduate Researcher* *Feb 2019 - June 2019*
  - Implemented and evaluated Wasserstein Auto-Encoders with adversarial training for diverse response generation in multimodal conversational AI.
  - *Reference:* Professor Eunho Yang, MLI Lab. eunho@kaist.ac.kr
- **Crazing Lab** Seoul, Korea  
*ML Engineering Intern* *June 2018 - Aug 2018*
  - Developed a real-time computer vision pipeline integrating RealSense depth sensors with OpenCV, extending YOLOv3 to process 16-bit depth maps through CUDA optimization for human pose estimation.
  - *Reference:* JaeYoung Lee, Senior Robotics Engineer. jy@wom.ai
- **Users & Information Lab (PI: Alice Oh)** Daejeon, Korea  
*Undergraduate Researcher* *Jan 2018 - June 2018*
  - Developed an LSTM-based sequence-to-sequence model with attention mechanism for generating natural language descriptions from source code changes.
  - *Reference:* JinYeong Bak, PhD candidate. jy.bak@kaist.ac.kr

## INVITED TALKS

---

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

## ACADEMIC SERVICE

---

- **Review:** ★: Outstanding 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, 2024
  - **Association for Computational Linguistics (ACL):** 2023
  - **Empirical Methods in Natural Language Processing (EMNLP):** 2023
  - **ACL Rolling Review (ARR):** Oct'23\*, Dec'23, June'24
- **Mentoring:**
  - **Academic Mentoring:** Aryan Nagpal (Fall 2023), Dipika Boro (Spring 2022), Manu Hegde (Fall 2021)
  - Additional service: KAIST International Student Mentor (2018-2020), KAIST International Student Representative (2017-2020)*
- **Teaching:**
  - **Teaching Assistant:** Foundations of Computer Science, UMass Lowell (Fall 2024)