

Natural language processing (NLP) community has moved towards a new paradigm for general models that are trained on broad data and are adaptable to a wide range of tasks. These models are called foundation models [1]. Almost all state-of-the-art NLP models use foundation models. While their large scale makes them powerful, all current models inherit the same biases, and their computational complexity makes them less accessible to individuals and academia [1]. In addition, current models are less interpretable [2], unable to understand situations that deviate somewhat from their training data or the designer's assumptions, and unable to generalize to new tasks [3]. To bridge the gap between humans and models, there is a tendency to build the next generation of models capable of generalizing to the unseen tasks and datasets [4] which helps the models benefit from cross-task and instance-level generalization [5].

As a Ph.D. student, I'm looking forward to working on generalization and enhancing the intelligence level in NLP systems. I had some initial research in this field, and I think pursuing a Ph.D. degree would be an excellent opportunity to contribute to this area.

Background During high school, I was introduced to Mathematical Olympiad, and I enjoyed solving challenging math problems. My experience convinced me to pursue a degree in STEM to apply my math aptitude to solve real-world problems. Also, it raised the problem of creating a system to solve mathematical questions in my mind, which was my initial motivation to enter the field of AI. With a ranking of 148th out of over 70,000 participants, I was admitted to the Amirkabir University of Technology, one of the top engineering schools in Iran.

Electrical engineering was my first choice as a major field. However, I preferred to study computer science and focus on AI. I was admitted to a double degree program and started computer engineering. Of course, arranging classes and taking more courses than other students each semester resulted in several conflicts in tests and projects and impacted my grades in some courses. Nonetheless, I think the experience was worthwhile. Knowledge of control systems has influenced my prospect of industrial projects and optimized systems. Furthermore, I owe my security concerns in Artificial Intelligence and NLP to study computer networks and security.

During my study, I grasped every opportunity to research. I was interned at the ITRC research center. Then, I was introduced to the ANS lab at the University of Tehran. I worked as a member of the AI group for two years in this lab. My main project was implementing a 3d eye tracker. The motivation of the project was to make high-accuracy eye-tracking available for low-cost devices. Finally, after testing different approaches, we used the U-Net architecture for pupil segmentation and modeled the gaze vector with pupil ellipse geometry.

Generalization Currently, I am working on the natural instruction expansion project¹. The motivation of this project is to increase the generalization to the unseen tasks. We collected a dataset of over 1700 tasks with their natural instruction. Reading the instructions usually builds humans' knowledge about the task, and the better the instructions explain the task, the better we perform it. How about the NLP models? In the previous version of this project, it was observed

¹ <https://github.com/allenai/natural-instructions-expansion>

that providing language instructions can improve the performance and generalization to novel tasks [6]. In this study, we covered a wide variety of tasks, domains, and reasoning skills to analyze the generalization and the structure and relation between the tasks.

One of the remarkable outcomes of the project for me was understanding the lack of generalization in NLP. Current NLP systems still have a long way to go to perform like humans. For example, we can understand a text or speech; then, we do the tasks based on the given context. In comparison, the current systems might outperform a particular task and completely fail on a novel task, which suggests their limited language understanding and reasoning abilities. Here, we should separate task-specified performance from intelligence. Most of the current impressive systems result from their creator's intelligence and large-scale datasets rather than the system's intelligence. An intelligence system comparable to an average human should be able to do the unseen tasks having the same amount of prior knowledge and experience as humans [7]. To achieve this goal, we should be able to evaluate the prior knowledge and experience of a system. Also, to avoid biases and ensure fairness, we should be able to debias the models.

Question-answering Question answering tasks have long been a measure of understanding in humans and computers. While we can ask and answer questions in different formats, most question-answering systems are format-specified. Although focusing on a narrow task can improve the results, due to the lack of generalization and biases in the datasets, it can't be a proper measure of understanding. In the UnifiedQA² project, we introduce a cross-format question-answering system. Previous work on this project has shown that these formats and boundaries are artificial, and training on a format can improve the results of others. Moreover, this system has a better performance in generalization to unseen tasks than the format-specified models [8]. As future work, I'm interested in diving deeper into this area to see how changing the training set while keeping it limited can improve performance and generalization to unseen tasks.

Studying Ph.D. Tackling different areas allowed me to make an informed decision for my Ph.D. NLP is my main area of interest. Changing the world through the way we can communicate with ourselves and computers more effectively is a step toward my dreams. Specifically, I am interested in question-answering, knowledge graphs, and commonsense reasoning. Pursuing a Ph.D. will be the best option for me to work in these fields and become an NLP researcher.

² <https://github.com/allenai/unifiedqa/>

References

- [1] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. v. Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, E. Brynjolfsson, S. Buch, D. Card, R. Castellon and e. al., "On the Opportunities and Risks of Foundation Models," *CoRR*, vol. abs/2108.07258, 2021.
- [2] M. Danilevsky, K. Qian, R. Aharonov, Y. Katsis, B. Kawas, and P. Sen, "A Survey of the State of Explainable AI for Natural Language Processing," in *1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 2020.
- [3] R. T. McCoy, J. Min, and T. Linzen, "BERTs of a feather do not generalize together: Large variability in generalization across models with similar test set performance," in *Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*, 2020.
- [4] D. Yogatama, C. d. M. d'Áutume, J. Connor, T. Kociský, M. Chrzanowski, L. Kong, A. Lazaridou, L. Y. Wang Ling, C. Dyer, and P. Blunsom, "Learning and Evaluating General Linguistic Intelligence," *CoRR*, vol. abs/1901.11373, 2019.
- [5] Q. Ye, B. Y. Lin, and X. Ren, "CrossFit: A Few-shot Learning Challenge for Cross-task Generalization in NLP," in *Conference on Empirical Methods in Natural Language Processing*, 2021.
- [6] S. Mishra, D. Khashabi, C. Baral and H. Hajishirzi, "Cross-Task Generalization via Natural Language Crowdsourcing Instructions," *CoRR*, vol. abs/2104.08773, 2021.
- [7] F. Chollet, "On the Measure of Intelligence," *CoRR*, vol. abs/1911.01547, 2019.
- [8] D. Khashabi, S. Min, T. Khot, A. Sabharwal, O. Tafjord, P. Clark and H. Hajishirzi, "UnifiedQA: Crossing Format Boundaries With a Single QA System," in *EMNLP*, 2020.