

The majority of the progress in Natural Language Processing (NLP) over the last few years have been achieved in English and other high-resource languages. It has however proved difficult to extend these same kinds of performance to low-resource language settings which have very limited data. Due to the large gap that currently exists between dominant and high-resource language technologies, my research goal is to bridge this gap by improving the state of modeling low-resource languages. This involves designing innovative ways to improve sample efficiency, representation quality, and ease of deployment in low-resourced settings. I am also very excited to explore multitask and multi-modal domains, so we can learn more efficient and robust shared representation and capture correspondences between modalities. These are the interesting problems I plan to tackle by pursuing a PhD at The University of Maryland. Below, I talk more about my past work along with the future directions I foresee pursuing.

**Previous Work.** I found the ability to control image generation with natural language very interesting which is the idea behind OpenAI’s DALL·E, however, these state-of-the-art models require massive amounts of computation and parameters to train. During the summer of 2021, I participated in the Hugging Face Community Event and worked on a group project ‘DALL·E mini’ [1] which is a JAX/FLAX implementation of OpenAI’s version but is about 30 times smaller. In the project, our goal was to implement and open-source a Zero-Shot Text-To-Image generation model and show that one can still achieve reasonable performance on this multi-modal task with far more accessible means of compute and data. Over the course of the project, I was involved in creating a unified data pipeline that preprocesses and combines multiple datasets. One challenge was that our available machine (TPU) storage (1TB) became insufficient for holding all the data. To combat this, I worked on the image encoding pipeline, to encode the RGB images into codewords using the VQ-GAN Encoder. This approach reduced the necessity of having images while training the generative model. It also reduced the computational overhead at training time. The project has received the ‘Best Project Award’ at the event and we have published a [technical report](#) as documentation of our work. Besides extending my knowledge about applied ML and large-scale experiments, I believe this experience has also offered me a peek into the exciting world of research and heavily impacted my research trajectory, as it honed my interests in exploring ways that involve making NLP technologies more sample efficient and accessible to people without large computational budgets.

**Future Research.** My current interests and experiences have drawn me towards two main research directions:

- **NLP Beyond English:** How models can learn effectively in a small data regime for low-resource languages?
- **NLP in Multimodal-Multitask Setting:** How can we move towards human-like linguistic generalization?

**NLP Beyond English.** Previously, I have worked on NLP projects ([gpt2-bengali](#), [sahajBERT](#)) for my native language Bengali. From these projects, I realized that high-performing results are often harder to achieve in very low-resource settings. Except for a handful of high-resourced languages, most world languages are still underserved by language technologies [2]. Recently, learning frameworks such as active learning are gaining ground because it reduces required training data while still maintaining good performance. One big challenge with adopting active learning for deep neural networks is that they fail to reliably produce uncertainty estimates, on which most query strategies rely. This is however something bayesian neural models will do better at [3, 4]. I intend to continue to explore frameworks such as active learning and other sample efficient methods in the NLP domain to understand how it helps to robustly learn from small data regimes while attaining good performance.

**NLP in Multimodal-Multitask Setting.** I have also found the recent works in the multitask [5, 6, 7] and multimodal [8, 9] settings very exciting and they have opened up many interesting research directions. The current multitask models especially with their ability to solve diverse sets of NLP tasks makes me wonder if they would take us toward human-like linguistic generalization, however, one of their big limitations is, these models can only solve the tasks that they were trained on. To go beyond, they need to be able to learn and generalize to unseen tasks. Also, I believe that the current methods for developing NLP models might have constrained them to never truly achieve language understanding despite how much we scale them. These methods treat text in isolation, neglecting the fact that language gets its meaning when agents interact in a complex multi-modal environment. Bearing this in mind, I believe that multimodality will help the models have the impression to recruit experiences of other sensory systems such as visual and auditory for instance. A big challenge however is to make multi-modal models use all modalities effectively, as they often fail to do so (e.g. focusing only on text). I plan to continue this line of analysis to further understand the empirical abilities and devise ways to overcome the limitations of these models. Through this direction, I hope to further explore my nascent interests in robustness. For instance, I will explore how we can make these neural models less sensitive to real-world perturbations so they are equitable and ethically be deployed for broader impact upon society.

**Why PhD.** All these problems I care about have influenced my decision to undertake a PhD. I am deeply passionate about the fundamental, open-ended, and thorough nature of research. This is especially present in the topics of my interest, and It is definitely the right interdisciplinary fit, where I could exercise basic scientific curiosity and build impactful tools for practitioners. From my own limited experience so far, I particularly enjoyed the collaboration in my projects with other researchers and the scientific inquiries that challenged me and deepened my understanding; however, I also learned to embrace and even enjoy the struggles such as chaotic experiments and late nights with collaborators.

**Future Plans.** My career aspiration is to continue to do research as a professor and mentor students in the process. This choice is particularly informed by my positive experiences while holding the instructor position of a student club for more than a year during my undergrad. As an instructor, I conducted workshops and mentored students in the following courses: Machine Learning, Artificial Intelligence, and Object-Oriented Programming. I've given my very first [live talk](#) on a machine learning topic (text-to-image generation) very recently before an English audience. I find that I enjoy teaching - both interaction with other students, and the opportunity to deepen my understanding of the material. A few months back, I received Delta Analytics Global Teaching Fellowship 2021. As a fellow, I was introduced to the pedagogical principles for teaching machine learning and building community, through which I came to believe that we share our successes and one's personhood is evaluated based on their contributions to the community. All These experiences have honed my leadership, teamwork, and communication skills, all of which will help my future career in academia thrive.

**Why UMD.** I'm particularly excited to work with Professors Jordan Boyd-Graber, Marine Carpuat, Hal Daumé III, and Rachel Rudinger, as we are generally interested in problems of building models that can break language barriers, can learn through interactions or multi-modal signals, and can learn in low data regimes. I'm also excited to work with Professors Tom Goldstein and Soheil Feizi due to our overlap in interests in performing both empirical and theoretical studies to find more robust and reliable models. I truly appreciate that the UMD NLP group (CLIP) includes students and faculties from diverse backgrounds and departments. Following the work of these professors has led me to see a clear fit for my skills and interests at UMD, and I am confident that it is a great place for me to pursue a PhD.

## References

- [1] Boris Dayma, Suraj Patil, Pedro Cuenca, Khalid Saifullah, Tanishq Abraham, Phúc Lê Khc, Luke Melas, and Ritobrata Ghosh. Dall-e mini, 7 2021.
- [2] Damián Blasi, Antonios Anastasopoulos, and Graham Neubig. Systematic inequalities in language technology performance across the world’s languages, 2021.
- [3] Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. In *International Conference on Machine Learning*, pages 1183–1192. PMLR, 2017.
- [4] Andreas Kirsch, Joost van Amersfoort, and Yarin Gal. Batchbald: Efficient and diverse batch acquisition for deep bayesian active learning. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.
- [5] Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, and Richard Socher. The natural language decathlon: Multitask learning as question answering. *CoRR*, abs/1806.08730, 2018.
- [6] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint arXiv:1910.10683*, 2019.
- [7] Victor Sanh, Albert Webson, Colin Raffel, Stephen H. Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Teven Le Scao, Arun Raja, Manan Dey, M Saiful Bari, Canwen Xu, Urmish Thakker, Shanya Sharma Sharma, Eliza Szczechla, Taewoon Kim, Gunjan Chhablani, Nihal Nayak, Debajyoti Datta, Jonathan Chang, Mike Tian-Jian Jiang, Han Wang, Matteo Manica, Sheng Shen, Zheng Xin Yong, Harshit Pandey, Rachel Bawden, Thomas Wang, Trishala Neeraj, Jos Rozen, Abheesht Sharma, Andrea Santilli, Thibault Fevry, Jason Alan Fries, Ryan Teehan, Stella Biderman, Leo Gao, Tali Bers, Thomas Wolf, and Alexander M. Rush. Multitask prompted training enables zero-shot task generalization, 2021.
- [8] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation, 2021.
- [9] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision, 2021.