Research in Computer Vision. I have worked in the Red Lab since my first semester in NSE at MIT. The lab focuses on building and deploying computer vision models for boiling diagnostics and automation in industrial applications. My responsibility in the lab is building accurate and generalizable computer vision models to process and semantically segment high-speed video data obtained from boiling. The project aims to deploy the custom-made models into physical boiling setups to enable autonomous experimentation, which can be applied in industrial setups. This can enhance the safety of various physical systems and prevent catastrophic accidents without endangering human life. Before this time, achieving this level of autonomy was difficult because of the complexity of the high-speed video data obtained from boiling processes, especially in cryogenics. Conventional image processing techniques failed to generalize on the massive datasets obtained from the high-speed cameras, limiting their deployment into physical systems. Our contribution to this project was the introduction of accurate vision models to process complex boiling data accurately, allowing for safe, autonomous operation.

In my first few semesters, I implemented transfer learning techniques for our large-scale boiling data (over 20 videos with 2,000 frames per video). Since obtaining annotated frames was expensive, I initiated a transfer learning technique using U-Net CNNs to transfer information from pre-trained biological cells to bubbles in our boiling data. In this way, I could finetune the weights from the biological cells and adapt them to our boiling dataset using just a minimal number of annotated frames (3-5 frames). After executing the fine-tuning experiments on PyTorch hosted on CUDA, I applied the model to segment our high-speed video data with high accuracy, Jaccard index, and dice coefficient values exceeding 0.98 on thousands of frames, a breakthrough in vision models on cryogenic boiling datasets.

Although high computer vision assessment metrics were obtained from the model segmentation phase, I noticed the model had poor generalization capabilities. Specifically, during experimentation, the models finetuned for a specific operating condition (fluid type, heating conditions) could not accurately perform zero-shot generalization on videos from other conditions. At first, I tried to augment the training and validation datasets to include various annotated frames from different operating conditions. However, data augmentation did not solve the problem, as the model still performed poorly due to the varying complexities of the various frames, resulting in higher cross-entropy loss errors during training. To solve this problem, I collaborated with an MIT EECS professor to utilize vision transformers (ViT-H) for our boiling data by finetuning the stateof-the-art foundation model, Segment Anything Model (SAM) released last year by Meta AI. The motivation for finetuning SAM to our custom data is that since SAM is a foundation model trained on over 11 billion segmentation masks, it will be able to provide some contextual features that may be relevant to our zero-shot generalization bubble segmentation task. Furthermore, due to the ambiguity awareness of SAM, it will be able to differentiate occluded/overlapping bubbles after successful finetuning. Finally, since we now have a large database of camera images and segmentation masks across various conditions, we can use these to finetune the SAM to develop a foundation model with better generalization capabilities. My contributions included implementing the finetuned SAM from scratch and applying it to our boiling data for zero-shot generalization on boiling video data. I also quantified the uncertainty in the vision model masks by developing a randomized algorithm for sampling bubble locations across different radii and grid resolutions and comparing the measured parameters against ground truth masks. Our results indicated a maximum error of 7% across thousands of frames on the masks obtained from our models. The findings of this study are being curated for possible publication in the Pattern Recognition Journal.

Research in Natural Language Processing. Despite my experience with computer vision models, multimodal AI must embrace other modalities of data, such as text. This sparked my lifelong interest in natural language processing (NLP), and last Fall, I took a graduate machine learning course (6.7900) at MIT taught by Tommi S. Jaakkola, where I learned the underlying theory behind generative pre-trained transformers (GPTs). Beyond getting a B (4.0/5.0) in the course, I was inspired by the self-attention mechanism of GPTs, and I embarked on a project to develop my own version of the GPT2 model on PyTorch which I called streamlined GPT2 (SGPT2) with around 0.21 M parameters. To achieve this, I finetuned the GPT2 model on the Shakespeare dataset containing 40 K lines of Shakespeare text curated from diverse plays. The SGPT2 architecture included self-attention, multi-head attention mechanisms, and multi-layer perceptron to understand and generate novel text in Shakespeare's writing style. The efficient text encoding and decoding system facilitated the learning process of the model. The text encoding transformed strings into numerical indices (*stoi*), a format the neural network could readily interpret and learn from. These encoded integers served as the inputs for the network, which then embarked on making predictions. Upon generating its outputs, the model utilized the reverse process—decoding integers back into strings (*itos*)—to render its predictions into human-readable text. After successful training on 5000 iterations and obtaining a converged minimized cross-entropy loss, SGPT2's generative capabilities were demonstrated by utilizing a custom text generation function, which showed its ability to synthesize new textual content stylistically like Shakespeare.

STEM Outreach. Besides pushing the boundaries in vision and language processing research, I have always strongly emphasized outreach throughout my research career. While an undergrad, I initiated an outreach during my free time to teach and mentor students who wanted to proffer computer science solutions to engineering problems. In this capacity, I taught over 500 students in 2 years the fundamentals of computer programming in C++, MATLAB, and Python and how to write clean and efficient code to solve theoretical and practical problems in engineering. Furthermore, due to my near-perfect score (98%) in computer programming and data structures (ECE 271/281), I was selected as a TA for these courses. Students always gave me the highest reviews, 4.8/5.0, due to my attention to detail and ability to break down tough concepts into smaller bits that are easy to grasp. Finally, post-graduation from my undergrad, I founded Enrai, a non-profit organization of 10+ members dedicated to providing computer science and engineering students with hands-on research projects, which will improve their portfolio and make them highly competitive for graduate school applications. Since its inception in Summer 2022, 3 members have been admitted to MSc/PhD programs in the US with full funding.

Future Work. I am confident that Ontario Tech would align well with my future ambitions due to its diverse strength of faculty and students. I would be glad to work under Professor Annie En-Shiun Lee, whom I have been in contact with. Since I have spent my research career working on natural language processing and predicting machine learning models, I would happily explore and contribute more to these areas. I intend to expand the current LLMs to accommodate additional languages to target underrepresented localities. In this capacity, I will use my deep expertise in Python and its deep learning frameworks to interact with open-source LLMs to increase their capacity to understand and generate other languages. Additionally, I will bring my expertise in reproducing code and experiments from LLM papers by interacting with git and data repositories. Professor Lee and others have greatly contributed to these areas by employing classical regression models to investigate how factors like the finetuning corpus, domain similarity, and language similarity affect low-resource language (LRL) models. Their study indicated that domain similarity between the source and target languages was the most impactful factor affecting the model. I cannot wait to exploit these findings using domain similarity to expand current LLMs to accommodate languages like LRLs. Continuing my proposed work on NLP model adaptability under Prof. Lee's guidance would provide a productive environment to pursue my interests further and advance my research. Another benefit would be to bring my accrued experiences from working on NLP and predictive machine learning at MIT to Ontario Tech, allowing me to make further connections and comparisons to the tools and methods at my current disposal. Post-graduate school, I intend to work in an academic setting toward a professorship.