I am an undergraduate in CS, Math, and Statistics. I am <2 years in my degree but a 4th-year student due to the course overload I took. Starting January 2025, I am joining NVIDIA Toronto AI Lab as a researcher (w/ Gavriel State) to work on vision.

My academic interests mainly lie in research around learning algorithms and vision. To this extent, most of my past work has been around this area. Furthermore, I also enjoy working on large-scale software engineering problems. For the rest of this cover letter, I share some of my most recent (first-author only) works and my interests. I have worked on other software projects and non first-author works too.

This summer, I am involved in academic Computer Vision research at Qualcomm AI Research hosted by Guillaume Berger and Roland Memisevic. My primary focus lies in two areas: precise video generation, which is currently under submission, and video question answering, which is still a work in progress. I first authored a paper on a novel diffusion model designed for precisely guided video generation based on uniquely modifying classifier-free guidance. This paper also introduces a new, large dataset comprising 165K long videos. I conducted all the large-scale experiments necessary for this research, which not only included writing experiment code for our novel approach but also developing software to manage experiments on more than 200 GPUs.

Most recently at UofT, I was supervised by Prof. David Lindell. The preprint that I worked on was about developing new Diffusion Models that could handle traditional graphics problems and we proposed to pose an entire image signal processing pipeline (with prompts for instance: "What this image would look like in high exposure") as a diffusion model problem. Furthermore, at UofT I also worked on another preprint paper on generating spatial audio for images and videos to accompany the visual content. I also recently worked on a project with Prof. Pascal Tyrrell and Prof. Rahul Krishnan on (very recently accepted to PMLR) building new NeRF models. We particularly proposed (1) a replacement for ray tracing involving not sampling colors and densities like NeRFs do (2) a new rendering algorithm to replace volumetric rendering, and (3) using a diffusion model trained on occupancy grids to serve as a regularizer.

Last summer, I worked on academic Computer Vision research at Civo Ltd. hosted by Josh Mesout. We developed a new multimodal speech and vision model, I not only led the research idea but also implemented and ran the experiments myself. I also led the development of a machine-learning runtime for our product, ''Recite'', which increased performance by approximately 120-200% compared to other optimized runtimes.

I also believe I have a good understanding of software for machine learning. In the past, I have also worked very closely with CUDA and Triton, and I have also had one first-author and one second-author (oral) paper at the PyTorch Conference: one of which was about building specialized backprop kernels for NeRF training to potentially train NeRFs on small devices and the other was about building capabilities to AOT-compile NeRFs to WebAssembly code. These also required us to write new backprop kernels in CUDA and Triton. Furthermore, I am also one of the maintainers of Kubernetes, one of the most popular orchestrators for software at scale where I have contributed to building new capabilities in the project, most recently contributing to building ''Dynamic Resource Allocation'' capabilities which are popularly used by most large ML clusters.