# Kabir Moghe '26

## NVIDIA GPU Technology Conference
## Winter 2024

This past week, I spent several days at the NVIDIA GPU Technology Conference (GTC) in San Jose, California. The conference took place during a really interesting and relevant period of time, with interest around the rise of AI and its intersection with hardware and other software having been piqued across the entire tech-world for the past year. NVIDIA has largely been at the center of the hardware (i.e. graphical processing units or GPUs) responsible for the sharp acceleration of AI and computational power, and it alongside OpenAI and many other companies have driven the meteoric rise in LLMs, Generative AI (GenAI), and other AI-based technology we've seen as of late. With that in mind, as a CS and math major with interests in data science, machine learning, and real-world applications of AI, I found the GTC to be a confluence of these areas, providing opportunities to meet people with corporate AI experience, consult experts in their various fields, and even pick up my own practical knowledge and hands-on skills. I spent the first day at a day-long deep learning workshop, where I had the chance to re-establish my understanding of the theory behind neural networks, explore their various optimizations, and how they should be and are currently being used in the real world. On the second day, I attended a series of interesting talks on the implications of AI and GPUs across multiple fields. The first discussed how AI is being used in the Search for Extraterrestrial Intelligence (SETI), used to help decipher and monitor potential alien radio frequencies; the next had computational scientists discussing the pressing need for parallelization with algorithms in the field of genomics, for which GPUs have been extremely helpful, accelerating long pipelines to find certain gene sequences, use GenAI for drug discovery, and more. I then attended an amazing keynote delivered by NVIDIA CEO Jensen Huang, during which he discussed many computational and AI-related improvements the company and industry had made; most notably, he unveiled the company's new and more powerful "Blackwell" chip and the notion of "digital twins," or the idea that for many problems, we can use digital/virtual simulations to train AI and perform hefty computations that can then be implemented and/or used in the real world, such as predicting weather on a digital twin of the Earth ("Earth2") or training humanoid robots digitally to prepare them for the real world (at BostonDynamics). I spent the next few days meeting people and learning how they were using AI and also visited several booths at the conference's exhibition, where I learned more about ExaBiome's use of GenAI for drug discovery and explored how certain startups were optimizing GPU use for their customers. I also continued attending interesting talks about autonomous vehicles and computer vision, artists' incorporation — and fear — of the creative capabilities of GenAI, personalized AI-powered healthcare, and the lack of "reasoning-based" output with large language models. On the final point, the final few days also introduced me to many potential issues with GenAI and LLMs that experts are coming around to. In particular, the unpredictability and hallucinatory nature of LLMs poses problems when trying to produce repeatable output poses. The idea that LLMs are trained on data to replicate and follow human output also suggests that this form of AI lacks the ability to reason as we do, while many alternatively suggest that this training and replicating is, in essence, how we subconsciously learn. Regardless, these are all interesting themes the AI community is and has been exploring and also pose potential research questions for me moving forward. Overall, I had a super productive and eye-opening experience. I had the chance to pick up new technical material, broaden my network in software and AI, and engage with pioneers across AI-related fields. Most importantly, I feel that I got the chance to get a glimpse — beyond my own personal experience — into how this new wave of AI has begun touching every part of the real world.