

## **Daniel Guo CS 224G Individual Insight Report**

CS 224G has been a phenomenal experience and I'm so glad to have been part of the first iteration of the class. Throughout the quarter, I gained invaluable learnings about the end-to-end process of developing an AI-powered application, each step from ideation to implementation to demo day. The past 10 weeks presented several technical and non-technical challenges that offered unique growth opportunities and key insights that I'll carry into my future academic and professional endeavors.

The most important insight that I gained from this class was how the product-building process works, particularly how early-stage ideas are prone to pivots after further market research and need-finding. As our team experienced through our pivot from PharmD, our initial view about the prescription drug pipeline was naive and incomplete – only after interviewing several nearby pharmacists did we realize the nuances and complexity of how drugs are prescribed, ordered, and delivered to the patient. As a result, we realized that the issue was far beyond our bandwidth, especially for our CS224G project, and made the difficult but correct decision to pivot our idea to Mentore.

Another important insight that I learned is that it's incredibly important to be dynamic and delve deep into the technical aspects of a project with a growth mindset. Since our team was relatively inexperienced with building LLM applications prior to this quarter, the bulk of our efforts was spent becoming familiar with the OpenAI APIs, systems design, and user experience design. We took inspiration from existing applications like Perplexity and ChatGPT for design choices and we each had to develop the skills we needed along the way. This was made much easier through leveraging tools like ChatGPT, the internet, and forums like Stack Overflow. In particular, I learned how to use the BeautifulSoup and Selenium libraries to build a webscraper. Furthermore, I learned that truncating input sequences into OpenAI's Embeddings API and precomputing these embeddings provided faster and more accurate results for mentor-matching compared to naive implementations.

From the classroom, a lesson that stood out to me was Anoop Sinha's Large Model Quality and Evaluation lecture. Sinha's talk inspired me to delve deeper into the model performance of Gemini, and Google's more recent Gemma, and how it compares to the performance of other existing open-source architectures like Llama-2, GPT-4, and Qwen. I learned a ton about how instruction-level finetuning can enhance the performance of models and other techniques like RLHF can be used to align to human preferences. This is certainly something that I wish to continue exploring in my future endeavors in machine learning.

Overall, I thoroughly enjoyed this class and I'm super grateful for all the mentorship provided by the CS 224G teaching team. Although I was on the fence about joining the class last-minute, I don't regret that decision one bit – I have learned far more practical skills in this class than I have in the majority of my other CS classes at Stanford. Not only am I now a more mature full-stack app-builder, but I also have a cool LLM product to show for it! I'm excited to carry forward the insights and skills I've gained from this class to continue building AI-integrated products in the future.