General Behavioral

Greatest Strength

(1)

My greatest strength is my ability to pick up on knowledge and new technologies quickly.

- *Situation:* Last summer, I worked in a research group that was heavily focused on SQL and database management. However, I had little to no experience with these technologies. Furthermore, I was the only undergraduate in a group of Masters and PHD students, which added to my steep learning curve.
- *Action:* In response, I put in extra hours learning SQL and databases from online resources, as well as gained access to relevant course material from professors in the research group.
- *Result:* As a result, I was able to learn most of the relevant information and developed a strong database management foundation in under a week, which allowed me to hit the ground running in this role.

This is very important to my role as a (data scientist/business consultant) as we tend to work in different fields and being able to quickly obtain domain knowledge allows me to better understand problems and provide more efficient solutions.

(2)

My greatest strength is my communication skills, including knowing when to ask for help on a task as to not slow projects down and communicating to my team what I am working on.

- *Situation:* In my role as a research intern at SEELab, I worked under a PHD student to help complete the development of a specialized sensor LLM.
- *Task*: In this role, I have been given all sorts of tasks and have been given the opportunity to greatly expand my understanding of generative AI and the different aspects of large-scale projects.
- *Action*: However, with this came me having to understand code for which there was limited documentation and complete actions with very little instruction. I constantly kept my mentor updated with tasks I had completed and current progress, which allowed her to provide me suggestions more often. This in turn allowed me to complete tasks faster and take on new responsibilities.
- *Result*: This led to project efficiency greatly increasing, and a targeted deadline of December has now turned into a targeted deadline of August.

A major reason why is the efficiency that my communication generated; it kept the project's many components constantly moving.

Greatest Weakness

A weakness that I have is I can get caught up in the details and forget to think about the bigger picture.

- *Situation*: Last spring, I led a group of 3 other students in a class final project, which was an end-to-end data analysis task.
- *Issue:* I became very caught up with smaller details, and realized that we were at risk of missing the submission deadline. My teammates were also unresponsive at times and it was a struggle to coordinate times where everyone was free to connect.
- *Action*: I revised my plan of action by breaking down the project into major, necessary segments and worked to make each segment functional while also delegating work when possible. I constantly reminded myself of the bigger picture and made sure not to get too caught up in the details. Finally, after every essential segment was complete, I went back and addressed some of the finer details.
- *Result:* We ended up receiving a very good score on the project, and it was due to our combination of project as a whole working well, and our attention to some finer details. However, had I continued focusing on the details, our final product would have been a lot choppier and likely would not have received the grade that we did.

Now, when I complete projects, I make sure to give the details proper attention, but try to focus on the bigger picture first and come back to the non-essential details after.

A Time I Messed Up

- *Situation*: When first starting my role as a research intern at my lab, my PHD student mentor gave me access to a remote desktop that I could work on.
- *Issue*: However, I did not know that other people were also working on this desktop, and when setting up my environment I accidentally wrote over other people's files.
- *Action:* I realized this soon later when I saw some code that I had not written, and so I immediately brought this up to my mentor. While I was scared that my position may be revoked, I strongly believe in honesty and owing up to one's actions, especially in a workplace.
- *Result:* Luckily she was able to recover all of the files and no permanent damage was done. She appreciated my transparency, as if I had waited longer there may have been a chance that months of research would have been lost.

I learned to understand new technologies better when first working with them and to ask more clarifying questions in unfamiliar environments so as to not accidentally destroy anything.

Disagreement With Colleague

- *Situation*: Occurred on a group project for a data analysis class earlier this year. I had taken the lead and created a plan on how we were going to split up work and get the project done
- *Issue*: One person did not want to do what I had in mind, and thus did not complete any work on the project.
- *Action*: I sat down with them and we talked about what they wanted to contribute, and we were at a stalemate because it seemed that they didn't want to do any meaningful work. Reason was because they wanted to go a completely different direction for the project, but had never voiced their thoughts in group meetings so I had no idea that this was the issue.
- I worked to come to a compromise, incorporating both of our ideas in our final project plan.
- I discussed the importance of everyone contributing and I was able to change their task to be a little more what they wanted, and offered guidance for any issues they faced
- *Result*: I discussed the importance of everyone contributing, and they were much more eager to, now that the project was a bit more tailored to their interests. They ended up contributing meaningful work to the project, and came to me with questions and issues that they faced and I was able to help them.

PromoDrone

- *Situation*: Brought in to revise backend infrastructure of the company's technology
- *Task*: Given very specific directions in what they wanted to accomplish and what technologies I could use to do so. They wanted my final product to be fully usable on the Google Cloud Platform, but one of their tasks for me was not accomplishable on Google Cloud.
- *Action*: I exerted my problem-solving skills and found a way around this. I coded as much as I could on Google Cloud, then created my own neural network (CNN) to handle the parts that it could not. Finally, I integrated the two together and presented my final product on the Google Cloud Platform to the CEO and board members.
- *Result*: They loved my work and adopted it, and now PromoDrone's product will be commercially available later this year.

Growing up with a father working in hardware and an older sister in software engineering meant I was constantly surrounded by technology I didn't yet understand. My curiosity led me towards engineering, finding interest in coding and the systems they built. As I learned more, I found myself drawn to the data powering them. Wanting to follow in my family's footsteps, I started building systems of my own and realized how quickly complexity emerges, especially with real-world data and multi-layered machine learning models. Experiencing this difficulty firsthand made me realize how inaccessible advanced technology can be for both creators and users.

With experience blending machine learning with the fields of finance, physical sensors, and marketing, I've found my passion for building intelligent AI systems that are intuitive and usable even across domains. These experiences also revealed gaps in my knowledge that I hope to address by pursuing a Master's in Data Science at Harvard. Harvard's unique combination of strong interdisciplinary focus and human-centered engineering labs makes it the ideal place for me to deepen my data science foundation while learning to develop systems that help both developers and end-users lower the technical barriers limiting the adoption of advanced AI. I'm especially drawn to the work of Professors Glassman and Wood, whose research in human cognition and wearable robotics powered by AI exemplifies how thoughtful design can improve accessibility and bring the capabilities of modern technology closer to everyone. This work reflects the impact I intend to create: developing powerful machine learning systems across domains that are accessible to users in real-world settings.

One of the reasons I feel strongly prepared for graduate study is due to my success in my research work. At UC San Diego's Systems Energy and Efficiency Lab, I co-authored two publications - *SensorQA* (ACM SenSys 2025) and *SensorChat* (IMWUT 2025) - that explored how natural language interfaces could help everyday people interact with long-term sensor data. Surrounded by PhD students and challenges I had no idea how to tackle, there were times when I felt overwhelmed and uncertain how to progress. Having always excelled academically, even graduating as my high school's valedictorian, I wasn't used to being out of my depth. I faced these challenges by spending more time rereading research papers, asking questions, and seeking help from mentors. That persistence paid off. As the youngest and only undergraduate author on both projects, I led data analysis, model development, and evaluation - meaningful contributions that shaped the project's outcome. By leveraging Retrieval-Augmentation Generation with GPT and BERT embeddings, I directly advanced our model's rigorous evaluation workflow by enabling the use of exact-match accuracy as a performance metric and uncovered latent structure across our dataset that informed final model design.

That experience left me eager to understand how research translates into real-world impact, a perspective I found through my internships at BILL. During my first summer, I developed a clustering algorithm for categorizing application error messages. Since the number of clusters needed and their strength was unknown, I could not use existing algorithms such as k-means or dendrogram-based clustering. This led me to develop my own algorithm based on boosted cosine-similarity embeddings tuned for this specific use-case, saving the team hundreds of thousands in infrastructure costs. The following year, I was given full ownership of a new project: developing an internal AI coding assistant to speed up development time. Without a concrete roadmap set for me, I devised my own plan from initial research all the way to final implementation, analogous to how a graduate student would be expected to scope and execute on their own research. This project was successful, cutting development time in more than half for external accounting software integration. Using my tool, the team's software engineers who were unfamiliar with AI-based development were able to leverage its coding abilities with just a few clicks. An integration that usually takes a year to build was finished by these engineers in a record 5 months directly due to my work. I'm confident that I will succeed in Harvard's program because of all that these experiences have taught me: how to work independently and in a team, navigate ambiguity, and see projects through from concept to deployment.

Alongside research and industry experience, I've also found purpose in mentorship and contributions to student community. After tutoring a freshman data science class, I realized many students were struggling to land their first industry opportunity. This led me to take on the role of Consulting Director for UCSD's Data Science Student Society, where I created opportunities for students to apply their education and gain real-world experience. By forming partnerships with six startups, I facilitated industry experience for thirty students and mentored them through workshops focused on technical and communication development. This experience taught me how to become a better mentor, effective leader, and confident speaker. I intend on continuing mentorship initiatives at Harvard in the Variation or Microrobotics Labs by challenging undergraduates to approach problems from different angles and running reproducibility or debugging sessions for undergraduates beginning their research journey.

Harvard's MSE in Data Science will strengthen my foundation in machine learning and intelligent system design while giving me the tools to continue building meaningful and accessible AI systems. The curriculum places a strong emphasis on both learning through classrooms and practical application, a balanced learning style which will help me address my knowledge gaps and become a stronger, more analytical engineer. After completing the program, I plan to work in industry as a Machine Learning Engineer, developing systems that make it easier for developers to leverage the power of advanced AI while broadening the population of everyday people that can benefit from them. Through Harvard's MSE in Data Science, I will continue my commitment to improving accessibility and make lasting contributions to the future of applied machine learning.

1. Tajana Rosing, ucsd professor. Can provide more information about her if you feel it is needed, let me know.
2. I did research with her from Feb 2024 - Sept 2024. But we can say 2 years to make it seem like I did it for longer.
3. Ranking-wise - let's go with 5 percent
4. I have 2 related publications, SensorQA and SensorChat. Below are the abstracts
    - SensorQA: With the rapid growth in sensor data, effectively interpreting and interfacing with these data in a human-understandable way has become crucial. While existing research primarily focuses on learning classification models, fewer studies have explored how end users can actively extract useful insights from sensor data, often hindered by the lack of a proper dataset. To address this gap, we introduce SensorQA, the first human-created question-answering (QA) dataset for long-term time-series sensor data for daily life monitoring. SensorQA is created by human workers and includes 5.6K diverse and practical queries that reflect genuine human interests, paired with accurate answers derived from sensor data. We further establish benchmarks for state-of-the-art AI models on this dataset and evaluate their performance on typical edge devices. Our results reveal a gap between current models and optimal QA performance and efficiency, highlighting the need for new contributions.
    - My Contribution was with dataset analysis:
      4 DATASET ANALYSIS In this section, we provide quantitative and qualitative analysis of the SensorQA to better understand its characteristics. Examples of the collected Q&As in SensorQA are displayed in Fig. 2. SensorQA contains 5,648 question-answer pairs, with an average length of 10.43 words per question and 10.48 words per answer. The dataset has a total of 118,051 tokens, of which 1,709 are unique and primarily related to daily activities. The repetition of words makes it more challenging for AI agents to answer questions accurately, as they must differentiate between similar questions based on the specifics of the sensor data. To closely inspect the diversity of SensorQA, we profile the question and answer categories. We manually label 200 pairs, then train two BERT models [11] to classify the question and answer categories, respectively. The final profiling results are displayed in Table 3. SensorQA includes six distinct question categories and seven answer categories. The distribution of questions and answers is imbalanced, with a notable focus on time-related aspects of activities, as seen in the high number of questions in the "Time Compare" and "Time Query" categories. This pattern aligns with practical user interests but has not been observed in previous QA datasets for human activities [29, 30, 48]. In addition to time-related queries, SensorQA covers a wide range of other aspects, including action, location, counting,
        - Published in ACM Sensys 2025
        - 6th author (not very prestigious but highlight the fact that I was the youngest and the only undergrad (sophomore at the time) and everyone else was grad student or higher

SensorChat:
atural language interaction with sensing systems is crucial for addressing users' personal concerns and providing health-related insights into their daily lives. When a user asks a question, the system automatically analyzes the full history of sensor data, extracts relevant information, and generates an appropriate response. However, existing systems are limited to short-duration (e.g., one minute) or low-frequency (e.g., daily step count) sensor data. In addition, they struggle with quantitative questions that require precise numerical answers. In this work, we introduce SensorChat, the first end-to-end QA system designed for daily life monitoring using long-duration, high-frequency time series data. Given raw sensor signals spanning multiple days and a user-defined natural language question, SensorChat generates semantically meaningful responses that directly address user concerns. SensorChat effectively handles both quantitative questions that require numerical precision and qualitative questions that require

high-level reasoning to infer subjective insights. To achieve this, SensorChat uses an innovative three-stage pipeline including question decomposition, sensor data query, and answer assembly. The first and third stages leverage Large Language Models (LLMs) to interpret human queries and generate responses. The intermediate querying stage extracts relevant information from the complete sensor data history. Real-world implementations demonstrate SensorChat's capability for real-time interactions on a cloud server while also being able to run entirely on edge platforms after quantization. Comprehensive QA evaluations show that SensorChat achieves 93% higher answer accuracy than the best performing state-of-the-art systems on quantitative questions. Furthermore, a user study with eight volunteers highlights SensorChat's effectiveness in answering qualitative questions.

- My Contribution was with dataset supplementation and generation: In this evaluation section, we focus mainly on the SensorQA dataset [48] and quantitative questions. To the best of our knowledge, SensorQA is the first and only available benchmarking dataset for QA interactions that use long-term timeseries sensor data and reflect practical user interests. While we focus on SensorQA [48] in this section, we emphasize that SensorChat is broadly applicable and can be extended to other real-world sensing applications. To ensure the best alignment between the QA pairs and sensor information, we conduct offline encoders pretraining on the ExtraSensory multimodal sensor dataset [53], which servers as the sensor data source for SensorQA. During pretraining, all sensor samples are aligned by a time window of 20 seconds. Dataset Variants and Metrics We evaluate three versions of SensorQA [48] using various metrics to assess both the quality and accuracy of the generated answers. • Full answers refer to the original full responses in SensorQA. We evaluate the model’s performance on the full answers dataset using Rouge-1, Rouge-2, and Rouge-L scores [18]. Rouge scores measure the overlap of n-grams between the machine-generated content and the ground-truth answers, expressed as F-1 scores. Higher Rouge scores indicate greater similarity between the generated and true answers. • Short answers are the 1-2 key words extracted from the full answers by GPT-3.5-Turbo [63], offered with the original SensorQA dataset [48]. We use the exact match accuracy on the short answers to evaluate the precision of generated answers, as detailed in Sec. 3.2. • Multiple choices are generated by prompting GPT-3.5-Turbo [63] to create three additional choices similar to the correct short answer. An example QA can be "Which day did I spend the most time with coworkers? A. Friday, B. Monday, C. Thursday, D. Wednesday", with the correct answer being "D" or "D. Wednesday

  So i generated the dataset variants, as well as reproduced some baselines that we compared our result to. The two papers are related in the sense that the sensorchat product is trained using the sensorqa dataset.
- Published in imwut 2025
- I was 5th author, but again the youngest and only undergrad (durting my summer between soph and junior yr)

- Broader lab focus is networking, iot, sensor stuff. But my contributions were related to dataset work and working with language models such as bert and gpt 3 to do so
- I have other experience but nothing else related to the research that this professor does. Is it still worth mentioning in this letter? Let me know. The other thing I did was serve as a tutor in a program meant to teach research/ml foundations to high schoolers, and this program was related to the lab so technically oversaw by my professor.
- Postgrad plans: want to work in ml industry. Developing models, agents, agentic workflows, etc. want to be at forefront of ml work