Anthropic Applied AI Take-Home Assessment

This take-home reflects things that you would typically spend time on day-to-day in the role. It helps us understand your ability to:
1. Understand a prospect or customer's problem
2. Iterate on a potential solution
3. Evaluate a model's quality on a particular task
4. Communicate your recommendations to a prospect or customer

Below is an email from Dario, a prospect who is evaluating using Anthropic models to build a system that converts natural language queries to SQL but isn't satisfied with the model's quality and isn't sure how to improve or evaluate its performance.

In responding to this take-home, you should:
1. Evaluate the accuracy of Dario's current approach
2. Suggest and evaluate improvements to his approach
3. Respond to his email with your findings, any code you wrote, and instructions on how to run your code

Submission Guidelines:
- Submit your response within 1 week
- Format your submission as you would a real customer communication
- If you have questions during the assessment, please reach out to sarahharley@anthropic.com

Please spend no more than 4 hours on this. We do not want you to spend a long time optimizing for accuracy. While improving performance from baseline is important, we're more interested in seeing your overall approach, thought process, and recommendations to move this deal forward in a time-boxed way. As you work on this, you are welcome to search the internet or use any tools / software you might use in your day-to-day.

Email from a Prospect

Subject: Improving Our NBA Stats Query Tool

Hello,

I'm reaching out to continue the conversation about our NBA stats database project. To recap: we're trying to build an internal tool that lets our analytics team query our database using natural language questions, but we're running into accuracy issues with our current solution.

Here's what we have:
1. A database with NBA statistics ([link](#)), which we adapted from [here](#)
2. Some ground truth data (natural language queries + associated SQL) we created to evaluate our tool ([link](#))
3. Some sample code for querying a .sqlite database using Python ([link](#))

Here's the prompt we're trying to use:
"""
Convert this question to SQL:
{question}
"""

Could you help us evaluate how well our current setup is working and improve the prompt to get better performance?

Looking forward to your thoughts and recommendations on how we can make this more reliable.

Thanks,
Dario