

How to Make More Reliable Reports Using AI — A Technical Guide

How to Make More Reliable Reports Using AI — A Technical Guide

Over the past year, I've had the opportunity to work on various projects involving AI software development and consulting. One common request from clients has been the need to create reliable reports using AI. In this post, I will share my experiences and insights on how to make these reports more reliable and practical.

Covering the Basics

First, let's explore some quick wins that can significantly improve the reliability of your AI system with minimal effort. These are simple yet effective strategies that should always be at the top of your checklist when building AI systems.

1. Use Markdown

When formatting tables in your AI prompt, use markdown. Language models are trained on markdown text, making it easier for them to understand and respond in the same format.

2. Write Clear Prompts

Ensure that your prompts are clear and concise. Unclear instructions can confuse language models, leading to lower quality responses. You can even ask the language

model to help you rewrite the prompt for better clarity.

Optimize Models

To achieve better performance and reliability, it's important to optimize your choice of AI models. Consider the following steps:

1. Choose the Right Model

While top-tier language models like GPT-4o or Claude 3.5 may seem appealing, it's worth exploring other models that perform better for specific tasks. You can refer to online language model leaderboards to identify the best models for your needs.

2. Adjust Model Settings

Even with the right model, tweaking settings like the maximum number of tokens or temperature can improve its performance. Experiment with different settings to find what works best for your specific task.

3. Consider Long-Context Models

Some tasks, such as generating detailed reports, require longer context windows. In such cases, models like Gemini 1.5, which support up to 2 million tokens, can be a better fit.

Smart Prompting

One challenge in working with AI-generated reports is getting accurate and reliable responses. Here are some prompting techniques that can help:

1. Chain-of-Thought Prompting

Add a phrase like "Explain your reasoning step by step" to improve the accuracy and accuracy of responses. This simple tweak often leads to better results.

2. Using Prompting Frameworks

Experiment with prompting frameworks like ReAct, which help language models choose tools or agents more effectively. Frameworks like DSPy provide built-in methods to easily incorporate and test different prompting strategies.

Evaluations

To improve the reliability of your AI system, it's crucial to use evaluation pipelines. These pipelines enable you to measure the effectiveness of your prompts, inputs, and outputs and refine them accordingly. Here's a high-level overview of how evaluations work:

1. Set Up Evaluation Queries

Define a set of queries, inputs, and expected outputs for your AI system.

2. Design a Metric

Create a metric that measures how well the output aligns with the task objective. For example, a text-to-SQL program could use a metric that scores queries based on their ability to successfully run and provide relevant information.

3. **Score Outputs**

Use the metric to score your AI system's outputs and compare them to the expected outputs. This allows you to assess the performance of your system and identify areas for improvement.

Simplify Your System

Lastly, simplifying your AI system can significantly improve its reliability. Here are some tips:

1. **Minimize API Calls**

The more API calls your system relies on, the higher the probability of unreliable outputs. Minimize the number of calls to reduce the chances of errors.

2. **Streamline Components**

Simplify your system's workflow by reducing unnecessary intermediary steps. This will make your system more efficient and less prone to errors.

By following these strategies and incorporating them into your AI development process, you can create more reliable reports using AI. Remember, these are just a few key aspects to consider, and there are many other optimization techniques available. Continuously iterate and improve upon your AI system to maximize its reliability.

Thank you for reading and feel free to share your thoughts and feedback.