EVALUATION REPORT

PROMPTS AND THE MODEL'S RESPONSES

Customer Concern	**Base Model**	**Fine-tuning model(GPT-2)**
"I've heard that customizing your product for specific business needs can be challenging. How adaptable is it?"	**"I think it's a good question. I think you can do a lot of things with your custom product. You can make it."**	**"I appreciate your concern about customizations. Can you share what specific aspects of customization are most critical for your operations?"**
"I'm concerned about how reliable your product is during high-demand periods. What measures are in place to maintain performance?"	**"We have systems in place for high-demand periods. Our product is designed to handle various load levels. What are your specific concerns about performance during peak times?"**	**"I appreciate your concern about our ability to meet our needs. Can you share what specific aspects of our onboarding"**
"It seems like your product may not have strong reporting capabilities. How can you ensure it covers all our reporting needs?"	**"Our product includes reporting features. What specific reporting needs do you have? I can show you how our product's features can meet your reporting requirements."**	**"I appreciate your concern about reliability. Can you share which types of adjustments are most important for you"**
"Your product's mobile features seem a bit limited. How can you assure me that it's suitable for a mobile workforce?"	**"Our product has mobile features. Can you tell me what mobile features you need? I can explain how our product supports mobile use or discuss potential improvements."**	**"I appreciate your concern about accessibility. Can you share what aspects of a streamlined approach are"**

"I noticed your product integrates with fewer third-party applications. How can I ensure it fits with the tools we already use?	**"Our product supports some third-party integrations. What applications do you use? I can help you understand how our product works with your existing tools and explore integration options."**	**“I appreciate your focus on simplicity. Can you tell me which types of applications or frameworks are most important”**

PERFORMANCE ANALYSIS

In this report, we evaluate the performance of a base GPT-2 model and a fine-tuned variant using the **perplexity** metric. **Perplexity** is a common measure of a model's performance in natural language processing tasks, reflecting how well the model predicts a sample of text. Lower perplexity indicates better performance. This analysis will compare the perplexity scores of the base and fine-tuned models to assess the effectiveness of the fine-tuning process.

Model	Perplexity Score
Base GPT-2	0.0159
Fine-Tuned Model	0.0050

The substantial reduction in perplexity from 0.0159 to 0.0050 suggests that fine-tuning has greatly **improved** the model’s ability to predict text. A lower perplexity score implies that the fine-tuned model is better at generating coherent and contextually appropriate responses.