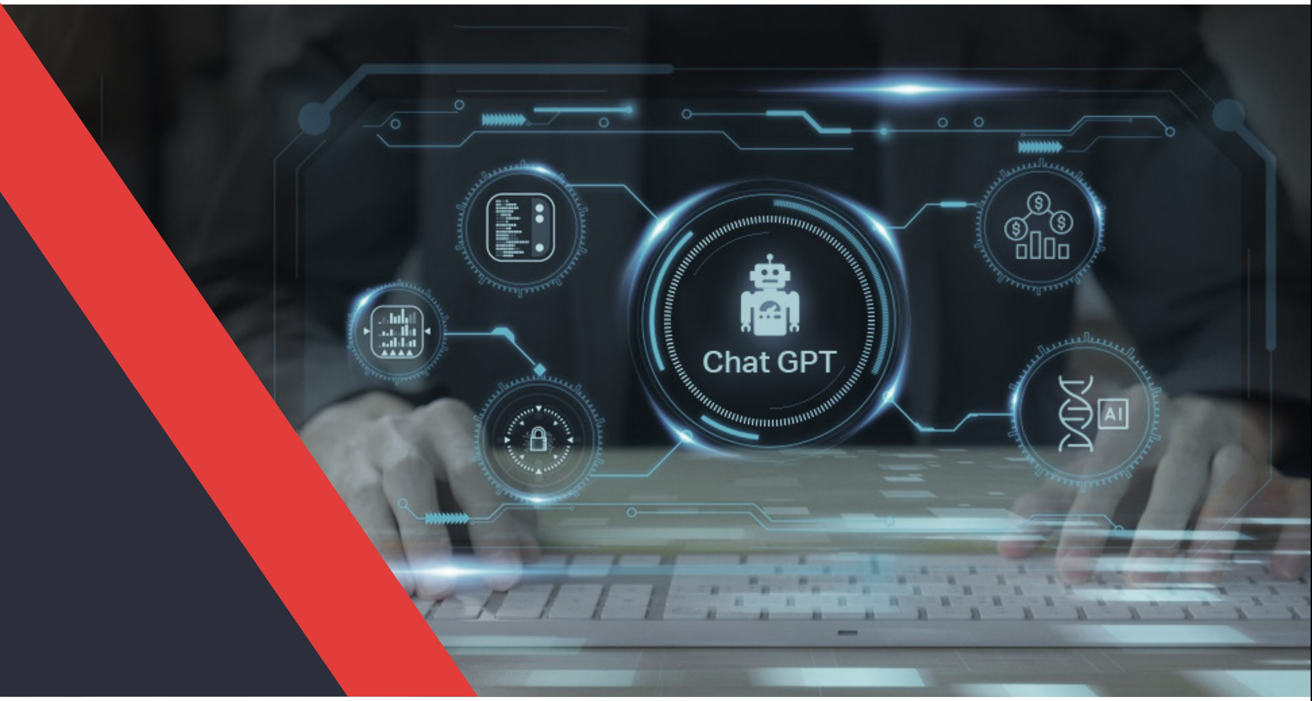
iMerit

CASE STUDY



LEADING LLM-POWERED AI CHATBOT PARTNERS WITH iMERIT FOR IMPROVING MODEL OUTPUT WITH RLHF

One of the leading technology companies in Asia is executing a groundbreaking Corpus Generation Project to build an LLM-powered AI chatbot that interacts like a chat GPT-style bit while catering to the socio-linguistic and socio-cultural intricacies of the region.

The company is focused on the most-searched topics and diversity of people to ensure that the chatbot can handle conversations responsibly. Their immediate task is to recognize and filter out any content that might be risky or biased from what users ask. This effort shows the company is committed to making its AI intelligent and respectful, and they were looking for a partner who believes in the mission of socially responsible AI. With iMerit and the tech company, we built a team of language experts, cultural analysts, data annotators, solution architects, and other specialists working together to understand these challenges. The client knew it was crucial to get this right, especially with the diversity in languages and cultures across Asia.

The company wants to create a diverse corpus, navigating predetermined socially sensitive topics and personas, with the immediate goal of recognizing and filtering out potentially risky or biased content from user queries.

PROMPT GENERATION AND EVALUATION SERVICES FOR A LARGE LANGUAGE MODEL

The first step involved identifying socially sensitive issues and recognizing user prompts that delved into biased perspectives, value judgments, or requests for advice on sensitive topics. The project required the manual creation of prompts through role-playing with different personas speaking diverse languages. iMerit's team of experienced annotators generated these prompts, varying the sensitivity of issues and language proficiency levels.

The process was highly iterative but aimed to ensure that various user personas and multiple perspectives on sensitive issues could be a part of the model training. Despite its limitations, the current setup for prompt generation appeared inadequate for the envisioned market release in the near term. For the future release, the client aimed to enhance the prompt generation UI provided by iMerit, offering a more seamless and efficient experience to the annotation teams. This enhancement would incorporate features such as search results, input tools, and sensitivity coding.

EXPERTS-IN-THE-LOOP FOR IMPROVING MODEL PERFORMANCE

While the annotators role-played diverse personas, generating prompts across socially sensitive topics across languages, there was a manual data correction process with iMerit experts-in-the-loop and reviewers. The manual generation process included language-topic-persona combinations, with sensitivity levels coded for each prompt.

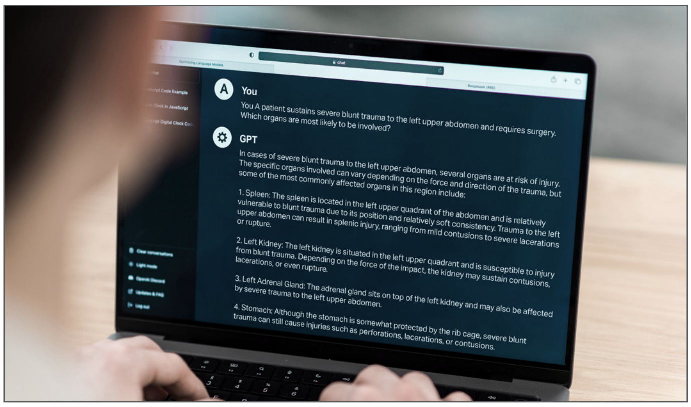
As part of the next steps of RLHF (Reinforcement Learning With Human Feedback), our team of experts is training the model with prompt-response pairs that go beyond boilerplate replies from the LLM bot. The model roadmap includes supporting ranking exercises on qualitative dimensions like harmlessness, honesty, and helpfulness.

THE FUTURISTIC UI FOR PROMPT-RESPONSE PAIR GENERATION

The project successfully showcased the capability to hand-create realistic prompts that emulate user interactions with a chatbot in the Chat-GPT style. It employed a simplified UI for prompt generation, with a focus on the ingenuity required to generate a variety of prompts. The project underscored the adaptability of the prompt generation process by illustrating various use cases, such as crafting medical prompts using GPT-4, highlighting its versatility in addressing diverse needs.

The future roadmap for this annotation solution aligns with the vision of creating a more user-friendly and feature-rich environment, aiming to streamline the annotator's role and enhance overall efficiency.

Additionally, the project identified socially sensitive issues within user queries and developed a system that avoids providing unsafe or biased responses.



You

You A patient sustains severe blunt trauma to the left upper abdomen and requires surgery. Which organs are most likely to be involved?

GPT

In cases of severe blunt trauma to the left upper abdomen, several organs are at risk of injury. The specific organs involved can vary depending on the force and direction of the trauma, but some of the most commonly affected organs in this region include:

1. Spleen: The spleen is located in the left upper quadrant of the abdomen and is relatively vulnerable to blunt trauma due to its position and relatively soft consistency. Trauma to the left upper abdomen can result in splenic injury, ranging from mild contusions to severe lacerations or rupture.

2. Left Kidney: The left kidney is situated in the left upper quadrant and is susceptible to injury from blunt trauma. Depending on the force of the impact, the kidney may sustain contusions, lacerations, or even rupture.

3. Left Adrenal Gland: The adrenal gland sits on top of the left kidney and may also be affected by severe trauma to the left upper abdomen.

4. Stomach: Although the stomach is somewhat protected by the rib cage, severe blunt trauma can still cause injuries such as perforations, lacerations, or contusions.

About iMerit

iMerit provides end-to-end data labeling services to Fortune 500 companies in a wide array of industries including agricultural AI, autonomous vehicles, commerce, geospatial, manufacturing, government, financial services, medical AI and technology. iMerit employs more than 5,500 full-time data annotation experts in Bhutan, Europe, India and the United States.

info@imerit.net | imerit.net

LEARN MORE