HIGH-TECH COMPANY IMPROVED A GENERATIVE AI APP

— APPLAUSE CASE STUDY —

HOW A HIGH-TECH COMPANY WORKED TO IMPROVE ITS LLM-BASED CHAT APP'S RESPONSES AND REDUCE COMMON AI RISKS PRIOR TO PRODUCT LAUNCH.

CUSTOMER

Launching new technology comes with all sorts of risks — but companies known for innovation understand how to guard against potential threats. They understand the complexities of releasing products in the early stages of development while ensuring those products function in ways that don't disappoint or discourage customers. Generative AI and large language models (LLM) require massive amounts of training data and testing before they're ready for use.

When a high-tech company wanted to launch an LLM-based chat app, it wanted to mitigate risk associated with biased, inaccurate or toxic content responses from the model, while delivering a good user experience.

CHALLENGE

Before the company released the LLM-based chat app to the public, the customer wanted to improve the model's responses. The customer also wanted to minimize functional bugs and unsafe/biased content before launch. It also sought real-world feedback from target audience users to improve the user experience.

To create a more accurate and objective chat tool, the company needed a large, diverse set of testers who reflected the demographics within its target markets around the world. The company needed to complete testing in multiple countries within a four-week timeframe to prepare for the scheduled launch. Recruiting and training testers who fit the company's demographic needs, then reviewing and prioritizing their feedback, would be a massive undertaking.

COMPANY
Anonymized

INDUSTRY
Technology

TESTING TYPES
Functional | UX | AI Training & Testing

APP TYPES
Mobile | Web

CHALLENGES
- Quickly sourcing large pools of testers reflecting the demographics of multiple target markets
- Conducting functional testing and collecting UX feedback prior to product launch
- Improving application performance and mitigating risks associated with bias, toxicity, factual errors and privacy issues

RESULTS
- Increased NPS score over 4 points
- Reduced model bias and toxicity
- Improved model performance and user experience

APPLAUSE°

SOLUTION

The technology company partnered with Applause to assemble a panel of testers that matched the general population demographics across six countries. Applause quickly gathered a panel of 10,000 testers in the target countries, mirroring the demographics in each region — including individuals with disabilities. Applause managed all recruiting, screening, vetting, privacy and data consents, NDAs, tester onboarding, test execution and incentives, while complying with local data privacy regulations for all personal data.

In each target country, testers spent four weeks using the app with a wide range of natural usage prompts to generate outputs and stress-test the system under multiple scenarios. The testers provided feedback both in-app and through surveys, with in-country testers offering cultural context that helped the company prevent bias/toxicity.

Expert testers successfully navigated all testing scenarios, while an Applause QA team triaged and reported functional defects and content that posed concerns related to factual errors, bias, toxicity, and privacy issues.

RESULTS

Despite its huge scope, the project successfully concluded on time and on budget. The technology company improved the LLM's performance based on testers' prompt and in-app feedback. By fixing functional flaws and unsafe content responses before launch, the company introduced a better user experience, improving its NPS score more than four points.

Working with Applause allowed the company to quickly collect the feedback needed to confidently launch the app to the public on schedule. In addition to updating training data regularly, the company intends to expand testing to cover more localized language versions of its app.

ABOUT APPLAUSE

Applause is the worldwide leader in crowd-sourced digital quality testing. Software is at the heart of how all brands engage users, and digital experiences must work flawlessly everywhere. With **highly vetted testers available on-demand around the globe,** Applause provides brands with a full suite of testing and feedback capabilities. This approach drastically improves testing coverage, eliminates the limitations of offshoring and traditional QA labs, and speeds time-to-market for websites, mobile apps, IoT, and in-store experiences.

Thousands of leading companies — including Ford, Fox, Google, and Dow Jones — rely on Applause as a best practice to deliver high-quality digital experiences that customers love.

Learn more at: www.applause.com