## BRAINSTATION CROSS-DISCIPLINE STUDENT TEAM

Amir Kiani
Data Scientist

Moises David Sosa Vera
Data Scientist

Emmanuel Atuahene
Software Engineer

Annie Yan
Data Scientist

Amy Chen
User Experience Designer

Chelsea Quindipan
Software Engineer

# 24 HR HACKATHON: 'INSIGHTS' INTO MAINTAINING TRUST IN AI-INTEGRATED ENVIRONMENTS

JULY 21, 2023

Presented by
The Extremers

# 24 HR HACKATHON: ‘INSIGHTS’ INTO MAINTAINING TRUST IN AI-INTEGRATED ENVIRONMENTS

## Our Vision Statement

I'm excited to share our team's approach in educating users about AI-powered Google products while ensuring user trust and avoiding misunderstandings. Through numerous conversations and discussions, we arrived at a fundamental idea: "Trust is how predictable, inclusive, and safe we feel towards something." For instance, just like we trust a lifelong friend to follow through on a favor, communicate with us, and respect our boundaries, we seek similar predictability and reliability from products and services.

Our initial focus was to understand the key reasons behind the general population's mistrust in the use of AI. Through extensive research, we identified three significant factors contributing to this mistrust: the lack of transparency, concerns about privacy invasion, and the difficulty in establishing ground truth for evaluating AI algorithm results in products. Addressing these concerns is pivotal in building trust and confidence in AI technologies.

On a high level, we wish to foster better understanding of AI through the presentation of educational insights that will empower users with the knowledge of how AI has been integrated into their day-to-day experience.

As a result, we are thrilled to introduce "Insights," a powerful widget integrated into the Google platform. "Insights" provides users with a unique glimpse behind the curtain, offering clear explanations on how AI enhances their experience, utilizes their data, and delivers the results they receive. With "Insights," we aim to foster transparency and empower users to confidently embrace AI-related features.

Let's delve deeper into the capabilities and benefits of this innovative widget.

Our vision centers around training or fine-tuning a sophisticated Large Language Model, such as Google's LaMDA, using Google product documentation and AI-related knowledge. This process allows us to develop an advanced widget that seamlessly interacts with Google's AI-powered tools, providing users with a clear understanding of AI's workings behind the scenes.

Upon interacting with our widget "Insights," users may choose the level of interactivity that best suits their comfort level. A passive hover-over feature will show basic information on how Google utilizes AI to enhance its products. Alternatively, click-based interaction will activate a more dynamic chatbot, capable of providing comprehensive explanations.

With "Insights" actively generating user data, we can collect valuable engagement metrics on the AI-powered tools. By analyzing these metrics using statistical methods (such as ANOVA and A|B testing), we aim to gauge the extent of users' interactions with the AI-powered tools. As we strongly believe that user engagement directly correlates with trust, this analysis will offer valuable information into the level of trust users have in our AI products.

By understanding this relationship, we can make actionable changes to the way users see, experience, and interact with Google products and AI technologies.