# DEBRIEF:

In this survey, you were tasked with labeling paintings by the art period they were created in. You were additionally given some advice to help you complete this task. People completing this survey were split into 2 groups and were given advice from two different version of an AI model. The first was more measured in giving advice, while the second was overly confident (both advised the same binary label, but they differed in how confident they were).

The purpose of this experiment was to investigate whether people would respond differently to advice of different confidence levels. This experiment is part of a broader effort to understand when advice from an AI agent will be used, and how to appropriately calibrate and/or train these models for correct levels of utilization.

This type of work is important for a few reasons. As AI technology has rapidly matured in the last decade, one of the biggest barriers to adoption is the limited trust people have in models they don't understand. This is particularly important in medicine, where doctors are often unwilling to trust an AI -- unlike when talking with people, an AI model's prediction cannot easily be verified or justified by asking the model why it made its prediction.

Continue