Regulatory experimentation case study

NSW Productivity and Equality Commission

Exploring new technology through regulatory experimentation

Testing the applications of generative artificial intelligence (AI) – Australian Securities and Investments Commission

Level: **Easy**

Industry: **Finance**

Location: **Australia**

Overview

The Australian Securities and Investments Commission (ASIC) sought to explore the capabilities and limitations of generative AI and its usefulness for their internal processes.

ASIC trained a generative AI large language model (LLM) and conducted an experiment to assess the quality of summaries produced by the LLM against those prepared by humans.

Key finding

AI-generated summaries did not capture the complexity and nuance as well as human-generated summaries.

Outcome

Provided valuable insights on the current limitations of a generative AI large language model (LLM).

Evaluation method

Quantitative assessment and qualitative debrief.

Background

The Australian Securities and Investments Commission (ASIC) wanted to explore the effectiveness of AI in summarising public submissions made to a parliamentary inquiry.

ASIC is Australia's independent integrated corporate, markets, financial services and consumer credit regulator. ASIC wanted to conduct an experiment on generative AI (a type of artificial intelligence that creates new content based on prompts provided by the user and informed by large datasets) with a focus on measuring the quality of the generated output rather than the performance of the models. The motivation stemmed from the need to understand the current capabilities and limitations of generative AI in an environment where the technology is rapidly evolving. ASIC also wanted to explore how it could potentially support or enhance internal processes. The trial was exploratory and not for regulatory use.

Intervention and outcome

AI-generated summaries were less effective than human-generated summaries, particularly in capturing nuance and context.

ASIC conducted a five week proof-of-concept trial using a specific generative AI large language model (LLM) to summarise a sample of public submissions made to a parliamentary inquiry. The trial aimed to optimise the prompts used to instruct the LLM and then compare the AI-generated summaries to human-generated ones. The final phase of testing involved a blind assessment of AI and human summaries based on a standardised criteria, focusing on the quality and accuracy of the generated outputs, followed by a qualitative debrief with the assessors.

ASIC found that the AI summaries performed lower on all criteria compared to human summaries, particularly in capturing nuance and context. These point-in-time results related to the use of certain prompts, using a specific LLM, for a specific use case and should not be extrapolated more widely. The trial provided valuable insights which can be applied to future AI experimentation to ensure ASIC has a continued understanding of the opportunities and uses of AI as the technology evolves, including its shortcomings.

Key steps for successful experiments

Experiment with novel and emerging technologies.

As a novel and emerging technology, generative AI offers fertile ground for impactful research. By emphasising the benefits of building evidence for AI's usefulness, the ASIC research team successfully gained senior buy-in and support to develop a safe and secure environment for rapid experimentation.

Consider how you might trial new technology to improve your regulatory processes and practices.

Using internal and external subject matter experts.

Partnering with a third-party AI expert organisation provided technical AI expertise. This enabled ASIC to blend external AI expertise with internal research and regulatory subject matter expertise, resulting in a more robust exploratory trial.

Consider seeking partnerships with external experts to bolster internal capabilities and gain access to specialised knowledge.

Using internal staff as participants

ASIC was interested in assessing the quality of the AI and human submission summaries, a task requiring specific internal expertise. ASIC was conscious that knowledge of the source of the summaries and the order summaries were assessed in could have introduced bias. To address this, ASIC randomised the order ASIC assessors viewed the summaries in and employed a blinded experiment which meant the assessors did not know where the summaries came from.