We supplement rule-based AI policymaking with an approach that leverages precedent “cases” to ground acceptable AI behavior.

Case Repositories: Towards Case-Based Reasoning for AI Alignment

Kevin Feng¹, Quan Ze (Jim) Chen¹, Inyoung Cheong¹, King Xia², Amy X. Zhang¹
¹University of Washington, ²Independent Attorney

Motivation
When we talk about “policy” for AI, we often think broad statements like “avoid dangerous and unethical responses.” However, it can be difficult to create and apply these policies in practice. How can we use case-based reasoning—long practiced in law, medicine, etc.—to supplement constitutional, rule-based AI policy?

Preliminary Results
We conducted 6 workshops with 20 legal experts and gathered 28 dimensions for how AI should respond to requests for legal advice. Below: a seed case with dimensions, and an LLM injecting details into the case based on the last dimension.

Repository Creation Process

Collect seed cases
Cases are inputs that a user may submit to an AI system. We source our seed cases from discourse in online communities.

Expert workshops
We conduct workshops with domain experts to elicit key dimensions they consider when evaluating AI responses to seed cases.

LLM-assisted case expansion
We develop an LLM-based method for rapidly expanding the case space through case generation guided by experts’ input.

Collect crowdsourced judgements
We collect public input by crowdsourcing judgements of appropriateness of AI responses to cases in our repository.

Future Work & Acknowledgements
Next, we will finalize our crowdsourcing task and pilot it before deployment. We will also derive approaches to handle inputs that are distant from existing cases.

This project was supported by the OpenAI “Democratic Inputs to AI” program. We thank OpenAI and others in the program for their feedback and support.