

Note that the sample size is small, and we did not interview AI researchers who are skeptical of existential risk from AI.¹⁰

3. A new database of empirical evidence for some claims about existential risk from AI

The full database can be accessed [here](#). It covers empirical evidence only, and includes evidence relating to specification gaming, goal misgeneralization and power-seeking (as well as deceptive alignment, self-improvement, and other claims relating to existential risk from AI).

The database draws significantly from existing databases on [specification gaming](#) (Krakovna, 2020) and [goal misgeneralization](#) (Shah et al., 2022a).

¹⁰We didn't have the resources to interview a representative sample, and decided that we would get the most relevant information from speaking with researchers who work on AI existential risk and so are familiar with the evidence.