

Nick Bostrom introduced the term [mindcrime](#) to describe a hypothetical situation in which an AI fully simulates a large number of conscious moral subjects such as humans and thereby makes them suffer. This might be morally catastrophic.

This could happen without being explicitly requested by the operators, even in a tool AI. For instance, if we asked the AI to run a detailed simulation of the effects of a certain policy, it might produce realistic simulations of all the humans affected by this policy, including their suffering. If this is scaled to simulating civilisations, the amount of suffering created might be unimaginably large.

Bostrom suggests that this could be a moral catastrophe and that precautions should be taken to avoid such a situation.

Further reading:

- [Arbital's page on mindcrime](#)

Related

- [Do AIs suffer?](#)
- [Are AIs conscious?](#)
- [What is an "s-risk"?](#)