There is a recent history of individuals and groups that have attempted to destabilize or destroy human civilization. For instance, the Unabomber believed that civilization falling would be better for the preservation of nature, and the Aum Shinrikyo religious group believed that humanity should fall[1] to be reborn again. Both intentionally killed people to advance these goals, but the amount of damage they could do was limited by their lack of access to technologies that enable mass destruction.

Actors that want to destroy civilization are quite rare, but unfortunately they exist, and a single one that succeeds is enough to spell disaster. As AI becomes a more powerful force multiplier, such a motivated individual or group might be tempted to leverage it to reach its aims. This is one of the most dangerous versions of AI misuse.

The pool of people with access to powerful AI is expected to grow. If we somehow got to the point where everyone had access to AI capable enough to kill everyone, and the effort involved were reduced to pulling a trigger, some people might be tempted to destroy the world in a moment of personal crisis.

Further reading:

- Q5 of Yoshua Bengio's FAQ

Alternative phrasings

- Who might want AI to kill everybody?

Related

- Isn't the real concern misuse?
- Why can't we just use a friendly AI to stop bad AIs?
- Might LLMs enable terrorists to build biological weapons?

Scratchpad

[1] David Thorstad points out that Aum Shinrikyo did not want to kill everyone but only non-believers, which is to say everybody except themselves. It seems likely that such selective killing would be harder to achieve than omnicide. Both cases would be very bad.