



01 JUN, 2023

AI chiefs warn of humanity's 'extinction risk'



The Australian, Australia

Page 1 of 1

AI chiefs warn of humanity's 'extinction risk'

NOAH YIM

AI could pose an existential threat to humanity, artificial intelligence luminaries have warned, in an extraordinary one-sentence open letter, urging that it be treated with the same level of urgency as pandemic and nuclear war.

"Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war," reads the statement, published by the US-based Centre for AI Safety.

It was signed by figures including Geoffrey Hinton and Yoshua Bengio – two of the three "godfathers" of modern AI – and Sam Altman, founder and CEO of OpenAI, the maker of ChatGPT. Other signatories include key players at AI frontrunners such as Google, its subsidiary AI lab Deepmind (which famously produced an algorithm to beat professionals at Go), Microsoft and Anthropic.

The letter is the most alarming and united expression of concern yet on the risk of AI from some of the field's top players. It comes after Professor Hinton's exit from Google to talk about the risks of AI more freely, and after Mr Altman testified before the US Senate urging regulation.



Australian Catholic University philosophy professor Simon Goldstein, currently on a fellowship at US-based Centre for AI Safety, was one of two people from Australian institutions whose name appeared on the original list of signatories.

"If you look at the track record in the history of this planet, when a new form of life comes into existence that is more intelligent than what existed before, that can lead to a transition to that life form taking over," he told The Australian.

The concerns of "extinction" raised in the letter are of a medium to long-term nature, and distinct from short-term, immediate fears such as misinformation or political bias.

"If the leading developers, you know, you've got Anthropic, you've got OpenAI, you've got Google, they agree it poses an extinction risk, then people need to know what these organisations think," said Tiberio Caetano, chief scientist at Australian AI ethics think tank Gradient Institute.

He told The Australian that he signed on to the statement after it was made public.

AI researchers say there are multiple ways the technology could pose an existential risk to humanity.

"If I'm going to play the world chess champion and you ask me, 'How are you going to lose this game?', my answer would be 'I have no idea'," said Dr Caetano.

"What I know is that I will lose the game, but I have no idea how. Is it going to be a checkmate with a queen, or with a knight, or a bishop? I don't know the answer.

"But the fact that this opponent is smarter than you and is a threat in and of itself, you can predict that there is a very high chance that the opponent wins the game."

WA Data Science Innovation Hub director Alex Jenkins was more doubtful of the statement, but said he would nonetheless support it.

"It's difficult to understand why they're focusing just on this when there are other real risks that are apparent right now," he said of the statement.

But he said the signatories were too prominent to ignore. "Whatever you think about that, it should be taken seriously because the people who have signed it are comprehensively the top people in AI," Mr Jenkins said.