AI doomsday warnings a distraction from the danger it already poses, warns expert

Focusing on doomsday scenarios in artificial intelligence is a distraction that plays down immediate risks such as the large-scale generation of misinformation, according to a senior industry figure attending this week's AI safety summit. Aidan Gomez, co-author of a research paper that helped create the technology behind chatbots, said long-term risks such as existential threats to humanity from AI should be "studied and pursued", but that they could divert politicians from dealing with immediate potential harms. "I think in terms of existential risk and public policy, it isn't a productive conversation to be had," he said. "As far as public policy and where we should have the public-sector focus – or trying to mitigate the risk to the civilian population – I think it forms a distraction, away from risks that are much more tangible and immediate." Gomez is attending the two-day summit, which starts on Wednesday, as chief executive of Cohere, a North American company that makes AI tools for businesses including chatbots. In 2017, at the age of 20, Gomez was part of a team of researchers at Google who created the Transformer, a key technology behind the large language models which power AI tools such as chatbots. Gomez said that AI – the term for computer systems that can perform tasks typically associated with intelligent beings – was already in widespread use and it is those applications that the summit should focus on. Chatbots such as ChatGPT and image generators such as Midjourney have stunned the public with their ability to produce plausible text and images from simple text prompts. "This technology is already in a billion user products, like at Google and others. That presents a host of new risks to discuss, none of which are existential, none of which are doomsday scenarios," Gomez said. "We should focus squarely on the pieces that are about to impact people or are actively impacting people, as opposed to perhaps the more academic and theoretical discussion about the long-term future." Gomez said misinformation – the spread of misleading or incorrect information online – was his key concern. "Misinformation is one that is top of mind for me," he said. "These [AI] models can create media that is extremely convincing, very compelling, virtually indistinguishable from human-created text or images or media. And so that is something that we quite urgently need to address. We need to figure out how we're going to give the public the ability to distinguish between these different types of media." The opening day of the summit will feature discussions on a range of AI issues, including misinformation-related concerns such as election disruption and erosion of social trust. The second day, which will feature a smaller group of countries, experts and tech executives convened by Rishi Sunak, will discuss what concrete steps can be taken to address AI risks. Kamala Harris, the US vice-president, will be among the attenders. Gomez, who described the summit as "really important", said it was already "very plausible" that an army of bots – software that performs repetitive tasks, such as posting on social media – could spread AI-generated

misinformation. "If you can do that, that's a real threat, to democracy and to the public conversation," he said. In a series of documents outlining AI risks last week, which included AI-generated misinformation and disruption to the jobs market, the government said it could not rule out AI development reaching a point where systems threatened humanity. A risk paper published last week stated: "Given the significant uncertainty in predicting AI developments, there is insufficient evidence to rule out that highly capable Frontier AI systems, if misaligned or inadequately controlled, could pose an existential threat." The document added that many experts considered such a risk to be very low and that it would involve a number of scenarios being met, including an advanced system gaining control over weapons or financial markets. Concerns over an existential threat from AI centre on the prospect of so-called artificial general intelligence – a term for an AI system capable of carrying out multiple tasks at a human or above-human level of intelligence – which could in theory replicate itself, evade human control and make decisions that go against humans' interests. Those fears led to the publishing of an open letter in March, signed by more than 30,000 tech professionals and experts including Elon Musk, calling for a six-month pause in giant AI experiments. Two of the three modern "godfathers" of AI, Geoffrey Hinton and Yoshua Bengio, signed a further statement in May warning that averting the risk of extinction from AI should be treated as seriously as the threat from pandemics and nuclear war. However Yann LeCun, their fellow "godfather" and co-winner of the ACM Turing award – regarded as the Nobel prize of computing – has described fears that AI might wipe out humanity as "preposterous". LeCun, the chief AI scientist at Meta, Facebook's parent company, told the Financial Times this month that a number of "conceptual breakthroughs" would be needed before AI could reach human-level intelligence – a point where a system could evade human control. LeCun added: "Intelligence has nothing to do with a desire to dominate. It's not even true for humans."