'It's not clear we can control it': what they said at the Bletchley Park AI summit

The global AI safety summit opened at Bletchley Park on Wednesday with a landmark declaration from countries including the UK, US, EU and China that the technology poses a potentially catastrophic risk to humanity. The so-called Bletchley declaration said: "There is potential for serious, even catastrophic, harm, either deliberate or unintentional, stemming from the most significant capabilities of these AI models." Here are some of the interventions from political and tech industry figures – as well as King Charles – on the day. Elon Musk The world's richest man and Tesla chief executive described AI as a threat to humanity. Musk, who co-founded the ChatGPT developer OpenAI, has launched a new venture called xAI and is attending both days of the summit, which is being held about 50 miles from London at the site which played host to top-secret codebreakers during the second world war. Describing AI as "one of the biggest threats to humanity", Musk said: "I mean, for the first time, we have a situation where there's something that is going to be far smarter than the smartest human. So, you know, we're not stronger or faster than other creatures, but we are more intelligent. And here we are, for the first time really in human history, with something that's going to be far more intelligent than us." In comments to the PA news agency on the summit sidelines, he said it was "not clear we can control such a thing", but "we can aspire to guide it in a direction that's beneficial to humanity". Mustafa Suleyman The co-founder of DeepMind, a British company that was acquired by Google and is now at the centre of the search company's AI efforts, said a pause in the technology's development might have to be considered over the next five years. Speaking to reporters at the summit, he said: "I don't rule it out. And I think that at some point over the next five years or so, we're going to have to consider that question very seriously." However, Suleyman said current AI models, such as the one powering ChatGPT, did not pose a serious threat. "I don't think there is any evidence today that frontier models of the size of [ChatGPT model] GPT-4 … present any significant catastrophic harms," he said. King Charles In a video message played to delegates at the beginning of the summit, the king described AI as "one of the greatest technological leaps in the history of human endeavour". He urged attenders to tackle the "challenges" of AI – such as protecting democracies – by taking the example of the climate crisis. He said governments, public sector, private sector and civil society had been joined together in a conversation about saving the environment and the same should be done for AI. "That is how the international community has sought to tackle climate change, to light a path to net zero, and safeguard the future of our planet. We must similarly address the risks presented by AI with a sense of urgency, unity and collective strength," he said. Michelle Donelan The UK technology secretary has attempted to strike a balance between risk and opportunity at the summit, an awkward task amid communiques warning of potential catastrophe and presentations on

bioweapon attacks. Asked if AI would disrupt jobs, she said: "I really do think we need to change the conversation when it comes to jobs … What AI has the potential to do is actually reduce some of those tedious administrative part of our jobs, which is particularly impactful for doctors, our police force, our teachers." Donelan added that the UK's education and skill sectors needed to help people adapt to any AI-related job changes. Věra Jourová The European Commission's vice-president for values and transparency said the UK was behind the US and EU in regulating AI by its "own decision". "They [the UK] take different paths," said Jourová, adding that the UK approach was to "focus on the possible risks" and then "regulate later". Rishi Sunak has ruled out bringing in AI legislation immediately, saying the UK government needed to understand the technology better before regulating it. Jourová said the UK's position did not surprise her because when the country was an EU member, its stance on regulation was one of relying on a sector taking "social responsibility". Matt Clifford At the start of each closed-door session, UK officials showed attenders their examples of how powerful AI models could make it easier for bad actors to wreak damage in a number of ways. During one session, Matt Clifford, who was in charge of organising the summit, showed delegates how large language models could make it easier for bedroom hackers to launch phishing attacks. "One of the things that's been really challenging about this debate for policymakers over the last year is sometimes it feels like just trading thought experiments," he said. "What's so great about what the frontier taskforce is doing, what the safety institute will do, is that it gets rid of these thought experiments. It just says, let's look at what these models can do right now."