UK summit aims to tackle thorny issues around cutting-edge AI risks

November 1 2023, by Kelvin Chan

US Vice President Kamala Harris, with husband Second Gentleman Douglas Emhoff, arrives at Stansted Airport for her visit to the UK to attend the AI safety summit at Bletchley Park, in Stanstead, England, Tuesday, Oct. 31, 2023. Credit: Joe Giddens/PA via AP

Digital officials, tech company bosses and researchers are converging

Wednesday on a historic estate near London to discuss and better understand the extreme risks posed by cutting-edge artificial intelligence.

So-called frontier AI refers to the latest and most powerful systems that take the technology right up to its limits, but could come with as-yet-unknown dangers. They're underpinned by foundation models, which power chatbots like OpenAI's ChatGPT and Google's Bard and are trained on vast pools of information scraped from the internet.

Some 100 people from 28 countries including China are expected to attend British Prime Minister Rishi Sunak's two-day AI Safety Summit, though the British government has refused to disclose the guest list.

One of the most high-profile delegates is U.S. Vice President Kamala Harris, who's scheduled to attend the summit's second day after making a separate speech on AI in London on Wednesday.

Tesla CEO Elon Musk is also scheduled to discuss AI with Sunak in a livestreamed conversation on Thursday night. The tech billionaire was among those who signed a statement earlier this year raising the alarm about the perils that AI poses to humanity.

European Commission President Ursula von der Leyen and executives from U.S. artificial intelligence companies such as Anthropic and influential computer scientists like Yoshua Bengio, one of the "godfathers" of AI, are also expected.

The meeting is being held at Bletchley Park, a former top secret base for World War II codebreakers that's seen as a birthplace of modern computing.

An exterior view shows the mansion house at Bletchley Park museum in the town of Bletchley in Buckinghamshire, England, Jan. 15, 2015. The U.K.-hosted AI Safety Summit is bringing politicians, computer scientists and tech executives to a site chosen for its symbolism: Bletchley Park, a name synonymous with codebreaking and the birth of computing. Credit: AP Photo/Matt Dunham, File

One of Sunak's major goals is to get delegates to agree on a first-ever communique about the nature of AI risks. He said the technology brings new opportunities but warns about frontier AI's threat to humanity, because it could be used to create biological weapons or be exploited by terrorists to sow fear and destruction.

Only governments, not companies, can keep people safe from AI's dangers, Sunak said last week. However, in the same speech, he also

urged against rushing to regulate AI technology, saying it needs to be fully understood first.

Harris, meanwhile, will stress the need to address the here and now, including "societal harms that are already happening such as bias, discrimination and the proliferation of misinformation."

She'll point to President Biden's executive order this week, setting out AI safeguards, as evidence the U.S. is leading by example in developing rules for artificial intelligence that work in the public interest.

A White House official gave details of Harris's speech, speaking on condition of anonymity to discuss her remarks in advance.