

Govts, tech firms vow to cooperate against AI risks at Seoul summit

May 22 2024, by Qasim Nauman and Hieun Shin



Attendees at the Ministers' Session of the AI Seoul Summit, where some of the world's biggest tech companies pledged to guard against the dangers of artificial intelligence.

More than a dozen countries and some of the world's biggest tech firms pledged on Wednesday to cooperate against the potential dangers of

artificial intelligence, including its ability to dodge human control, as they wrapped up a global summit in Seoul.

AI safety was front and center of the agenda at the two-day gathering. In the latest declaration, more than two dozen countries including the United States and France agreed to work together against threats from cutting-edge AI, including "severe risks".

Such risks could include an AI system helping "non-state actors in advancing the development, production, acquisition or use of chemical or biological weapons", said a joint statement from the nations.

These dangers also include an AI model that could potentially "evade human oversight, including through safeguard circumvention, manipulation and deception, or autonomous replication and adaptation", they added.

The ministers' statement followed a commitment on Tuesday by some of the biggest AI companies, including ChatGPT maker OpenAI and Google DeepMind, to share how they assess the risks of their tech, including what is considered "intolerable".

The 16 tech firms also committed to not deploying a system where they cannot keep risks under those limits.

The Seoul summit, co-hosted by South Korea and Britain, was organized to build on the consensus reached at the inaugural AI safety summit last year.

"As the pace of AI development accelerates, we must match that speed... if we are to grip the risks," UK technology secretary Michelle Donelan said.

"Simultaneously, we must turn our attention to risk mitigation outside these models, ensuring that society as a whole becomes resilient to the risks posed by AI."

The summit also saw a separate commitment—the so-called Seoul AI Business Pledge—from a group of tech companies including South Korea's Samsung Electronics and US titan IBM, to develop AI responsibly.

AI is "a tool in the hands of humans. And now is our moment to decide how we're going to use it as a society, as companies, as governments," Christina Montgomery, IBM's Chief Privacy and Trust Officer, told AFP on the sidelines of the summit.

"Anything can be misused, including AI technology," she added. "We need to put guardrails in place, we need to put protections in place, we need to think about how we're going to use it in the future."



AI ethics experts such as Rumman Chowdhury warn that artificial intelligence can be misused in a wide variety of ways.

Seeking consensus

AI's proponents have heralded it as a breakthrough that will improve lives and businesses around the world, especially after the stratospheric success of ChatGPT.

However, critics, rights activists and governments have warned that the technology can be misused in a wide variety of ways, including election manipulation through AI-generated disinformation such as "deepfake" pictures and videos of politicians.

Many have called for international standards to govern the development and use of AI. But experts at the Seoul summit warned that AI poses a huge challenge to regulators because it is rapidly developing.

"Dealing with AI, I expect to be one of the biggest challenges that governments all across the world will have over the next couple of decades," said Markus Anderljung, head of policy at the UK-based non-profit Center for the Governance of AI.

Jack Clark, co-founder of the AI startup Anthropic, said consensus on AI safety cannot be left to tech firms alone, and that government and academic experts are needed in the conversation.

"At this summit, I've actually been asking every single person I met with: What's safety to you? And I've had a different answer from each person," Clark told reporters. "And I think that illustrates the problem."

"You aren't going to arrive at a consensus by the companies alone, and if you did, I doubt it would be the correct one."

Also on the agenda in Seoul was ensuring that AI is inclusive and open to all.

It is not just the "runaway AI" of science fiction nightmares that is a huge concern, but also inequality, said Rumman Chowdhury, an AI ethics expert who leads the non-profit AI auditor Humane Intelligence.

"All AI is just built, developed and the profits reaped (by) very, very few people and organizations," she told AFP.

People in developing countries such as India "are often the staff that does the clean-up. They're the data annotators, they're the content moderators. They're scrubbing the ground so that everybody else can

walk on pristine territory".

© 2024 AFP

Citation: Govts, tech firms vow to cooperate against AI risks at Seoul summit (2024, May 22)
retrieved 22 June 2024 from <https://techxplore.com/news/2024-05-ai-firms-pledge-responsible-tech.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.