PRESS RELEASE: For Immediate Release

Leading Experts Will Gather in Tokyo to Discuss How to Make AI Safe

More information and photos are available here: https://tais2024.cc/

TOKYO, 15th March, 2024 - On April 5th and 6th, the Technical AI Safety conference will convene in Tokyo, bringing together leading experts in artificial intelligence (AI) safety and alignment from leading Japanese and international institutions and organisations. Computer scientists, globally recognised in the field of artificial intelligence, will present pioneering ideas and thought-provoking research findings on how to develop AI systems that are safe, beneficial, and aligned with human values.

Organised by AI Safety Tokyo in partnership with Noeon Research and AI Industry Foundation, the two-day conference will host some of the most distinguished names in the field. Speakers include Dan Hendrycks (Center for AI Safety, UC Berkeley), Miki Aoyagi (Nihon University), Noah Siegel (Google DeepMind), Hoagy Cunningham (Anthropic) and others. Full list of confirmed speakers is available at the conference website.

As latest AI systems evince human-level capabilities on specific tasks, scientists in the field face growing pressure to mitigate near-term harms and prevent catastrophic threats when working on new advances in artificial intelligence. The imperative to address risks associated with AI has also prompted policymakers and legislators to introduce laws and regulations aimed at establishing guardrails controlling the development of the ever-emerging technology. That includes Japan, where the ruling Liberal Democratic Party is reportedly working on a draft AI legislation, with hopes of implementing it by the end of this year.

Blaine Rogers, Director of AI Safety Tokyo at the AI Industry Foundation, said: “Robustness and alignment have long been barriers to using deep neural networks in safety-critical domains. I’m excited to capture the momentum that comes with international attention and foster collaboration between researchers inside and outside Japan.”

Ryan Kidd, Co-Director of the ML Alignment & Theory Scholars Program and a Board Member and Co-Founder of the London Initiative for Safe AI (LISA), said: "Establishing the safety, reliability, and explainability of frontier AI systems is critical to ensuring these revolutionary technologies have a positive and harmonious impact on society. I look forward to making strong connections with AI safety researchers in Japan and sharing insights for new research and education projects."

Jesse Hoogland, Executive Director of Timaeus, said: Japan has the opportunity to have a unique impact on AI safety if it leverages its strong existing research expertise. I'm excited to make contact with local researchers working on AI and learning theory.

Miki Aoyagi, Professor at the Department of Mathematics, College of Science & Technology, Nihon University, said: ”The development of AI is rapid. While attention tends to focus on its merits, it is important to also consider its drawbacks. TAIS 2024 is expected to host discussions on AI from various perspectives. I look forward to it.”

Dan Hendrycks, Founder and Executive Director of the Center for AI Safety, said: "International coordination on AI safety is crucial to ensure that AI is ultimately beneficial. I'm thrilled to attend TAIS 2024 and hope to help foster collaboration among global researchers."

James Fox, Research Director of the London Initiative for Safe AI (LISA), said: "I'm really excited to engage with AI safety researchers across the globe collected at TAIS 2024. I'm eager to help foster connections between multiple different communities and help to advance our collective goal of ensuring that AI is beneficial for all."

Matt MacDermott, PhD researcher at Imperial College London and a member of the Causal Incentives Group, said: "Safe development of AI is one of the most important challenges the world faces over the coming years and decades, so I'm excited to have the opportunity to talk about it with researchers from across the global."

Scott Emmons, PhD researcher at the University of California, Berkeley and the Center for Human-Compatible AI, said: "Global coordination will be key to ensuring the safe development of frontier AI systems. I'm excited to convene in Japan and help build international collaboration on mitigating AI risk."

Oliver Klingjeford, Co-Founder of the Meaning Alignment Institute, said: "I am very excited to see initiatives like TAIS, bringing attention to important problems in AI to researchers in places other than London and SF."

Andrei Krutikov, Chief Executive Officer of Noeon Research, said: "At Noeon Research, we aspire to build systems that are robust, reliable, and harmless, and in our quest for pushing the boundaries of what artificial intelligence can do, safety and alignment remain foundational principles. Therefore, we are proud to bring this conference together with our partners, setting the stage for a lively discussion of the latest ideas in AI safety with some of the brightest minds in the area."

"We are looking forward to welcoming our speakers in Japan, our home country and one of the leading global centres for artificial intelligence. We hope that, being held in Tokyo, TAIS 2024 will elevate the domestic debate surrounding AI safety, helping Japanese researchers, business leaders and policymakers facilitate responsible development of AI."

Tim Parker, PhD researcher at Institut de Recherche en Informatique de Toulouse, Université de Toulouse, said: "I am excited to attend TAIS 2024 in order to meet and discuss with AI safety researchers from all over the world, advancing the introduction of safe and beneficial AI."

Hoagy Cunningham, Member of the Interpretability Team at Anthropic, said: "Really excited to come and speak at TAIS 2024. Interpretability is one of the most exciting frontiers of research in the world at the moment so I'm excited to talk about this. I'm also super excited to hear from all of the other researchers, many of whom I already know are doing excellent work, and to meet and hear from Japanese researchers and understand how they view the problem of handling the advent of AI in a safe way."

Oskar John Hollinsworth, SERI MATS 4.0 Fellow, said: "Many traditional machine learning conferences and workshops suffer from relative ignorance of technical AI safety amongst their readerships. I am very excited to attend TAIS where researchers in these fields can deeply engage with each other's work."

Noah Y. Siegel, Senior Research Engineer at Google DeepMind, said: "Reducing catastrophic and existential risk from advanced AI systems will require international collaboration. I'm excited for this opportunity to build connections, and to share and learn about progress in the field."

We are able to arrange interviews with speakers, as well as with conference organisers and partners. Please contact simeon@noeon.ai to set up a meeting with the individual you would like to interview.

- ENDS -

Notes for editors:

1. Please contact Simeon Ganiushin (simeon@noeon.ai) to request media passes for the conference.
2. Conference partners are able to cover travel costs for selected media representatives. Please send an enquiry to Simeon Ganiushin (simeon@noeon.ai).
3. AI Safety Tokyo at the AI Industry Foundation (AIST at AIIF) is a Tokyo-based group of AI safety stakeholders advancing AI safety within the ambit of the AIIF. The AIIF’s mission is to get its executive AI leaders working with AI at a continuous and advanced level in all relevant domains. Learn more: https://aisafety.tokyo/ and https://www.aiindustryfoundation.org.
4. Noeon Research is a Tokyo-based company developing an alternative AI architecture for problem-solving. Learn more: https://noeon.ai/