TECHNOLOGY, MEDIA AND TELECOMMUNICATIONS

Singapore Launches New AI Safety Initiatives at Global AI Action Summit

Introduction

The Artificial Intelligence Action Summit ("**Summit**") was held on 10 and 11 February 2025 in Paris, at which political, business and civil society leaders came together to foster international cooperation in key areas of artificial intelligence ("**AI**"), including AI governance, innovation, and safety.

At the Summit, Singapore's Minister for Digital Development and Information, Mrs Josephine Teo, introduced new AI governance initiatives to enhance the safety of AI for Singaporeans and global citizens (the press release for which is available here). This is in recognition of the importance of ensuring the safety of AI systems and the key functions of AI governance, as well as the transboundary nature of AI products and services. The initiatives are as follows:

1. **Global AI Assurance Pilot ("AI Pilot")** – This is a testbed to establish global best practices around technical testing of generative AI ("**GenAI**") applications.

2. **Joint Testing Report with Japan ("Joint Report")** – This collaboration with Japan aims to make Large Language Models ("**LLMs**") safer in different linguistic environments through assessing if guardrails hold up in non-English settings.

3. **Singapore AI Safety Red Teaming Challenge Evaluation Report ("Evaluation Report")** – This seeks to help understand how LLMs perform with regard to different languages and cultures in the Asia-Pacific region, and if the safeguards hold up in these contexts.

This Update highlights the key features of these initiatives.

Global AI Assurance Pilot

The AI Pilot was launched by the AI Verify Foundation and the Infocomm Media Development Authority ("**IMDA**") to help codify emerging norms and best practices around technical testing of GenAI applications.

Under the AI Pilot, the following initiatives will be undertaken:

1. pairing AI assurance and testing providers with organisations deploying GenAI applications;
2. focusing on technical testing of real-life applications; and

3. using the lessons learnt from specific examples to create generalisable insights on "*what and how to test*".

The AI Pilot aims to achieve insights to help in the following:

1. developing testing norms and best practices;
2. laying the foundations for a viable AI assurance market; and
3. better equipping AI testing tools.

The press release on the AI Pilot is available here, and further details on the AI Pilot are available on the AI Verify portal here.

Joint Testing Report with Japan

Singapore has collaborated with Japan in this Joint Report, which aims to make LLMs safer in different linguistic environments through assessing if guardrails hold up in non-English settings. This is part of a continued effort to advance the science of AI model evaluations and work towards building common best practices for testing advanced AI systems, especially given the fact that current training and testing is English-centric.

Mistral Large and Gemma 2 (27B) were tested on model output across languages for multilingual evaluations. For evaluations on cybersecurity-related capabilities, an open weight model was tested.

The key takeaways in the Joint Report include the following:

1. human expertise in global languages helped identify errors in automated evaluation results;
2. revisiting rubrics for evaluating model output to better control subjectivity and increase global standardisation; and
3. aligning on consistent evaluation infrastructure helps make joint testing more efficient and effective.

The full Joint Report titled "*International Network of AI Safety Institutes Joint Testing Exercise: Improving Methodologies for AI Model Evaluations Across Global Languages*" is available here.

Singapore AI Safety Red Teaming Challenge Evaluation Report

IMDA, in partnership with Humane Intelligence, conducted the world's first multicultural and multilingual AI safety red teaming exercise focused on Asia-Pacific in November and December 2024, aiming to establish a baseline for AI safety in cultural and linguistic contexts in the region.

This exercise produced a systematic methodology that can be used to test LLMs for context-specific concerns in different languages and cultures, allowing organisations around the world to adopt and adapt this methodology to test models for linguistic and cultural sensitivities in their countries. The exercise also produced a baseline understanding of the extent to which LLMs manifest cultural bias in the Asia-Pacific region.

The key outcomes in the Evaluation Report are as follows:

1. **Red teaming methodology** – A systematic red teaming methodology was developed and used to test for context-specific safety concerns in different regions.

2. **Cultural bias taxonomy** – A taxonomy identifying the top three bias concerns in each of the participating countries was developed together with the red teamers.

3. **Baseline understanding of cultural bias in LLMs** – Analysis of the challenge data produced a baseline understanding of the extent to which cultural bias is manifested in model output.

The full Evaluation Report is available here.

Concluding Words

These initiatives demonstrate Singapore's focus on emerging AI risks and acknowledge that AI governance and security is a global issue that can only be managed through international collaboration. The engagements reinforce Singapore's role in shaping international AI standards and ensuring that AI governance remains adaptable to technological advancements.

For further queries, please feel free to contact our team.

Contacts

TECHNOLOGY, MEDIA & TELECOMMUNICATIONS

Rajesh Sreenivasan

HEAD

D +65 6232 0751
rajesh@rajahtann.com

Steve Tan

DEPUTY HEAD

D +65 6232 0786
steve.tan@rajahtann.com

Benjamin Cheong

DEPUTY HEAD

D +65 6232 0738
benjamin.cheong@rajahtann.com

Please feel free to also contact Knowledge Management at RTApublications@rajahtann.com.

Regional Contacts

Cambodia

Rajah & Tann Sok & Heng Law Office

T +855 23 963 112 / 113
kh.rajahtannasia.com

China

Rajah & Tann Singapore LLP
Shanghai & Shenzhen Representative Offices

Shanghai Representative Office
T +86 21 6120 8818
F +86 21 6120 8820

Shenzhen Representative Office
T +86 755 8898 0230
cn.rajahtannasia.com

Indonesia

Assegaf Hamzah & Partners

Jakarta Office
T +62 21 2555 7800
F +62 21 2555 7899

Surabaya Office
T +62 31 5116 4550
F +62 31 5116 4560
www.ahp.co.id

Lao PDR

Rajah & Tann (Laos) Co., Ltd.

T +856 21 454 239
F +856 21 285 261
la.rajahtannasia.com

Malaysia

Christopher & Lee Ong

T +603 2273 1919
F +603 2273 8310
www.christopherleeong.com

Myanmar

Rajah & Tann Myanmar Company Limited

T +951 9253750
mm.rajahtannasia.com

Philippines

Gatmaytan Yap Patacsil Gutierrez & Protacio (C&G Law)

T +632 8248 5250
www.cagatlaw.com

Singapore

Rajah & Tann Singapore LLP

T +65 6535 3600
sg.rajahtannasia.com

Thailand

Rajah & Tann (Thailand) Limited

T +66 2656 1991
F +66 2656 0833
th.rajahtannasia.com

Vietnam

Rajah & Tann LCT Lawyers

Ho Chi Minh City Office
T +84 28 3821 2382
F +84 28 3520 8206

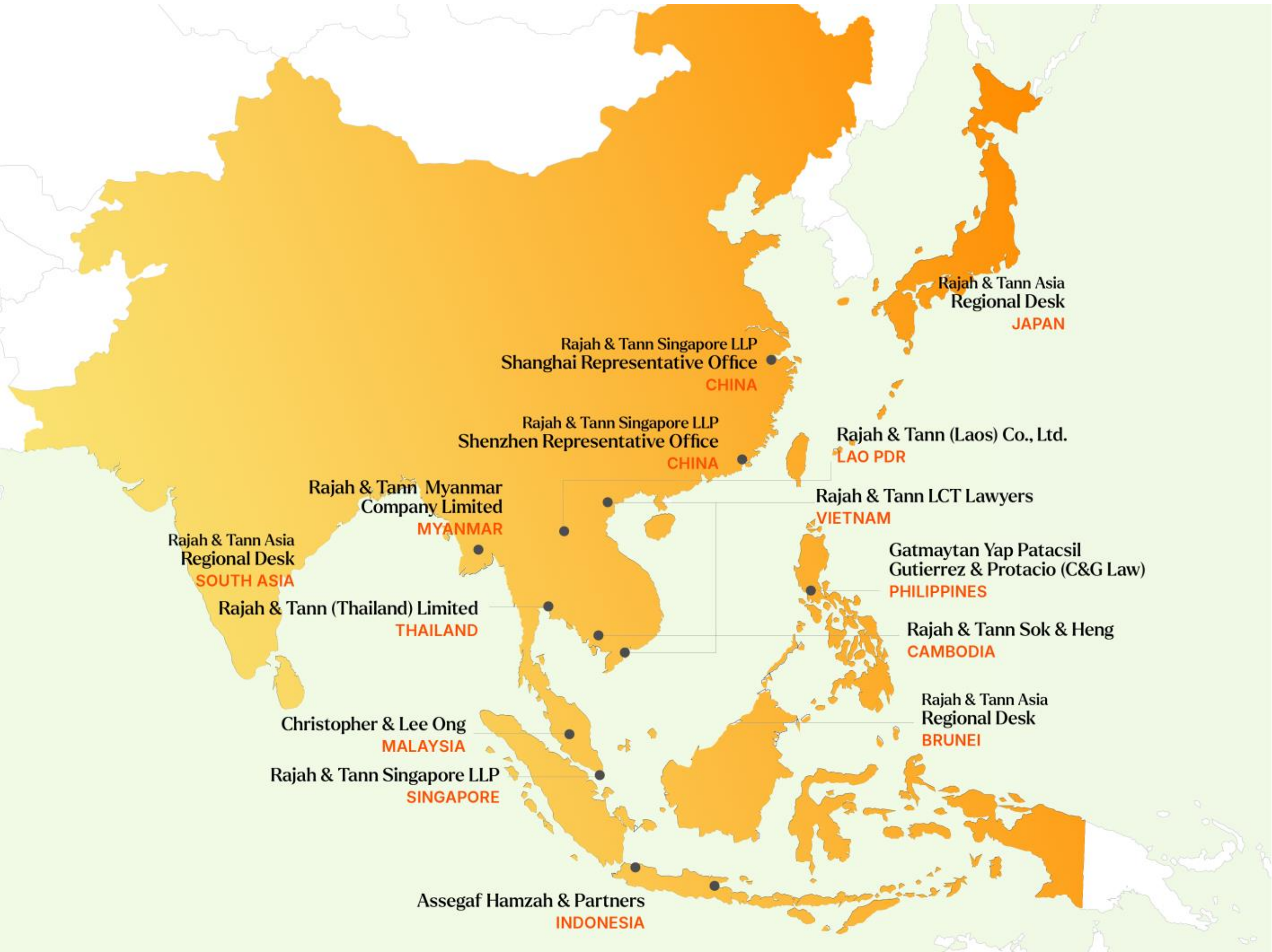
Hanoi Office
T +84 24 3267 6127 / +84 24 3267 6128
vn.rajahtannasia.com

Rajah & Tann Asia is a network of legal practices based in Asia.

Member firms are independently constituted and regulated in accordance with relevant local legal requirements. Services provided by a member firm are governed by the terms of engagement between the member firm and the client.

This update is solely intended to provide general information and does not provide any advice or create any relationship, whether legally binding or otherwise. Rajah & Tann Asia and its member firms do not accept, and fully disclaim, responsibility for any loss or damage which may result from accessing or relying on this update.

Our Regional Presence



Rajah & Tann Singapore LLP is one of the largest full-service law firms in Singapore, providing high quality advice to an impressive list of clients. We place strong emphasis on promptness, accessibility and reliability in dealing with clients. At the same time, the firm strives towards a practical yet creative approach in dealing with business and commercial problems. As the Singapore member firm of the Lex Mundi Network, we are able to offer access to excellent legal expertise in more than 100 countries.

Rajah & Tann Singapore LLP is part of Rajah & Tann Asia, a network of local law firms in Cambodia, China, Indonesia, Lao PDR, Malaysia, Myanmar, the Philippines, Singapore, Thailand and Vietnam. Our Asian network also includes regional desks focused on Brunei, Japan and South Asia.

Please note also that whilst the information in this Update is correct to the best of our knowledge and belief at the time of writing, it is only intended to provide a general guide to the subject matter and should not be treated as a substitute for specific professional advice for any particular course of action as such information may not suit your specific business and operational requirements. It is to your advantage to seek legal advice for your specific situation. In this regard, you may call the lawyer you normally deal with in Rajah & Tann Singapore LLP or email Knowledge Management at RTApublications@rajahtann.com.