

Úřad pro umělou inteligenci přispívá ke třetímu společnému testování Mezinárodní sítě institutů pro bezpečnost umělé inteligence

Úřad pro umělou inteligenci se jako člen sítě aktivně účastní a přispívá hlubokými technickými odbornými znalostmi k hodnocení obecných modelů umělé inteligence. Byl jedním ze tří členů sítě, kteří prováděli hodnocení agentů pro složku kybernetické bezpečnosti.



AdobeStock © Supatman

S cílem doprovázet svolání Mezinárodní sítě institutů pro bezpečnost UI ve Vancouveru (Kanada) zveřejnila síť [výsledky svého třetího společného testování](https://www.aisi.gov.uk/work/international-joint-testing-exercise-agentic-testing). Cvičení se zaměřilo na hodnocení „agentů“ – třídy pokročilých programů UI, které samostatně uvažují, plánují, používají nástroje a provádějí úkoly. Cílem je pokročit v globálním porozumění tomu, jak bezpečně a spolehlivě testovat tyto vznikající agentické systémy, které představují nová rizika v důsledku omezeného lidského dohledu.

Tento třetí test se zaměřil na dvě prioritní rizikové oblasti:

- Únik citlivých informací a podvody
- Kybernetická bezpečnost

Toto cvičení vychází z poznatků ze dvou předchozích společných testovacích cvičení provedených sítí v San Francisku (listopad [2024](https://www.nist.gov/system/files/documents/2024/11/21/Improving%20International%20T))

<https://www.nist.gov/system/files/documents/2024/11/21/Improving%20International%20T>

[esting%20of%20Foundation%20Models-%20%20%20A%20Pilot%20Testing%20Exercise%20from%20the%20International%20Network%20of%20AI%20Safety%20Institutes.pdf](https://sgaisi.sg/wp-api/wp-content/uploads/2025/03/International-Network-of-AI-Safety-Institutes-Joint-Testing-Exercise-Improving-Methodologies-for-AI-Model-Evaluations-Across-Global-Languages.pdf)) a v Paříži (únor 2025).
(<https://sgaisi.sg/wp-api/wp-content/uploads/2025/03/International-Network-of-AI-Safety-Institutes-Joint-Testing-Exercise-Improving-Methodologies-for-AI-Model-Evaluations-Across-Global-Languages.pdf>) Cílem těchto cvičení je umožnit síti dále zdokonalovat osvědčené postupy pro testování pokročilých systémů UI.

Tradiční metody hodnocení se ukázaly jako nedostatečné pro zachycení složitosti chování autonomních agentů. Zúčastnění členové za tímto účelem shromáždili své kolektivní technické a jazykové odborné znalosti. Důraz tohoto hodnocení nebyl kladen pouze na výsledky testování, ale také na zlepšení metodik, přičemž bylo uznáno, že malé změny v koncepci hodnocení mohou mít na výsledky významný dopad.

Toto společné úsilí představuje důležitý krok v rozvoji vědy o hodnocení agentů a představuje zásadní investici do bezpečného a důvěryhodného vývoje pokročilých systémů UI.

Toto je strojový překlad vygenerovaný službou eTranslation provozovanou Evropskou komisí. ☐ Jeho účelem je pomoci vám porozumět obsahu této stránky. ☐ [Přečtěte si ☐ prosím podmínky](https://ec.europa.eu/info/use-machine-translation-europa-exclusion-liability_en) (https://ec.europa.eu/info/use-machine-translation-europa-exclusion-liability_en). Text v původním jazyce ☐ [je k dispozici na stránce se zdrojovým textem](https://digital-strategy.ec.europa.eu/en/news/ai-office-contributes-third-joint-testing-exercise-international-network-ai-safety-institutes) (<https://digital-strategy.ec.europa.eu/en/news/ai-office-contributes-third-joint-testing-exercise-international-network-ai-safety-institutes>).

Source URL:

<https://digital-strategy.ec.europa.eu/news/ai-office-contributes-third-joint-testing-exercise-international-network-ai-safety-institutes>

© European Union, 2026 - [Utváření digitální budoucnosti Evropy](https://digital-strategy.ec.europa.eu/cs) (<https://digital-strategy.ec.europa.eu/cs>) - PDF generated on 29/10/2026

Reuse of this document is allowed, provided appropriate credit is given and any changes are indicated (Creative Commons Attribution 4.0 International license).
For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.

Source URL:

<https://digital-strategy.ec.europa.eu/news/ai-office-contributes-third-joint-testing-exercise-international-network-ai-safety-institutes>