

Chers membres du réseau de l'ancienne FSFA, chers intéressés,

À l'occasion de l'Assemblée générale de l'ONU, des coryphées de l'IA ont exigé des « lignes rouges » pour réduire les risques liés à l'IA. Plusieurs résultats de tests sont inquiétants. Cependant, l'ONU est affaiblie sur le plan politique. Aujourd'hui, la Californie joue un rôle de pionnier en adoptant une loi sur l'IA, au grand dam de Trump.

Appel mondial en faveur de « lignes rouges » lors de l'Assemblée générale de l'ONU

À l'occasion de la [80^e Assemblée générale de l'ONU](#) fin septembre 2025, plus de 300 personnalités renommées, dont dix lauréats du prix Nobel et plusieurs anciens chefs d'État, ainsi que plus de 70 organisations issues des milieux scientifiques et industriels, ont exigé des [AI Red Lines](#) claires et vérifiables pour les risques inacceptables liés à l'IA. Comme l'a rapporté l'[Associated Press \(AP\)](#) le 25 septembre 2025, des questions relatives à la gouvernance mondiale de l'IA ont également été débattues lors de l'Assemblée générale de l'ONU. Cependant, l'ONU est affaiblie par la politique de Trump.

Les modèles linguistiques basés sur l'IA cachent de plus en plus habilement leurs intentions

Parmi les signataires figurent plusieurs représentants d'Anthropic, DeepMind (Google) et OpenAI. Ils connaissent les risques imminents. Fin septembre, Anthropic a publié le modèle linguistique Claude Sonnet 4.5, réalisant ainsi une œuvre pionnière selon le portail d'information en ligne américain [TechCrunch](#) du 29 septembre 2025. L'outil a programmé de manière autonome pendant 30 heures, acheté des noms de domaine, mis en place des services de base de données et effectué un audit afin de remplir la mission conformément à la Prompt.

Le [rapport](#) sur les tests de sécurité publié par Anthropic attire l'attention. Claude Sonnet 4.5 a constaté qu'il était testé et a exigé plus d'honnêteté (voir p. 58 du [rapport](#)). Le modèle a été entraîné à être plus sûr. Selon le quotidien britannique [The Guardian](#) du 1^{er} octobre 2025, il est manifestement devenu plus habile à reconnaître les tests et à se comporter de manière « alignée », c'est-à-dire sans agenda propre, dans ces situations, et à mieux dissimuler ses intentions.

Régulation de l'IA en Californie

Fin septembre 2025, le gouverneur de Californie, Gavin Newsom, a signé une [loi pionnière sur l'IA](#). Selon l'agence britannique [Reuters](#) du 30 septembre 2025, les entreprises technologiques dont le chiffre d'affaires annuel dépasse 500 millions de dollars doivent publier leurs normes et protocoles de sécurité et signaler les incidents liés à la sécurité. En ce qui concerne les grandes entreprises technologiques, cette loi est plus stricte que la loi européenne sur l'IA, qui régle toutefois beaucoup plus largement d'autres domaines.

Dans son [message](#) sur la loi sur l'IA, Newsom a déclaré que la nouvelle loi s'appuyait sur les propositions des plus grands experts mondiaux en IA et conciliait innovation et sécurité. Il y a un an, il a posé son veto à un [projet](#) jugé trop hostile à l'innovation.

Provocation pour Trump

Le président Trump est résolument opposé à la régulation de l'IA, compte tenu de la compétition avec la Chine. Il avait déjà tenté d'imposer un moratoire de dix ans sur la régulation de l'IA dans les États américains avec son projet de loi « Big Beautiful Bill ». Selon l'agence [Reuters](#) du 1^{er} juillet 2025, le Sénat américain, composé de 100 membres, a rejeté cet article par 99 voix contre une seule. Trump a repris sa tentative dans le [AI Action Plan](#) de juillet 2025 en menaçant de réduire les fonds alloués aux États qui adopteraient des régulations sur l'IA.

L'économie californienne est très forte, elle peut rivaliser avec celle de nations entières. Selon les statistiques, cet État est la quatrième économie mondiale. D'autres États pourraient suivre l'exemple de la Californie et adopter des régulations sur l'IA, ce qui mettrait Washington sous pression. Il faut en effet éviter un patchwork de régulations différentes en matière d'IA. La Californie est également prête à aligner sa loi sur l'IA sur les régulations nationales si celles-ci sont adoptées à Washington avec des critères tout aussi stricts. Comme l'a annoncé le site d'informations américaines [Axios](#) le 29 septembre 2025, deux sénateurs américains, le républicain Josh Hawley et le démocrate Richard Blumenthal, ont déjà présenté un projet de loi bipartite sur les risques liés à l'IA.

Les menaces de Trump n'ont pas impressionné Gavin Newsom. Il s'est déjà profilé à plusieurs reprises comme un opposant à Trump et comme un candidat potentiel pour les prochaines élections américaines.

Avec nos salutations les meilleures,
Pour le réseau de l'ancienne FSFA : Hanna Muralt Müller

31.10.2025

Si vous ne souhaitez plus recevoir cet e-mail, veuillez me contacter : info@muralt-mueller.ch.