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

L'AI Act de l'UE, la première législation transnationale au monde sur l'utilisation de l'intelligence artificielle, sera mise en œuvre par étapes selon un calendrier. Il convient tout d'abord de définir des exigences techniques concrètes, mesurables et vérifiables à partir des dispositions légales. Ce n'est pas facile.

L'ETHZ développe un modèle de mise en œuvre avec COMPL-AI
En collaboration avec l'institut bulgare d'IA INSAIT, à la fondation duquel l'ETHZ et l'EPFL ont participé en 2022, la spin-off LatticeFlow AI de l'ETHZ a présenté COMPL-AI, un modèle permettant de mesurer si une application d'IA examiné répond aux exigences du AI Act de l'UE. COMPL-AI est le premier instrument au monde permettant d'évaluer les exigences réglementaires.

Douze grands modèles linguistiques éminents – dont ChatGPT d'OpenAI, Llama de Meta, Claude d'Anthropic – ont été examinés avec COMPL-AI. Des déficits ont été constatés pour tous, notamment en ce qui concerne les critères de transparence, de sécurité, de protection des données, d'équité et aussi d'autres exigences du AI Act de l'UE (voir ETH News du 21.10.2024). L'étude a été publiée sur la plateforme arXiv de l'Université Cornell.

La Commission européenne a salué l'outil mis à disposition sous forme d'open source comme un premier pas vers la mise en œuvre. COMPL-AI peut également être téléchargé sur GitHub. Le professeur Martin Vechev, de l'ETHZ, cofondateur et directeur scientifique d'INSAIT, a encouragé d'autres groupes de recherche et praticiens à développer l'outil open source. La méthodologie élaborée avec COMPL-AI peut également être adaptée à d'autres législations comparables. L'outil permet désormais à tous les fournisseurs de vérifier eux-mêmes leurs modèles d'IA et de prendre des mesures pour qu'ils soient conformes aux futures exigences légales (voir informations sur le site de LatticeFlow).

La confiance dans l'IA – un thème important à l'ETHZ et à l'EPFL
Une application d'IA fiable et digne de confiance résulte de l'interaction entre la technique et les principes et valeurs éthiques. Les conditions préalables à une IA digne de confiance sont donc un thème interdisciplinaire très important pour les deux écoles polytechniques fédérales, l'ETHZ et l'EPFL. Il ne s'agit pas seulement d'appliquer des principes. Une IA doit aussi être techniquement vérifiable, en particulier dans les domaines d'application délicats, comme par exemple les diagnostics en médecine.

Dans le cadre de la Swiss AI Initiative et du Swiss National AI Institute (SNAI), fondé en octobre 2024, plus de 650 chercheurs de l'ETHZ, de l'EPFL et de dix autres hautes écoles suisses développent actuellement un grand modèle de langage suisse comme base open source pour les développements ultérieurs des entreprises et des start-ups. Non seulement les codes sources, mais aussi les données d'entraînement et les paramètres importants seront librement accessibles (voir ETH News du 31.3.2025).

Selon une ETH News du 28.3.2025, le mathématicien Juan Gamella a développé des mini-laboratoires dans lesquels des algorithmes et des modèles d'IA peuvent être vérifiés dans un environnement de test avec des données de mesure réelles. Parfois, les applications d'IA ne fonctionnent pas comme prévu dans des conditions réelles et doivent être améliorées. À l'instar des constructeurs aéronautiques qui testent d'abord en tunnel aérodynamique un modèle miniature d'avion conçu par ordinateur, les chercheurs peuvent rendre leurs applications d'IA plus sûres en les testant dans des mini-laboratoires. En outre, les mini-laboratoires pourraient également être utilisés pour explorer les relations entre les causes et les effets, c'est-à-dire les liens de causalité, et les modéliser dans des algorithmes.

Des chercheurs de l'EPFL – comme ils l'expliquent dans leur news du 10.4.2025 – ont développé un outil révolutionnaire grâce auquel les informations d'une IA deviennent plus robustes et donc plus fiables, même si les données d'entraînement contiennent, comme c'est souvent le cas, de nombreuses données erronées ou prêtant à confusion. Cet outil est le résultat de plusieurs années de recherche, décrit dans une publication du professeur Rachid Guerraoui et d'autres auteurs. Les chercheurs sont convaincus que la Suisse peut jouer un rôle de pionnier avec cette innovation pour une IA plus sûre.

Avec nos salutations les meilleures,
Pour le réseau de l'ancienne FSFA : Hanna Muralt Müller 29.4.2025