LUXEMBOURG INSTITUTE OF SCIENCE AND TECHNOLOGY | LIST

COMMUNIQUÉ DE PRESSE

BELVAL – 22 MARS 2024

LANCEMENT DU AI SANDBOX DU LIST ET DE SON LEADERBOARD CONTRE LES BIAIS ÉTHIQUES

Le Luxembourg Institute of Science and Technology (LIST) a dévoilé ses « AI regulatory sandboxes », visant à faire progresser les activités de recherche et de développement dans le domaine de l'intelligence artificielle à Amsterdam lors de la conférence AIMMES 2024.

En s'appuyant sur son expérience en matière de collaboration avec les autorités de régulation et de contrôle de la conformité, le LIST mène des activités de recherche et de développement axées sur les sandboxes réglementaires en matière d'intelligence artificielle (IA). Ces sandboxes fournissent des environnements de test encadrés où les technologies émergentes en IA peuvent être soumises à des essais dans un cadre garantissant la conformité réglementaire.

16 LLM (Large Language Models) pour évaluer 7 biais éthiques

Les sandboxes réglementaires en matière d'IA jouent un rôle majeur en contribuant aux discussions en cours sur la réglementation de l'IA, en particulier compte tenu de l'AI Act de l'Union européenne. Le projet d'accord souligne l'importance de développer et d'utiliser les systèmes d'IA de manière à promouvoir la diversité, l'égalité et l'équité, tout en abordant et en évitant les impacts discriminatoires et les préjugés interdits par le droit de l'Union ou le droit national.

Francesco Ferrero, directeur du département IT for Innovative Services au LIST, indique : « L'AI Act souligne l'importance du développement inclusif et de l'égalité d'accès aux technologies de l'IA tout en atténuant les impacts discriminatoires et les préjugés. Notre sandbox IA s'aligne étroitement sur ces objectifs, en fournissant une plateforme pour tester et affiner les systèmes d'IA dans un cadre centré sur la conformité. Il ne s'agit pas du sandbox réglementaire envisagé par la loi sur l'IA, qui sera mis en place par l'agence qui supervisera la mise en œuvre de la réglementation, mais c'est un premier pas dans cette direction. »

Ce leaderboard, le premier au monde à porter sur les biais sociaux, couvre 16 LLM, y compris les variations, et les évalue sur sept biais éthiques : l'âgisme, les discriminations LGBTIQ+, politiques, religieuses, le racisme, le sexisme et la xénophobie. La plateforme assure la transparence en présentant les performances de chaque modèle en fonction des différents préjugés. Cet outil peut intégrer différentes suites de tests éthiques. Actuellement, elle intègre une adaptation de LangBiTe dans le cadre d'une collaboration avec l'UOC (Universitat Oberta de Catalunya).

Jordi Cabot, Head of the Software Engineering RDI Unit au LIST, qui a dirigé l'équipe à l'origine de la sandbox, explique : « Le leaderboard a été conçu pour offrir de la transparence et faciliter l'engagement des utilisateurs. Les utilisateurs peuvent accéder à des informations détaillées sur les biais, à des exemples de tests réussis et infructueux, et même contribuer à la plateforme en suggérant de nouveaux modèles ou tests. »

Faire progresser l'équité

Sur la base des connaissances acquises lors de l'élaboration du leaderboard, le LIST souligne l'importance du contexte dans le choix des LLM et l'importance des modèles de grande taille qui présentent des biais plus faibles. Des difficultés ont été rencontrées lors des tentatives d'évaluation, notamment des divergences dans les réponses des LLM et la nécessité d'expliquer les processus d'évaluation.

Francesco Ferrero conclut : « Nous pensons que la capacité à apporter des explications est essentielle pour favoriser la confiance et faciliter le retour d'information en vue d'une amélioration continue. En tant que communauté, nous devons aborder les enjeux de manière collaborative afin de faire prendre conscience des limites inhérentes à l'IA, d'inspirer une utilisation responsable des grands modèles de langage et d'autres outils d'IA générative et, au fil du temps, de contribuer à accroître leur fiabilité. Cela est d'autant plus important que

les modèles les plus performants sont des "boîtes noires" secrètes, qui ne permettent pas à la communauté de chercheurs d'examiner leurs limites. »

Le LIST s'engage à faire progresser la recherche sur l'IA et à favoriser un environnement propice à l'équité, à la transparence et à la responsabilité dans les technologies de l'IA.

Ce travail a été partiellement financé par le Luxembourg National Research Fund (FNR) via le programme PEARL, le gouvernement espagnol et le projet TRANSACT.

Pour plus d'informations, visitez le site LIST AI Sandbox.

A propos du LIST

Le Luxembourg Institute of Science and Technology (LIST) est une Organisation de Recherche et de Technologie (RTO) dépendant du Ministère de l'Enseignement Supérieur et de la Recherche dont la mission est de développer des prototypes de produits/services compétitifs et orientés marché à destination d'acteurs publics et privés.

Avec ses plus de 700 employés, dont 77% sont des chercheurs ou experts en innovation du monde entier, le LIST est actif dans les domaines de l'informatique, des matériaux, des ressources spatiales et de l'environnement, et travaille sur l'ensemble de la chaîne de l'innovation : recherche fondamentale et appliquée, incubation et transfert de technologies.

En transformant les connaissances scientifiques en technologies, données et outils intelligents, le LIST :

- Eclaire les citoyens européens dans leurs choix
- Soutient les pouvoirs publics dans leurs décisions
- Booste les entreprises dans leur développement

Pour plus d'informations sur le Luxembourg Institute of Science and Technology, rendez-vous sur : https://www.list.lu/

CONTACT PRESSE :

LIST

Paramita Chakraborty

Communication Officer

Tel: (+352) 275 888 2237

Email: communication@list.lu