

L'initiative LINGUA de Microsoft sélectionne le luxembourgeois pour renforcer l'inclusion dans l'IA



Microsoft annonce les lauréats de son LINGUA Open Call, une initiative destinée à renforcer la présence des langues sous-représentées dans l'intelligence artificielle (IA). Parmi les projets retenus figure LuxVLD (Luxembourgish Vision-Language Dataset for Education and Digital Inclusion), mené par le SnT (Interdisciplinary Centre for Security, Reliability and Trust) de l'Université du Luxembourg.

LuxVLD développera des ensembles de données de haute qualité pour soutenir les systèmes d'IA en luxembourgeois. Ce projet souligne l'engagement du Luxembourg en matière d'IA et reflète l'ambition de Microsoft de voir toutes les langues représentées dans le monde numérique.

La diversité linguistique est menacée

Selon l'UNESCO, plus de 2 500 langues risquent de disparaître d'ici la fin du siècle. Avec l'avancée des technologies numériques, les langues disposant de contenus en ligne limités sont davantage exposées au risque d'exclusion des bénéfices de l'IA. Aujourd'hui, près de la moitié de l'ensemble des contenus web est en anglais, alors que seulement 5,1 % de la population mondiale

a l'anglais comme langue maternelle. Le luxembourgeois, parlé par environ 400 000 personnes, soit à peine 0,005 % de la population mondiale, fait face à des défis de sous-représentation numérique. Ce déséquilibre affecte la qualité des résultats produits par les systèmes d'IA.

Des recherches récentes du Microsoft AI for Good Lab montrent que les pays où les langues sous-représentées dominent, adoptent l'IA 20 % plus lentement, même lorsque la connectivité et les compétences numériques sont comparables. Sans intervention, cet écart continuera de se creuser.

Présentation de l'initiative LINGUA

Pour relever ce défi, Microsoft a lancé LINGUA, un appel à projets européen soutenant des projets qui créent des ensembles de données ouverts et de haute qualité pour les langues sous-représentées. Ces ensembles de données permettent d'entraîner des modèles d'IA multilingues et d'aider les communautés à développer des technologies linguistiques qui reflètent leur réalité culturelle et linguistique. Dans le cadre du programme LINGUA, les projets sélectionnés bénéficient d'un soutien financier pour la création d'ensembles de données, ainsi que d'un accompagnement technique grâce à la collaboration de Microsoft avec le consortium de recherche APERTUS.

Les lauréats de LINGUA couvrent 16 langues réparties dans 10 pays, représentant plus de 65 millions de locuteurs. Les projets s'étendent de l'islandais et du basque au maltais, au romani ou encore au luxembourgeois, et incluent des ensembles de données textuels, vocaux et multimodaux. LINGUA fait partie d'EU Digital Unlock et est développé en coordination avec le projet APERTUS, un grand modèle de langage (LLM) entièrement ouvert conçu par l'EPFL et l'ETH Zurich, en consultation avec le Conseil de l'Europe. L'initiative contribue ainsi à renforcer la souveraineté numérique de l'Europe et à préserver son patrimoine linguistique.

Renforcer le luxembourgeois dans l'IA

Parmi les projets sélectionnés, LuxVLD vise à améliorer la disponibilité d'ensembles de données multimodaux pour le luxembourgeois, une langue au cœur de l'identité du pays mais historiquement sous-représentée dans les environnements numériques. En soutenant le développement de ces ensembles de données, Microsoft et le SnT contribuent à renforcer la présence du luxembourgeois dans la prochaine génération de systèmes d'IA utilisés dans l'éducation, la recherche, les services publics et la communication quotidienne. LuxVLD s'inscrit dans la stratégie nationale d'IA du Luxembourg, qui souhaite positionner le pays comme une société numérique de premier plan grâce à une IA centrée sur l'humain.

« Je suis fier de voir cette initiative prendre forme. Celle-ci renforce la place du luxembourgeois dans le monde de l'IA », déclare Pierre-Adrien Grange, Country Manager de Microsoft Luxembourg. « En permettant aux systèmes d'IA de raisonner directement en luxembourgeois, LuxVLD stimulera l'innovation éducative, favorisera l'inclusion numérique et posera les bases des futures applications luxembourgeoises alimentées par l'IA ».

Le groupe de recherche CVI² de l'Université du Luxembourg, dirigé par la professeure Djamila Aouada, a obtenu une reconnaissance européenne en faisant partie des dix équipes pionnières sélectionnées pour l'initiative LINGUA de Microsoft. Cette distinction met en lumière le travail innovant de CVI² dans l'IA et l'inclusion digitale, positionnant le Luxembourg à l'avant-garde de la recherche soutenant la diversité linguistique.

« Grâce à l'initiative LINGUA, Microsoft s'engage à faire en sorte que la richesse linguistique reste au cœur de l'avenir de l'IA au Luxembourg et en Europe », conclut Pierre-Adrien Grange.

Construire un avenir de l'IA au service de toutes les langues

« Nous avons appris que l'adaptation de l'IA se fait beaucoup plus rapidement lorsque les personnes peuvent utiliser leur propre langue », explique Marijke Schroos, General Manager de Microsoft Belux. « En soutenant des projets tels que LuxVLD, nous contribuons à garantir que les locuteurs luxembourgeois puissent participer pleinement à la révolution de l'IA sans avoir à abandonner leur langue. C'est ce type de progrès technologique que nous voulons voir : une innovation qui inclut réellement tout le monde. »

En élargissant les ressources de données ouvertes pour les langues sous-représentées, LINGUA soutient un écosystème d'IA plus représentatif et plus équitable. Ces efforts contribuent au développement de modèles multilingues tels qu'APERTUS, EuroLLM et SmoLLM3, garantissant que la diversité linguistique et culturelle de l'Europe soit pleinement reflétée dans les technologies futures.

« Le projet LuxVLD pose les fondations d'un modèle vision-langage luxembourgeois capable de comprendre, raisonner et interagir en luxembourgeois à travers le texte et l'image. Cette ambition ouvre la porte à une nouvelle génération d'applications d'IA centrées sur l'humain, développées pour le Luxembourg et en luxembourgeois », déclare la professeure Djamila Aouada, directrice adjointe du SnT et responsable du groupe de recherche CVI².

« Wingos en est un exemple immédiat, qui bénéficiera d'un modèle vision-langage sensible au luxembourgeois pour soutenir une éducation inclusive et personnalisée. Le soutien de Microsoft à notre projet reconnaît son impact durable, en particulier pour les futurs étudiants et apprenants. Tous les résidents du Luxembourg pourront participer à notre campagne de collecte de données

pour le LuxVLD et contribuer à façonner l'avenir de l'IA en luxembourgeois.

Cette annonce renforce l'engagement de Microsoft à créer une IA qui bénéficie à tous, dans toutes les langues, en amplifiant les voix des communautés européennes à l'ère numérique.