

L'EPFL près de lancer un modèle de langage IA destiné au «bien commun»

TECHNOLOGIE Hier, l'EPFL a donné des précisions sur ses ambitions en matière d'intelligence artificielle et la mise à disposition imminente de son modèle de langage. Martin Jaggi, du Centre IA de la haute école, détaille le projet et ses motivations

ANOUSH SEYDTAGHIA

ChatGPT, Meta AI, DeepSeek, Claude ou encore Gemini de Google auront bien-tôt face à eux un nouveau concurrent très particulier: un modèle de langage 100% suisse, développé par l'EPFL et l'EPFZ. Mercredi, les deux institutions ont livré des détails sur l'avancement de leur projet majeur en intelligence artificielle. D'ici quelques semaines, le modèle de langage sera lancé sur le marché, offrant une alternative de choix aux internautes, développeurs, entreprises ou institutions publiques, notamment ceux qui recherchent une solution souveraine. Voici l'essentiel à savoir avec l'expertise du professeur Martin Jaggi, du Centre IA de l'EPFL.

Pourquoi concevoir un modèle de langage?

Il existe aujourd'hui des dizaines, voire des centaines de modèles de langage sur le marché, développés avant tout par les géants américains ou chinois du secteur.



«Pour l'immense majorité des langues, notre modèle sera extrêmement compétitif»

MARTIN JAGGI, PROFESSEUR À L'EPFL

Pourquoi lancer un nouveau modèle suisse? «L'immense majorité de ces modèles sont à visée commerciale, ils ont été conçus souvent en secret par les géants de la tech, on sait très peu de choses d'eux, par exemple les données sur lesquels ils ont été entraînés, répond Martin Jaggi. Tout cela génère des

risques pour les utilisateurs. Il est aussi très important à notre sens de proposer un modèle totalement ouvert, qui sera en open source et dont les poids (*weights*), soit les paramètres appris durant l'entraînement, seront publics.» Certains modèles, tels ceux de Meta, se veulent «open source», mais n'offrent en revanche pas de transparence sur les données d'entraînement.

L'idée est aussi de tendre vers une certaine universalité. Le futur modèle maîtrisera plus de 1500 langues, selon les deux EPF. «On ne sera sans doute pas aussi bon que les modèles des géants américains en anglais. Mais pour l'immense majorité des langues, notre modèle sera extrêmement compétitif», ajoute le professeur de l'EPFL. Le modèle a été entraîné sur des données à 60% en anglais, à 40% dans d'autres langues. On connaîtra ses données d'entraînement.

Au final, les deux EPF insistent sur la notion de «bien commun», avec un modèle transparent, ouvert et utilisable par tous. Aucun modèle américain, chinois ou même européen ne répond aujourd'hui à ces critères.

Comment a-t-il été créé?

Plusieurs ingrédients ont été utilisés. D'abord, l'expertise des ingénieurs des deux EPF, notamment des scientifiques

qui œuvrent dans les «AI Centers» des institutions. Ensuite, les données: les ingénieurs ne se sont basés que sur des éléments non protégés par des droits d'auteur (une différence avec les pratiques des géants de la Silicon Valley), sans que cela n'affecte la qualité du modèle, selon eux. L'entraînement a été effectué sur plus de 15 000 milliards de ce que l'on appelle des *tokens*, soit des unités représentant un mot ou une partie de mot.

Et enfin, il a la puissance de calcul. Le modèle a été entraîné sur le supercalculateur Alps du Centre suisse de calcul scientifique à Lugano, l'une des plateformes d'IA les plus avancées du monde équipée de plus de 10 000 super-puces Grace Hopper du géant américain Nvidia. «Il est certain que sans ce superordinateur, le plus puissant en Europe à ce jour, nous n'aurions pas pu développer un tel modèle de langage», assure Martin Jaggi. Alps a été inauguré à l'automne 2024 et son développement a coûté quelque 80 millions de francs.

Et ce n'est pas terminé. «Nous estimons que notre modèle est au niveau de ceux, commerciaux, disponibles sur le marché. Et nous prévoyons d'ores et déjà de le mettre à jour régulièrement pour rester dans la course», poursuit le professeur de l'EPFL.

■ A qui sera-t-il destiné?

La réponse est simple: à tout le monde, et gratuitement. «Nous imaginons que l'administration publique suisse, les tribunaux ou des entreprises helvétiques seront très intéressés par notre modèle. Mais n'importe qui pourra l'utiliser, même sans avoir de lien avec la Suisse. Il est vraiment destiné à tout le monde», affirme Martin Jaggi. Libre aussi à chacun d'analyser le modèle et de le modifier librement. On peut imaginer que des géants de la tech américains, chinois, voire européens, vont l'analyser de près.

Il sera possible de le télécharger pour le faire fonctionner localement, par exemple sur un ordinateur, comme c'est déjà le cas avec des modèles relativement légers. Et n'importe quel internaute pourra faire de même. Par contre, il n'y aura pas, dès le début, une interface de type ChatGPT pour l'utiliser. Mais des entreprises ou collectivités publiques pourront très rapidement créer ces interfaces pour le démocratiser.

Le modèle sera publié a priori à la fin de cet été – aucune date précise n'est pour l'heure communiquée. Deux tailles (8 milliards et 70 milliards de paramètres) seront mises à disposition. Selon les EPF, la version à 70 milliards de paramètres «devrait figurer parmi les modèles entièrement ouverts les plus puissants du monde». ■