

Premier aperçu de l'impact écologique d'une IA

TECHNOLOGIE L'entreprise française Mistral AI a partagé des chiffres sur le cycle de vie de son grand modèle de langage et plaide pour la création d'un standard mondial. Mais elle s'est bien gardée de partager l'étude qui lui a permis d'obtenir ces résultats

GRÉGOIRE BARBEY

Quel est le véritable impact environnemental des grands modèles de langage comme ChatGPT, Claude ou encore DeepSeek? Cette question obsède les spécialistes du sujet, tant les entreprises qui conçoivent et commercialisent ces logiciels ont érigé l'opacité en règle de conduite. Et lorsqu'elles communiquent des chiffres, elles ne partagent pas la méthode qui leur a permis d'obtenir un tel résultat.

Sam Altman, le patron d'OpenAI, écrivait par exemple en juin sur son blog qu'une requête moyenne traitée par ChatGPT consommerait environ 0,34 wattheure et l'équivalent d'un quinzième d'une cuillère à café d'eau. Qu'est-ce qu'une requête moyenne? Comment est-il parvenu à ce résultat? Autant de questions qui restent sans réponse.

Une analyse complète du cycle de vie

Mardi 22 juillet, l'entreprise française Mistral AI a jeté un pavé dans la mare, annonçant s'engager pour la création d'un standard environnemental mondial pour l'intelligence artificielle, tout en partageant les résultats d'une analyse complète du cycle de vie de son grand modèle de langage Large 2.

L'étude englobe la création du modèle et dix-huit mois d'utilisation. Elle a été réalisée en collaboration avec le cabinet de conseil Carbone 4 et avec le soutien de l'Agence française de la transition écologique. Pour «garantir sa robustesse», elle a été revue par les cabinets Resilio et Hubblo, spécialisés dans les audits environnementaux du secteur numérique.

Ainsi, l'entraînement et l'inférence – c'est-à-dire l'utilisation du modèle pour produire une réponse – représentent 85,5% de l'émission de gaz à effet de serre et 91% de la consommation d'eau sur l'ensemble du cycle de vie de Mistral Large 2. Le matériel, ce qui inclut la fabrication, le transport et la fin de vie des serveurs, totalise, quant à lui, 11% des émissions et 5% de la consommation d'eau.

Mistral AI indique que son modèle aurait donc produit durant la période sous revue 20400 tonnes de CO₂ et consommé 281000 m³ d'eau et l'équivalent de 660 kg d'antimoine en ressources matérielles – il s'agit de l'unité de référence en la matière. Sur son site, l'entreprise française fait un effort évident de pédagogie pour détailler l'importance de ces informations. Mais elle n'a pas jugé opportun de partager l'étude.

«C'est difficile de tirer quoi que ce soit de ces chiffres, car ils ne sont pas étayés»

HUGUES FERREBOEUF, CHEF DE PROJET NUMÉRIQUE AUPRÈS DE THE SHIFT PROJECT

Un choix qui interpelle Hugues Ferreboeuf, chef de projet numérique auprès du groupe de réflexion The Shift Project. «C'est difficile de tirer quoi que ce soit de ces chiffres, car ils ne sont pas étayés et il est difficile de savoir à quoi ils correspondent précisément», indique-t-il. Une critique également formulée sur LinkedIn par Sasha Luccioni, chercheuse spécialisée dans l'empreinte écologique de l'IA chez Hugging Face. Bien qu'elle salue la démarche de Mistral AI, elle dit regretter le fait que les informations communiquées fassent abstraction de la dimension énergétique.

«L'entreprise mentionne que son modèle consommerait l'équivalent de 1,14 g de CO₂ pour le traitement d'une requête, explique Hugues Ferreboeuf. Le problème, c'est que ce chiffre n'a pas la même signification selon le lieu où sont situés les serveurs.» En France, dont la production d'électricité est largement décarbonée, ce chiffre représenterait une consommation énergétique d'environ 20 Wh. «Ce serait énorme!» lâche le spécialiste.

Un premier pas positif

En revanche, si ce chiffre correspond à l'utilisation de centres de données basées aux États-Unis, la consommation énergétique serait plutôt autour de 2 Wh. «Ce serait plus conforme aux ordres de grandeur que nous avons en tête jusqu'ici», précise Hugues Ferreboeuf.

«Un wattheure en France ou aux États-Unis est toujours un wattheure, ce qui permet ensuite de tirer des conclusions sur l'impact environnemental en tenant compte des lieux d'émission», explique le spécialiste du Shift Project. Etant donné que Mistral AI a conclu des partenariats avec les entreprises américaines Microsoft et Oracle pour diffuser ses modèles, de telles précisions s'avèrent d'autant plus cruciales.

Bien que la démarche de l'entreprise française ne soit pas encore de nature à pleinement contenter les experts, son engagement en faveur de la création de standards mondiaux va dans le bon sens. «C'est essentiel d'avoir des référentiels communs pour comparer des pommes avec des pommes plutôt que des carottes avec des tomates», image Hugues Ferreboeuf. Ce d'autant plus que bon nombre d'éléments dans la chaîne de valeur de l'intelligence artificielle ne sont pas encore bien définis. «Nous savons à peu près ce que consomme un processeur graphique de Nvidia, mais c'est difficile d'extrapoler ce chiffre à l'ensemble d'un centre de données, car il s'agit d'un système très complexe», conclut le spécialiste. ■