Pour une IA européenne souveraine, explicable et sobre

Description

Si l'engouement pour les applications d'intelligence artificielle générative n'est pas près de retomber, leurs enjeux et leurs écueils se font de plus en plus visibles, notamment parce qu'elles sont entraînées très majoritairement à partir de corpus en anglais. De ces travers se dessine une stratégie IA européenne, et également française.

Rappelons tout d'abord que les grands modèles de langages (Large Language Models – LLM) sont des systèmes informatiques « *artificiellement intelligents* », incapables de penser ou de raisonner. Ces systèmes reposent sur des « *probabilités de succession de mots basées sur l'analyse du contexte de leur utilisation* », explique Laurence Devillers, professeure en informatique à l'université Paris-Sorbonne et chercheuse au Limsi (Laboratoire d'informatique pour la mécanique et les sciences de l'ingénieur) du CNRS (Centre national de la recherche scientifique) (voir *La rem* n°65, p.27). Ces systèmes s'appuient sur de volumineux corpus de textes à partir desquels ils « apprennent » statistiquement quelle probabilité ont deux mots d'être corrélés, en fonction d'un nombre de paramètres variant de quelques millions à dorénavant un trillion, avec ChatGPT-4, sorti en mars 2023.

Une IA explicable et interprétable

Pendant de nombreuses années, la recherche en matière d'intelligence artificielle (IA), d'apprentissage machine et de réseau de neurones (voir *La rem* n°30-31, p.75), a privilégié « *la performance de ces algorithmes* […] *au détriment de leur compréhension* », précise Marine Lhuillier, ingénieure Recherche et Développement en informatique. Beaucoup de ces IA sont qualifiées de « boîtes noires ». Un seul exemple pour l'illustrer : Sundar Pichai, PDG de Google, a récemment souligné, lors d'une interview à CBS, qu'une IA expérimentale « *avec très peu de requêtes en bengali, peut maintenant traduire tout le bengali* […] *alors qu'elle n'a pas été formée à le connaître* ». Personne, chez Google, ne sait comment cette IA s'y est prise.

Pour la chercheuse Ikram Chraibi Kaadoud, chef de projet européen en IA de confiance et gestion éthique à l'Inria (Institut national de recherche en informatique et en automatique) Bordeaux, « *quand un expert humain prend une décision, il peut expliquer sur quelles connaissances, à partir de quels faits, et quelles inférences il a utilisées pour arriver à sa conclusion. On parle d'explicabilité pour un système d'IA quand il peut lui aussi décrire comment a été construite sa décision. Dans certains cas, cette explication peut-être très complexe, voire impossible à appréhender par un humain ; en effet, un système de raisonnement automatisé peut enchaîner un très grand nombre de données, d'inférences qui dépassent de loin la capacité de nos cerveaux. Et c'est d'ailleurs bien pour ça que nous faisons appel à des machines qui ne sont pas intelligentes mais qui savent traiter des masses gigantesques d'informations*

». L'IA explicable (ou XAI pour eXplainable Artificial Intelligence) et l'interprétabilité des réseaux de neurones artificiels sont, depuis 2017, des domaines de recherche qui attirent les chercheurs, toujours plus nombreux.

Si l'interprétabilité permet à un expert en apprentissage machine de comprendre « comment » l'algorithme prend ses décisions, l'« explicabilité » permet à quiconque de comprendre « pourquoi » l'algorithme a pris telle ou telle décision. L'interprétabilité passe par l'analyse des données sur lesquelles l'algorithme a été entraîné ou par l'examen de la représentation interne des données, et l'explicabilité répondra à la question « pourquoi » l'algorithme a pris telle décision et pas une autre. Si un modèle est explicable, c'est qu'il est nécessairement interprétable ; en revanche, un modèle interprétable est loin d'être toujours explicable. L'Union européenne a finalisé, en décembre 2023, le règlement sur l'intelligence artificielle (AI Act), qui entrera en vigueur au plus tôt en 2025, et qui exigera des systèmes d'IA notamment de fournir tout à la fois aux utilisateurs et aux autorités compétentes une documentation détaillant leur fonctionnement ; elle imposera, en outre, à ceux utilisés dans des domaines à risques élevés, comme la santé, la justice ou la sécurité publique, d'être à la fois explicables et interprétables.

Une IA souveraine

Stable Diffusion est un modèle d'apprentissage automatique générant des images numériques photoréalistes à partir de descriptions en langage naturel. « *Le monde est dirigé par des PDG blancs. Les femmes sont rarement médecins, avocates ou juges. Les hommes à la peau foncée commettent des crimes, tandis que les femmes à la peau foncée font cuire des hamburgers.* » C'est la conclusion d'une enquête menée par Bloomberg en 2023, destinée à évaluer l'ampleur des biais dans l'IA générative, et qui a demandé à Stable Diffusion « *de créer des représentations de travailleurs pour quatorze emplois – 300 images pour chacun des sept emplois généralement considérés comme « bien rémunérés » aux États-Unis et des sept autres considérés comme « mal rémunérés » – ainsi que trois catégories liées à la criminalité* ». Sasha Luccioni, chercheuse au sein de la start-up franco-américaine Hugging Face qui développe une plateforme ouverte de traitement automatique des langues, et coauteur de l'étude sur les préjugés dans les modèles d'IA générative texte-image menée par Bloomberg en 2023, admet que ces IA « *projettent essentiellement une vision unique du monde, au lieu de représenter divers types de cultures ou d'identités visuelles* ». Ces biais proviennent tout à la fois des médias et des langues à partir desquels ces IA sont préentraînées. Le projet de science ouverte et participative Bloom (pour BigScience Large Open-science Open-access Multilingual Language Model), piloté par Hugging Face, a ainsi été entraîné sur un corpus de textes en quarante-six langues, du français au basque en passant par le mandarin et vingt langues africaines (voir *La rem* n°65, p.27).

C'est également l'objet d'un projet « *de base de données expérimentale destinée à combattre les biais culturels des IA majoritairement anglo-saxonnes* » mené en France par le ministère de la culture, avec notamment la direction interministérielle du numérique (Dinum), l'Institut national de l'audiovisuel (Ina), la Bibliothèque nationale de France (BnF), l'Inria et le CNRS. Nommée Villers-Cotterêts, cette start-up d'État incubée par la Dinum va rassembler en un même endroit de grands corpus de textes dans la langue de Molière et les valoriser auprès des entreprises développant des grands modèles de langages, et dont certaines

avaient d'ailleurs déjà sollicité la BnF ou l'Ina en ce sens. Une initiative intéressante pour tenter de dépasser les problématiques de contenus protégés par le droit d'auteur ou les documents comportant des données personnelles, qui empêchent parfois d'utiliser des textes en français pour entraîner ces grands modèles de langage. Une ambition partagée avec le projet PIAF (Pour une IA francophone) dont l'objet est « *de construire ce(s) jeu(x) de données francophones pour l'IA de manière ouverte et contributive* », porté par Etalab, un département de la Dinum, et qui bénéficie du financement du Programme d'investissements d'avenir, piloté par le secrétariat général pour l'investissement et la Caisse des dépôts.

« *La France dispose de données extrêmement riches issues des institutions culturelles mais toutes ne sont pas publiques* », expliquait Romain Delassus, le chef du service numérique du ministère de la culture, lors d'une réunion de présentation du projet. Et d'ajouter que « *un opérateur national pourrait négocier au nom des institutions patrimoniales* », par exemple sur les questions de droit d'auteur.

En 2018, Google Research a développé et publié en open source BERT (Bidirectional Encoder Representations from Transformers), un modèle de traitement automatique du langage naturel. BERT a, pour la première fois, introduit le traitement de la polysémie des mots dans un grand modèle de langage, ce qui a considérablement amélioré les performances des IA génératives, nécessitant toutefois d'être adapté à chaque langue. Ce fut chose faite avec le français, d'abord en 2019 par l'équipe ALMAnaCH (Automatic Language Modelling and Analysis & Computational Humanities) de l'Inria avec le modèle CamemBERT, préentraîné sur un corpus de 138 gigaoctets de texte, puis en 2020 par des chercheurs du laboratoire d'informatique de Grenoble avec le modèle FlauBERT, préentraîné sur un corpus de 71 gigaoctets de texte, dont l'intégralité de Wikipédia en français, plusieurs années du journal *Le Monde*, des ouvrages francophones du projet Gutenberg ou encore des transcriptions des débats du Parlement européen. La « *guerre du contenu sémantique* », pour reprendre la formule de Laurence Devillers, est avant tout culturelle. Depuis sa publication en 2019, CamemBERT est devenu le modèle de traitement automatique de langage naturel le plus utilisé par les entreprises françaises, téléchargé plus de 22 millions de fois. Ainsi, Enedis s'en sert pour catégoriser automatiquement les milliers de messages quotidiens de ses clients, dont le tri était auparavant manuel, ce qui lui permettrait d'économiser 3 millions d'euros par an et d'apporter sans doute plus rapidement des réponses.

Toutefois, la souveraineté numérique ne concerne pas uniquement la source des données et leur langue, mais également les outils. Depuis octobre 2023, sous l'impulsion de Stanislas Guerini, ministre de la transformation et de la fonction publiques, environ mille agents volontaires – notamment de la Caisse nationale de l'Assurance Maladie (Cnam), de la Caisse nationale d'assurance vieillesse (Cnav), de la Mutualité sociale agricole (MSA), de l'Agence nationale des titres sécurisés (ANTS), de la Gendarmerie, ou encore de certaines préfectures et tribunaux – testent trois IA génératives du marché (ChatGPT d'OpenAI, LLaMA de Meta et Bloom de Hugging Face) pour les aider à rédiger des réponses aux commentaires des usagers déposés sur la plateforme Services Publics +, dans la rubrique « Je donne mon avis ». Après deux mois d'expérimentation, le temps de réponse moyen est passé de sept à trois jours et le taux de satisfaction des usagers a gagné dix points. Si bien que la Dinum développe Albert, une IA générative « *souveraine, libre et ouverte, créée par et pour des agents publics* », dont le déploiement est prévu dans les prochains

mois, toujours auprès d'agents volontaires, au sein de France services, un bouquet d'accès aux services publics à destination des Français. Une manière de se passer complètement des systèmes IA américains.

Une IA sobre

Une autre question surgit quant à l'impact environnemental et à la consommation énergétique de ces IA génératives. À l'occasion de la publication de son rapport annuel, Microsoft a indiqué que sa consommation mondiale d'eau avait crû de 34 % entre 2021 et 2022. Selon Shaolei Ren, chercheur à l'université de Californie, Riverside, il fait peu de doute que « *la majorité de la croissance* [de la consommation d'eau] *est due à l'IA* ». Même constat du côté de Google, qui a fait état d'une augmentation de 20 % de sa consommation d'eau au cours de la même période. En étudiant la géographie de la consommation d'eau de ces entreprises, il est d'ailleurs possible d'en déduire où se situent leurs installations informatiques dédiées à l'IA. En revanche, lorsque l'on interroge Bard, l'IA générative de Google, à propos de l'emplacement exact de ses serveurs ou encore de l'endroit où l'IA a été préentraînée, l'outil répond que « *la réponse n'est pas publique* ». « *C'est ça le pire : on n'a aucune information, alors même qu'on assiste à une course des entreprises pour déployer des IA génératives partout, sans se poser la question de l'impact* », déplore Sasha Luccioni de Hugging Face.

L'empreinte hydrique et l'empreinte carbone des IA génératives est, au mieux, un sujet occulté et, au pire, un secret bien gardé. Pourtant, « *plus de 99 % des cas d'usage peuvent être couverts par des modèles plus petits, moins coûteux et plus spécialisés* », explique Clément Delangue, le président d'Hugging Face. Ce qui fait dire à Jean-Baptiste Bouzige, président d'Ekimetrics, spécialiste européen en science des données : « *En sortant du fétichisme des grands modèles, les Européens peuvent anticiper l'avenir et se doter d'un avantage compétitif important. De plus, penser en termes de cas d'usage, c'est apporter une réponse à une des grandes préoccupations des investisseurs. À savoir la rentabilité de l'IA générative.* » Un argument dont la convergence des intérêts, environnementaux et financiers, pourrait satisfaire tout le monde.

Sources :

- Allauzen Alexandre, Besacier Laurent, Schwab Didier, « FlauBERT à la rescousse du traitement automatique du français », CNRS Sciences informatiques, ins2i.cnrs.fr, 16 janvier 2020.
- Enard Marie-Agnès, Guitton Pascal, Viéville Thierry, « IA explicable, IA interprétable : voyage dans les archives Binaires », lemonde.fr/blog/binaire, 6 novembre 2022.
- Li Pengfei, Yang Jianyi, Islam Mohammad A., Ren Shaolei, « Making AI Less "Thirsty": Uncovering and Addressing the Secret Water Footprint of AI Models, ArXiv, october 29, 2023, arxiv.org/abs/2304.03271
- Stokel-Walker Chris, « The Generative AI Race Has a Dirty Secret », wired.co.uk, February 10, 2023.
- Nicoletti Leonardo, Bass Dina, « Humans are biased. Generative AI is even worse », bloomberg.com, June 14, 2023.
- O'Brien Matt, Fingerhut Hannah, « Artificial intelligence technology behind ChatGPT was built in Iowa – with a lot of water », apnews.com, September 9, 2023.
- Manenti Boris, « Pourquoi ChatGPT est une bombe environnementale », nouvelobs.com, 18

septembre 2023.
- Bouzige Jean-Baptiste, « Opinion | IA : pour une Europe souveraine et sobre », lesechos.fr, 26 octobre 2023.
- Madelaine Nicolas, « Une initiative française pour veiller aux biais antidémocratiques de l'IA », lesechos.fr, 13 novembre 2023.
- « L'Union européenne trouve un accord pour encadrer le développement de l'intelligence artificielle », *Le Monde* avec AFP, lemonde.fr, 9 décembre 2023.
- Dèbes Florian, « Intelligence artificielle : la France se lance dans la bataille culturelle des données », lesechos.fr, 12 décembre 2023.
- Biseul Xavier, « IA dans la fonction publique : l'État ouvre son CamemBERT », zdnet.fr, 14 décembre 2023.
- Microsoft, « 2022 Environmental Sustainability Report », microsoft.com, consulté le 12 février 2024.

Categorie

1. Techniques

date créée
17 avril 2024
Auteur
jacquesandrefines