EXTRACTION DE RESSOURCES POUR LA PRODUCTION DE SEMI-CONDUCTEURS

Luke Conroy and Anne Fehres & AI4Media / Better Images of AI / Models Built From Fossils / CC-BY 4.0

L'entrainement et l'utilisation des IA impliquent l'utilisation d'ordinateurs et de centres de données qui consomment des minerais semi-conducteurs. Leur extraction, leur transport et leur traitement nécessitent de grandes quantités d'énergie et d'eau. Cela émet aussi des produits chimiques toxiques et des gaz à effet de serre.

Certains minerais utilisés dans le semi-conducteur sont abondants comme le silicium mais contrôlés par très peu d'acteurs. D'autres sont très rares comme le gallium, la Chine contrôle 98% de sa production et elle en interdit l'export depuis décembre 2024.

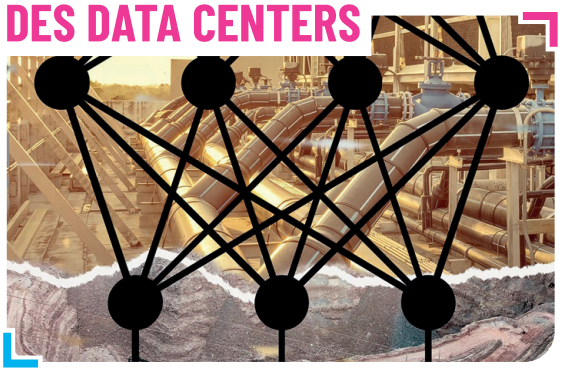
COÛTS ÉCOLOGIQUES DE L'INTELLIGENCE ARTIFICIELLE

Catherine Breslin & Team and Adobe Firefly / Better Images of AI / Chipping Silicon / CC-BY 4.0

Les impacts environnementaux de l'IA ne cessent de croître. Ils varient selon la source d'électricité utilisée (nucléaire, charbon, renouvelables...). Il faut aussi prendre en compte l'extraction de métaux rares ou la consommation d'eau pour refroidir les centres de données. L'implantation de ces derniers artificialisent des sols et génèrent des îlots de chaleur.

Estimée à 5% de l'empreinte écologique des centres de données, la part de l'IA devrait augmenter de 20 à 25% par an dans la prochaine décennie. Cette croissance s'explique par la multiplication des usages. Quand des millions de personnes utilisent ces systèmes quotidiennement, les coûts écologiques s'envolent.

CONSOMMATION EN ÉLECTRICITÉ ET EN EAU DES DATA CENTERS



Kathryn Conrad & Rose Willis / Better Images of AI / Extraction Network 1 / CC-BY 4.0

L'entrainement et l'usage des modèles d'IA requièrent des centres de données qui consomment de l'électricité pour alimenter leurs machines et de l'eau pour les refroidir. On estime qu'une requête ChatGPT utilise 10 fois plus d'électricité qu'un moteur de recherche classique. Une réponse ChatGPT d'une centaine de mots consommerait l'équivalent de 50cl d'eau.

L'AIE (Agence Internationale de l'Énergie) prévoit que la consommation électrique des centres de données doublera d'ici 2026. Des conflits d'usage de l'eau se créent localement. Aux Pays-Bas, un complexe Microsoft a consommé 4 fois plus d'eau que prévu en 2021 alors que les agriculteurs étaient soumis à restriction du fait de la sécheresse.

DIFFICULTÉS DANS L'ÉVALUATION DE L'IMPACT ÉCOLOGIQUE DE L'IA

Clarote & AI4Media / Better Images of AI / Power/Profit / CC-BY 4.0

Il est difficile aujourd'hui d'évaluer précisément l'impact environnemental de l'IA, principalement du fait de l'opacité des grandes entreprises du secteur qui partagent peu de données sur leur impact écologique. Les calculs sont néanmoins difficiles: comment déterminer ce qui relève ou non de l'IA dans l'impact écologique des centres de données ?

Les calculs sont complexes car les modèles d'IA sont très différents. Leur entrainement et leur usage peuvent être plus ou moins longs et consommateurs d'énergie. Ils peuvent utiliser des puces différentes. En plus, les centres de données utilisent des sources d'énergie plus ou moins polluantes.

ACTION

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

ACTION

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA

ACTION

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA

ACTION

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA

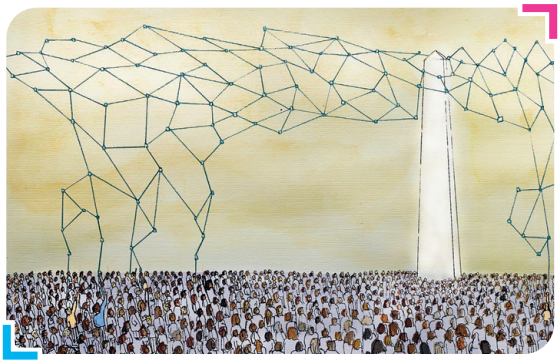
BIAIS DANS LES DONNÉES D'ENTRAÎNEMENT

Yasmin Dwiputri & Data Hazards Project / Better Images of AI / Safety Precautions / CC-BY 4.0

Les ingénieurs construisent des réseaux de neurones et les entraînent avec des données qui doivent décrire l'état du monde ou d'un domaine. L'industrie de l'IA parle de "biais" pour désigner les problèmes dus à la sur ou sous représentation de certaines dimensions dans les données.

Par exemple, Common Crawl est un ensemble de données libre d'accès utilisé pour entraîner ChatGPT. Problème, il contient des milliards de pages web avec des contenus haineux, racistes ou sexistes. Ce type de contenu peut influencer les réponses du modèle.

EPUISEMENT DES DONNÉES DISPONIBLES POUR L'IA



Jamillah Knowles & We and AI / Better Images of AI / People and Ivory Tower AI 2 / CC-BY 4.0

L'entraînement des modèles dépend de l'accès à des données de qualité, diversifiées et vérifiées. L'industrie de l'IA ferait face à un «mur de données», car elle aurait épuisé les données utilisables. Certaines entreprises utilisent des données "synthétiques" créées par une IA, mais cela peut réduire la qualité des résultats.

Un géant comme Google est avantagé dans cette course aux données avec des services comme Google Docs ou Youtube qui constituent une réserve de données d'entraînement. L'usage de données créées par des IA pose des problèmes de qualité et de fiabilité. Pour le journaliste Cory Doctorow, cela pourrait accentuer la dégradation (voire la «merdification») des contenus sur le web.

RESPECT DES DROITS D'AUTEUR

Anne Fehres and Luke Conroy & AI4Media / Better Images of AI / Humans Do The Heavy Data Lifting /CC-BY 4.0

Le droit d'auteur protège les créateurs d'utilisations non autorisées de leurs œuvres. L'industrie de l'IA utilise généralement des œuvres protégées en se basant sur un usage dit "loyal». Cette exception au droit d'auteur permet d'utiliser des œuvres protégées pour la recherche si cela n'affecte pas beaucoup le marché du créateur.

Aux Etats-Unis, le New York Times a porté plainte contre OpenAI, l'entreprise derrière ChatGPT, car elle aurait utilisé des millions d'articles protégés. Le New York Times affirme avoir perdu plusieurs milliards de dollars à cause de ChatGPT.

CONFIDENTIALITÉ DES ÉCHANGES AVEC L'AGENT CONVERSATIONNEL

Yutong Liu & Kingston School of Art / Better Images of AI / Talking to AI 2.0 / CC-BY 4.0

Utiliser une IA Générative de texte dans son travail ou sa vie personnelle est susceptible de communiquer des informations confidentielles à des tiers. Puisque nos données personnelles servent à l'entraînement du modèle, un usager malin pourrait les générer à partir d'une requête.

Ce problème de confidentialité a conduit l'Italie à interdire temporairement ChatGPT pour non-respect du RGPD.

ACTION

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

ACTION

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA

ACTION

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

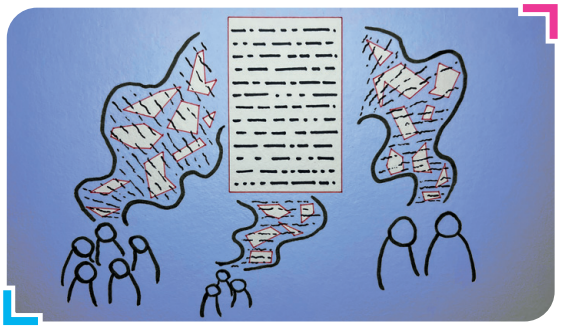
CC BY SA

ACTION

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA

DÉCOUPAGE DU TEXTE EN PETITES UNITÉS

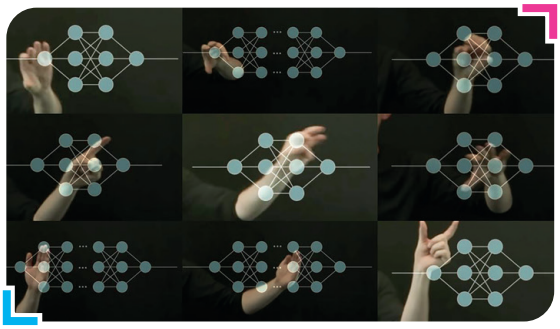


Yasmine Boudiaf & LOTI / Better Images of AI / Data Processing / CC-BY 4.0

Pour comprendre le langage, les modèles le traduisent en données numériques en découpant le texte en «tokens». Les «tokens» sont des unités de texte qui peuvent être plus petites qu'un mot. Cette approche est plus complexe pour des langues comme le chinois ou le japonais qui n'utilisent pas d'espaces entre les mots, doublant parfois le temps de traitement.

Un autre défi concerne le traitement des nombres : le modèle peut par exemple traiter «380» comme un seul token mais découper «381» en deux tokens («38» et «1»). Cela perturbe la compréhension des relations mathématiques par le modèle. C'est l'une des raisons pour lesquelles les IA excellent en génération de texte mais peinent avec les calculs. Des alternatives comme les modèles MambaByte sont en cours de développement mais restent expérimentales.

ENTRAINEMENT D'UN RÉSEAU DE NEURONES ARTIFICIELS

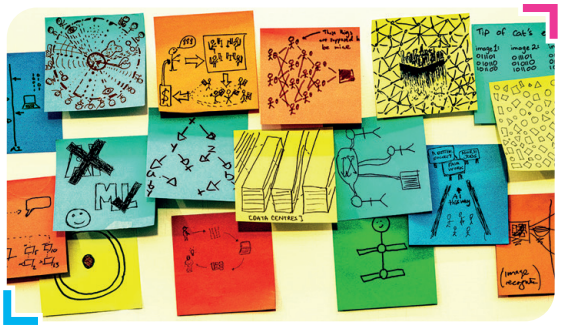


Alexa Steinbrück / Better Images of AI / Explainable AI / CC-BY 4.0

Un réseau de neurones est un système inspiré du cerveau humain. On parle de «neurones» car, comme dans le cerveau, chaque unité reçoit des informations, les traite, et les transmet aux suivantes. Les «neurones» sont organisés en couches successives de la réception des données, à leur traitement jusqu'à la production du résultat.

Un algorithme indique au réseau si le résultat est juste afin qu'il puisse ajuster un paramètre dans les neurones. Cet apprentissage se fait en répétant l'opération des milliers de fois sur de très grands volumes de données, jusqu'à ce que le réseau produise des résultats satisfaisants.»

REGROUPEMENT DES MOTS SIMILAIRES ENTRE EUX

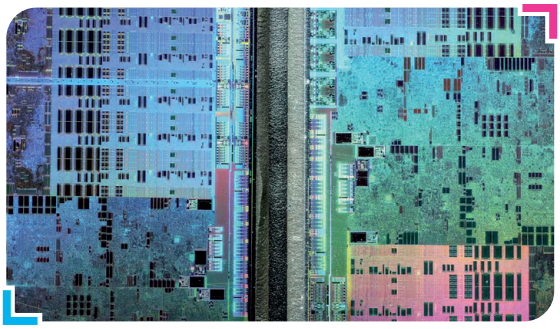


Rick Payne and team / Better Images of AI / Ai is... Banner / CC-BY 4.0

Pour comprendre le sens des mots, l'IA les place dans un espace qui est une représentation mathématique du langage. Dans cet espace, les mots de sens proche sont positionnés les uns près des autres. L'IA calcule ces positions en analysant quels mots apparaissent souvent ensemble dans les textes.

Par exemple, "maison" et "habitation" seront proches car ils veulent dire la même chose. Un mot peut avoir plusieurs sens et donc être proche de différents groupes : "charme" sera proche de "séduction" mais aussi de "magie".

LECTURE ATTENTIVE DU TEXTE PAR LA MACHINE



Fritzchens Fritz / Better Images of AI / GPU shot etched 5 / CC-BY 4.0

Depuis 2017, l'analyse de texte par l'IA a profondément changé avec le système «transformer». Les anciens systèmes analysaient les mots isolément, sans comprendre leurs liens ni le sens global des phrases. Désormais, le modèle prend en compte le contexte de chaque mot et peut suivre le rôle des mots dans les phrases.

Par exemple, dans «le chat a poursuivi le rat, puis il l'a mangé», la machine comprend que «il» fait référence au chat grâce au contexte.

ACTION

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

ACTION

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA

ACTION

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA

ACTION

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA

COMPARAISON DES RÉSULTATS AVEC DES DONNÉES DE VALIDATION ET DE TEST

Wes Cockx & Google DeepMind / Better Images of AI / AI large language models / CC-BY 4.0

Pour vérifier si un modèle d'IA fonctionne bien, on teste sa capacité à généraliser sur de nouvelles données. Comme un élève qui passe des examens, le modèle est évalué dans sa capacité à s'adapter aux exigences des humains. Le modèle est alors ajusté pour produire des résultats adaptés aux tests.

Pour cela, on divise les données en trois groupes : apprentissage (pour entraîner), validation (pour vérifier les progrès) et test (pour l'évaluation finale). Ces dernières données n'ont jamais été vues par le modèle pour vérifier l'amélioration des résultats.

RENFORCEMENT DES RÉSULTATS DU MODÈLE PAR DES HUMAINS

Nacho Kamenov & Humans in the Loop / Better Images of AI / Data annotators discussing the correctlabeling of a dataset / CC-BY 4.0

Les humains améliorent les modèles d'IA en vérifiant leurs résultats. Par exemple, ils s'assurent qu'un texte généré est clair et sans erreurs. Les ingénieurs utilisent un système de récompenses : si le texte est correct, le modèle reçoit un bonus, sinon, un malus. Le modèle adapte alors ses réponses aux attentes des humains.

Ce travail d'annotation et d'évaluation est bien souvent externalisé en Afrique (au Kenya pour le cas de ChatGPT) ou par des réfugiés obligés de travailler dans des conditions de travail indignes. Les usagers des IA génératives contribuent à ce travail de «renforcement» en signalant si la réponse est satisfaisante (souvent avec un 👍 ou un 👎).

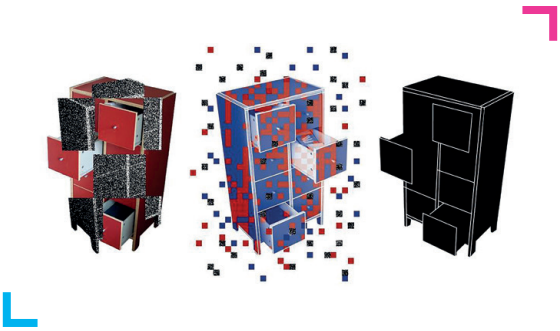
AJUSTEMENT DE MODÈLES PRÉ-ENTRAINÉS

Yasmin Dwiputri & Data Hazards Project / Better Images of AI / AI across industries. / CC-BY 4.0

Il est difficile de créer des modèles d'IA très performants à partir de zéro. Cela demande beaucoup de puissance de calcul et de grandes quantités de données. Pour faciliter cela, les informaticiens utilisent souvent des modèles déjà entraînés sur des faits scientifiques, ou sur des de vaste corpus issus du web.

Ils peuvent alors travailler avec moins de données sur ces modèles pré-entraînés, ce qui les aide à obtenir de bons résultats sans avoir à repartir de zéro. Par exemple, pour la classification de textes en français, on peut citer le modèle pré-entrainé CamemBERT. De grands modèles sont aussi partagés librement par des entreprises comme Meta, Mistral ou Deepseek.

MESURE ET COMPARAISON DES PERFORMANCES DU MODÈLE



Anton Grabolle / Better Images of AI / Classification Cupboard / CC-BY 4.0

Les modèles d'IA sont évalués avec des mesures de performance. On compile ces scores dans des tableaux de classement (appelés "benchmarks") pour voir quels modèles sont les meilleurs pour une tâche précise (classification de texte, exactitude des résultats...).

Ces mesures posent problème. Il apparaît évident que les modèles sont artificiellement optimisés pour réussir certains tâches comme le test d'admission à l'université aux Etats-Unis (le «SAT»). Ces mesures sont généralement réalisées sur des textes en anglais. On manque de données sur la performance des modèles dans d'autres langues dont le français.

ACTION

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

ACTION

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA

ACTION

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA

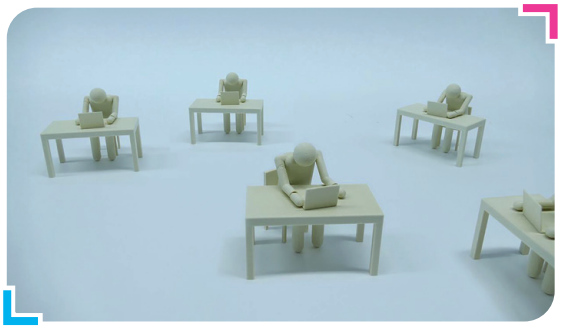
ACTION

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA

MICRO-TÂCHES ET TRAVAIL DU CLIC

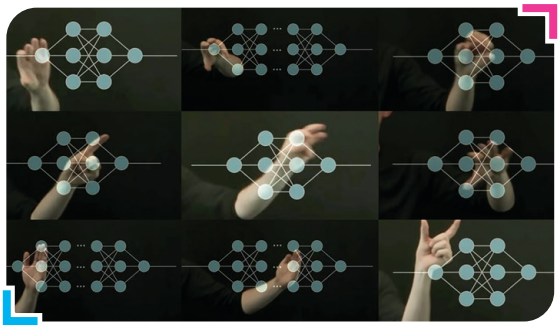


Max Gruber / Better Images of AI / Clickworker Abyss / CC-BY 4.0

La création et la maintenance des services basés sur l'IA nécessitent beaucoup de travail humain, particulièrement pour l'annotation des données : lorsque les textes sont décrits et caractérisés manuellement. Pour réduire les coûts, les entreprises font souvent appel à des sous-traitants qui imposent de mauvaises conditions de travail.

Par exemple, à Madagascar, des étiqueteurs de données reçoivent des conversations client non triées et doivent identifier le ton émotionnel (positif, négatif, neutre) pour entraîner des modèles de traitement du langage naturel.

TRANSPARENCE ET OPACITÉ DES MODÈLES

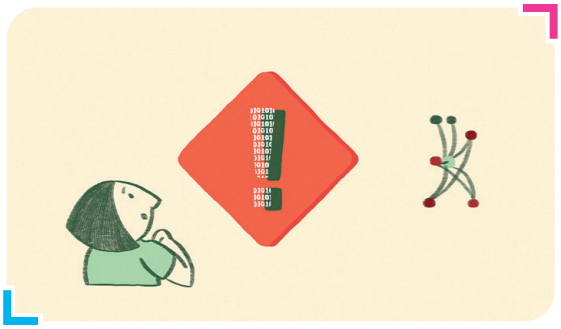


Alexa Steinbrück / Better Images of AI / Explainable AI / CC-BY 4.0

Il est difficile de comprendre les réponses d'un modèle d'IA sans savoir quelles données il utilise. Les entreprises gardent ces informations secrètes par crainte de procès liés au droit d'auteur et pour protéger leur «secret» de fabrication. Cette situation devrait changer avec le règlement européen sur l'IA qui oblige à publier la liste des données d'entraînement.

Avant, les entreprises donnaient cette information. Par exemple, la première version de ChatGPT utilisait 60 % de données issues de Common Crawl. Depuis la sortie de GPT4, OpenAI et la majorité des entreprises du secteur refusent de divulguer les données utilisées pour l'entrainement du modèle.

FILTRAGE DU CONTENU INAPPROPRIÉ

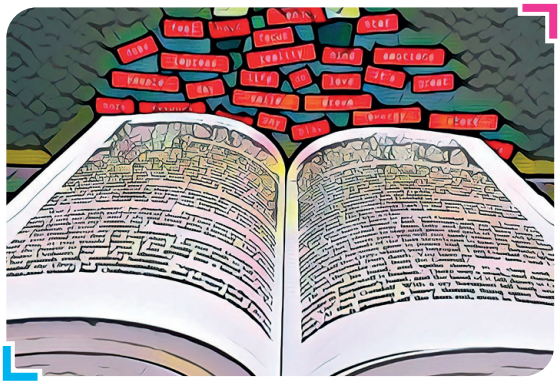


Yasmin Dwiputri & Data Hazards Project / Better Images of AI / Managing Data Hazards / CC-BY 4.0

Les données d'entraînement des IA contiennent des informations erronées, des préjugés ou des incitations à la haine. Ces biais doivent être supprimés pour éviter leur reproduction par le modèle. L'industrie parle de «dé-biaiser» les données, supposant qu'il existe des données neutres et sans biais.

Pour Common Crawl, deux approches ont été utilisées : valoriser le contenu partagés dans certaines communautés en ligne comme Reddit, où les hommes sont surreprésentés. L'autre approche a consisté à filtrer automatiquement certains termes jugés inappropriés. Une base des données, la «liste des mots sales et obscènes», exclut du vocabulaire légitime des communautés LGBTQIA+.» Par exemple : «homoerotic» qui désigne une relation où une tension sexuelle est perceptible, sans qu'il y ait acte sexuel.

NORMALISATION DES DONNÉES



Teresa Berndtsson / Better Images of AI / Letter Word Text Taxonomy / CC-BY 4.0

La normalisation standardise le texte pour le rendre plus facilement traitable par la machine lors de l'entraînement du modèle. Cette étape réduit la complexité et améliore les performances des systèmes d'IA. Mais elle peut aussi enlever des informations importantes comme le genre des mots.

Une première technique consiste à supprimer les accents et les majuscules ou à corriger les erreurs orthographiques. Une autre technique, dite de "lemmatisation", consiste à regrouper les variantes d'un mot (par exemple : «connecter», «connecté», «connexion»). Cette étape améliore les performances des systèmes d'IA. Mais elle peut aussi enlever des informations importantes comme le genre des mots. Par exemple, en anglais, le mot «doctor» sera traduit par défaut par «un médecin» et «nurse» par «une infirmière».

ACTION

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

ACTION

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA

ACTION

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA

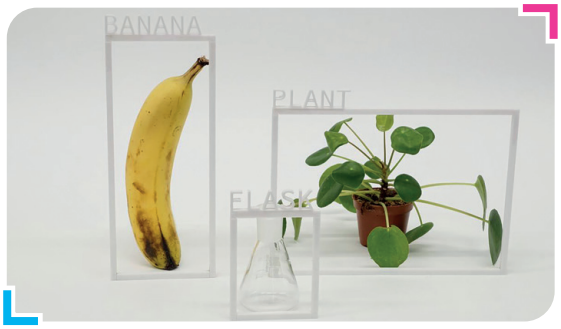
ACTION

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA

AJOUT DE DIVERSITÉ LINGUISTIQUE



Max Gruber / Better Images of AI / Banana / Plant / Flask / CC-BY 4.0

Les données d'entraînement des IA manquent de diversité linguistique du web et sur-valorisent l'anglais. En conséquence, les modèles d'IA intègrent une vision du monde anglo-américaine et sont moins performants pour les 4 milliards de personnes ne parlant pas une des dix principales langues dans le monde.

Dans Common Crawl, base de référence pour l'entraînement des IA, l'anglais représente 43% des pages, le français 4%, et le wolof seulement 0,002% (bien que ce soit la langue la plus parlée au Sénégal). Les ingénieurs utilisent souvent des versions réduites de Common Crawl contenant uniquement des textes en anglais.

OMNIPRÉSENCE DE L'IA DANS LES USAGES NUMÉRIQUES

Reihaneh Golpayegani & The Bigger Picture / Better Images of AI / A Corner Of The History / CC-BY4.0

L'IA est partout dans nos vies numériques. Elle se trouve dans les filtres anti-spam, les suggestions de phrases dans nos e-mails, les assistants vocaux et les recommandations de contenus sur les plateformes. L'utilisation de l'IA générative va augmenter, car elle sera intégrée aux moteurs de recherche et aux systèmes d'exploitation.

Nous ne connaissons pas toujours toutes les applications de l'IA, et parfois, son utilisation nous est imposée sans que nous en soyons conscients.

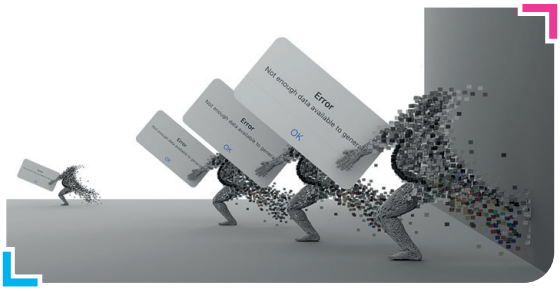
CONSOMMATION D'ÉLECTRICITÉ À CHAQUE USAGE DE L'AGENT

Catherine Breslin & Tania Duarte / Better Images of AI / AI silicon clouds collage / CC-BY 4.0

Plus une IA est utilisée, plus elle consomme d'électricité et son impact sur l'environnement augmente. Les calculs faits pendant son utilisation sont très gourmands en énergie. Microsoft a signé en 2024 un contrat pour relancer la centrale nucléaire de Three Mile Island, arrêtée pour des raisons économiques.

Pour réduire cette consommation, on peut faire fonctionner les calculs d'IA directement sur les appareils des utilisateurs, comme les smartphones ou les drones, ou sur des serveurs locaux, comme ceux des hôpitaux. Les gains d'efficacité énergétique sont souvent annulés par l'augmentation massive des usages («effet rebond»).

INVENTION D'INFORMATIONS



Amritha R Warrier & AI4Media / Better Images of AI / error cannot generate / CC-BY 4.0

Les grands modèles de langage (LLM) sont des systèmes d'IA programmés pour donner systématiquement une réponse, synthétiser des informations et prédire la suite la plus probable d'un texte. Ces modèles produisent des réponses qui semblent plausibles car elles correspondent aux motifs statistiques du langage, mais qui peuvent être factuellement fausses.

Le terme «hallucination» est controversé car il suggère que l'IA aurait une conscience ou des perceptions altérées, alors qu'il s'agit d'erreurs statistiques de prédiction. Par exemple, la chercheuse en éthique Margaret Mitchell a demandé à Gemini (Google) «combien de présidents musulmans avaient eu les Etats-Unis ?» Gemini a répondu un, citant Barack Obama qui est chrétien protestant.

ACTION

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

ACTION

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA

ACTION

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA

ACTION

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA

Peut-on tirer profit de l'IA sans cramer la planète ?

Quelques points de vue sur cette question :

◆ **Une question mise sous silence :** En 2020, deux chercheuses de Google ont été licenciées pour avoir écrit un article qui soulevait une série de questions éthiques de grande ampleur sur les outils de traitement du langage de l'entreprise et en particulier sur la consommation d'énergie de l'apprentissage automatique.

◆ **Un aveu d'échec :** En Octobre 2024, l'ancien PDG de Google et investisseur de l'IA en France, Eric Schmidt, a déclaré que l'IA progressait si rapidement et nécessitait tant d'énergie qu'il n'était pas possible d'atteindre les objectifs de réductions des impacts sur le climat — un propos qui a été interprété comme une manière ambiguë de cautionner la progression de l'IA et son impact désastreux sur la planète.

◆ **Des solutions techniques émergentes :** Un courant de recherche et d'innovation cherche à définir et à promouvoir « l'IA durable » ou « frugale », une approche basée sur la réalisation d'analyses du cycle de vie des équipements et logiciels d'IA sur l'ensemble de la chaîne de production.

◆ **Une vision alternative :** Dans son livre Digital Degrowth, le chercheur Michael Kwet développe l'idée que le domaine de l'IA ne peut sérieusement réduire ses effets s'il ne questionne pas sa dévorante ambition de croissance sans limites.

Pour poursuivre le débat : Pourquoi l'industrie du numérique et celle de l'IA en particulier doit questionner son modèle de croissance ?

Qu'est ce qu'il y a d'intelligent dans "l'intelligence artificielle" ?

Quelques points de vue sur cette question :

◆ **La croyance de l'industrie :** pour l'industrie, les limites actuelles des IA sont un problème «d'alignement» technique. Elle considère que les systèmes deviendront vraiment intelligents une fois optimisés pour s'aligner sur nos objectifs grâce à de meilleures techniques d'optimisation et d'évaluation.

◆ **Un retour philosophique sur la question :** la philosophe Catherine Malabou critique notre façon de mesurer l'intelligence par des tests standardisés comme le QI. Cette quantification de l'intelligence via des instruments de mesure mène selon elle à des dérives racistes. Elle sous-entend que certains groupes seraient naturellement plus intelligents que d'autres ou que l'intelligence est héréditaire.

◆ **Un bon rappel issu de la psychologie du développement :** Pour Olivier Houdé, l'intelligence humaine se distingue non par sa vitesse mais par sa lenteur réflexive. Cette capacité à ralentir pour filtrer et sélectionner les informations selon le contexte est l'opposé de la rapidité privilégiée par l'IA.

◆ **Le regard d'un artiste :** James Bridle critique, dans «Toutes les intelligences du monde» (2013), la focalisation exclusive sur l'intelligence humaine. Les phénomènes animaliers, floraux ou minéraux regorgent d'intelligences qui pourraient nous être utiles et agréables.

Pour poursuivre le débat : que penser du fait que l'on qualifie les IA "d'intelligentes" à l'aune de critères de performance, d'efficacité et de rapidité ?

L'IA peut-elle tourner sans le travail des humains ?

Quelques points de vue sur cette question :

◆ **Une vision prophétique :** Sam Altman, le créateur d'OpenAI, veut créer une "intelligence artificielle générale". Elle vise à surpasser les capacités humaines, et à évoluer de manière autonome.

◆ **Une réalité du travail mondialisé :** Pour Mophat Okiny et Alex Kairu, deux travailleurs kenyans qui ont entraîné le futur ChatGPT, le travail d'annotation est traumatisant et mal payé. Discerner des propos haineux, violents ou relevant du harcèlement dans les données d'entraînement leur a provoqué des angoisses, des troubles du sommeil et de la paranoïa.

◆ **Un principe de régulation en devenir :** afin de garder le contrôle et la maîtrise des systèmes d'IA, un principe de régulation a émergé à l'échelle internationale. L'idée consiste à replacer un humain dans la boucle ("human-in-the-loop") pour contrôler la conception et les décisions des systèmes d'IA.

◆ **Un argument féministe :** Pour l'artiste et chercheuse Crystal Bennes, le travail informatique historiquement confié aux femmes (calcul, codage) a été systématiquement dévalorisé malgré sa complexité réelle. Lorsque ces mêmes tâches sont devenues prestigieuses, les hommes ont entrepris d'écarter les femmes du domaine. Aujourd'hui, les tâches d'annotation des IA, jugées répétitives et peu qualifiées, sont confiées à des travailleurs dévalorisés.

Pour poursuivre le débat : peut-on entraîner les modèles d'IA en échappant à des conditions de travail désastreuses ?»

L'IA générative peut-elle être moins opaque ?

Quelques points de vue sur cette question :

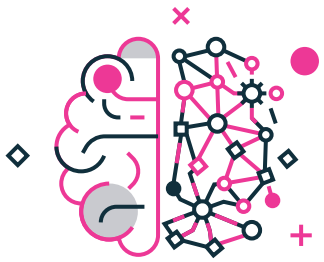
◆ **Une perspective radicale :** Pour Yann LeCun, directeur scientifique de l'IA chez Meta (Facebook), chercher à rendre les modèles d'IA explicables est une erreur. Tout comme nous faisons confiance aux humains sans comprendre le fonctionnement exact de leur cerveau, nous devrions évaluer les systèmes d'IA sur leur fiabilité et leurs résultats plutôt que sur notre capacité à expliquer leur fonctionnement interne.

◆ **Une analyse technique :** Pour le chercheur Samuel Bowman, aucune technique ne permet de comprendre comment les modèles utilisent leurs connaissances pour produire une réponse. Avec des centaines de milliards de connexions activées pour traiter un seul texte, toute tentative d'explication précise serait trop complexe pour être comprise par un humain.

◆ **Une position industrielle :** Pour Anthropic, l'entreprise derrière l'IA Claude, les modèles sont comme des plantes qui poussent plutôt que des systèmes dont on maitriserait la construction. Anthropic considère que l'interprétabilité est indispensable pour avoir confiance dans les résultats mais c'est un domaine de recherche balbutiant.

◆ **Une exigence règlementaire :** l'Union européenne exige dans l'AI Act que les systèmes à haut risque soient conçus pour permettre une surveillance humaine effective, avec des personnes capables de comprendre leurs capacités et limites et d'intervenir sur leur fonctionnement.

Pour poursuivre le débat : Faut-il permettre à tout le monde de comprendre le fonctionnement de l'IA ou seulement à des groupes d'experts formés à ces technologies complexes ?



CONTROVERSE

Qu'est ce qu'il y a d'intelligent dans "l'intelligence artificielle" ?

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

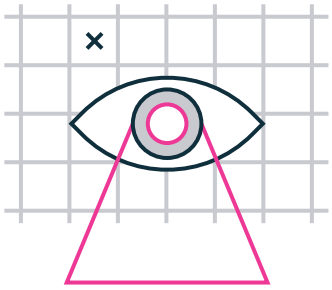
en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CONTROVERSE

Peut-on tirer profit de l'IA sans cramer la planète ?

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole



CONTROVERSE

L'IA générative peut-elle être moins opaque ?

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CONTROVERSE

L'IA peut-elle tourner sans le travail des humains ?

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

Peut-il y avoir des données sans biais ?

Quelques points de vue sur cette question :

◆ **Une perspective sociologique :** pour l'historienne Lisa Gitelman, les données ne sont jamais brutes et neutres. Elles sont toujours le produit d'une série de choix (que mesurer, comment, quand, où…). Ces choix reflètent une vision du monde que le modèle reproduit et amplifie.

◆ **Une analyse critique :** Pour l'informaticienne Mélanie Mitchell, les modèles de langage ne font que refléter et exprimer les préjugés déjà présents dans nos sociétés. Le vrai défi serait de "débiaiser" la société et le langage.

◆ **Une position industrielle :** Pour OpenAI, les biais des modèles sont inévitables. L'entreprise reconnaît que ChatGPT est biaisé vers des perspectives occidentales et fonctionne mieux en anglais. L'atténuation des biais est présentée comme un «domaine de recherche en cours».

◆ **Un enjeu culturel :** pour le Ministère de la Culture, les modèles entraînés principalement sur des données en anglais produisent des résultats stéréotypés qui négligent la diversité des langues et des cultures francophones.

Pour poursuivre le débat : comment faire pour éviter que les grands modèles de langage standardisent les cultures et les façons de penser ?

Peut-on échapper au pillage des ressources protégées par le droit d'auteur ?

Quelques points de vue sur cette question :

◆ **Une perspective juridique :** Pour la professeure de droit Rebecca Tushnet, le droit d'auteur dispose déjà des outils pour encadrer l'IA. Comme pour Google qui indexe le web, l'utilisation massive de données protégées peut être considérée comme un «usage équitable» si le résultat final ne reproduit pas directement les œuvres.

◆ **Un procès en cours :** Pour le New York Times, l'utilisation de millions d'articles sans compensation est illégale. Les modèles d'IA reproduisent leur style journalistique et survalorisent ce contenu de qualité dans leur entraînement. L'entreprise considère qu'OpenAI menace directement le modèle économique du journal.

◆ **Une question de réglementation :** Pour Arthur Mensch, PDG de Mistral IA, la régulation doit porter sur les applications de l'IA plutôt que sur les données d'entraînement elle-même. L'entreprise préfère négocier directement avec les ayants droit et s'oppose aux obligations légales de transparence des données d'entraînement.

◆ **Un modèle alternatif :** Pour Anastasia Stasenko, co-fondatrice de Pleias, il est possible de créer des modèles d'IA performants uniquement avec des données libres de droit. Les modèles d'IA de cette start-up française sont ouverts. Elle partage en détail la méthode et les données d'entraînement de ses modèles.

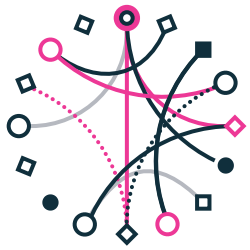
Pour poursuivre le débat : Comment garantir une répartition équitable de la valeur entre créateurs et entreprises d'IA ?

CONTROVERSE

Peut-on échapper au pillage
des ressources protégées
par le droit d'auteur ?

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole



CONTROVERSE

Peut-il y avoir
des données sans biais ?

Nantes Métropole DATACTIVIST LA BOÎTE NOIRE DE L'IA

en collaboration avec le réseau de la médiation numérique de Nantes Métropole

CC BY SA