

FUTURA

Google veut utiliser vos données publiques pour entraîner son IA

Podcast écrit et lu par Adèle Ndjaki

[Générique d'intro, une musique énergique et vitaminée.]

Google qui utilise vos données publiques pour nourrir son IA, c'est l'actu de la semaine, dans Vitamine Tech.

[Fin du générique.]

On ne le présente plus, Google est le moteur de recherche le plus utilisé au monde. Partie intégrante du net, au fil des années, l'entreprise a su se diversifier et proposer de nombreux services comme Google Translate, Google Maps, Google News et bien d'autres. Et dernièrement, c'est Bard, une intelligence artificielle similaire à ChatGPT, qui a fait son arrivée chez le géant de la tech. Sauf que voilà, pour aider son IA à se démarquer face à une concurrence redoutable, Google s'est récemment réservé le droit de l'entraîner en exploitant vos données publiques. Une annonce tout sauf anodine.

[Une musique électronique calme.]

Google et le respect de la politique de confidentialité, c'est loin d'être une grande histoire d'amour. Le géant de la Silicon Valley a plusieurs fois été attaqué en justice pour avoir collecté les données publiques ou privées de ses utilisateurs. Le 31 décembre 2021, par exemple, l'entreprise avait été condamnée par la CNIL pour avoir empêché ses utilisateurs de refuser les cookies aussi facilement qu'ils pouvaient les accepter. Une combine qui lui avait tout de même coûté 150 millions d'euros. On se souviendra aussi que Google a été accusé en 2022 de tromper ses utilisateurs, en leur faisant croire que la géolocalisation était désactivée tandis que leur appareil continuait de collecter des informations. Il n'est donc pas rare pour la firme de se faire rappeler à l'ordre mais il semblerait que les sanctions ne suffisent pas à la faire rentrer dans le rang. Car figurez-vous que depuis ce 1er juillet 2023, Google se réserve désormais le droit de collecter toutes les informations que vous mettez en ligne publiquement. Le géant de la Silicon Valley a mis à jour sa politique de confidentialité et y a inséré un passage dans lequel on peut y lire que les informations publiques pourront être utilisées pour aider à former ses intelligences artificielles. Si auparavant, l'entreprise avait déclaré que les données seraient utilisées, je cite, pour « les modèles de langage » plutôt que pour les « modèles d'IA », la nouvelle politique de Google ajoute dorénavant, aux côtés de Google Translate, Bard et Cloud AI. Depuis le lancement de ChatGPT, le chatbot conversationnel d'OpenAI, il semble que pas un seul matin ne se passe sans qu'une nouvelle IA ne soit créée. D'après les estimations de Fortune Business Insights, le marché galopant de l'intelligence artificielle devrait s'élever à 266 milliards de dollars d'ici à 2027.

Les grandes entreprises s'empressent donc de rejoindre le mouvement en développant leurs propres chatbots, comme Google avec Bard ou le géant chinois Baidu avec Ernie. Mais il ne suffit toutefois pas de créer un chatbot pour dominer le secteur. Les nombreuses « hallucinations » dont Bard a été victime depuis son lancement, de fausses informations entièrement fabriquées, démontrent que le chatbot a encore besoin d'entraînement avant d'être jugé pleinement opérationnel. Et pour cela, Google compte bien mettre à profit le moindre morceau d'information publique sur lequel son IA peut se faire la dent. Mais d'ailleurs, qu'entend Google par « information publique » ? Comprenez ici que l'entreprise devrait pouvoir se servir de photos, de vidéos, d'articles, de commentaires, bref, d'un vaste ensemble d'écrits et de médias publiés sur Internet à la vue de tous. Vos échanges de mails et vos posts privés sur Facebook sont donc hors d'atteinte, mais pas les photos du vous adolescent que vous aviez postées sur Skyblog il y a quinze ou vingt ans. Car non, Google n'a pas défini de limite de temps concernant les données qu'il peut collecter. Attention donc si certains de vos dossiers embarrassants traînent encore sur Internet ; il est peut-être temps d'envisager un petit coup de propre.

[Virgule sonore, une cassette que l'on accélère puis rembobine.]

[Une musique de hip-hop expérimental calme.]

Alors avant de paniquer et de vous demander où vous avez bien pu poster ces photos de votre enterrement de vie de garçon ou de jeune fille, laissez-moi vous rassurer. À moins que vous ne viviez aux États-Unis, vous êtes pour l'instant hors de danger puisque c'est bien uniquement aux US que s'appliquent pour l'instant les nouvelles règles de confidentialité de Google. Bard, l'IA de Google n'est d'ailleurs pas accessible pour le moment sur le territoire français, même s'il existe des stratagèmes pour y accéder. L'entreprise, qui voulait lancer son intelligence artificielle au sein de l'Union européenne à la mi-juin, s'est vu refuser cette requête faute d'avoir indiqué comment elle comptait se conformer à la RGPD, le Règlement Général sur la Protection des Données. Cette réglementation, mise en place pour protéger les données personnelles des utilisateurs européens, est pour l'instant indicative de la posture qu'adopte l'Union européenne face aux géants de la tech et aux IA qui foisonnent sur le web. La confidentialité reste prioritaire et doit être protégée – au moins du regard des entreprises privées. Ceci étant dit, comment Google compte distinguer un contenu américain d'un contenu guatémaltèque ou français reste un mystère. Google se contentera-t-il de collecter les données publiques stockées sur des serveurs américains ? Ou excluera-t-il spécifiquement de sa liste tous les médias, personnes ou entités identifiés comme appartenant à un autre pays ? Difficile de le savoir, avec les informations mises à disposition à ce jour. Afin de ne pas laisser de place au doute, un peu de vigilance s'impose. Pour protéger au mieux vos informations, évitez de partager des choses impulsivement en ligne. Demandez-vous si vous accepteriez que tout le monde puisse voir ces photos ou ces textes que vous publiez. Sur vos réseaux sociaux, pensez à vérifier si vos publications sont uniquement accessibles à un cercle réduit de personnes ou si elles sont partagées publiquement. Pensez à conserver une liste des sites sur lesquels vous êtes inscrits et à supprimer vos comptes quand vous n'en avez plus l'utilité. Vous avez également le droit de demander aux sites d'effacer vos données de ses serveurs. Et enfin refuser d'accepter les cookies des sites sur lesquels vous vous baladez sont des idées parmi tant d'autres pour éviter que vos informations soient collectées.

[Virgule sonore, un grésillement électronique.]

C'est tout pour cet épisode de Vitamine Tech. Si ce podcast vous plaît, pensez à vous y abonner et à nous laisser un commentaire sur votre app de prédilection, ça semble peu de chose mais ça nous aide vraiment beaucoup. Cette semaine, je vous recommande notre dernier épisode de Bêtes de Science, où Gaby Fabresse vous parlera d'un animal plus ingénieux qu'il n'y paraît : le termite. Pour le reste, je vous souhaite une excellente journée ou une très bonne soirée et je vous dis à la semaine prochaine dans Vitamine Tech.

[Un glitch électronique ferme l'épisode.]