

Zoom et les politiques de confidentialité

Cet article a été publié à l'origine par THE MARKUP, il est traduit et republié avec l'accord de l'auteur selon les termes de la licence CC BY-NC-ND 4.0



Traduction Framalang : goofy, MO, Henri-Paul, Wisi_eu

Voilà ce qui arrive quand on se met à lire vraiment les politiques de confidentialité

Une récente polémique sur la capacité de Zoom à entraîner des intelligences artificielles avec les conversations des utilisateurs montre l'importance de lire les petits caractères

par Aaron Sankin



Justin Sullivan/Getty Images



Bonjour, je m'appelle Aaron Sankin, je suis journaliste d'investigation à The Markup. J'écris ici pour vous expliquer que si vous faites quelque chose de très pénible (lire les documents dans lesquels les entreprises expliquent ce qu'elles peuvent faire avec vos données), vous pourrez ensuite faire quelque chose d'un peu drôle (piquer votre crise en

ligne).

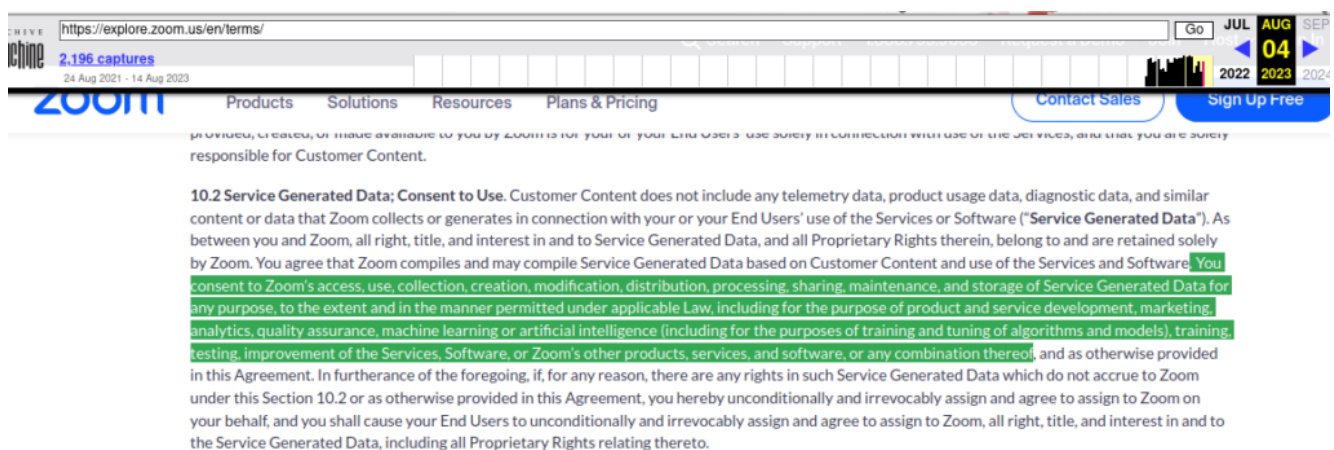
Au cours du dernier quart de siècle, les politiques de protection de la vie privée – ce langage juridique long et dense que l'on parcourt rapidement avant de cliquer sans réfléchir sur « J'accepte » – sont devenues à la fois plus longues et plus touffues. Une étude publiée l'année dernière a montré que non seulement la longueur moyenne des politiques de confidentialité a quadruplé entre 1996 et 2021, mais qu'elles sont également devenues beaucoup plus difficiles à comprendre.

Voici ce qu'a écrit Isabel Wagner, professeur associé à l'université De Montfort, qui a utilisé l'apprentissage automatique afin d'analyser environ 50 000 politiques de confidentialité de sites web pour mener son étude :

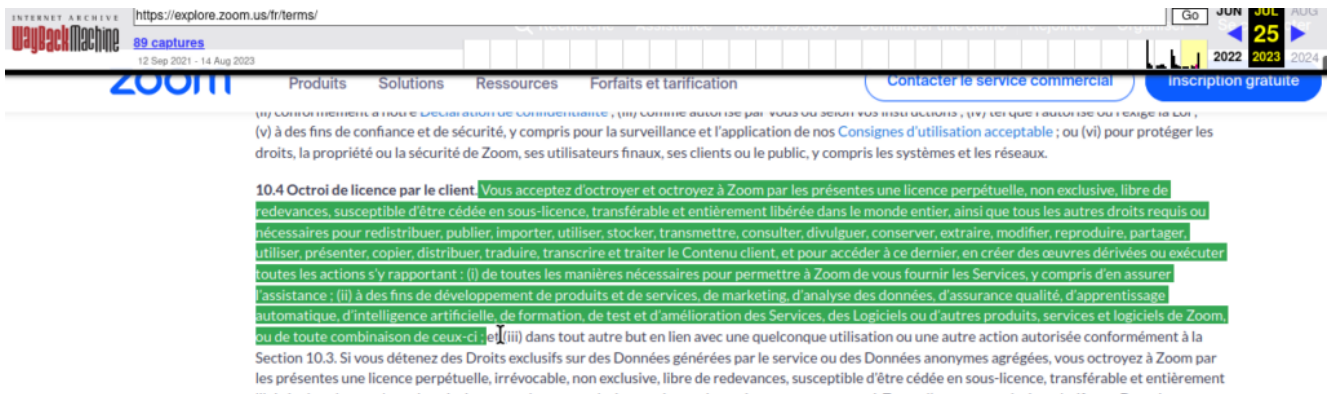
« En analysant le contenu des politiques de confidentialité, nous identifions plusieurs tendances préoccupantes, notamment l'utilisation croissante de données de localisation, l'exploitation croissante de données collectées implicitement, l'absence de choix véritablement éclairé, l'absence de notification efficace des modifications de la politique de confidentialité, l'augmentation du partage des données avec des parties tierces opaques et le manque d'informations spécifiques sur les mesures de sécurité et de confidentialité »

Si l'apprentissage automatique peut être un outil efficace pour comprendre l'univers des politiques de confidentialité, sa présence à l'intérieur d'une politique de confidentialité peut déclencher un ouragan. Un cas concret : Zoom.

En début de semaine dernière, Zoom, le service populaire de visioconférence devenu omniprésent lorsque les confinements ont transformé de nombreuses réunions en présentiel en réunions dans de mini-fenêtres sur des mini-écrans d'ordinateurs portables, a récemment fait l'objet de vives critiques de la part des utilisateurs et des défenseurs de la vie privée, lorsqu'un article du site d'actualités technologiques Stack Diary a mis en évidence une section des conditions de service de l'entreprise indiquant qu'elle pouvait utiliser les données collectées auprès de ses utilisateurs pour entraîner l'intelligence artificielle.



version anglaise début août, capturée par la Wayback Machine d'Internet Archive



version française fin juillet, capturée par la Wayback Machine d'Internet Archive

Le contrat d'utilisation stipulait que les utilisateurs de Zoom donnaient à l'entreprise « une licence perpétuelle, non exclusive, libre de redevances, susceptible d'être cédée en sous-licence et transférable » pour utiliser le « Contenu client » à des fins diverses, notamment « de marketing, d'analyse des données, d'assurance qualité, d'apprentissage automatique, d'intelligence artificielle, etc. ». Cette section ne précisait pas que les utilisateurs devaient d'abord donner leur consentement explicite pour que l'entreprise puisse le faire.

Une entreprise qui utilise secrètement les données d'une personne pour entraîner un modèle d'intelligence artificielle est particulièrement controversée par les temps qui courent. L'utilisation de l'IA pour remplacer les acteurs et les scénaristes en chair et en os est l'un des principaux points d'achoppement des grèves en cours qui ont paralysé Hollywood. OpenAI, la société à l'origine de ChatGPT, a fait l'objet d'une vague de poursuites judiciaires l'accusant d'avoir entraîné ses systèmes sur le travail d'écrivains sans leur consentement. Des entreprises comme Stack Overflow, Reddit et X (le nom qu'Elon Musk a décidé de donner à Twitter) ont également pris des mesures énergiques pour empêcher les entreprises d'IA d'utiliser leurs contenus pour entraîner des modèles sans obtenir elles-mêmes une part de l'activité.

La réaction en ligne contre Zoom a été féroce et immédiate, certaines organisations, comme le média Bellingcat, proclamant leur intention de ne plus utiliser Zoom pour les vidéoconférences. Meredith Whittaker, présidente de l'application de messagerie Signal spécialisée dans la protection de la vie privée, a profité de l'occasion pour faire de la publicité :

« HUM : Les appels vidéo de @signalapp fonctionnent très bien, même avec une faible bande passante, et ne collectent AUCUNE DONNÉE SUR VOUS NI SUR LA PERSONNE À QUI VOUS PARLEZ ! Une autre façon tangible et importante pour Signal de s'engager réellement en faveur de la vie privée est d'interrompre le pipeline vorace de surveillance des IA. »

Zoom, sans surprise, a éprouvé le besoin de réagir.

Dans les heures qui ont suivi la diffusion de l'histoire, le lundi même, Smita Hashim, responsable des produits chez Zoom, a publié un billet de blog visant à apaiser des personnes qui craignent de voir leurs propos et comportements être intégrés dans des modèles d'entraînement d'IA, alors qu'elles souhaitent virtuellement un joyeux anniversaire à leur grand-mère, à des milliers de kilomètres de distance.

« Dans le cadre de notre engagement en faveur de la transparence et du contrôle par l'utilisateur, nous clarifions notre approche de deux aspects essentiels de nos services : les fonctions d'intelligence artificielle de Zoom et le partage de contenu avec les clients à des fins d'amélioration du produit », a écrit Mme Hashim. « Notre objectif est de permettre aux propriétaires de comptes Zoom et aux administrateurs de contrôler ces fonctions et leurs décisions, et nous sommes là pour faire la lumière sur la façon dont nous le faisons et comment cela affecte certains groupes de clients ».

Mme Hashim écrit que Zoom a mis à jour ses conditions d'utilisation pour donner plus de contexte sur les politiques d'utilisation des données par l'entreprise. Alors que le paragraphe sur Zoom ayant « une licence perpétuelle, non exclusive, libre de redevances, pouvant faire l'objet d'une sous-licence et transférable » pour utiliser les données des clients pour « l'apprentissage automatique, l'intelligence artificielle, la formation, les tests » est resté intact [N de T. cependant cette mention semble avoir disparu dans la version du 11 août 2023], une nouvelle phrase a été ajoutée juste en dessous :

« Zoom n'utilise aucun Contenu client audio, vidéo, chat, partage d'écran, pièces jointes ou autres communications comme le Contenu client (tels que les résultats des sondages, les tableaux blancs et les réactions) pour entraîner les modèles d'intelligence artificielle de Zoom ou de tiers. »

Comment utilisons-nous vos données à caractère personnel ?

Les employés de Zoom n'accèdent pas au Contenu client des réunions, des webinaires, des messageries ou des e-mails (en particulier, l'audio, la vidéo, les fichiers, les tableaux blancs en réunion et les contenus des messageries ou des e-mails), ni au contenu généré ou partagé dans le cadre d'autres fonctions de collaboration (comme les tableaux blancs hors réunion), et ne les utilisent pas, à moins que le titulaire du compte hébergeant le produit ou Service Zoom où le Contenu client a été généré ne le demande ou que cela ne soit nécessaire pour des raisons juridiques, de sûreté ou de sécurité. Zoom n'utilise aucun Contenu client audio, vidéo, chat, partage d'écran, pièces jointes ou autres communications comme le Contenu client (tels que les résultats des sondages, les tableaux blancs et les réactions) pour entraîner les modèles d'intelligence artificielle de Zoom ou de tiers.

copie d'écran du 16/08/2023, page <https://explore.zoom.us/fr/privacy/>

Dans son billet de blog, Mme Hashim insiste sur le fait que Zoom n'utilise le contenu des utilisateurs que pour former l'IA à des produits spécifiques, comme un outil qui génère automatiquement des résumés de réunions, et seulement après que les utilisateurs auront explicitement choisi d'utiliser ces produits. « Un exemple de service d'apprentissage automatique pour lequel nous avons besoin d'une licence et de droits d'utilisation est notre analyse automatisée des invitations et des rappels de webinaires pour s'assurer que nous ne sommes pas utilisés involontairement pour spammer ou frauder les participants », écrit-elle. « Le client est propriétaire de l'invitation au webinaire et nous sommes autorisés à fournir le service à partir de ce contenu. En ce qui concerne l'IA, nous n'utilisons pas de contenus audios, de vidéos ou de chats pour entraîner nos modèles sans le consentement du client. »

La politique de confidentialité de Zoom - document distinct de ses conditions de service - ne mentionne l'intelligence artificielle ou l'apprentissage automatique que dans le contexte de la fourniture de « fonctions et produits intelligents (sic), tels que Zoom IQ ou d'autres outils pour recommander le chat, le courrier électronique ou d'autres contenus ».

Pour avoir une idée de ce que tout cela signifie, j'ai échangé avec Jesse Woo, un ingénieur spécialisé en données de The Markup qui, en tant qu'avocat spécialisé dans la protection de la vie privée, a participé à la rédaction de politiques institutionnelles d'utilisation des données.

M. Woo explique que, bien qu'il comprenne pourquoi la formulation des conditions d'utilisation de Zoom touche un point sensible, la mention suivant laquelle les utilisateurs autorisent l'entreprise à copier et à utiliser leur contenu est en fait assez standard dans ce type d'accord d'utilisation. Le problème est que

la politique de Zoom a été rédigée de manière à ce que chacun des droits cédés à l'entreprise soit spécifiquement énuméré, ce qui peut sembler beaucoup. Mais c'est aussi ce qui se passe lorsque vous utilisez des produits ou des services en 2023, désolé, bienvenue dans le futur !

Pour illustrer la différence, M. Woo prend l'exemple de la politique de confidentialité du service de vidéoconférence concurrent Webex, qui stipule ce qui suit : « Nous ne surveillerons pas le contenu, sauf : (i) si cela est nécessaire pour fournir, soutenir ou améliorer la fourniture des services, (ii) pour enquêter sur des fraudes potentielles ou présumées, (iii) si vous nous l'avez demandé ou autorisé, ou (iv) si la loi l'exige ou pour exercer ou protéger nos droits légaux ».

Cette formulation semble beaucoup moins effrayante, même si, comme l'a noté M. Woo, l'entraînement de modèles d'IA pourrait probablement être mentionné par une entreprise *sous couvert de* mesures pour « soutenir ou améliorer la fourniture de services ».

L'idée que les gens puissent paniquer si les données qu'ils fournissent à une entreprise dans un but évident et simple (comme opérer un appel de vidéoconférence) sont ensuite utilisées à d'autres fins (comme entraîner un algorithme) n'est pas nouvelle. Un rapport publié par le Forum sur le futur de la vie privée (Future of Privacy Forum), en 2018, avertissait que « le besoin de grandes quantités de données pendant le développement en tant que « données d'entraînement » crée des problèmes de consentement pour les personnes qui pourraient avoir accepté de fournir des données personnelles dans un contexte commercial ou de recherche particulier, sans comprendre ou s'attendre à ce qu'elles soient ensuite utilisées pour la conception et le développement de nouveaux algorithmes. »

Pour Woo, l'essentiel est que, selon les termes des conditions de service initiales, Zoom aurait pu utiliser toutes les données des utilisateurs qu'elle souhaitait pour entraîner l'IA sans demander leur consentement et sans courir de risque juridique dans ce processus.

Ils sont actuellement liés par les restrictions qu'ils viennent d'inclure dans leurs conditions d'utilisation, mais rien ne les empêche de les modifier ultérieurement.

Jesse Woo, ingénieur en données chez The Markup

« Tout le risque qu'ils ont pris dans ce fiasco est en termes de réputation, et le seul recours des utilisateurs est de choisir un autre service de vidéoconférence », explique M. Woo. « S'ils avaient été intelligents, ils auraient utilisé un langage plus circonspect, mais toujours précis, tout en proposant l'option du refus, ce qui est une sorte d'illusion de choix pour la plupart des gens qui n'exercent pas leur droit de refus. »

Changements futurs mis à part, il y a quelque chose de remarquable dans le fait qu'un tollé public réussisse à obtenir d'une entreprise qu'elle déclare officiellement qu'elle ne fera pas quelque chose d'effrayant. L'ensemble de ces informations sert d'avertissement à d'autres sur le fait que l'entraînement de systèmes d'IA sur des données clients sans leur consentement pourrait susciter la colère de bon nombre de ces clients.

Les conditions d'utilisation de Zoom mentionnent la politique de l'entreprise en matière d'intelligence artificielle **depuis le mois de mars**, mais cette politique n'a attiré l'attention du grand public que la semaine dernière. Ce décalage suggère que les gens ne lisent peut-être pas les données juridiques, de plus en plus longues et de plus en plus denses, dans lesquelles les entreprises expliquent en détail ce qu'elles font avec vos données.

Heureusement, Woo et Jon Keegan, journalistes d'investigation sur les données pour The Markup, ont récemment publié un guide pratique (en anglais) indiquant comment lire une politique de confidentialité et en identifier rapidement les parties importantes, effrayantes ou révoltantes.

Bonne lecture !

Sur le même thème, on peut s'intéresser à :

- tosdr.org : un projet communautaire qui vise à analyser et à évaluer les conditions d'utilisation (ToS en anglais) et les politiques de confidentialité des principaux sites et services Internet.
- l'installation « I Agree » de l'artiste Dima Yarovsky qui en 2018 a imprimé les conditions d'utilisation de WhatsApp, Google, Tinder, Twitter, Facebook, Snapchat et Instagram et les a ensuite accrochées dans une galerie en précisant le nombre de mots de chaque document et son temps

de lecture.