

COMMUNIQUÉ DE PRESSE

L'IA générative menace l'histoire de l'Holocauste (rapport UNESCO)

Paris, le 18 juin 2024 – L'UNESCO alerte sur le fait que l'IA pourrait déformer le récit historique de l'Holocauste et alimenter l'antisémitisme si des garde-fous éthiques ne sont pas rapidement mis en place. Selon un rapport publié mardi par l'Organisation, l'IA générative peut non seulement permettre à des acteurs malveillants d'alimenter la désinformation et les récits haineux, mais aussi produire par inadvertance des contenus faux ou trompeurs sur l'Holocauste.

- [Téléchargez le rapport de l'UNESCO : « AI and the Holocaust: rewriting history »](#)

« Si nous permettons que les faits terribles de l'Holocauste soient dilués, déformés ou falsifiés par une utilisation irresponsable de l'IA, nous risquons une propagation fulgurante de l'antisémitisme et une diminution de notre compréhension des causes et des conséquences de ces atrocités. Il est urgent de mettre en œuvre la Recommandation de l'UNESCO sur l'éthique de l'IA afin que les jeunes générations se nourrissent de faits avérés et non fabriqués », prévient Audrey Azoulay, Directrice générale de l'UNESCO.

Publié en partenariat avec le Congrès juif mondial, ce rapport souligne que les informations déformées sur l'Holocauste ont trouvé un nouveau moyen de se propager à travers les contenus générés par l'IA et risquent d'atteindre les apprenants, qui utilisent de plus en plus l'IA générative lors de leurs recherches en ligne. Alors que 80% des jeunes âgés de 10 à 24 ans utilisent désormais l'IA plusieurs fois par jour à des fins éducatives et ludiques notamment, des mesures rapides sont nécessaires pour doter ces nouvelles technologies d'un cadre éthique.

L'IA peut absorber et amplifier les préjugés de la société

L'IA générative a besoin de grandes quantités de données pour être développée. Ces données, qui proviennent souvent d'internet, sont susceptibles de contenir des informations trompeuses ou néfastes. Les systèmes d'IA héritent alors de préjugés humains qu'ils risquent de renforcer et génèrent de fausses informations sur certains événements. L'Holocauste, source d'une désinformation foisonnante, en est un exemple. Selon le rapport, en raison du manque actuel de contrôle, de régulation et de modération des développeurs d'IA, les outils d'IA générative peuvent également se nourrir de données provenant de sites web négationnistes.

Il a également été démontré que l'IA permet à des acteurs malveillants de déformer des contenus liés à l'Holocauste, en créant des témoignages fabriqués de toutes pièces ou en modifiant des documents historiques. Les fausses images et contenus audio créés à l'aide de l'IA générative sont particulièrement efficaces chez les jeunes, notamment sur les réseaux sociaux. L'application Historical Figures permettait

aux utilisateurs de discuter avec des Nazis tels qu'Adolf Hitler et Joseph Goebbels et prétendait que des individus tels que Goebbels n'étaient pas intentionnellement impliqués dans l'Holocauste et avaient essayé d'empêcher les violences infligées aux Juifs.

Des faits historiques « hallucinés » en raison de données manquantes

Les modèles d'IA générative sont susceptibles d'inventer ou d'« halluciner » des événements, des personnalités et même des faits historiques lorsqu'ils ne trouvent pas suffisamment de données. Le rapport souligne que ChatGPT et Google's Bard ont tous deux généré des contenus détaillés sur des événements liés à l'Holocauste qui n'ont jamais eu lieu. À ce titre, ChatGPT a entièrement fabriqué le concept des campagnes d'« Holocauste par noyade », au cours desquelles les Nazis auraient noyé des Juifs dans des rivières et des lacs, et Google's Bard a produit de fausses citations de témoins pour étayer des récits déformés des massacres de l'Holocauste.

Outre le risque de manipulations et d'« hallucinations » qui nuisent aux faits avérés et à la confiance dans les experts, le rapport souligne que la propension de l'IA à simplifier une histoire complexe, en privilégiant une sélection limitée de sources et d'événements, est dangereuse. Sur les moteurs de recherche, qui sont alimentés par l'IA, 60 à 80 % des images qui apparaissent en tête de liste proviennent d'un seul site de l'Holocauste, Auschwitz-Birkenau.

Une solution : prendre des mesures fortes pour mettre en œuvre les principes éthiques de l'UNESCO

L'UNESCO appelle les gouvernements à accélérer la mise en œuvre de sa [Recommandation sur l'éthique de l'IA](#), première et unique norme mondiale dans ce domaine, adoptée à l'unanimité par ses États membres en 2021. Plus de cinquante pays sont actuellement en train d'intégrer cette Recommandation dans leur législation.

L'UNESCO appelle les entreprises technologiques à mettre elles aussi en œuvre cette norme, afin d'assumer leurs responsabilités et de s'assurer que des principes tels que l'équité, la transparence, les droits humains et la vigilance requise sont intégrés dans les applications dès la phase de conception. En février 2024, [huit premières entreprises technologiques se sont engagées](#) auprès de la Directrice générale de l'UNESCO à adopter cette approche éthique.

Les entreprises technologiques doivent par ailleurs travailler en étroite collaboration avec la communauté juive, les survivants de l'Holocauste et leurs descendants, les éducateurs, les experts en antisémitisme et les historiens lorsqu'elles développent de nouveaux outils d'IA.

L'UNESCO appelle enfin les systèmes éducatifs à contribuer à la préservation des faits relatifs à l'Holocauste en dotant les jeunes d'une culture numérique, d'une pensée critique et d'une bonne compréhension de l'histoire de ce génocide.

En savoir plus

- [Rapport de l'UNESCO : « AI and the Holocaust: rewriting history »](#)
- [Rapport de l'UNESCO : « Bias Against Women & Girls in Large Language Models »](#)
- [Recommandation de l'UNESCO sur l'éthique de l'intelligence artificielle](#)
- [Orientations pour l'intelligence artificielle générative dans l'éducation et la recherche](#)
- [Prévenir l'antisémitisme par l'éducation](#)
- [La lutte contre les discours de haine](#)

À propos de l'UNESCO

Avec 194 États membres, l'Organisation des Nations unies pour l'éducation, la science et la culture contribue à la paix et à la sécurité en promouvant la coopération multilatérale dans les domaines de l'éducation, de la science, de la culture, de la communication et de l'information. L'UNESCO coordonne un réseau de plus de 2000 sites inscrits au patrimoine mondial, de réserves de biosphère et de géoparcs mondiaux ; de plusieurs centaines de villes créatives, apprenantes, inclusives et durables ; et de plus de 13 000 écoles associées, chaires universitaires, centres de formation et de recherche. Basée à Paris, l'Organisation dispose de bureaux dans 54 pays et emploie plus de 2 300 personnes. Sa Directrice générale est Audrey Azoulay.

« Les guerres prenant naissance dans l'esprit des hommes, c'est dans l'esprit des hommes que doivent être élevées les défenses de la paix » – Acte constitutif de l'UNESCO, 1945.

Plus d'information : www.unesco.org

CONTACT PRESSE

Clare O'HAGAN

+33 (0) 145 68 17 29

c.o-hagan@unesco.org