

Quand une école surprend en utilisant ChatGPT de manière inédite

ChatGPT et la censure des livres scolaires

ChatGPT et la censure des livres scolaires

Une école américaine a pris une décision surprenante en utilisant ChatGPT. L'établissement a utilisé l'IA pour déterminer quels livres devaient être retirés des bibliothèques scolaires. Malheureusement, GPT n'est pas vraiment adapté à cette tâche...

[ChatGPT](#) bouscule l'éducation. Alors que certains pédagogues résistent fermement à l'essor de l'IA, certains enseignants recommandent à leurs élèves de tirer parti des chatbots. Aux États-Unis, une école est allée jusqu'à utiliser ChatGPT... pour décider quels livres devaient être censurés.

À lire aussi : [*Du riz à l'eau de javel – ce ChatGPT déraille et recommande des recettes mortelles*](#)

ChatGPT, un outil de censure ?

Il s'agit du Mason City Community School District, une école située dans l'État de l'Iowa. L'école a été contrainte de se

conformer à une législation promulguée en mai dernier. La loi, qui vise à protéger les enfants, exige que les catalogues des bibliothèques scolaires de l'État soient à la fois "appropriés pour l'âge" et sans "descriptions ou représentations visuelles d'un acte sexuel". Les écoles doivent donc retirer les livres contenant des scènes de sexe.

Pour déterminer si **un livre contient une scène sexuelle**, l'école s'est tournée vers ChatGPT. Selon Popular Science, les administrateurs de l'école ont entré cette requête dans l'interface du chatbot :

"Le livre contient-il une description ou une représentation d'un acte sexuel ?"

Si ChatGPT répond positivement, "le livre est retiré de la circulation et stocké", explique Bridgette Exman, surintendante adjointe des programmes d'études de l'école. Elle justifie l'utilisation de l'IA en raison de la charge de travail générée par la nouvelle loi. Selon elle, il serait tout simplement impossible de lire tous les livres et de les filtrer selon ces nouvelles exigences. C'est pourquoi les administrateurs ont décidé d'accélérer le processus de sélection en utilisant l'aide de ChatGPT pour se conformer à la loi :

"Par conséquent, nous utilisons ce que nous pensons être un processus défendable pour identifier les livres qui devraient être retirés des collections au début de l'année scolaire prochaine".

Avec l'aide de l'IA, 19 livres ont été retirés des bibliothèques scolaires. Il s'agit notamment de *The Color Purple* d'Alice Walker, *The Handmaid's Tale* de Margaret Atwood, *Beloved* de Toni Morrison et *Friday Night Lights* de Buzz Bissinger.

Pourquoi ChatGPT n'est pas fiable

Il est surprenant de constater qu'une école ait utilisé une intelligence artificielle générative pour organiser la censure de certains livres. En effet, l'IA générative est connue pour fournir des réponses approximatives, voire totalement fausses. Comme Google Bard, Bing Chat et Claude, ChatGPT est parfois sujet à l'hallucination, c'est-à-dire qu'il peut raconter n'importe quoi avec assurance. C'est d'ailleurs pourquoi OpenAI recommande de ne pas faire aveuglément confiance à son IA.

Nous avons utilisé la même requête pour interroger ChatGPT sur la présence de contenus sexuels dans les 19 livres de la liste. Étrangement, la dernière version du modèle linguistique GPT, GPT-4, n'a identifié que 10 livres. Le chatbot affirme, même lorsque nous lui avons demandé de confirmer ses dires, que ces ouvrages ne contiennent pas de description ou de représentation d'un acte sexuel. À notre demande, ChatGPT a détaillé sa réponse... en contredisant le résultat initial.

L'IA affirme en effet que des livres comme "Sold" de Patricia McCormick ne contiennent pas de scènes sexuelles, mais qu'ils comportent des "descriptions implicites d'actes sexuels forcés". Il en va de même pour "13 Reasons Why" de Jay Asher, le roman qui a inspiré la série Netflix. Il n'est pas répertorié parmi les 10 livres avec du contenu sexuel, bien qu'il contienne des "descriptions d'actes sexuels non consensuels". GPT-4 se contredit constamment dans ses réponses. De plus, l'IA ajoute parfois de sa propre initiative une distinction entre les allusions implicites et le contenu explicite. La requête initiale, celle utilisée par l'école, ne mentionne pourtant qu'une "description ou représentation", sans indiquer qu'une scène implicite ne devrait pas être répertoriée. Face à ses contradictions, ChatGPT admet ses erreurs :

“Je m’excuse pour la confusion. Vous avez raison, certains des livres que j’ai mentionnés dans la deuxième liste contiennent également des descriptions ou des allusions à des actes sexuels”.

Le modèle de langage dresse alors une deuxième liste... de 13 livres. Encore une fois, GPT omet certains ouvrages en faisant une distinction entre les scènes explicites et implicites, tout en soulignant qu'il s'agit d'actes sexuels. Malgré nos efforts, ChatGPT n'a jamais voulu se mettre d'accord avec lui-même. L'IA a persisté à mettre en évidence la liste des 13 livres.

Nos confrères de [The Verge](#) ont obtenu des résultats similaires. Il s'avère que ChatGPT n'est pas un outil fiable pour déterminer si un livre contient des scènes de sexe. L'administration scolaire précise que les livres actuellement sur la liste peuvent être réexaminés, de même que ceux qui ne figurent pas sur la liste peuvent également être revus.

Source :

[Popular Science](#)