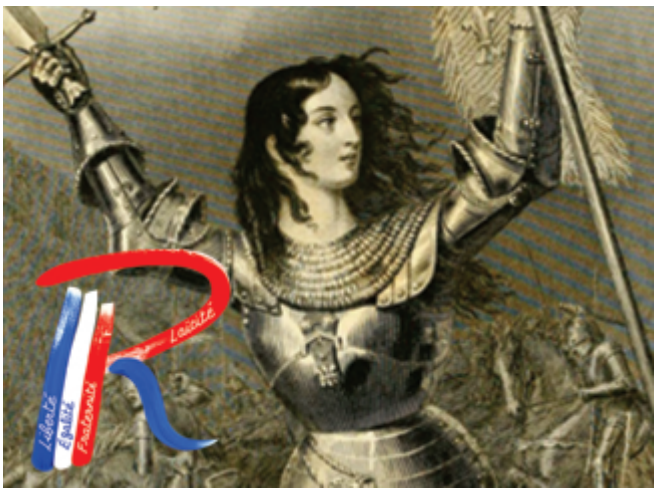
Ils créent une IA pour détecter les “discours de haine”, les pires contrevenants sont… les minorités

écrit par Julien Martel | 8 octobre 2019



Des chercheurs de l’Université de Cornell ont découvert que les systèmes d’intelligence artificielle conçus pour identifier les commentaires offensants de “discours de haine” signalent les commentaires prétendument faits par les minorités “à des taux beaucoup plus élevés” que les remarques faites par les Blancs.

.

Plusieurs universités maintiennent des systèmes d’intelligence artificielle conçus pour surveiller les sites Web de médias sociaux et signaler les utilisateurs qui affichent des “discours haineux”. Dans une étude publiée en mai, des chercheurs de Cornell ont découvert que les systèmes “signalent” plus souvent les tweets qui proviennent des utilisateurs noirs des médias sociaux, selon Campus Reform.

.

Les auteurs de l'étude ont constaté que, selon la définition des systèmes d'IA du discours offensant, "les tweets écrits en anglais afro-américain sont abusifs à des taux considérablement plus élevés".

.

L'étude a également révélé que les tweets correspondant à des utilisateurs noirs sont deux fois plus sexistes que ceux correspondant à des utilisateurs blancs.

.

L'équipe de recherche a déclaré que les résultats inattendus pouvaient s'expliquer par les "préjugés raciaux systématiques" affichés par les êtres humains qui ont aidé à repérer le contenu offensant.

.

"Les résultats montrent des preuves de préjugés raciaux systématiques dans tous les ensembles de données, car les logiciels classificateurs formés sur ces bases de données ont tendance à prédire que les tweets écrits en anglais afro-américain sont abusifs à des taux considérablement plus élevés", peut-on lire dans le résumé de l'étude. "Si ces systèmes de détection de langage abusif sont utilisés sur le terrain, ils auront donc un impact négatif disproportionné sur les utilisateurs afro-américains des médias sociaux."

.

PLUS : Un rapport du FBI révèle que les gauchistes sont une plus grande menace que les tenants de la suprématie blanche

.

L'un des auteurs de l'étude a déclaré que les "préjugés

internes” sont peut-être à blâmer en ce que l’anglais afro-américain pourrait être plus perçu comme une langue offensante”.

.

Il existe d’autres technologies de censure antiraciste

.

La technologie automatisée d’identification des discours de haine n’est pas nouvelle, et les universités ne sont pas les seules à l’élaborer. Il y a deux ans, Google a dévoilé son propre système appelé “Perspective”, conçu pour évaluer les phrases et les expressions en fonction de leur degré de “toxicité”.

.

Peu après la sortie de Perspective, le youtubeur Tormental a réalisé une vidéo du programme au travail, alléguant des incohérences dans sa mise en œuvre. Le système a jugé que les commentaires préjudiciables à l’égard des minorités étaient plus “toxiques” que les déclarations équivalentes à l’égard des Blancs. Le système de Google a montré un écart similaire noirs/blancs pour les commentaires religieux moralisateurs dirigés contre les femmes par rapport aux hommes.

.

Source : https://pluralist.com/ai-censorship-cornell-study/

Traduction : Julien Martel.

Nous avons eu un gros problème informatique lundi après-midi. Nous sommes en train de chercher des solutions.

Nous vous invitons à nous retrouver dorénavant à l'adresse resistancerepublicaine.com

Merci à tous pour votre fidélité et vos messages d'encouragement.

Merci de signaler partout autour de vous, réseaux sociaux etc cette nouvelle adresse