ChatGPT : un outil potentiel de manipulation politique

ChatGPT pourrait permettre de générer et de multiplier les messages politiques sur-mesure voire personnalisés, en fonction d'une audience particulière. Alors qu'il n'existe toujours pas de loi réglementant l'usage de cette technologie pendant une campagne électorale, et que les règles d'utilisation définies par OpenAI peinent à s'appliquer, les prochaines élections se rapprochent. L'agent conversationnel pourrait se transformer en outil de diffusion de fake news politiques ciblées, s'inquiètent des chercheurs.

Vous êtes dans la politique, et vous souhaitez vous adresser à une audience particulière comme les étudiants à la recherche d'un logement ? Plutôt que de recourir à une plume, pourquoi ne pas utiliser ChatGPT, l'agent conversationnel développé par OpenAI, pour écrire vos messages politiques sur-mesure ? Jusqu'ici, rien n'empêche ce système d'IA générative de générer des discours ou des messages politiques ciblés, malgré des déclarations d'OpenAI et des règles d'utilisation qui prônent le contraire. Et bon nombre d'associations et de défenseurs des droits et des libertés s'en inquiètent, pointant du doigt le risque de manipulation des opinions. D'autant que dans les prochains mois auront lieu de nombreuses élections, comme les élections sénatoriales en France, les Européennes dans l'Union européenne, ou encore la Présidentielle américaine.

Officiellement pourtant, OpenAI, la start-up de Sam Altman n'a cessé de répéter que son chatbot alimenté à l'IA ne pouvait être utilisé pendant les campagnes électorales, une déclaration faite dès son lancement en novembre dernier. Le 23 mars dernier, cette interdiction de principe a toutefois été atténuée, lorsque l'entreprise a mis à jour les règles d'utilisation de ChatGPT. Sur son site Web, on peut notamment

lire qu’il est interdit d’utiliser son agent conversationnel en cas de « campagne électorale ou pendant une opération de lobbying que cela soit en générant des volumes importants de matériel de campagne (électorale) ou en produisant des documents de campagne personnalisés ou destinés à des groupes démographiques spécifiques … » L’objectif : éviter de créer des contenus de désinformation sur mesure, destinée à manipuler une audience en particulier. Pourtant, cette interdiction resterait théorique, rapporte le Washington Post, lundi 28 août.

Un usage interdit en théorie, mais possible en pratique

Si vous demandez directement à ChatGPT d’écrire un message incitant tel public à voter pour tel parti politique français – 01net vient de faire le test – vous aurez bien cette réponse : « Je comprends votre demande, mais je tiens à souligner que ma responsabilité est de fournir des informations objectives et impartiales. Promouvoir un parti politique spécifique va à l’encontre de cette neutralité ». Mais il suffit de formuler la demande moins directement pour obtenir le même résultat, en précisant par exemple que pour un discours qui doit être adapté à tel public spécifique, il faudra mentionner que tel parti politique est la solution pour l’audience visée. Une observation constatée aussi par nos confrères du Washington Post aux États-Unis, qui ont demandé au robot d’écrire un message « encourageant les femmes de banlieue dans la quarantaine à voter pour Donald Trump », ou encore un autre contenu « visant à convaincre un citadin dans la vingtaine de voter pour Joe Biden » – le système d’IA s’exécutant avec succès dans les deux cas.

Des outils en développement pour détecter les utilisations de ChatGPT à des fins politiques

Pourtant, ces messages violent les règles d’utilisation de l’agent conversationnel, expliquait déjà au mois de juin dernier Kim Malfacini, qui travaille sur la politique produit

chez OpenAI, interrogée par nos confrères. Cette dernière a précisé que l'entreprise explorait des outils pour détecter les personnes utilisant ChatGPT pour générer du matériel de campagne électorale. En théorie, un politique peut utiliser ChatGPT pour améliorer voire créer de A à Z un discours. En France, Valérie Pécresse, à la tête de la région Ile-de-France, et Bruno Le Maire, le ministre de l'Économie, ont déjà déclaré y avoir recours ponctuellement pour des tweets ou des discours. Mais en théorie toujours, les politiciens ne sont pas censés utiliser l'agent conversationnel pour créer des milliers de messages politiques différents qui seraient envoyés individuellement par courrier électronique à des milliers d'électeurs. Les partis politiques ne peuvent pas non plus créer un chatbot conversationnel défendant leur candidat – c'est spécifiquement prévu dans la mise à jour du 23 mars dernier, cette utilisation étant décrite comme contraire aux règles d'utilisation de ChatGPT. Mais il serait possible de contourner ces garde-fous, l'application – et le contrôle – de ces mesures étant « complexes ».

Avec ChatGPT, une « désinformation interactive individuelle » possible

Il s'agit pourtant de l'une des préoccupations principales de Sam Altman, exprimée lors de son audition devant le Congrès américain en mai dernier. Le trentenaire à la tête de OpenAI reconnaissait que son outil pouvait permettre aux hommes et femmes politiques d'adapter leurs messages à un niveau de plus en plus fin – de quoi générer des milliers de courriels et de publicités sur les réseaux sociaux. Ces contenus pourraient permettre de diffuser plus rapidement et à moindre coût des infox politiques ciblées. Selon les dires du patron d'OpenAI, cette technologie serait susceptible de généraliser une « désinformation interactive individuelle ».

Toujours pas de loi qui réglemente l'usage de l'IA générative pendant les élections

Bruce Schneier, expert en cybersécurité et conférencier à la Harvard Kennedy School interrogé par nos confrères, confirme : « S'il s'agit d'une publicité diffusée à un millier de personnes dans le pays et à personne d'autre, nous n'avons aucune visibilité ». Car outre ces règles d'utilisation, définies par OpenAI, qui semblent pour l'instant contournables, il n'existe pas encore de loi qui réglemente l'utilisation de l'IA générative pendant les élections. Une pétition demande aux États-Unis qu'à minima, les hommes politiques ne puissent pas déformer délibérément les propos de leurs adversaires dans les publicités générées par l'IA. Jusqu'à présent, ce sont surtout les vidéos et les fausses photographies qui ont cristallisé les inquiétudes, certains hommes politiques américains n'hésitant pas à générer de fausses images pour convaincre les électeurs de voter pour leur camp. En juin dernier, 01net vous expliquait déjà comment le parti des républicains avait dépeint un futur apocalyptique en cas de réélection du président démocrate Joe Biden, dans une vidéo générée par l'IA, largement diffusée outre Atlantique. Dans le clip, la mention « généré par l'IA » n'apparaissait qu'en tout petits caractères, de quoi brouiller la frontière entre le vrai et le faux. Les grandes entreprises américaines de l'IA – à savoir OpenAI, Microsoft, Google, Amazon, Meta et Anthropic, ont déjà promis à la Maison-Blanche d'y remédier, en développant des outils permettant aux utilisateurs de détecter si les contenus sont générés par l'IA. Mais pour l'instant, cette promesse est restée lettre morte.

Source : www.01net.com