

Découverte par OpenAI de son récent instrument Examiner pour repérer les illustrations produites par l'IA

Article sur l'intelligence artificielle

La validation des contenus en ligne constitue une préoccupation cruciale avec la prolifération de l'intelligence artificielle, qui permet de générer, notamment, des images fictives de personnes.



Date de publication : 08/05/2024 08:42

Durée de lecture : 1 min



ChatGPT, l'instrument d'intelligence artificielle développée par la firme américaine OpenAI. (JEAN-MARC BARRERE / HANS LUCAS / AFP)

Un antidote à ses propres créations. OpenAI, la jeune pousse à l'origine des célèbres outils d'intelligence artificielle génératrice ChatGPT et Dall-E, offre désormais la possibilité à des chercheurs de tester un instrument conçu pour identifier si des illustrations numériques ont été produites par des programmes d'IA. D'après des essais en interne sur une édition précédente de cet instrument, "il repère correctement environ 98% des illustrations de Dall-E 3. Moins de 0,5% des images non générées par des logiciels d'IA ont été erronément associées à Dall-E 3", a souligné la société californienne.

La start-up, bénéficiant d'un important soutien financier de la part de Microsoft, a toutefois noté que son logiciel était moins efficace lorsque des illustrations créées avec Dall-E 3 étaient ultérieurement modifiées ou lorsqu'elles provenaient d'autres modèles.

Un signe distinctif pour déceler les contenus produits par IA

OpenAI a également révélé qu'elle allait à présent apposer des marques d'eau sur les images d'IA, afin de respecter les standards de la Coalition for Content Provenance and Authenticity (C2PA), une démarche de l'industrie technologique qui établit des normes techniques pour déterminer la provenance et l'authenticité des contenus numériques.

Le mois dernier, le groupe Meta (Facebook, Instagram) a annoncé qu'il débiterait l'étiquetage des contenus conçus par l'IA à partir du mois de mai en utilisant la norme C2PA. Son homologue Google s'est également joint à l'initiative.

La validation des contenus en ligne est devenue une problématique majeure avec l'essor de l'IA génératrice, permettant de produire une variété de contenus sur simple demande, incluant des fausses images de personnes ou de faux enregistrements sonores à des fins de tromperie, par exemple.

[Découvrir les avis](#)

Partage : le contenu sur les réseaux sociaux



Les nouvelles à 18h30

Chaque soir, recevez les actualités essentielles

France Télévisions récupère votre courriel pour vous envoyer la newsletter “Les nouvelles à 18h30”. À tout moment, vous pouvez annuler votre abonnement en cliquant sur le lien en bas de ces courriels. Pour plus d’informations, [consultez notre politique de confidentialité](#).

Les étiquettes associées à cet article