

Google éradique les commentaires indésirables avec Perspective

Google et Jigsaw lancent aujourd'hui une nouvelle API, [Perspective](#), ayant comme objectif l'aide à la détection des **commentaires indésirables**. Intégrée par les développeurs sur les sites d'éditeurs indépendants, cette technologie permettra de signaler automatiquement des messages potentiellement tendancieux, ou insultants, qu'il sera alors possible de filtrer.

Les commentaires postés sur **Wikipedia** et le site du **New York Times** ont été épluchés par Perspective pendant plusieurs mois. En évaluant le degré de toxicité des commentaires, il a permis de faciliter le travail des modérateurs. **Machine learning** oblige, une amélioration du taux de réussite de cet outil est à attendre...

... ou pas. Rappelons en effet que dans d'autres secteurs, le machine learning affiche régulièrement des ratés. Sur Gmail par exemple, où le filtre antispam peut parfois 'devenir fou', au point de classer tout ou partie de messages acceptables en indésirables. Du jour au lendemain.

La mort de l'originalité ?

Le principal problème de l'IA vient toutefois de son amour **de l'uniformité**. Un site web traitant d'un sujet particulièrement original a toutes les chances d'être considéré comme suspect. De même lorsque Gmail reçoit un message contenant une pièce jointe d'un format peu courant. Faudra-t-il ne plus faire preuve d'originalité, dans le but de ne pas être considéré comme impertinent par Perspective ?

Certes, Perspective n'a pour seul objectif que d'aider les modérateurs à filtrer les messages indésirables. Pas de les remplacer. Gageons toutefois que de nombreux éditeurs mettront en place des systèmes classant **automatiquement** des commentaires sans suite si le taux de toxicité de Perspective dépasse un certain seuil.

À lire aussi :

[Le patron d'Apple Tim Cook s'inquiète à son tour de la montée des fausses news](#)

[Les fausses news mises au rencard en France, à la veille des élections](#)

[Facebook veut mettre un frein aux hoax et fausses informations](#)

Crédit photo : © VisualHunt