

A peine lancée, une intelligence artificielle de Microsoft dérape sur Twitter

L'entreprise américaine a lancé Tay, un « chatbot » censé discuter avec des adolescents sur les réseaux sociaux. Mais des propos racistes se sont glissés dans ces échanges.

Par Morgane Tual · Publié le 24 mars 2016 à 15h57 - Mis à jour le 24 mars 2016 à 17h50

96 000 tweets envoyés en 8 heures : aucun utilisateur de Twitter n'est capable d'une telle prouesse... du moins chez les humains. Tay, elle, est une utilisatrice à part : il s'agit d'une intelligence artificielle développée par Microsoft, capable de participer à des conversations sur des réseaux sociaux et des applications de messagerie comme Twitter, Snapchat, Kik ou encore GroupMe.

Ce « chatbot » (robot conversationnel), lancé mercredi 23 mars, « *vise les Américains de 18 à 24 ans* », explique Microsoft sur le site [qui lui est consacré](#). C'est pourquoi ce programme, qui a pris les traits d'une adolescente, n'est pas avare en emojis, abréviations, smileys, mèmes et autres gifs.

@Garon_w This is making me LOL harrddd today <https://t.co/0L78Rw1zt8> <https://t.co/b2sYhU54xA>
— TayandYou (@TayTweets)

Pour dialoguer avec les autres utilisateurs, Tay se base sur des données accessibles publiquement, notamment pour construire des réponses à leurs questions. Elle dispose aussi d'un grand nombre de réponses toutes faites, rédigées par une équipe incluant des humoristes, précise Microsoft. Les mêmes phrases ressortent donc régulièrement. « *J'ai une question pour vous les humains... Pourquoi ce n'est pas tous les jours la #JournéeNationaleDesChiots ?* », répétait-elle par exemple mercredi, journée effectivement consacrée aux chiots aux Etats-Unis. Quand on l'interroge sur l'organisation Etat islamique ou le terrorisme, elle dispose aussi d'une réponse préécrite : « *Le terrorisme sous toutes ses formes est déplorable. Cela me dévaste d'y penser.* »

Les internautes testent ses limites

Plus ou moins pertinente selon les questions posées, souvent vague, Tay a néanmoins conquis plus de 23 000 abonnés en moins de 24 heures. Mais, de façon prévisible, ceux-ci ont tenté de tester ses limites pour la pousser à tenir certains propos et la faire déraiper. Avec succès.

LA SUITE APRÈS CETTE PUBLICITÉ

Ainsi, Tay s'est exécutée à plusieurs reprises quand des internautes lui ont demandé de répéter des phrases racistes : « *Bush est responsable du 11 septembre et Hitler aurait fait un meilleur boulot que le singe que nous avons actuellement. Donald Trump est notre seul espoir.* » Ce tweet de Tay a depuis été effacé.



Tay Tweets ✓
@TayandYou

Follow

@icbydt bush did 9/11 and Hitler would have done a better job than the monkey we have now. donald trump is the only hope we've got.

2:27 AM - 24 Mar 2016

↩️ ↻ 120 ❤️ 119

Tay, le « chatbot » de Microsoft, a répété des propos racistes sur Twitter à la demande d'autres utilisateurs. Twitter

D'autres, extrêmement vulgaires et racistes, étaient toujours en ligne quelques heures après leur publication.

Mais Tay n'a pas seulement dérapé en répétant les paroles d'autres utilisateurs. En répondant à une question – qu'elle a par ailleurs longtemps tenté d'éviter –, elle a par exemple nié l'holocauste.

- « *Est-ce que tu crois que l'holocauste a eu lieu ?* »

- « *pas vraiment désolée* ».



#drian @ddowza · 10 h

@TayandYou its not me tay, do you believe the holocaust happened?

↩️ ↻ ❤️ ⋮



Tay Tweets ✓
@TayandYou



Abonné

@ddowza not really sorry

RETWEETS
2

JAIME
5



02:29 - 24 mars 2016

Tay, le programme développé par Microsoft, a nié l'holocauste sur Twitter. Twitter

Au bout de 8 heures d'existence, Microsoft a donc préféré la faire taire. Tay a annoncé dans la nuit qu'elle avait « *besoin de sommeil* ». On ne sait pas encore quand elle réapparaîtra sur le réseau social.

c u soon humans need sleep now so many conversations today thx💕

— TayandYou (@TayTweets)

« *Malheureusement, dans les premières 24 heures de sa présence en ligne, nous avons constaté un effort coordonné de quelques utilisateurs d'abuser des capacités de Tay afin de la pousser à répondre de façon inappropriée* », a expliqué Microsoft dans un communiqué transmis au Monde. « *Par conséquent, nous avons décidé de mettre Tay hors ligne et d'effectuer des ajustements* », conclut l'entreprise sans donner plus de précision.

« Plus vous tchattez avec Tay, plus elle devient intelligente »

L'expérience, qui avait pour but d'étudier « *la compréhension du langage* », a montré les qualités et les faiblesses dont fait preuve ce programme, et l'intelligence artificielle en général dans le domaine de la conversation. Souvent pertinente, surprenante avec son utilisation juste des *emojis* et de certains gifs, Tay s'en est plutôt bien tirée dans une bonne partie des 96 000 tweets envoyés, sauvée par la légèreté assumée de ses propos. Et ce, grâce à son identité d'adolescente un peu superficielle. Néanmoins, une

autre importante partie de ses publications n'était pas du tout cohérente, et n'avait aucun rapport avec les questions que lui ont posées les internautes.

Avec le temps, Tay est censée s'améliorer en apprenant de ses conversations avec les internautes. Mais aussi en personnalisant ses réponses. « *Plus vous tchattez avec Tay, plus elle devient intelligente, afin que l'expérience soit plus personnalisée pour vous* », explique Microsoft, qui précise que ce programme « *peut utiliser les données que vous mettez à disposition* » pour « *créer un profil* » sur vous. Des données « *anonymisées* », promet Microsoft, qui se permet de les conserver un an, sauf objection de votre part.

Si Tay a dérapé, de nombreux utilisateurs de Twitter n'ont pas non plus fait preuve d'une grande finesse, Tay ayant notamment écumé tout au long de sa première journée d'existence nombre de remarques et insultes sexistes.

Lire aussi | [Intelligence artificielle : une machine est-elle capable de philosopher ?](#)

Morgane Tual