Le nouvel algorithme qui censure la propagande djihadiste, un outil efficace?

08/09/2016 à 11h02

Le logiciel eGLYPH peut reconnaître et censurer les contenus djihadistes sur Internet, mais est-ce la bonne stratégie?

En juin, une organisation américaine de lutte contre l'extrémisme a annoncé la création d'eGLYPH, un logiciel qui repère et élimine les contenus pro-terrorisme en ligne. Selon le Counter Extremism Project, cet algorithme «réduira fortement la capacité des extrémistes et des terroristes à utiliser internet pour radicaliser, recruter et inciter à la violence.»

Depuis plusieurs années, les gouvernements français, anglais, allemands et américains tentent de travailler sur la question de la propagande djihadiste avec les géants du Web et les réseaux sociaux. Twitter a récemment fait des efforts: de février à août, le réseau social a suspendu 235.000 comptes qui faisaient l'apologie du terrorisme. Selon Mark Wallace, le directeur du Counter Extremism Project, qui était ambassadeur auprès de l'ONU sous George W. Bush, cet algorithme pourrait avoir un effet positif. Interviewé par The Verge, il explique: «Si un groupe extrémiste sait qu'au moment où ils essayent de poster une vidéo en ligne, elle sera immédiatement effacée et ne pourra pas devenir virale, alors poster ce genre de choses n'aura peut être plus de sens car ils ne pourront plus atteindre leurs buts en matière de propagande.»

Le logiciel eGLYPH a été développé par Hany Farid de Dartmouth University, qui a aussi créé PhotoDNA, un logiciel qui permet

de repérer la pédopornographie. Les deux algorithmes fonctionnent de manière similaire, mais eGLYPH permet aussi de repérer des vidéos et de l'audio. Le logiciel est encore en phase de test mais sera bientôt mis gratuitement à disposition des réseaux sociaux, qui ont déjà commencé à utiliser d'autres systèmes de détection automatique.
Sauf que contrairement à la pédopornographie, certains contenus de propagande djihadiste méritent peut être d'être vus. Par exemple, si les médias évitent en général les images violentes, beaucoup ont montré des extraits de propagande de Daech. Pour contourner ce problème, Hany Farid explique que le logiciel peut s'adapter afin que les médias ou certaines organisations aient accès à ces contenus.

(…) Slate.fr