

L'expression de la haine sur les réseaux sociaux

Mathieu GUIDERE (INSERM / PARIS VIII)

Résumé

Les réseaux sociaux ont connu au cours des dernières années une expansion sans précédent qui a donné lieu à des usages divers et variés mais aussi à des dérives graves tels que la diffusion de la haine à grande échelle. En effet, le web social a été investi par des minorités radicales, allant des suprématistes jusqu'aux antisémites, en passant par les « haters » et autres « trolls », qui passent leur temps à déverser leur haine sur les autres. Mais comment identifier un contenu comme incitatif à la haine ? Cet article propose une étude du contenu des messages et des documents ayant trait à l'apologie du terrorisme et de l'antisémitisme. Après un exposé théorique et méthodologique, il présente les résultats d'une analyse semi-automatique des contenus haineux avec une exploration des traits psycholinguistiques associés aux profils étudiés. L'objectif est de montrer l'intérêt d'un recours à l'intelligence artificielle dans le traitement automatique des verbalisations en vue du repérage et de la prévention.

Mots clés

Réseaux sociaux, haine digitale, apologie du terrorisme, analyse du contenu, traitement du langage, intelligence artificielle.

Abstract

Social networks have experienced an unprecedented expansion in recent years which has given rise to varied uses but also to serious abuses such as the dissemination of hatred on a large scale. Indeed, the social web has been taken over by radical minorities, ranging from supremacists to anti-Semites, via "haters" and other "trolls", who spend their time pouring their hatred on others. The question is: how do you identify content as an incitement to hatred? This article proposes a content analysis of messages and documents relating to the apology of terrorism and anti-Semitism. After a theoretical and methodological framework, it presents the results of a semi-automatic analysis of hate content with an exploration of the psycholinguistic traits associated with a selection of profiles. The objective is to show how artificial intelligence can be used in the automatic processing of language for tracking and preventing digital hatred.

Key Words

Social networks, digital hatred, apology for terrorism, content analysis, language processing, artificial intelligence.