Facebook wil computers inzetten tegen wangedrag

17-02-2017 08:56



Facebook wil kunstmatige intelligentie gebruiken om berichten, foto's en video's op te merken die beledigend, discriminerend, misleidend of haatdragend zijn. Wat onder die begrippen verstaan moet worden, zal zo veel mogelijk per land bepaald worden.

Topman Mark Zuckerberg heeft dat, uiteraard via Facebook, bekend gemaakt. Volgens Zuckerberg kunnen computers op basis van kunstmatige intelligentie zelfstandig berichten, foto's en video's doorzoeken en aan de hand van een serie kenmerken beoordelen dat de inhoud wellicht ontoelaatbaar is. Vervolgens moeten Facebook-medewerkers beslissen of de inhoud wel of niet geblokkeerd moet worden. Van deze beslissingen leert de software vervolgens, zodat het aantal berichten dat ten onrechte wordt doorgelaten of geblokkeerd in de loop van de tijd steeds verder afneemt.

Het gaat nog wel een tijdje duren voordat het systeem zo accuraat is dat het buiten de testomgeving ingezet kan worden. Tot die tijd blijft Facebook werken op basis van meldingen van gebruikers, maar ook dat is verre van waterdicht.

Zuckerberg schrijft dat met 1,8 miljard maandelijkse gebruikers één gezamenlijke lijst met regels - die dus wereldwijd geldt - niet werkt. "Vanuit Californië kunnen we niet alle culturele normen en waarden overzien die in verschillende delen van de wereld gelden." Daarom wil hij gebruikmaken van lokale regels. Ook komen er persoonlijke filters, zodat gebruikers zelf kunnen aangeven in hoeverre ze bijvoorbeeld beelden van naakt of geweld willen blokkeren.