KROPSREAKTIONER KAN AFSLØRE HADEFULDE YTRINGER

Et hold forskere fra Danmark og Tyskland undersøger kulturelle forskelle på, hvordan folk opfatter hadefulde ytringer på internettet. Det kan medvirke til at gøre beslutningstagere bedre til at definere, hvad hadefulde ytringer er, og gøre algoritmer bedre til at spotte dem på sociale medier.

Foto: Colourbox

Facebooks algoritmer fjernede 27 millioner hadefulde beskeder alene i løbet af de sidste tre måneder af 2020. Kunstig intelligens med en politihat på er en smart og hurtig måde at eliminere betændte debatter og ytringer på, men det er langt fra en ufejlbarlig metode. Det siger Olivier Niebuhr, som er lektor ved Center for Industriel Elektronik på Syddansk Universitet i Sønderborg.

»Digitale algoritmer kan spotte hadefulde kommentarer på Facebook og andre sociale medier, men algoritmerne fokuserer ofte kun på specifikke nøgleord og har ikke modtageren for øje. Derfor har vi forsket i hadefulde ytringer ud fra kulturelle forskelle, så det i sidste ende bliver muligt at designe systemer, der træffer mere oplyste og bedre beslutninger.«

Hvad er hadefulde ytringer?

En rapport udarbejdet af Institut for Menneskerettigheder definerede i 2017 hadefulde beskeder som "Stigmatiserende, nedsættende, krænkende, chikanerende og truende ytringer, der fremsættes offentligt mod et individ eller en gruppe baseret på individets eller gruppens køn, etnicitet, religion, handicap, seksuelle orientering, alder, politiske observans eller sociale status."

Problemet er bare, at vi opfatter disse ting forskelligt. Oliver Niebuhrs forskning viser, at der er stor forskel på, hvad vi opfatter som "et sprogligt angreb" alt efter, hvor vi kommer fra eller hvilken uddannelse eller job vi har. For eksempel er der forskel på, hvad danskere og tyskere opfatter som stærkt stigmatiserende, krænkende, nedsættende eller truende.

»Hvis vi kigger på afsenderne, så har tyskere stor berøringsangst i forhold til emner som muslimer og ikke mindst Holocaust. I forhold til disse emner, holder danskerne sig ikke tilbage på helt samme måde.«

Oliver Niebuhr fortæller videre, at hans forskning til gengæld viser, at når det kommer til "udlændinge" generelt, så har tyskerne færre forbehold med hensyn til, hvor krast man kan formulere sig.

»Vi har også fundet andre kulturelle forskelle i forhold til, hvad danskere og tyskere lægger vægt på i forhold til deres definition af, hvad der er hadefulde beskeder, og hvad der ikke er. Det er absolut "no-go" for danskere at sammenligne personer

Om forfatteren
Sune Holst er journalist,
Tek kommunikation
SDU
suneholst@tek.sdu.dk

Oliver Niebuhr og hans forskning er en del af det dansk/tyske XPEROHS-projekt, som er støttet af Villum Fonden. En del af projektets partnere rådgiver tyske beslutningstagere i forbindelse med hadefulde ytringer. Foto: Sune Holst

med et dyr, for eksempel at sige, at "Han er en stor gris!" Det har tyskerne ikke helt samme tilbageholdenhed i forhold til sammenlignet med danskere.«

Kulturelle forskelle er vigtige at holde sig for øje

Oliver Niebuhr mener, at det er essentielt ikke blot at have fokus på specifikke nøgleord, når det kommer til definition af hadefulde ytringer, men i lige så høj grad på modtageren.

»Det er vigtigt at analysere den indre stemme. Det er umuligt at læse en tekst uden, at sætningens melodi spiller i dit hoved. Og den melodi – eller indre stemme – gør en enorm forskel.«

SDU-forskeren forklarer, at den indre stemme faktisk kan forvandle noget, som umiddelbart virker til at være en hadefuld ytring til noget helt andet. Det handler om betoning. Hvis du læser en tekst på en ironisk eller sarkastisk måde, så tager det brodden af de værste og modtageren opfatter det knapt så krænkende, nedsættende eller truende.

»Hvis vi vil give politiske beslutningstagere guidelines til at definere, hvad der er hadefulde ytringer, og hvad der ikke er, så kan vi ikke ignorere faktorer som sætningens melodi og den indre stemme.«

Fysiologiske signaler lyver ikke

Forskerholdet har i forbindelse med deres undersøgelse målt deltagernes fysiologiske reaktioner, når de læser hadefulde ytringer.

Som de første i verden har forskerne monitoreret, hvordan hjernebølger og vejrtrækningsmønstret ændrer sig, når forsøgspersonerne læser en hadefuld besked op. Bliver hænderne svedige, stiger pulsen og udvider pupillerne sig?

»Det er interessant, for vi kan se, at der er forskel på, hvad folk siger og hvordan deres krop reagerer.«

Oliver Niebuhr fortæller, at forskerne fandt en korrelation imellem, hvad der kom ud af forsøgspersonernes mund, og hvad deres kroppe fortalte i forbindelse med specifikke sætningsstrukturer.

»Det bonede specielt ud i forbindelse med det, vi har valgt at kalde "indirekte hadefulde ytringer". Det kunne være en sætning, som begynder med: "Jeg har intet imod muslimer, men……". Sætninger som disse bliver ikke ratet specielt hadefulde af forsøgspersonerne, men de reagerer fysiologisk på dem.«

Forskeren mener derfor, at det giver mening at kigge på de fysiologiske signaler, da de er direkte indikationer, som ikke er underlagt et socialt filter.

»Jeg vurderer, at det er et mere præcist svar du får.«

Håber på en fremtid uden had på sociale medier

Målet med forskningen er at assistere politiske beslutningstagere og sociale medier med at bekæmpe hadefulde ytringer på nettet.

»Konklusionen er, at det simpelthen ikke er nok at kigge på specifikke nøgleord. Det giver ikke en fyldestgørende evaluering. Du bliver nødt til at kigge på modtageren og konteksten.« ■