Tinder uses prompt feature to reduce offensive language

Match Group has introduced a tool designed to stop offensive messages on the dating app Tinder. The feature uses AI to detect potentially problematic language and triggers an "Are you sure?" prompt to warn users and offer the chance to change their wording. Early testing of the feature resulted in a 10% reduction in offensive language, while those receiving the prompt were also less likely to be reported for inappropriate messaging the following month, the company said.

Temps de lecture : minute

---

31 May 2021

Big Tech has faced growing criticism and scrutiny over online content during the last year, as global lockdowns have forced many onto social media platforms for communication and information. The echo chambers that have formed online as a product of users seeking like-minded people and political views have not only caused global concern for our information consumption, but have manifested into discriminatory violence – attacks on Asian people, 5G towers and the US Capitol to name a few.

As a result, regulatory bodies, governments, social advocates and even individuals within the tech community have called for more stringent measures to stop fake news and hate speech.

Why does this matter?

Tinder's solution is to discourage harmful rhetoric before it happens. The

strategy banks on creating greater awareness around language and the impact it can have – with the prompts being used to disrupt casual language use and make users conscious of their words. Internal adoption of technology would also provide Big Tech the opportunity to address its own diversity and inclusion issues that have worsened through the pandemic.

Other tech examples

Tinder is introducing this prompt after numerous cases of inappropriate language and imagery directed towards women on its platform. Other technology firms are also implementing changes to address similar issues. Snapchat is re-evaluating its camera technology to be more sensitive to all skin tones as research continually suggests a racial bias within facial-recognition software. Google is also addressing language concerns by launching a writing assistant that suggests more inclusive terms such as “chairperson” instead of “chairman” and, at the same time, has unveiled a more inclusive camera on its Pixel products to depict skin tones more accurately.

Additionally, inclusive technology is increasingly entering the workplace and offers an opportunity to educate employees about social issues. Vantage Point, for example, uses VR training programmes to place people in scenarios where they experience racial or sexist discrimination – integrating photo-realistic characters and tonality to immerse participants.

Bottom line

These solutions are a positive step, but attention must be paid to underlying technology to ensure such delicate issues are handled appropriately. Google has previously apologised for racist results produced by its Vision AI labelling software. However, when the

company's own AI ethics researcher expressed concerns about research into marginalised groups she was promptly relieved of her position. This raises concerns about responsible development of inclusive solutions and whether those with questionable internal practices should be the ones to dictate language online.

Lateral thought

While increased awareness of language and its cultural significance is an important step towards creating inclusive environments, if adoption of these kinds of tools becomes mainstream, what effects could that have on how we use language?

Language systems share the same cultural diversity and significance as the communities who speak them – such as the use of African-American Vernacular English in Black communities. If our online communications become shaped by algorithms that aren't vetted for cultural bias and programmed for diverse speech, non-standard language users may face erasure of their dialects.

Article by Sara Trett & Fred Fullerton