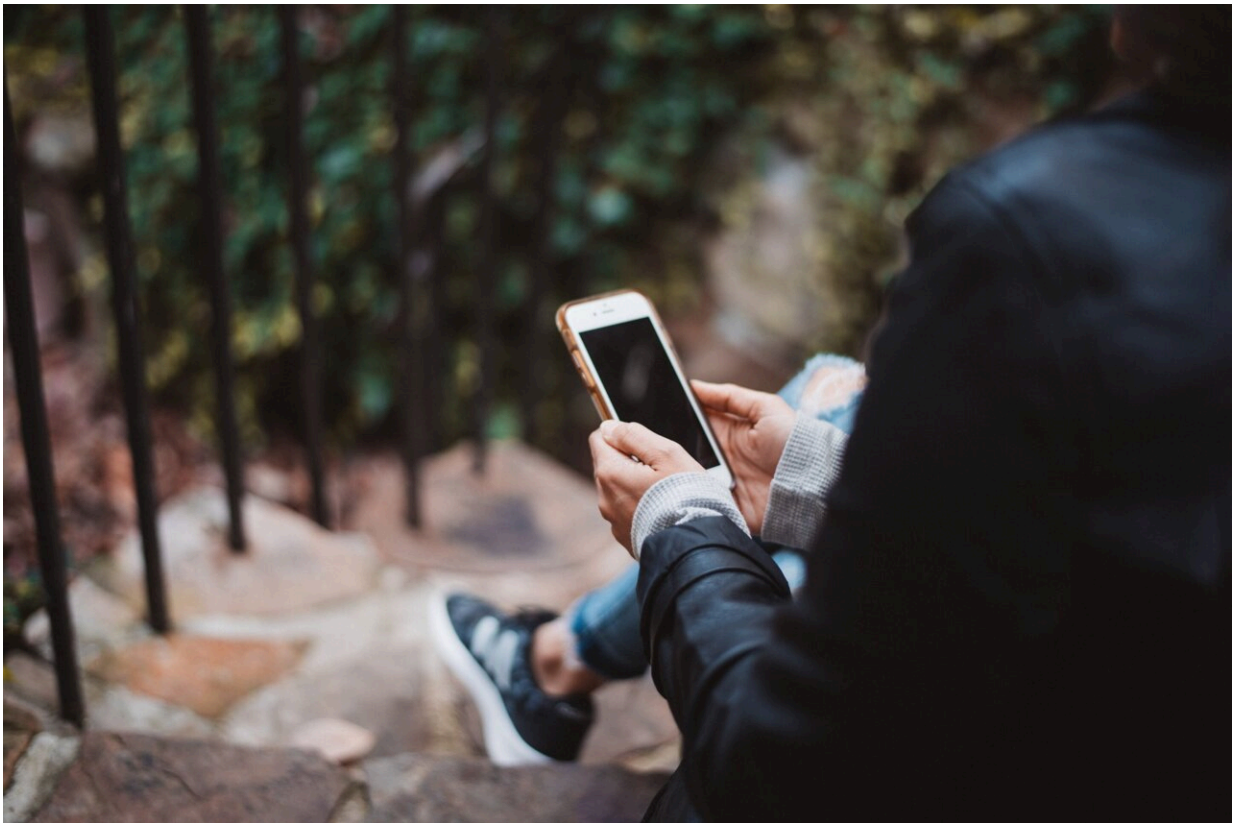


'Responsible' AI can protect children on social media, say researchers

February 7 2023



Credit: Unsplash/CC0 Public Domain

New research by academics at Warwick Business School at the University of Warwick argues that "responsible" AI can prevent children from seeing both harmful and legal but harmful content online.

This new research presents a framework which uses so called "responsible" AI to help with content moderation. The study is published in the journal *Competitive Advantage in the Digital Economy (CADE 2022)*.

The proposed system can sift through vast quantities of language and images to compile "dictionaries" of insights around each of the most harmful areas that threaten children and [young people](#), including [hate speech](#), [cyber bullying](#), suicide, anorexia, child violence and child sex abuse.

The increasing prevalence of social media has led to increased calls for oversight when moderating online content for [vulnerable groups](#), such as children. However, this presents difficulties when moderating the sheer volume of online content.

The system uses [natural language](#) processing algorithms with a layer of knowledge, which can then allow the technology to understand language more like humans do.

This means the technology that can understand context of comments, the nuances of speech, social bonds between individuals their age and relationships.

The new research suggests a system for content moderation that takes into account the context of social harms and behaviors, which may demand human interpretation, combined with an AI system that can trawl through the massive amounts of information social networks produce.

Shweta Singh, professor at Warwick Business School, University of Warwick, is one of the authors of the research and plans to give evidence to Parliament on online safety for children. She commented, "So far,

lawmakers have largely allowed [social media platforms](#) to mark their own homework—prompting outrage from both carers of vulnerable youngsters and defenders of free speech.

"Legislators need a better understanding of the technology they are seeking to govern. Tech businesses have little incentive to impose it responsible AI, with recent whistle-blowers speaking out against Meta's moderation methods and their harmful impact.

"If regulators understood what is possible—'intelligent' technology that reads between the lines and sifts benign communication from the sinister—they could demand its presence in the laws they're seeking to pass."

More information: S. Singh et al, A Framework for integrating responsible AI into social media platforms, *Competitive Advantage in the Digital Economy (CADE 2022)* (2022). [DOI: 10.1049/icp.2022.2051](https://doi.org/10.1049/icp.2022.2051)

Provided by University of Warwick

Citation: 'Responsible' AI can protect children on social media, say researchers (2023, February 7) retrieved 26 April 2024 from <https://techxplore.com/news/2023-02-responsible-ai-children-social-media.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.