
Working towards a safer internet

For all its positives, the internet is a place where hate speech, harassment, extremism and misinformation thrive. The sheer amount of harmful content, and the speed with which it travels, make it difficult to keep track of and understand the problem. Furthermore, policy makers are faced with the challenging task of shielding users from harm while protecting freedom of speech.

In response to these issues, the Turing's public policy programme set up an **Online Safety Team** – a group of researchers who are using data science and AI to provide regulators and policy makers with the tools and evidence they need to detect and tackle toxic online content. In March 2022, the team launched the **Online Harms Observatory**, a collaboration with the Department for Digital, Culture, Media & Sport (DCMS) that identifies and monitors harmful content in real time using a custom-built dashboard. Its first application, which has been used for research

commissioned by Ofcom, is tracking abuse on Twitter directed at Premier League footballers. Underpinned by algorithms developed at the Turing, the Observatory is set to be rolled out fully to regulators and policy makers in the coming months.

Also this year, the Turing's **Bertie Vidgen**, who played an instrumental role in setting up the Online Safety Team, provided direct input to the government's **Online Safety Bill** – a landmark piece of legislation which is currently going through Parliament, aimed at protecting internet users and holding tech companies to account. Bertie was a specialist advisor to the **Joint Committee on the Draft Online Safety Bill**, the result of which was a 200-page report with 127 recommendations, of which the government accepted 66.

As harmful content touches ever more lives, the team will continue to work closely with stakeholders such as DCMS and Ofcom to help create a safer online environment for all.

“The internet should feel like a safe space, not a battleground. The Turing's Online Safety Team is pioneering the use of data science and AI to detect, understand and counter online hate and abuse.”



Bertie Vidgen
Outgoing Head of Online Safety
The Alan Turing Institute

Research highlights of the year



Related programmes and teams
Public policy
AI for science and government