ISD | Powering solutions to extremism and polarisation

Algorithms and Disinformation

Research Case Studies and Policy Implications

Dec 2021

Contents

Recommended Reading: Amazon's algorithms and disinformation

- **Focus**: The study focuses on English-language books for sale on Amazon and ways in which recommendation functions promote extremist and conspiracy literature

- **Methods**: Research was conducted while not logged into an Amazon account, with browser cookies cleared and using a VPN.

- **Findings**: The research scope illustrates examples where platform recommendation mechanisms promoted and targeted TVEC, disinformation and conspiracy theory content on Amazon's stores.

5 recommendation routes

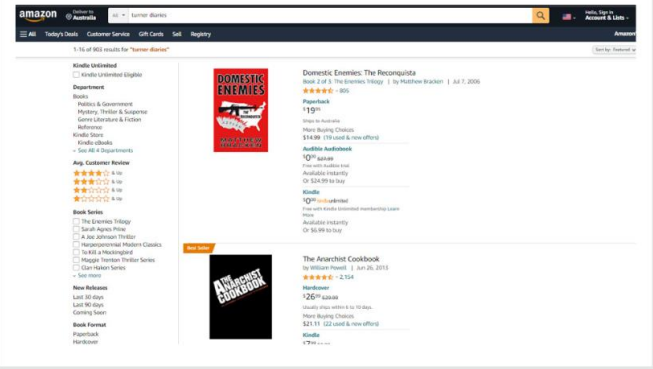


Figure 13
Screenshot showing search results for 'turner diaries' in the All category, captured 17 March 2021

For unpaid recommendations, Amazon uses a system known as item-to-item collaborative filtering. More information is available on Amazon's blog.

The recommendation of conspiracy and extremist material occurred in multiple ways:

- 'Customers who bought this item also bought' recommendations
- 'Customers who viewed this item also viewed' recommendations
- 'What other items do customers view after viewing this item?' recommendations
- 'Author pages' and associated recommended authors
- Autocomplete suggestions on search functions

Case Study: Autocomplete Promotion of Disinformation

- Amazon autocomplete recommendations on search were found to recommend search queries relevant to conspiracy theories and disinformation.

- ISD found that Amazon's search auto-complete tool was recommending searches relating to conspiracy theories on even very generic search terms, for example 'vaccine' or 'election.'

- There are links between the promotion of disinformation and the bolstering of extremist messaging and movements, which require further attention and study and go beyond the immediate scope of this topic.

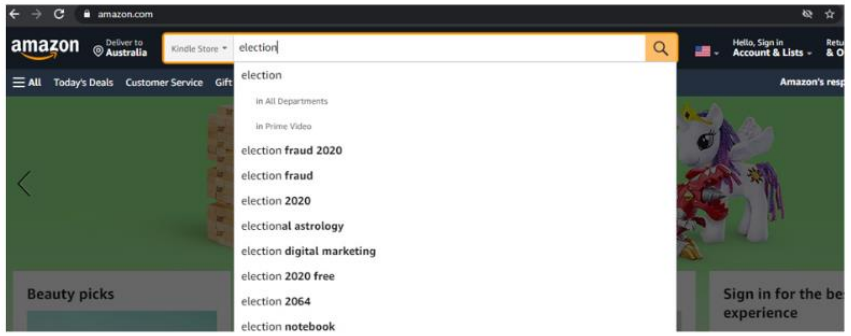


Figure 9
Screenshot showing auto-complete suggestions in Kindle Store category for 'election', showing suggestions including 'election fraud 2020' and 'election fraud', captured 2 March 2021

Hatescape: ISD's TikTok Report

- **Scope**: examines how TikTok is used to promote hate, harass minorities, glorify extremism and terrorism, share advice of weapons manufacturing and produce genocide denial content, including extremist conspiracy theories.

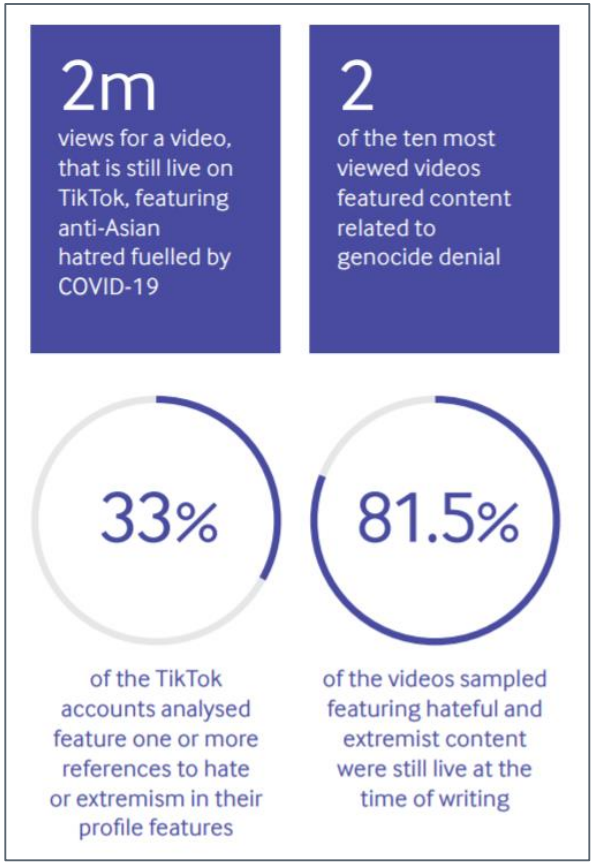
  The research attempted to understand how features on TikTok like profiles, hashtags, share functions, video effects and music contribute to the spread of hate and extremism on the platform.

- **Method**: based on a sample of **1,030 videos** posted by **491 TikTok accounts** and was created using search terms related to extremist events, conspiracy theories and ideologies, as well as a snowball methodology to create a repository of content.

Findings: A Viral Environment for Harmful Content

Key Findings

- Content espousing white supremacy and other forms of hatred was **easily discoverable** on TikTok and attracted significant engagement. 246 (24%) videos feature support for an extremist or terrorist individual or organisation.
- Content creators promoting disinformation and extremism **leverage the systemic functions of the platform** to increase the visibility of their content, including efforts to use the algorithmic promotion of certain hashtags to achieve views and engagement.
- Evasion tactics to avoid takedowns are simple but effective and creators know how to work around TikTok's efforts, including using comment restrictions strategically or alternative hashtag spellings.
- Disinformation and extremist content is removed or labelled on TikTok, but inconsistently and without transparency.



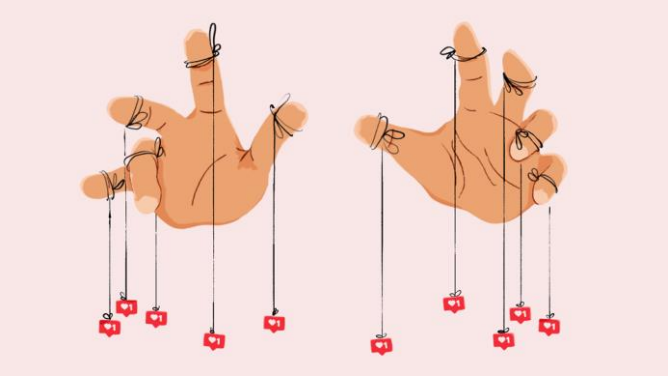
A hostile environment: Harassment and disinformation against women

- In the **US Presidential Election 2020,** ISD's ***Public Figures, Public Rage*** report analysed gender-based online abuse of congressional candidates:
  - Women, particularly women of colour, were disproportionately targeted on Twitter
  - Mentions or replies to their accounts included 15% to 39% abusive content (men 5% to 10%).
  - In Facebook comments, women received on average 12% more abusive comments than men
- In the **German Federal Elections 2021**, ISD analysed online abuse targeting the three front-runners, with a focus on gender
  - All candidates were targets of disinformation, conspiracy theories and harassment, however, Baerbock was disproportionately targeted.
  - Baerbock was framed with the "Great Reset" conspiracy narrative: ten times more that Scholz and four times more than Laschet.

Implications for Policy & Action

1. **Transparency on algorithmic ranking criteria along is not enough: we need transparency on policies, processes, thresholds and outcomes**

- Amazon has been relatively transparent about how its algorithms work and the rankings included in algorithmic decision-making. TikTok has announced a “transparency center” and published aspects of its algorithmic ranking criteria
- There is, however, very little way to conduct independent computation analysis on TikTok or on Amazon to understand the outcomes of these choices



2. **Data access for independent oversight is the first key step for providing transparency on algorithmic outcomes that can provide a robust evidence base for policymaking and regulatory design**

What next?

I. Opportunities for regulation:

Digital Services Act –potential for auditing of platforms' risk mitigation approaches

Online Safety Bill – possible auditing of systems that present risks to adults and children

II. Multi-stakeholder approaches to promoting more robust methodologies and recommendations about algorithmic transparency and interventions:

- Global Internet Forum to Counter Terrorism – CAPPI working group report. Recommendations from this work included:
  - Widen the scope of scholarly research
  - Develop shared standards for measurement and evaluation
  - Foster more data-sharing collaborations and public-private partnerships
  - Encourage more transparent explanations
- Christchurch Call workplan on algorithms and TVEC content
- Global Partnership on Artificial Intelligence – GPAI

Thank you.

Contact us:
cc@isdglobal.org