Research Methods and Professional Practice January 2022

Home / / My courses/ / RMPP_PCOM7E January 2022 / / Unit 1 / / Collaborative Learning Discussion 1 /

/ Initial Post /

« Collaborative Learning Discussion 1

Freya Basey

Initial Post

9 days ago

3 replies

Last now

Case Study: Malicious Inputs to Content Filters

Blocker Plus is an internet content filter intended for use in institutions, such as schools and libraries, as well as domestic settings to protect children from harmful online content (ACM, N.D.). The makers employed machine learning to maintain the content blacklist based on user feedback and this resulted in manipulation of the blacklist by activists.

Whilst the premise of the system supports the ethical use of computers and intends to protect children, the maker’s implementation has proved problematic and does not align to the ACM Code of Ethics (ACM, 2018). The product is not sufficiently protected from misuse leading to discrimination and suppression of information. This reinforces the importance of principle 2.5 as powerful computing concepts, such as machine learning, should be used with due care. Whilst there is evidence that testing and risk assessment were used, unethical decisions have still resulted in complaints and the deception of stakeholders.

Another applicable code of conduct is that of The Chartered Institute for IT (BCS, 2021). Blocker Plus is intended to support the public interest, however it has failed in this through reducing access to vital public health information and impacting the digital inclusion of children. In terms of duty to the relevant authority, professional judgment and due diligence has not been properly exercised which has resulted in further poor decisions to take advantage of consumer ignorance and continue to use a vulnerable model based on user feedback.

Further to professional codes of conduct, there are also legal and social issues to be considered. This content filter is intended for institutions to comply with the US Children’s Internet Protection Act (CIPA) (ACM, N.D.). The product achieves this aim but, by restricting legitimate material, it could disadvantage children through enforcing prejudices and reducing important learning opportunities (Thoreson, 2016). Whilst legislation in this area varies by state, restricting access to this legitimate material goes against human rights guaranteed by the International Covenant on Civil and Political Rights (ICCPR).

This case study presents professional, legal and social issues that could impact both the producer of and consumers of this product. Whilst the impacts are not intentional, this could still leave the Blocker Plus organisation and the institutions using the product open to reputational damage and possible litigation on the basis of the harm caused to end consumers.

References

ACM (2018) ACM Code of Ethics and Professional Conduct. Available from: **https://www.acm.org/code-of-ethics#h-2.3-know-and-respect-existing-rules-pertaining-to-professional-work** [Accessed 25 January 2022].

ACM (N.D.) Case: Malicious Inputs to Content Filters. Available from: **https://ethics.acm.org/code-of-ethics/using-the-code/case-malicious-inputs-to-content-filters/** [Accessed 25 January 2022].

BCS (2021) Code of Conduct for BCS Members. Available from: **https://www.bcs.org/media/2211/bcs-code-of-conduct.pdf** [Accessed 25 January 2022].

Thoreson, R. (2016) Discrimination Against LGBT Youth in US Schools. Available from: **https://www.hrw.org/report/2016/12/08/walking-through-hailstorm/discrimination-against-lgbt-youth-us-schools** [Accessed 26 January 2022].

Reply

3 replies

1 Post by **Simon Miller**

3 days ago

Peer Response

Freya succinctly conveys the breadth and range of implications this case represents. Part of the challenge of this case are the parties impacted and involved as well as the complexity of impacts, from societal to legal and professional. While it’s outside of the scope of this post to cover everything in its entirety, I’ve instead decided to review a handful of challenges this case represents.
While the application for machine learning has increased tremendously, there are some areas where it’s use should be handled with caution, or otherwise be avoided entirely (Horvitz and Mulligan, 2015). In this case, using machine learning to manage content availability in an education setting is problematic for a number of reasons. In corporate settings content availability has largely been handled by hardware or cloud based firewalls and while smart algorithms and even machine learning can be applied to these systems there is still a heavy reliance on human interaction to oversee the execution of content restriction (Abu Al-Haija and Ishtaiwi, 2021). If organizations still manage content through human driven firewall rules perhaps the same level of oversight should be measured for platforms managing the availability of content to children in educational settings. Nevermind the labyrinthian complexity that comes with working with highly varied school boards, municipalities, counties, provinces or states. There are questions that need to be answered at almost every level from household to federal.

References:

Abu Al-Haija, Q. and Ishtaiwi, A. (2021). Machine Learning Based Model to Identify Firewall Decisions to Improve Cyber-Defense. *International Journal on Advanced Science, Engineering and Information Technology*, 11(4), p.1688. Available at: https://www.researchgate.net/profile/Qasem-Abu-Al-Haija/publication/354227799_Machine_Learning_Based_Model_to_Identify_Firewall_Decisions_to_Improve_Cyber-

Defense/links/612e4be338818c2eaf72ad54/Machine-Learning-Based-Model-to-Identify-Firewall-Decisions-to-Improve-Cyber-Defense.pdf [Accessed 31 Jan. 2022].

Horvitz, E. and Mulligan, D. (2015). Data, privacy, and the greater good. *Science*, 349(6245), pp.253–255. Available at: http://erichorvitz.com/data_privacy_greater_good.pdf [Accessed 31 Jan. 2022].

Reply

2

Post by **Steph Paladini**

11 hours ago

Re: Initial Post

Hello Simon and Freya,

the questions you raised here are very relevant and go even further the content filtering for specific purposes (like parental control, say, or corporate practice).

Just think about the increasing debate about content just being removed from social media. In some cases, the intentions are absolutely laudable (stopping fake news etc) but in other cases, the use of algorithms can create more issues than it solves and ends up being counterproductive.

The discussion here of course could be very lengthy, so I am just flagging this point for further reflection.

This article is a good starting point:

Seargeant, P. and Tagg, C., 2019. Social media and the future of open debate: A user-oriented approach to Facebook's filter bubble conundrum. *Discourse, Context & Media*, *27*, pp.41-48.

Best wishes,

Steph

Reply

3

Post by **Jan Küfner**

now

peer response

It is a common fact that artificial intelligence (AI) is very dependent on the training material that is provided to it. By allowing that training content is provided by user makes the AI vulnerable. This attack was not proactively considered, but only hot fixed by some extent. One could claim that the company also violated BCS Code of Conduct article 2a, since they seem to lack security competence by not doing proactive threat modelling and preventing the issue from occurring in the first place. One could also argue that the company does not have due regard for the wellbeing of others by not properly fixing the issue and by accepting that some information remains still blocked although it should not be blocked. (BCS 2021)

BCS (2021) Code of Conduct for BCS Members. Available from: https://www.bcs.org/media/2211/bcs-code-of-conduct.pdf [Accessed 25 January 2022]

Reply Edit Delete

Maximum rating: -

Add your reply

Your subject

Type your post

Dateien auswählen Keine ausgewählt

Submit

Use advanced editor and additional options

OLDER DISCUSSION

Initial Post

NEWER DISCUSSION

Initial Post