Peer Response

Reply to Post 1

I agree that there have been rising complaints of legitimate content being blocked inappropriately by Blocker Plus, violating the ACM principles. Techno-regulation can be undemocratic, especially when transparency and accountability are not upheld (van den Berg and Keymolen, 2017, pp.192). The circulation of illegitimate or damaging content online or the exclusion of valid and helpful content stems from the collective obligation of users, moderators, and automated tools like machine learning. However, these components are not necessarily to blame for inappropriate blocking of helpful content because content moderation is overall expansive and complex (Zeng and Kaye, 2022, pp.81). Most systems face uncertain operating conditions characterized by sudden changes in resource availability or unpredictable user behavior (Weyns et al., 2022, pp.1). Sometimes false positives or negatives may be evident even when human moderators are used.

Blocker Plus processes user-generated data, but the filter itself may categorize some items as harmful, leading to their automated removal, or classify them as unharmful, making the items visible online. On the other hand, the filter's machine learning technology may also incorrectly categorize some items as uncertain and potentially harmful making them subject to human review, after which they are blocked or published. Content can be subjected to human review based on the user's grievance against automated deletion and violation of the ACM principles. Automation is characterized by 'scientific' impartiality that only benefits platform companies that

consider their decisions non-negotiable (Gorwa et al., 2020, pp.12). Moreover, artificial intelligence filters often bypass traditional checks and balances covered by the law, creating barriers to oversight and hiding essential value tradeoffs and choices (Elkin-Koren, 2020, pp.1). Human review rectifies the past outcomes of automated filtering and provides necessary feedback to improve future performance. Therefore, as much as the machine learning approach is used to moderate internet content, human assessments should be incorporated to prevent the inappropriate removal of helpful content.

References List

Elkin-Koren, N. (2020). Contesting algorithms: Restoring the public interest in content filtering by artificial intelligence. *Big Data & Society*, 7(2), p.205395172093229. doi:10.1177/2053951720932296.

Gorwa, R., Binns, R. and Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1), p.205395171989794. doi:10.1177/2053951719897945.

van den Berg, B. and Keymolen, E. (2017). Regulating security on the Internet: control versus trust. *International Review of Law, Computers & Technology*, 31(2), pp.188–205. doi:10.1080/13600869.2017.1298504.

Weyns, D., Gheibi, O., Quin, F. and Van Der Donckt, J. (2022). Deep Learning for Effective and Efficient Reduction of Large Adaptation Spaces in Self-adaptive Systems. *ACM Transactions on Autonomous and Adaptive Systems*, 17(1-2), pp.1–42. doi:10.1145/3530192.

Zeng, J. and Kaye, D. Bondy V. (2022). From content moderation to visibility moderation: A case study of platform governance on TikTok. *Policy & Internet*. doi:10.1002/poi3.287.

Reply to Post 2

Indeed, Rogue Services hosts several malicious services and websites which violate the ACM principles. The ACM code of ethics is the foundation for solutions in case violations happen because it comprises principles created as declarations of responsibility stemming from the awareness that maximizing public good should be prioritized (*The Code Affirms an Obligation of Computing Professionals to Use Their Skills for the Benefit of Society.*, 2018). However, Rogue facilitates the harm caused by their customers by permitting the hosting of malicious software, violating Principles 1.1 and 1.2. (*Case: Malware Disruption*, 2018). For instance, during its Malware disruption, Rogue violated Principle 2.8 since the ISP were always aware that their devices were hosting code that caused unauthorized viruses and did not prioritize the public good, violating Principle 3.1. (*Case: Malware Disruption*, 2018). Although Rogue cloud service providers are founded in jurisdictions with lax cybercrime regulation and can offer private hosting and data storage services at a steep fee, such services are often abused by cybercriminals to store and circulate criminal information, including child abuse materials, for commercial reasons to evade law enforcement agencies scrutiny.

Compromised websites like Rogue's are often vulnerable to attackers distributing malicious content or hosting phishing pages to steal confidential data from their victims.

Although Rogue's unethical conduct was deliberate, some targeted web hosting providers are controlled by users with no security background and are unable to identify these threats or the need to contact external professional security services to address the issues (Canali et al., 2013). Reliability is the most crucial constituent of web hosting services (Li, 2022). However, most providers cannot detect the simplest warning signs of malicious activity on their websites (Canali et al., 2013). Although the ACM principles can help providers like Rogue Services deal with contemporary cyber threats, they are often ignored. The code's principles apply to various ethical challenges, including malware disruption (Gee, 2018). Therefore, Rogue Services could mitigate the impacts of the malware disruption if it had observed the ACM recommendations.

References List

Canali, D., Balzarotti, D. and Francillon, A. (2013). The role of web hosting providers in detecting compromised websites. *Proceedings of the 22nd international conference on World Wide Web - WWW '13*. [online] doi:10.1145/2488388.2488405.

Case: Malware Disruption. (2018, July 10). ACM Ethics. https://ethics.acm.org/code-of-ethics/using-the-code/case-malware-disruption/.

Gee, S. (2018). *ACM Revises Code of Ethics*. [online] www.i-programmer.info. Available at: https://www.i-programmer.info/news/99-professional/11997-acm-revises-code-of-ethics.html [Accessed 26 Sep. 2022].

Li, L. (2022). *7 Best Shared Web Hosting Services in 2022*. [online] Management Library. Available at: https://managementhelp.org/shared-web-hosting-services [Accessed 26 Sep. 2022].

The Code affirms an obligation of computing professionals to use their skills for the benefit of society. (2018). Acm.org. https://www.acm.org/code-of-ethics#.