Vulnerable by Design: Reconsidering User Vulnerability and Recommender Systems

Megan Nyhan[1,4,*], Josephine Griffith[2], Qin Ruan[1], Tai Tan Mai[3], Ruihai Dong[1] and Susan Leavy[1,4]

[1]University College Dublin, Dublin, Ireland
[2] University of Galway, Galway, Ireland
[3] Dublin City University, Dublin, Ireland
[4]Insight SFI Research Centre for Data Analytics, Ireland

Abstract

Recommender systems are invaluable in filtering vast amounts of information online. However, there are ethical challenges related to their objectives and design that have the potential to make some users vulnerable. Within emergent AI policy and regulation, *vulnerable users* have been given safeguarding measures to protect them against manipulation or exploitation. Vulnerable users are primarily defined as children and adults with particular characteristics. However, this definition focuses attention on the cause of vulnerability being the user's characteristics rather than the design of recommender systems. However, all users regardless of personal characteristics, may be considered vulnerable to negative effects associated with recommender algorithms. This paper examines three threads of vulnerability within recommender systems: vulnerability derived from specific user characteristics, the vulnerabilities of the recommender systems themselves and vulnerability caused by the nature of interactions between users and recommendation algorithms. This paper argues that while it is essential to offer more protection and assistance to users who are considered vulnerable by virtue of certain characteristics, it is also important to acknowledge the possibility of all users being rendered vulnerable by features of the recommendation algorithms themselves. This reconsideration of the concept of vulnerability serves to highlight the importance of researching the effects of recommender algorithms on user groups that are currently understudied.

Keywords

Recommender Systems, Vulnerable Recommender Systems, Vulnerable Users

1. Introduction

Recommender systems within large-scale online platforms are profoundly influential in society, given their role in governing the dissemination of content online. Ethical challenges that these systems pose at a societal and user level are well documented [1, 2, 3]. However, there is a need for further research on how different user groups are affected by the design of recommendation algorithms.

Measures to ensure the safeguarding of large-scale online platforms are being developed within the EU, with special consideration given to user groups commonly termed "vulnerable". These considerations are based on characteristics such as age, disability, gender, mental incapacity, physical ability, racial or ethnic origins and sexual orientation [4, 5, 6, 7]. However, research has shown that design features of recommendation algorithms can negatively affect many user groups who are outside these pre-defined categories. Indeed, following on from Riefa [8] and adopting Fineman's theoretical framework of vulnerability [9], the conceptualisation of vulnerability in the context of technology assumes that certain characteristics of user groups are the cause of negative effects of otherwise ethical systems, when in fact it is the design of the system that renders the users vulnerable. For instance, while the

Joint Proceedings of BIR 2024: 14th International Workshop on Bibliometric-enhanced Information Retrieval and IR4U2 2024: 1st Workshop on Information Retrieval for Understudied Users

*Corresponding author.
megan.nyhan@ucdconnect.ie (M. Nyhan); josephine.griffith@universityofgalway.ie (J. Griffith); qin.ruan@ucdconnect.ie (Q. Ruan); tai.tanmai@dcu.ie (T. T. Mai); ruihai.dong@ucd.ie (R. Dong); susan.leavy@ucd.ie (S. Leavy)
0000-0002-0877-7063 (M. Nyhan); 0000-0002-1560-1867 (J. Griffith); 0000-0001-5822-9260 (Q. Ruan); 0000-0001-6657-0872 (T. T. Mai); 0000-0002-2509-1370 (R. Dong); 0000-0002-3679-2279 (S. Leavy)

consumption of dieting videos may not pose harm to many users, individuals with diagnosed eating disorders who are excessively exposed to such content could become susceptible to negative impacts [10], potentially rendering them vulnerable. However, users who suffer from eating disorders under the current definition of vulnerability are not explicitly considered to be vulnerable within the context of AI systems. Given the pace of change in the design of recommendation algorithms on online platforms and their widespread use throughout society, there is a clear need therefore, for extensive research on the effect of recommender algorithms on different and understudied user groups to identify design practices that may affect them negatively.

2. The Impact of Recommender Systems on User Engagement and their Ethical Challenges

Recommender systems are vital for the effective filtering of information online with many people interacting with them daily. They have been widely deployed within various domains including electronic services, streaming platforms and social media (e.g. Netflix, Amazon, Instagram, YouTube and Facebook). At its core, recommender systems are AI systems that pair users with items (e.g. movies, books, songs, social media posts) based on previous implicit or explicit interactions between the user and the system employing several techniques to do so, the most popular being collaborative filtering and content-based filtering.

Recommender systems face a range of ethical challenges, which can be exacerbated by their design and objectives, including maximising user engagement and retention. These challenges can encompass issues relating to breaches of user privacy, behaviour manipulation, recommendation of inappropriate or harmful content, autonomy and personal identity theft, bias and marginalisation, and social effects caused by filter bubbles and information echo chambers. While the drive to maximise engagement can contribute to these problems, it is important to recognise that the ethical landscape is complex and multifaceted, often involving the interplay of multiple factors beyond solely engagement metrics. For instance, recommender systems are considered to be one of the major culprits for the creation of filter bubbles on social media [11]. Filter bubbles emerge when recommender systems selectively curate the content that an individual is exposed to due to their previous interactions, preferences and online behaviour, often limiting the user to information that aligns with their interests, beliefs and preferences [12]. Recommender systems have also faced criticism for recommending harmful and problematic content online [13]. For example, results of one study (see [14]) have shown that the chance of encountering hateful content relating to gender, ethnicity, political views, terrorism, and religion, of users between the ages of 15 and 30, had tripled between 2013-2015 [15]. This increase is considered to be a direct result of user interactions with recommender systems leading users to content that they may not have encountered on their own.

Along with design features, inherent vulnerabilities of recommender systems in turn render users vulnerable to issues concerning privacy and security. If gathered data is mishandled, this could result in privacy breaches for users and identity theft, financial loss, phishing attacks and unauthorised access to accounts. Data poisoning is a method used by attackers where they inject false or misleading data into a data set with the intention of influencing the outcomes of recommender systems trained on that data. This can lead to incorrect decisions made by the algorithm potentially causing individual and societal harm [16, 17]. Collaborative filtering is susceptible to various attacks, including profile injection [18], which involves the introduction of fake users or false ratings on items with malicious intent. As outlined in Biden's Executive Order [19], AI systems need to be created with robust security measures to prevent malicious attacks. This can be done by employing privacy-preserving techniques, businesses regularly updating the algorithms used, and continuously monitoring and employing independent auditors. Due to the inherent vulnerabilities of recommender algorithms, users are exposed to potential risks, in turn, rendering them vulnerable.

3. Three Threads of Vulnerability

Similar to Riefa [8], this paper views vulnerability through the lens of Martha Fineman's vulnerability theory. Fineman argues that society should acknowledge that vulnerability is inescapable [9]and not limited to *certain groups*, but is a fundamental aspect of the human condition. Currently, AI policies have implemented assistive and protective measures for vulnerable users of AI systems, mostly focusing on the vulnerable personal characteristics of users. AI policy has also acknowledged how systems themselves can be vulnerable. This paper evaluates concepts of vulnerability in existing AI policy (i.e. the European Commission's Digital Services Act and the European AI Act, the G7 Hiroshima summit's Proceedings, Biden's Executive Order Safe, Secure, and Trustworthy Artificial Intelligence, and the Online Safety Bill) and evalutes them in the context of a broader conceptualisation of vulnerability.

3.1. Characteristic-based Vulnerability

Riefa [8] argues that the notion of vulnerability is widely understood, transcending various disciplines and encompassing several factors including age, gender, locality and socio-economic factors along with personal or *special* characteristics such as mental incapacity and physical disability. Emergent AI policy and regulation reflects this concept of vulnerability. For example, users under the EU AI Act are protected when they are members of a vulnerable group who have faced "historical patterns of discrimination … certain age groups … persons with disabilities … or persons of certain racial or sexual orientation" [7]. This definition is also reflected in the Digital Services Act (DSA), as its exploration of vulnerability encompasses gender, race or ethnic origins, religion, disability, age (specifically minors and children), and sexual orientation [5]. This, as a result,focuses the definition of vulnerability on the personal factors in relation to users. The UK Online Safety Act also focuses on a definition of vulnerability, including several protection measures for vulnerable adult users, children, a member of a class or group with a certain characteristic, and women and girls [4]. Finally, in Biden's Executive Order, vulnerable users are described as protected groups [19]. Building on these approaches to user vulnerability, further research into the effects of recommender algorithms on different user groups would serve to broaden the conceptualisation of groups of people afforded protection and question the source of vulnerability, from a characteristic-based approach to one which critically evaluates the design of recommender algorithms.

3.2. Recommender System Vulnerability

Vulnerability in the context of recommender algorithms can also refer to the system itself. For instance, within Biden's Executive Order, vulnerability is considered in the context of openness to flaws and cyberattacks [19]. Conceiving the vulnerability of recommender algorithms in this way emphasises the need for strong security measures but does not consider potential large-scale negative societal effects as also evidence of system vulnerability.

3.3. Vulnerable by Design

Within the EU's AI Act, there is a focus on prohibiting some AI practices that "manipulate persons through subliminal techniques beyond their consciousness or exploit vulnerabilities of specific vulnerable groups such as children or persons with disabilities" [7]. The act accounts for vulnerability in a broader sense, stating that it takes into account that potentially harmed or adversely impacted persons are in a vulnerable position in relation to being a user of an AI system. This is by far the most inclusive definition of vulnerability. However, it does focus on user characteristics. It calls for the specific protection of vulnerable adults and children, against the exposure to malicious attacks and harmful content rallying for more prohibitions covering manipulative practices [7]. However, when users interact with recommendation algorithms they may be rendered vulnerable by their design. For example, as explored by Hasan, the use of recommendations, along with psychological vulnerabilities, including low-self esteem, loneliness, depression, shyness, low-self control or deficient self-regulation,

sensation seeking, social anxiety, and locus of control may lead to excessive use of video streaming services [20]. Excessive use of video streaming services, for example YouTube, has resulted in health (and other) issues amongst children and adolescents. This has been found to be a direct result of the platform's constrained autonomy (due to content recommendation) and the youth's natural need for companionship and relatability [21]. The results of a study conducted by Scully et al., found that an individual's dissatisfaction with their own body image is significantly related to time spent engaged in social comparisons, specifically when engaged with female content creators whilst using online platforms [21].

4. Uncovering Vulnerabilities

Reconsidering the definition of vulnerability as something that can be caused by the design of an algorithm, rather than something inherent within a person, prompts the need for further research to explore the effects of recommender systems on previously understudied users. It disrupts assumptions that susceptibility to the negative effects of recommender algorithms may be limited to those who are already considered vulnerable in other policy domains.

Given the protections afforded to vulnerable users within the existing and emergent digital policy, it follows that further research into the vulnerabilities caused by interaction with recommender systems will result in an extension of those protections to new user groups. While there has been much research on the overall societal effects of recommender systems such as polarisation, for instance, there is a clear need for further research focusing on the effects of these algorithms on an individual and user-group level.

5. Conclusion

This paper highlights the need to reconsider vulnerability beyond an alignment with user characteristics. Re-framing the concept of vulnerability in the context of recommender systems as something that can be caused by the system rather than the person necessitates new research on the susceptibility of presently under-studied users. Uncovering new paradigms of vulnerability to the negative effects of recommender systems would therefore lead to a more comprehensive and adaptable application of existing protections within digital policy and regulation.

6. Acknowledgements

This work was conducted with the financial support of the Research Ireland Centre for Research Training in Digitally-Enhanced Reality (d-real) under Grant No. 18/CRT/6224. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: Grammar and spelling check, and reword. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] S. Milano, M. Taddeo, L. Floridi, Recommender systems and their ethical challenges, AI & Society 35 (2020) 957–967.

[2] E. Karakolis, P. F. Oikonomidis, D. Askounis, Identifying and addressing ethical challenges in recommender systems, 2022 13th International Conference on Information, Intelligence, Systems & Applications (IISA) (2022) 1–6.
[3] D. Paraschakis, Algorithmic and ethical aspects of recommender systems in e-commerce, Proceedings of the 10th ACM conference on Recommender Systems (2018).
[4] Parliament of the United Kingdom, Online safety act 2023 (2023). URL: https://www.legislation.gov.uk/ukpga/2023/50/pdfs/ukpga_20230050_en.pdf.
[5] The European Commission, Regulation of the european parliament and of the council on a single market for digital services (digital services act) and amending directive 2000/31/ec (2020). URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0825.
[6] OECD, G7 Hiroshima Process on Generative Artificial Intelligence (AI), 2023. URL: https://www.oecd-ilibrary.org/content/publication/bf3c0c60-en.
[7] The European Commission, Laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts (2021). URL: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF.
[8] C. Riefa, Protecting vulnerable consumers in the digital single market, European Business Law Review 33 (2022).
[9] M. Albertson Fineman, Vulnerability and Inevitable Inequality, Oslo Law Review 4 (2017) 133–149. URL: http://www.idunn.no/doi/10.18261/issn.2387-3299-2017-03-02. doi:10.18261/issn.2387-3299-2017-03-02.
[10] J. Stray, A. Halevy, P. Assar, D. Hadfield-Menell, C. Boutilier, A. Ashar, C. Bakalar, L. Beattie, M. Ekstrand, C. Leibowicz, C. Moon Sehat, S. Johansen, L. Kerlin, D. Vickrey, S. Singh, S. Vrijenhoek, A. Zhang, M. Andrus, N. Helberger, P. Proutskova, T. Mitra, N. Vasan, Building human values into recommender systems: An interdisciplinary synthesis, ACM Trans. Recomm. Syst. (2023). doi:10.1145/3632297.
[11] Q. M. Areeb, M. Nadeem, S. S. Sohail, R. Imam, F. Doctor, Y. Himeur, A. Hussain, A. Amira, Filter bubbles in recommender systems: Fact or fallacy—a systematic review, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 13 (2023) e1512.
[12] N. Helberger, K. Karppinen, L. D'acunto, Exposure diversity as a design principle for recommender systems, Information, Communication & Society 21 (2018) 191–207.
[13] D. O'Callaghan, D. Greene, M. Conway, J. Carthy, P. Cunningham, Down the (white) rabbit hole: The extreme right and online recommender systems, Social Science Computer Review 33 (2015) 459–478. doi:10.1177/0894439314555329.
[14] M. Kaakinen, A. Oksanen, P. Räsänen, Did the risk of exposure to online hate increase after the November 2015 Paris attacks? A group relations approach., Computers in Human Behavior 78 (2018) 90–97. doi:10.1016/j.chb.2017.09.022, place: Netherlands Publisher: Elsevier Science.
[15] M. Yesilada, S. Lewandowsky, A systematic review: The youtube recommender system and pathways to problematic content (2021).
[16] H. Huang, J. Mu, N. Z. Gong, Q. Li, B. Liu, M. Xu, Data poisoning attacks to deep learning based recommender systems, in: Proceedings 2021 Network and Distributed System Security Symposium, Internet Society, 2021. doi:10.14722/ndss.2021.24525.
[17] M. Fang, N. Z. Gong, J. Liu, Influence function based data poisoning attacks to top-n recommender systems, in: Proceedings of The Web Conference 2020, WWW '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 3019–3025. doi:10.1145/3366423.3380072.
[18] C. A. Williams, B. Mobasher, R. Burke, Defending recommender systems: Detection of profile injection attacks, Service Oriented Computing and Applications 1 (2007) 157–170. doi:10.1007/s11761-007-0013-0.
[19] T. W. House, FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence, 2023. URL: https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/.
[20] M. R. Hasan, A. K. Jha, Y. Liu, Excessive use of online video streaming services: Impact of

recommender system use, psychological factors, and motives, Computers in Human Behavior 80 (2018) 220–228. doi:https://doi.org/10.1016/j.chb.2017.11.020.

[21] M. Scully, L. Swords, E. Nixon, Social comparisons on social media: online appearance-related activity and body dissatisfaction in adolescent girls, Irish Journal of Psychological Medicine 40 (2023) 31–42.