Cascade! Human in the loop shortcomings can increase the risk of failures in recommender systems

Wm. Matthew Kennedy
Oxford Internet Institute
University of Oxford
Oxford, United Kingdom
matt.kennedy@oii.ox.ac.uk

Nishanshi Shukla
Western Governors University
Millcreek, UT, United States
nishanshi.shukla@gmail.com

Cigdem Patlak
Independent
Irvine, CA, United States
cigdem.patlak@gmail.com

Blake Chambers
Independent
Boston, MA, United States
blakejwc@alum.mit.edu

Theodora Skeadas
Humane Intelligence
Boston, MA, United States
Theodora@humane-intelligence.org

Tuesday
ARTIFEX Labs
United States
Tuesday@artifex.fun

Kingsley Owadara
Pan-Africa Center for AI Ethics
Lagos, Nigeria
Kowdara@yahoo.com

Aayush Dhanotiya
Amazon Inc.
Seattle, WA, United States
aayush.dhanotia@gmail.com

ACM Reference Format:
Wm. Matthew Kennedy, Nishanshi Shukla, Cigdem Patlak, Blake Chambers, Theodora Skeadas, Tuesday, Kingsley Owadara, and Aayush Dhanotiya. . Cascade! Human in the loop shortcomings can increase the risk of failures in recommender systems. In *Proceedings of Workshop on Socially Responsible Recommender Systems (FAccTRec@RecSys).* ACM, New York, NY, USA, 3 pages.

1 Introduction

Recommender systems are among the most commonly deployed systems today. Systems design approaches to AI-powered recommender systems (that is, systems that are both data-driven and employ computationally expensive algorithms such as neural networks) have done well to urge recommender system developers to follow more intentional data collection, curation, and management procedures [3, 4, 18]. So too has the "human-in-the-loop" paradigm been widely adopted (at least nominally), primarily to address the issue of accountability [6, 21, 22].

However, in this paper, we take the position that human oversight in recommender system design also entails novel risks that have yet to be fully described. These risks are "codetermined" [24] by the information context in which such systems are often deployed. Furthermore, new knowledge of the shortcomings of "human-in-the-loop" practices to deliver meaningful oversight of other AI systems suggest that they may also be inadequate for achieving socially responsible recommendations. We review how the limitations of human oversight may increase the chances of a specific kind of failure: a "cascade" or "compound" failure. We then briefly explore how the unique dynamics of three common deployment contexts can make humans in the loop more likely to fail in their oversight duties. We then conclude with two recommendations.

FAccTRec@RecSys,
.

2 The cascade problem

Among the best known failure modes of recommender systems is the "data cascade" failure. Cascade failures might be thought of as compound failures, in that they are caused by several apparently trivial failures bootstrapping into a larger, more significant failure. These properties make them exceedingly difficult to identify, much less mitigate, in production systems. Likewise, they subject users of that system to a prolonged series of low-value recommendations, which may in fact accelerate the compounding effect of a cascade failure. Cascade failures occur for three reasons primarily: low data quality, poor system design, and insufficient oversight (including human oversight). It is the latter category on which this paper focuses [13].

3 Current human-in-the-loop approaches may be insufficient

The scientific literature on human oversight of recommendation systems is growing but not yet mature. Where human-in-the-loop approaches have been discussed or applied in the context of improving recommender systems, it has usually taken the form of human evaluation of the information retrieved via user searches or the information quality of the search space itself. As a result, the literature has tended to focus on search-based recommender systems [21] or as a proposed measure to resolve longstanding technical challenges to recommender system design – for instance the "long tail" problem [6] that requires systems to produce recommendations for items or content that itself is not highly subscribed [25].

With time, however, we are beginning to understand that "human in the loop" practices vary widely in their configuration and indeed in their quality [19]. This variance is often a result of human component failures. Human in the loop failures take many forms that are consequential to recommender systems. For instance, recent research from the field of algorithmic decision-making has demonstrated that humans in "in the loop" governance functions

provided "correct" oversight only about half the time - lapses primarily caused by human motivation to ensure compliance with their organization's goals rather than responsible AI principles [7].

Considering that humans in the loop in recommender systems are involved "upstream" in data quality evaluation functions, humans in the loop may be inadequate to regularly perform such data quality duties. AI-powered recommender systems may pose special challenges. For instance, humans tasked with evaluating the data quality or information retrieval components of a conversational recommender system that employs LLMs to elicit user preferences will encounter highly context-specific natural language. Their assessments may introduce new errors, inadvertently pushing the system into lower value areas of its search space, or causing the system to make low-quality associations (Squires 2006). If these kinds of degradations of the information context are not identified and mitigated quickly, they can cause recommender systems to deliver recommendations that may at first appear appropriate but that eventually lead the system (and thence the user) down the path of compounding diminishing returns [23]. In the worst case, such failures can lead to model collapse [15].

4 Examples

To concretize this conceptual exploration, we briefly review three different 'information contexts' in which recommender systems are often deployed and where incorrect solutions to long-tail problems are likely to have more substantial consequences: education, social media, and e-commerce.

4.1 Example 1: education

As education technology application developers embrace AI in their offerings, several claim to have developed 'personalized' or 'adaptive' learning solutions that renovate previous intelligent tutoring systems (ITS) to respond to individual student needs [9]. However, many edtech researchers are critical of these claims, pointing out that 'personalization' is rarely truly personalized and should rather be thought of as 'pre-programmed' [9, 10]. They also warn that classroom use of such systems can subtly steer what kinds of learning content a student encounters, affecting student mastery, advancement, and therefore educational attainment [20]. The stakes in these environments are high, and the margin for error is small. A misstep in what gets recommended or what gets withheld can shape a student's academic path in ways that are difficult to detect or reverse [11]. Furthermore, human oversight of such systems is difficult as application developers rarely provide real-time monitoring, and teacher interaction usually only comes after the fact, through dashboards, summary reports, or assessment.

4.2 Case 2: content recommender systems in social media

Recommender systems can facilitate the sharing of content from conspiracy-oriented channels, hateful content, and divisive content [1] which fuel misinformation, political division, and radicalization. Moderation algorithms make millions of content and account removal decisions daily, but many of these decisions are incorrect, informed by poor language and context understanding and therefore improper prioritization, a consequence of under-resourced human oversight. In some cases, this underresourcedness leads to unbalanced deployment of algorithmic content moderation tools. For example, a human rights due diligence of Meta's impacts in Israel and Palestine by Business for Social Responsibility found that "proactive detection rates of potentially violating Arabic content were significantly higher than proactive detection rates of potentially violating Hebrew content...[likely because] there was an Arabic hostile speech classifier but not a Hebrew hostile speech classifier" [2]. Additionally, human oversight resources are concentrated on high-value accounts, a concept referred to as 'analog privilege' [12].

4.3 Case 3: recommender systems in e-commerce

Recommender systems are widely deployed to e-commerce platforms and heavily influence user perceptions of the goods or services available for purchase and, in some cases, also producing "personalization" in pricing, though this practice is increasingly prohibited [8, 14]. In either case, recommender systems in e-commerce are critical parts of a platform's sales funnel, aiding in keeping users engaged and moving towards a transaction by surfacing useful results to queries and in helping sellers market their products to likely buyers [16]. Of course, such systems perennially encounter the long tail problem, and have been criticized for appearing to recommend already items that are already popular, making it harder for smaller sellers or more niche, ethical, or sustainable options to surface, especially on very large platforms. These systems also serve the interests of platform providers themselves, and precisely how proprietary search-based recommender systems retrieve potential recommendations is not always clear to consumers, sellers, or regulators. Many of the largest e-commerce platform providers have been accused of deploying recommender systems that unfairly shape commercial information contexts to favor their own platform or the goods that provide them with greater commercial benefits instead of items that may actually better match user searches and preferences [17]. Over time, this can subtly shape market perception and consumer behavior. More germane to this paper, this places humans in the loop in a sustained dilemma where they must choose which values–their organization's or those that comprise responsible AI–to uphold when they observe conflicts. There is now emerging evidence to suggest that, when faced with such a dilemma, humans in the loop are just as likely to align to one as with the other [7].

5 Recommendations for the field

Clearly, there is more work to be done mapping out this particular area of risk. We hope, however, that our paper establishes the importance of doing this work. To that end, we conclude with two of our own recommendations for the field of socially responsible recommender system design. Firstly, we urge recommender system designers to consider more deeply the role they are expecting humans to play in delivering high-value recommendations in all kinds of recommender systems. Further and more intentional review of the human components of these systems is critical in order to avoid their becoming "moral crumple zones" [5]. Second, we urge systems designers to investigate the limitations of humans in

evaluation functions for recommenders operating over very large search spaces that require opaque ML/AI algorithms to process. Given the critical role humans in the loop have been expected to play in ensuring data and search quality (and therefore recommendation quality), greater attention should be paid to the potential shortcomings and misallocations of human oversight capabilities.

References

[1] Cody Buntain, Richard Bonneau, Jonathan Nagler, and Joshua A. Tucker. 2021. YouTube Recommendations and Effects on Sharing Across Online Social Platforms. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 11 (April 2021), 26 pages. doi:10.1145/3449085

[2] Business and Social Responsibility. 2022. Human Rights Due Diligence of Meta's Impacts in Israel and Palestine. https://www.bsr.org/en/reports/meta-human-rights-israel-palestine

[3] Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. 2023. Bias and Debias in Recommender System: A Survey and Future Directions. *ACM Trans. Inf. Syst.* 41, 3, Article 67 (Feb. 2023), 39 pages. doi:10.1145/3564284

[4] Juan Ignacio del Valle and Francisco Lara. 2023. AI-powered recommender systems and the preservation of personal autonomy. *AI Soc.* 39, 5 (July 2023), 2479–2491. doi:10.1007/s00146-023-01720-2

[5] Madeleine Clare Elish. 2019. Moral Crumple Zones:Cautionary Tales in Human-Robot Interaction. *Engaging Science, Technology, and Society* 5 (2019), 40–60. doi:10.17351/ests2019.260

[6] Zuohui Fu, Yikun Xian, Shijie Geng, Gerard de Melo, and Yongfeng Zhang. 2021. Popcorn: Human-in-the-loop Popularity Debiasing in Conversational Recommender Systems. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management* (Virtual Event, Queensland, Australia) *(CIKM '21)*. Association for Computing Machinery, New York, NY, USA, 494–503. doi:10.1145/3459637.3482461

[7] Alexia Gaudeul, Ottla Arrigoni, Vasiliki Charisi, Marina Escobar Planas, and Isabelle Hupont Torres. 2025. *The Impact of Human-AI Interaction on Discrimination.* Technical Report KJ-01-24-180-EN-N (online). European Commission Joint Research Center, Luxembourg (Luxembourg). doi:10.2760/0189570(online)

[8] Axel Gautier, Ashwin Ittoo, and Pieter Cleynenbreugel. 2020. AI algorithms, price discrimination and collusion: a technological, economic and legal perspective. *European Journal of Law and Economics* 50, 3 (2020), 405–435. https://EconPapers.repec.org/RePEc:kap:ejlwec:v:50:y:2020:i:3:d:10.1007_s10657-020-09662-6

[9] Wayne Holmes, Stamatina Anastopoulou, Heike Schaumburg, and Manolis Mavrikis. 2018. Technology-enhanced Personalised Learning: Untangling the Evidence. Robert Bosch Stiftung GmbH. http://oro.open.ac.uk/56692/1/TEPL_en.pdf

[10] Irina Jurenka, Markus Kunesch, Kevin R. McKee, Daniel Gillick, Shaojian Zhu, Sara Wiltberger, Shubham Milind Phal, Katherine Hermann, Daniel Kasenberg, Avishkar Bhoopchand, Ankit Anand, Miruna Pîslar, Stephanie Chan, Lisa Wang, Jennifer She, Parsa Mahmoudieh, Aliya Rysbek, Wei-Jen Ko, Andrea Huber, Brett Wiltshire, Gal Elidan, Roni Rabin, Jasmin Rubinovitz, Amit Pitaru, Mac McAllister, Julia Wilkowski, David Choi, Roee Engelberg, Lidan Hackmon, Adva Levin, Rachel Griffin, Michael Sears, Filip Bar, Mia Mesar, Mana Jabbour, Arslan Chaudhry, James Cohan, Sridhar Thiagarajan, Nir Levine, Ben Brown, Dilan Gorur, Svetlana Grant, Rachel Hashimshoni, Laura Weidinger, Jieru Hu, Dawn Chen, Kuba Dolecki, Canfer Akbulut, Maxwell Bileschi, Laura Culp, Wen-Xin Dong, Nahema Marchal, Kelsie Van Deman, Hema Bajaj Misra, Michael Duah, Moran Ambar, Avi Caciularu, Sandra Lefdal, Chris Summerfield, James An, Pierre-Alexandre Kamienny, Abhinit Mohdi, Theofilos Strinopoulous, Annie Hale, Wayne Anderson, Luis C. Cobo, Niv Efron, Muktha Ananda, Shakir Mohamed, Maureen Heymans, Zoubin Ghahramani, Yossi Matias, Ben Gomes, and Lila Ibrahim. 2024. Towards Responsible Development of Generative AI for Education: An Evaluation-Driven Approach. arXiv:2407.12687 [cs.CY] https://arxiv.org/abs/2407.12687

[11] Wm. Matthew Kennedy and Daniel Vargas Campos. Forthcoming 2025. A Vernacularized Taxonomy of Harms for AI in Education. In *Handbook on Critical Studies of AI in Education*, Wayne Holmes (Ed.). Edward Elgar, London.

[12] Maroussia Lévesque. 2024. Analog Privilege. *New York University Journal of Legislation and Public Policy* 26 (08 2024). Issue 3. https://ssrn.com/abstract=4528278

[13] Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M Aroyo. 2021. "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 39, 15 pages. doi:10.1145/3411764.3445518

[14] Peter Seele, Claus Dierksmeier, Reto Hofstetter, and Mario D. Schultz. 2019. Mapping the Ethicality of Algorithmic Pricing: A Review of Dynamic and Personalized Pricing. *Journal of Business Ethics* 170, 4 (2019), 697–719. doi:10.1007/s10551-019-04371-w

[15] Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross J. Anderson, and Yarin Gal. 2024. AI models collapse when trained on recursively generated data. *Nat.* 631, 8022 (July 2024), 755–759. https://doi.org/10.1038/s41586-024-07566-y

[16] Daria Sorokina and Erick Cantu-Paz. 2016. Amazon Search: The Joy of Ranking Products. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Pisa, Italy) *(SIGIR '16)*. Association for Computing Machinery, New York, NY, USA, 459–460. doi:10.1145/2911451.2926725

[17] Ilan Strauss, Tim O'Reilly, and Mariana Mazzucato. 2024. Amazon's Algorithmic Rents: The economics of information on Amazon. *UC Law Science and Technology Journal* 15 (2024). Issue 2. https://repository.uclawsf.edu/hastings_science_technology_law_journal/vol15/iss2/5

[18] Jonathan Stray, Alon Halevy, Parisa Assar, Dylan Hadfield-Menell, Craig Boutilier, Amar Ashar, Chloe Bakalar, Lex Beattie, Michael Ekstrand, Claire Leibowicz, Connie Moon Sehat, Sara Johansen, Lianne Kerlin, David Vickrey, Spandana Singh, Sanne Vrijenhoek, Amy Zhang, McKane Andrus, Natali Helberger, Polina Proutskova, Tanushree Mitra, and Nina Vasan. 2024. Building Human Values into Recommender Systems: An Interdisciplinary Synthesis. *ACM Trans. Recomm. Syst.* 2, 3, Article 20 (June 2024), 57 pages. doi:10.1145/3632297

[19] Mark Tsagas. 2024. Human oversight of AI systems may not be as effective as we think — especially when it comes to warfare. Retrieved 18 July 2025 from https://theconversation.com/human-oversight-of-ai-systems-may-not-be-as-effective-as-we-think-especially-when-it-comes-to-warfare-230322

[20] UNESCO. 2023. *Guidance on Generative AI in Education and Research.* Technical Report. UNESCO.

[21] Dmitry Ustalov, Natalia Fedorova, and Nikita Pavlichenko. 2022. Improving Recommender Systems with Human-in-the-Loop. In *Proceedings of the 16th ACM Conference on Recommender Systems* (Seattle, WA, USA) *(RecSys '22)*. Association for Computing Machinery, New York, NY, USA, 708–709. doi:10.1145/3523227.3547373

[22] Roanne van Voorst. 2024. Challenges and Limitations of Human Oversight in Ethical Artificial Intelligence Implementation in Health Care: Balancing Digital Literacy and Professional Strain. *Mayo Clinic Proceedings: Digital Health* 2, 4 (01 Dec 2024), 559–563. doi:10.1016/j.mcpdig.2024.08.004

[23] Sandra Wachter, Brent Mittelstadt, and Chris Russell. 2024. Do large language models have a legal duty to tell the truth? *Royal Society Open Science* 11, 8 (2024).

[24] Laura Weidinger, Maribeth Rauh, Nahema Marchal, Arianna Manzini, Lisa Anne Hendricks, Juan Mateos-Garcia, Stevie Bergman, Jackie Kay, Conor Griffin, Ben Bariach, Iason Gabriel, Verena Rieser, and William Isaac. 2023. Sociotechnical Safety Evaluation of Generative AI Systems. arXiv:2310.11986 [cs.AI] https://arxiv.org/abs/2310.11986

[25] Zhipeng Zhao, Kun Zhou, Xiaolei Wang, Wayne Xin Zhao, Fan Pan, Zhao Cao, and Ji-Rong Wen. 2023. Alleviating the Long-Tail Problem in Conversational Recommender Systems. In *Proceedings of the 17th ACM Conference on Recommender Systems* (Singapore, Singapore) *(RecSys '23)*. Association for Computing Machinery, New York, NY, USA, 374–385. doi:10.1145/3604915.3608812