Social Influence for Societal Interest: A Pro-Ethical Framework for Improving Human Decision-Making Through Multi-Stakeholder Recommender Systems

Matteo Fabbri [1]

[1] *IMT School for Advanced Studies, Lucca, Italy (matteo.fabbri@imtlucca.it)*

Keywords
AI Ethics, Recommender Systems, Social Influence, Explainability

[Full paper published in *AI & Society* (2022): https://doi.org/10.1007/s00146-022-01467-2]

In the contemporary digital age, recommender systems (RSs) play a fundamental role in managing information in online platforms: from social media to e-commerce, from travels to cultural consumptions, automated recommendations influence the everyday choices of users at an unprecedented scale. RSs profile individuals to target them with suggestions about content they may like with the aim of reducing information overload. However, the presence of algorithmic bias poses pressing questions as regards the ethical and social implications of RSs. Moreover, the information filtering processes managed by RSs can influence users' political behaviour, especially when the recommendations concern the selection of news [1]. For these reasons, users should understand not only "what recommenders recommend" [2], but also why they recommend what they recommend. In other words, it is essential to explain the relationship between the design of RSs and their influence on users' choices.

Whilst a wide array of research centred on explainability in RSs has been produced during the last decade[1], there still is a significant lack of scholarship addressing the ethical and social implications of RSs. In this article, I trace a conceptual framework for directing the influence of automated recommendations towards societal interest. To reach this aim in the context of RSs, it is useful to consider the problem of incentives for explainable recommendations, both on the users' and on the companies' side. In fact, the ultimate impact of RSs concerns all the multiple stakeholders involved in the recommendation process. According to the multi-stakeholder recommender systems (MRSs) framework proposed by [5], there are four stakeholders in a recommendation: users, providers, the system and society. Users are "the parties to whom the recommendation is targeted"; providers are the subjects "who make the options available", who sell their product or service through the platform; the system refers to "the interests of the platform on which the recommendations are generated"; and finally society is the collective stakeholder upon which "recommendations made by a system can have systemic effects […] for example by altering or reinforcing existing social norms", or by modifying a social environment [*ibidem*]. The ontology of MRSs can provide the grounding for an approach to recommendations that addresses their influence in the perspective of social good.

However, policies centred on explainability within a multi-stakeholder framework may be ineffective, because the firms who design and own RSs do not have appropriate incentives to make their functioning transparent. Moreover, from a user-centred perspective, this approach could be considered paternalistic, as it would put the designer's ethical evaluations above the user's interest. Indeed, a user may prefer to receive recommendations about items or contents which are closer to their expected preferences than to

[1] To have an idea of the evolution of this research in about a decade, compare [3] and [4].

EWAF'23: European Workshop on Algorithmic Fairness, June 07–09, 2023, Winterthur, Switzerland

socially preferable outcomes. In this case, a recommendation policy aimed at fostering social good may not be tolerant towards users' attitudes and choices. Therefore, two main objections can be raised here: on the one side, the lack of incentives for private companies to modify their RSs may undermine the effectiveness of the policy; on the other side, the impact of the policy on users' range of choices may limit their freedom to an even greater extent, because they will be exposed to pre-determined contents that are not linked to their preferences.

To answer these objections, I frame the approach presented above according to an argument put forward by [6] about the relationship between toleration and paternalism in the design of digital technologies. He argues that "one form of paternalism, based on pro-ethical design, can be compatible with toleration […], by operating only at the informational and not at the structural level of a choice architecture". Within this framework, the designer does not aim at orienting the user to *de facto* pre-determined choices, but he/she forces the user to make a choice before the latter is able to enjoy the service provided the technology. In the case of RSs, this kind of informational nudging would imply that the system may ask users questions about the contents that are going to be recommended or the categories through which the recommendation is informed.

From the perspective of social good, this approach to RSs has the advantage of making the users aware of the potential implications of their preferences without constraining or biasing the range of contents which they are exposed to. Therefore, this framework would balance paternalism with toleration through enhancing users' awareness without limiting their autonomy. Pro-ethical design applied to RSs can also contribute to address the question about incentives for private firms: in fact, informational nudging would allow companies to gather data about users' explicit preferences, thereby making recommendations more targeted and precise. In this regard, I discuss the application of beneficent informational nudging to the case of conversational recommender systems (CRSs), which rely on user-system dialogic interactions. Subsequently, through a comparison with standard recommendations, I outline the incentives for platforms and providers in adopting this approach and its benefits for both individual users and society (*see table*).

	Conversational recommendations	**Standard recommendations**
Users	1. Improved relevance derived from users' explicit input and feedback 2. Increased interpretability and transparency as explanations are embedded by design in the system	1. Focus on accuracy metrics and exploitative feedback effects 2. Problems with explicability due to black-box models, proprietary constraints, and lack of direct user-system interaction
Providers	1. Incentive for micro-targeted ads as individual users provide explicit ready-made personal data on preferences and interests 2. Potential higher profitability (as users are more likely to be interested in and eventually buy the products advertised)	1. Targeting often relying on low-quality implicit data 2. Higher risk of irrelevant or repetitive ads that fail to increase product sales
System	1. Profiling based on explicit data voluntarily provided by users (less potential privacy breaches) 2. Increased likelihood of diversifying the recommendation policies thanks to more granular data	1. Profiling often based on implicit data such as digital traces, click-through rates and browsing history 2. Increased likelihood of keeping the same recommendation policy (often exploitation)
Society	1. Informational nudging both ex ante and ex post 2. Increased individual and collective awareness of the socio-technical structure and ethical implications of the process thanks to dialogic explanations	1. Only ex post informational nudging 2. Limited understanding of the distribution of the interests at stake and its connection with the structure of the system (unaccountable social influence)

Comparison of conversational and standard RSs from the perspective of their impact on the different stakeholders (Table taken from the full paper: https://link.springer.com/article/10.1007/s00146-022-01467-2/tables/2).

References

(References of the full paper: https://link.springer.com/article/10.1007/s00146-022-01467-2#Bib1)

[1] Mittelstadt, B. (2016). Automation, algorithms, and politics| auditing for transparency in content personalization systems. *International Journal of Communication*, *10*, 12.

[2] Jannach, D., Lerche, L., Kamehkhosh, I., & Jugovac, M. (2015). What recommenders recommend: an analysis of recommendation biases and possible countermeasures. User Modeling and User-Adapted Interaction, 25(5), 427-491.

[3] Tintarev, N., & Masthoff, J. (2007, April). A survey of explanations in recommender systems. In *2007 IEEE 23rd international conference on data engineering workshop* (pp. 801-810). IEEE.

[4] Zhang, Y., & Chen, X. (2018). Explainable recommendation: A survey and new perspectives. *arXiv preprint arXiv:1804.11192*.

[5] Milano, S., Taddeo, M., & Floridi, L. (2021). Ethical aspects of multi-stakeholder recommendation systems. *The information society*, *37*(1), 35-45.

[6] Floridi, L. (2016). Tolerant paternalism: Pro-ethical design as a resolution of the dilemma of toleration. *Science and Engineering Ethics*, *22*(6), 1669-1688.