Towards Responsible Recommender Systems

Przemysław Kazienko, Wrocław University of Science and Technology, Poland
Erik Cambria, Nanyang Technological University, Singapore

Abstract—Recommender systems have transformed our digital experiences in many regards. We enumerate their six positive effects on the economy and humans, such as greater user satisfaction, time savings, broadening user horizons, or positive behavioral nudging. However, it is crucial to acknowledge the potential downsides inherent in their design. One significant concern is that these algorithms often prioritize the interests of the company deploying them, aiming to maximize profits and user engagement rather than solely focusing on enhancing user experience. Therefore, we also list and consider two use cases and six negative long-term impacts on humans, including addiction, reduced ability to think critically, less autonomy, or weakened human relationships caused by more and more human-like virtual assistants. Despite the undeniable utility of recommender systems, it is imperative to approach them critically, advocating for transparency, ethical considerations, and user empowerment to ensure that they serve as tools for enrichment rather than exploitation. To accomplish this, the idea and challenges of the Responsible Recommender System (RRS) are presented. RRS extends common recommender systems with components related to individual human values and goals as well as widely accepted well-being and lifestyle guidelines.

Recommender systems (RSs) mark a significant leap forward in how we navigate and interact with online content. Through the utilization of sophisticated algorithms, these systems have revolutionized our online experiences by offering personalized recommendations tailored to individual preferences and behaviors. While undoubtedly advantageous in numerous aspects, it is crucial to recognize the inherent drawbacks in their design. One notable concern is that these algorithms often prioritize the interests of the deploying company, aiming to maximize profits and user engagement rather than solely focusing on enhancing user experience. Consequently, there is a risk that the recommendations provided may not consistently align with users’ best interests, potentially leading to the formation of echo chambers, filter bubbles, or even manipulation of user behavior.

Despite their undeniable usefulness, it is essential to approach RSs with a critical perspective, advocating for transparency, ethical considerations, and user empowerment to ensure that they function as tools for enrichment rather than exploitation. In this paper, we discuss these issues under the lens of recent developments in RSs and artificial intelligence (AI) and propose recommendations for the future. In particular, the remainder of the paper is organized as follows: firstly, we illustrate the goals of business RSs; secondly, we discuss both positive and negative impacts of RSs on society; next, we provide recommendations for future developments of RSs; finally, we offer concluding remarks.

GOALS OF BUSINESS RECOMMENDER SYSTEMS

Most of the evaluations and metrics of RSs focus on validation of the optimization process while finding the items of user interest in very large item collections. In
most recommendation cases, we commonly have two players: (1) users who are advised with some items and (2) item providers. In business RS applications, users are a weaker partner, as they do not design, control, or understand the limitations of the recommendation engines. Reciprocity or fairness mechanisms could address these concerns, but they are relatively rarely considered in the context of corporate RSs.

Companies are driven by business objectives, such as greater profit, market size occupied, as well as user time or attention gained. For them, human needs and satisfaction are, then, only auxiliary goals. This means that companies must respect their customers, but only to the extent that it enables them to achieve their objectives and earn money. Accordingly, they employ RSs to support their customers in decision-making processes resulting in purchases, reviews, etc., which directly impact their income (Fig. 1). Then, can we claim that RSs just simply provide a recommendation list, satisfying users? They primarily satisfy companies themselves, then users.

Moreover, the evaluation measures commonly used in scientific papers are being adapted in practical implementations to the business models used. Obviously, it does not mean that companies are against their customers. They just focus on their business. In summary, commercial RSs are driven by business interests rather than the interests of their users. Therefore, they are commonly beyond the control and awareness of customers.

**POSITIVE RS IMPACT ON ECONOMY AND HUMANS**

The main goals of recommender systems and their positive impact on the economy and humans (Fig. 1) are:

- **P11:** Positive economic impact on companies using RSs [1], i.e., personalization provided by RSs can be a competitive advantage over market rivals;
- **P12:** Better user experience, which directly results from support of RSs in navigation and communication with the system;
- **P13:** Better user satisfaction can be one of the measured feature of RSs making users utilize them [2];
- **P14:** Time-saving is a consequence of P12, i.e., faster navigation through online services and large item collections [3];
- **P15:** Broadening of horizons – some RS suggestion of items beyond known user preferences [4]. It is also addressed using quite well known concepts and measures: diversity, coverage, novelty, unexpectedness, and serendipity [5];
- **P16:** Nudging users towards their positive decisions and behavior, e.g., related to unhealthy eating [6] or news diversity.

**POTENTIAL LONG-TERM NEGATIVE RS IMPACT ON HUMANS**

Most research related to recommender systems focuses on better immediate recommendations, i.e., more precisely infer about user needs while directly fulfilling business goals. However, users exposed to recommender systems can experience some long-term effects, including negative ones (Fig. 1). In particular, we would like to highlight six of them that we believe require further investigation:

- **N11:** Information bubble – greater user confinement;
- **N12:** Addiction to very good prompts that excellently meet user needs; excessive use of RSs;
- **N13:** Lower critical thinking ability due to better and better recommendations (why seek anything else?);
- **N14:** Less user autonomy to make their own choices.
- **N15:** Replacing human advisors who become inferior and more expensive;
- **N16:** Weaker human relationships as more and more user needs are filled by systems.

All the abovementioned effects may lead to other effects like decreased well-being or physical and mental health.
There are also other adverse effects on humans, which often directly result from the contradictory goals of sellers and customers, like nudging user moods to induce unplanned purchases [7]. The phenomenon commonly called information bubble (NI1) stems from human to consume increasingly similar items over time even without any recommendation. However, RSs can reinforce this effect, even if the user may sometimes feel bored or less satisfied [8]. On the other hand, there are RS solutions going in the opposite direction, i.e., broadening user horizons (PI3). Unfortunately, however, are not commonly used in commercial applications.

User autonomy (NI4) refers to the user’s ability to undertake their free choices [9], i.e., possibly without any manipulation. This is also closely related to user control over their decisions and recommendation mechanisms [10]. The loss of such control may partially result from being in the information bubble (NI1), excessive RS use (NI2), and loss of criticism (NI3).

**TOWARDS HUMAN-LIKE RS**

Recently, companies and other stakeholders can collect large amounts of data about their users. More data potentially leads to more accurate and friendly RSs. This is additionally supported by multimodal RSs that process diverse information about the environment, recommended items, and also user social networks. Besides, RSs can make use of the general data about users like their personality traits, cognitive abilities or more temporal affective states like emotions [11]. As a result, the current RSs, and future RS in particular, will increasingly resemble human advisors.

It will be boosted by a great progress in generative AI, which enhance RS interaction capability, and leading to conversational RS (CRS) [12]. This, in turn, is likely to be soon combined with virtual or augmented reality making the user experience with RS even more immersive and mimicking human beings [13].

This means that the development of LLM and other multimodal generative AI systems and virtual reality will make RSs more capable of personalizing multimodal and immersive interaction, while benefiting from the general human profile (personality traits, beliefs, general taste), temporal state (emotions, mood, attention) and more comprehensive behavioral user data. As a result, RSs will be developed towards human-like assistants rather than simple recommendation list generators (Fig. 2). Such interactive RSs may substitute for real human advisors (NI5). This means that users will tend to replace their natural interpersonal communication with virtual assistants, thus reducing their interpersonal contacts, which may eventually cause greater isolation from real life (Fig. 3).

Moreover, such RSs may know the user better than other people. Going further, users may establish a kind of relationship with virtual advisors. The case of Replika is a good and recent example of such a virtual engagement [14]. Perhaps future RSs will become more like our friendsimitating and substituting for human friends? This, however, may be harmful and dangerous to the users [15] and negatively affect their interpersonal relationships (NI6). At the same time, this may dovetail with another general social phenomenon: less face-to-face contacts that directly decrease life satisfaction [16].
Simultaneously, high quality and effectiveness of RSs may provide excessive use and addiction (NI2) to various digital services such as online video streaming services [17]. It means that users are becoming increasingly dependent on RS suggestions, thus, losing their critical thinking ability (NI3), and are even susceptible to changing their personal identity [18]. This is also due to the exhibit of confirmation bias by many people, that is, the tendency to actively seek information that confirms initial preferences. However, typical RSs do little to avoid this bias, since they suggest items that are in line with user’s preferences.

USE CASES
To demonstrate the problem of contradictory business goals of RS providers and human objectives we would like to consider two examples: (1) streaming services, especially VoD (Video on Demand) like Netflix or Disney+, and (2) dating services like Tinder.

Recommendations of movie series
Typically, streaming services that provide movie series jump from one episode to another without any break. This is further supported by cliffhangers at the end of the episode, i.e., suspending the action of the film so that the viewers are left in a very exciting or scary moment to encourage them to continue watching. This will be even more challenging for users in the future due to more immersive virtual reality solutions and personalization exploiting human emotions. As a result, the user can be kept in the service for many hours. However, is spending a long time like this the intention and need of the user? When launching the service, would the user consciously choose such a scenario? It means that, in many cases, user and business goals may be opposed. This is closely related to the negative long-term impacts: NI1 and NI2.

Dating services and long-term impact
If the business model of dating services relies on monthly subscriptions, then RS in such a service can be optimized to achieve business goals, i.e., to retain paying users. This may result in matching people to short-term human relationships rather than long-term ones. This, in turn, may not be in line with the long-term goals of many users seeking a more permanent relationships.

BUSINESS VS. USER AND SOCIETY-WIDE GOALS
In general, commercial RSs focus on (1) meeting users’ temporary needs (by suggesting potentially useful items), (2) while simultaneously accomplishing business goals. On the other hand, we should be aware of risks of negative impact (NI1-NI6), and challenges arising from human-like RSs. Therefore, in the landscape of rapid and continuous development of recommender systems, we would like also to consider: (3) the life goals and values of each individual, and (4) general, agreed society-wide recommendations related to health and well-being, e.g., physical and mental health guidelines. This leads us to the concept of Responsible Recommender System (RRS) (Fig. 4).

IDEA OF RESPONSIBLE RECOMMENDER SYSTEM (RRS)
Previous work referring to the term Responsible Recommender System actually focused on some of its features such as trustworthiness, fairness, accountability, explainability, and transparency. Most of these topics are covered among others, by papers presented at the series of The FAccTRec Workshop: Responsible Recommendation collocated at the RecSys conference. All of them are very important components of RRS. In this paper, however, we go further and postulate to additionally respect new issues not much considered before (Fig. 4), in particular:

1) human values and user personal goals especially related to long-term impact on them, and
2) societal recommendations related to lifestyle or health like physical activities, sleep, the need for breaks in online activity, etc.

All of the above should be taken into account while maintaining user autonomy [9] and non-conflict with business objectives. Preserving user autonomy is a crucial component of RRS mitigating the long-term negative impact NI4. It enforces that the RS designer should strive to keep the user free: free to change their choices, free to make new ones, or even to disable the system altogether.

Our RRS concept meets the postulates of digital humanism focusing on norms in order to make the technology more ethical and value-driven [19]. Please note that we focus primarily on RS’s impact on individual users. However, the social effect can be considered as well.

1Note that we are not analyzing any particular service, but only want to indicate some of the potential risks and directions for further studies.
Responsible Recommender System (RRS)

Business goals
- Commercial provider

Decisions Impact
- NI1: Information bubble
- NI2: ...
- Personal goals
- Long-term impact on user
- Explainability, transparency
- Trustworthiness, fairness, accountability

FIGURE 4. Idea of Responsible Recommender System (RRS) in a market environment. It respects: (1) individual user goals and values, especially long-term ones, (2) societal goals, e.g., lifestyle or health recommendations, and (3) business goals. RRS also (4) preserves trustworthiness, fairness, accountability, and explainability leading to better transparency.

CHALLENGES OF RESPONSIBLE RECOMMENDER SYSTEMS

New research lines. New measures

Defining, identifying, and measuring long-term impact on humans is a great challenge for research. It means we need to establish new procedures and measures suitable for (1) different kinds of recommender systems like content (news, videos) delivery or online trade, (2) different personalities, e.g., people who are more susceptible to influence and manipulation, (3) different social groups and cultures sharing the same values, e.g., religious ones.

Controlling emotions. Another challenge is to analyze and quantify the emotional engagement of users. It is important since RS providers may be tempted to exploit user emotions to achieve their business goals. This, however, may be in opposition to the user goals, values, and even health restrictions (e.g., for people with heart diseases). As a result, users should both be aware and consciously decide to what extent they want their emotions to be evoked.

All the above tasks also require new measures that would enable the quantifying impact or emotions.

New methods and technologies

Implementation of the RRS concept requires new methods and reasoning architectures that would integrate contradictory goals: business vs. personal vs. societal ones. In particular, we need to develop how to combine general and uncontroversial health recommendations in the form of incentives with personal preferences while maintaining benefits for RS operators.

Additionally, there is a demand for new methods to identify future, especially non-obvious consequences of recommendations on humans, e.g., many user features can be derived from their activities, e.g., simple Facebook likes, which was observed many years ago [20]. Moreover, these consequences (possible impact and risks) should be presented to the user in an understandable and editable form – preferably in the interaction, which again requires new technologies.

Information ecology

RSs especially if combined with generative AI models provide users with successive portions of new information. This, in turn, leads to information overload and potential information pollution, resulting in harm to the user, which also requires new solutions and maybe even legal regulations [21].

Overall, some of the risks with a long-term effect, perhaps should be addressed through solutions similar to smoking and perhaps even investing in the stock market, e.g., increase awareness among users regarding the potential hazards of using RSs.

WHAT NEXT

AI and RSs are poised to revolutionize the fabric of societal interactions, offering unprecedented opportunities to reshape the way individuals connect, communicate, and collaborate [22]. While they can facilitate connections by matching individuals with shared interests and preferences, there is concern that these mediated interactions may lack the depth and authenticity of face-to-face communication.

Over-reliance on algorithmic recommendations may contribute to the formation of echo chambers, limiting exposure to diverse perspectives and potentially weakening interpersonal bonds. Moreover, the increasing integration of AI and RSs into daily life raises questions about individual autonomy.

2Please note that existing nudging solutions (PI6), focus on single positive nudging objectives rather than on integration and alignment with other targets, e.g., through multi-criteria inference.
While these systems aim to streamline decision-making by offering tailored suggestions, there is a risk of subtle influence on user behavior. This raises ethical concerns about the manipulation of human autonomy and the potential erosion of free will in the face of algorithmic determinism. On the one hand, AI and RSs provide access to vast amounts of information and support data-driven decision-making, empowering individuals to make more informed choices. Simultaneously, the reliance on algorithmic recommendations may reduce individuals’ inclination to critically evaluate information and exercise independent judgment. Moreover, the lack of transparency in recommendation processes may hinder users’ understanding of how decisions are made, potentially undermining trust in the information presented.

In the near future, it will be increasingly important to raise awareness among users about the potential risks associated with RSs, drawing parallels to other contexts such as smoking and investing in the stock market. Just as education campaigns have been instrumental in informing the public about the health hazards of smoking and the financial risks of stock market investments, a similar approach is needed to highlight the potential pitfalls of relying blindly on recommendation algorithms. By elucidating the possible consequences, including the formation of echo chambers, filter bubbles, and the manipulation of user behavior, individuals can make more informed decisions about their online interactions. Moreover, fostering a culture of critical thinking and digital literacy can empower users to navigate RSs responsibly, mitigating the adverse effects and ensuring that they serve as tools for enrichment rather than exploitation.

A key element of RRSs will be explainability. When users are presented with recommendations, they often desire insight into the rationale behind the suggestions. Explainability offers transparency into the underlying mechanisms and criteria employed by the system to generate these recommendations. Firstly, explainable RSs foster trust by providing users with visibility into how recommendations are formulated. Users are more likely to trust recommendations if they comprehend the reasoning behind them, which consequently enhances engagement and satisfaction. Secondly, explainability aids users in understanding why specific recommendations are presented to them, resulting in a more meaningful and personalized experience. With a grasp of the factors influencing recommendations, users can better evaluate and interpret them, thereby enabling more informed decision-making.

Moreover, explainability plays a crucial role in mitigating undesired biases. By unveiling the factors considered by the RS, users can identify and address potential biases related to demographics, preferences, or historical interactions. RRSs will empower users to provide feedback and adjust their preferences based on the recommendations they receive. When users comprehend the reasoning behind recommendations, they can offer valuable feedback, thereby contributing to the improvement of recommendations over time. Additionally, in industries such as finance or healthcare, regulatory requirements often mandate transparency and accountability in recommendation systems. Explainability ensures compliance with regulations and standards by offering clear explanations of recommendations and decision-making processes. Finally, RRSs will facilitate the identification and resolution of potential ethical concerns, such as privacy violations or manipulation of user behavior.

**TAKEAWAY MESSAGE**

So, are RSs friends, foes, or frenemies? They can be friends when they provide transparent and personalized recommendations that genuinely enhance user experiences. For instance, on streaming platforms like Netflix or Spotify, RSs recommend movies, shows, or music tailored to users’ tastes and preferences, helping them discover content they might enjoy. They can be foes when they prioritize the interests of businesses or other stakeholders over those of users, which can lead to a proliferation of sponsored content or biased recommendations, potentially undermining user trust and satisfaction. Given the pervasive influence of RSs on various aspects of human life, it is crucial to study their mechanisms, effects, and implications more comprehensively.

Understanding how RSs operate, the algorithms they employ, and the biases they may exhibit is essential for addressing potential ethical concerns, ensuring transparency, and promoting user empowerment. Finally, studying the impact of RSs on society can inform the development of regulatory frameworks, awareness campaigns, industry standards, and best practices to mitigate negative consequences and maximize the benefits of these systems for individuals and communities alike. We are convinced that the concept of Responsible Recommender System (RRS) presented in this paper will inspire the emergence of new commercial solutions that support both the well-being, goals and values of users and pursue necessary business objectives.
Acknowledgment

This work was financed by (1) the National Science Centre, Poland, project no. 2021/41/B/ST6/04471; (2) the statutory funds of the Department of Artificial Intelligence, Wroclaw University of Science and Technology; (3) the Polish Ministry of Education and Science within the programme "International Projects Co-Funded"; (4) the European Union under the Horizon Europe, grant no. 101086321 (OMINO). However, the views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency. Neither the European Union nor European Research Executive Agency can be held responsible for them.

REFERENCES

17. M. R. Hasan, A. K. Jha, and Y. Liu, “Excessive use of online video streaming services: Impact of...


**Przemysław Kazienko, Ph.D.** is a full professor and leader of ENGINE - the European Centre for Data Science and two research groups: HumaNLP and Emoginition at The Department of Artificial Intelligence, Wroclaw University of Science and Technology, Poland. He authored over 300 research papers, including over 50 in journals with impact factor related to recommender systems, personalization and subjective tasks in NLP, Large Language Models (LLM), affective computing and emotion recognition, social/complex network analysis, deep machine learning, sentiment analysis, DSS in medicine, finances, and telecommunication, knowledge management, information retrieval, and data security. He also led over 50 projects, including large European ones, chiefly in cooperation with companies with total local budget over €10M. He gave 30 keynote/invited talks for international audience and served as a co-chair of over 20 international scientific conferences and workshops. He is an IEEE Senior Member and a member of the Editorial Board of several scientific journals. Contact him at kazienko@pwr.edu.pl.

**Erik Cambria** is currently a Professor at Nanyang Technological University, where he also holds the appointment of Provost Chair in Computer Science and Engineering, and Founder of several AI companies, such as SenticNet, offering B2B sentiment analysis services, and finaXai, providing fully explainable financial insights. Prior to moving to Singapore, he worked at Microsoft Research Asia (Beijing) and HP Labs India (Bangalore), after earning his PhD through a joint program between the University of Stirling (UK) and MIT Media Lab (USA). Today, his research focuses on neurosymbolic AI for interpretable, trustworthy, and explainable affective computing in domains like social media monitoring, financial forecasting, and AI for social good. He is an IEEE Fellow, Associate Editor of various top-tier AI journals, e.g., Information Fusion and IEEE Transactions on Affective Computing, and is involved in several international conferences as keynote speaker, program chair and committee member. Contact him at cambria@ntu.edu.sg.