How does explainability affect perceived transparency, trust, acceptance and usefulness?

Yifan Ding[1], Setia Hermawati[2]

[1]School of Computer Science, the University of Nottingham, UK
[2]Human Factors Research Group, Department of Mechanical, Materials and Manufacturing Engineering, Faculty of Engineering, the University of Nottingham, UK

SUMMARY

The effects of explainability on perceived trust, transparency, acceptance and usefulness were explored in a within subjects' study (n=15) using an online shopping recommender system as a context. The study investigated three levels of explainability (low, medium, high) and perceived transparency, trust, acceptance and usefulness were obtained using standardised questionnaires. The results showed that explainability significantly affected perceived transparency ($p<0.001$), trust ($p<0.001$), acceptance ($p<0.05$) and usefulness ($p<0.05$).

KEYWORDS

Explainability, transparency, trusts, acceptance and usefulness, recommendation system

Introduction

Online shopping has taken over as the main option for conveniently and securely from the comfort of their homes. Online retailers offer a diverse range of products, including books, electronics, clothing, furniture, and appliances, which leads consumers to face challenges in finding the exact items they need (Viappiani et al., 2007). Recommendation systems, which provide personalised service support based on user preferences, have become a primary market demand in online shopping (Lim et al., 2023). According to the study by Zhu and Li (2014), recommendation systems have garnered significant attention within prominent e-commerce companies such as Tmail.com and Amazon.com. A recommendation system is as an information filtering mechanism designed to address the challenge of information overload by sieving out essential details from products. Recommendation systems typically learn from user past actions to estimate their current interest in a given item and provide them with individualised service support (Zhang et al., 2021).

Users are also more likely to want to know why they get particular recommendations (Ozok et al., 2010). However, the majority of online recommendation systems, according to research by Sinha and Swearingen (2002), operate like a "black box," preventing consumers from fully comprehending the logic or rationale behind the recommendations. According to Johnson and Johnson (1993), explanations are crucial in how users and complex systems interact, with one of their main goals being to show causal links and enable meaningful, predictable, and effective user-system interaction. Furthermore, users are more likely to accept and use recommendations if they believe that the system can offer useful suggestions and if the interface and actions are easy to use (Bansah and Darko, 2022).

The majority of studies on recommender systems focus on enhancing algorithm of product search in online directories and pointing customers to the best alternatives (Bridge et al., 2005). On the other hand, studies on UI (UI) and understanding user views have received scant attention. It is crucial that recommendation systems be transparent so that consumers understand how recommendations are made and why they are relevant to their interests (Ozok et al., 2010). Users may begin to question the validity of recommendations if they are unable to understand the techniques or logic employed. There has been increasing interest in more user-centric assessment metrics and many studies have examined the effect of UI on perceived trust, transparency, usefulness, and acceptance. However, they predominantly focused on analysing these factors separately (Herlocker et al., 2000) and there is a lack of comprehensive research that examines them collectively. Furthermore, upon closer inspection, they often used arbitrary/non-standardised measures to assess perceived trust, transparency, usefulness, and acceptance. This research seeks to inform best practices on explainability of UIs by incorporating different levels of explainability into the graphical UI of online shopping recommendation systems and examine their impact on users perceived trust, transparency, usefulness and acceptance, by using standardised measures. This work provides helpful advice for interface designers in the domain of online shopping recommendation systems.

Methodology

A literature review was conducted to collect and identify guidelines and recommendations pertinent to improving explainability in the domain of several online recommendation system. The relevant studies were then selected to inspire the design of the UI for explainability. For example, Hernandez-Bocanegra and Ziegler (2023) recommended providing a reason behind recommendation of a certain product and parameters that substantiated claims supporting the usefulness of recommended items. Cosley et al. (2003) suggested the use of visual representations, such as star ratings, to be included in a recommended content. Caro-Martínez et al. (2021) advocated the need to amplify feedback from other shoppers, for instance, by using charts to depict the most accessed recommendations.

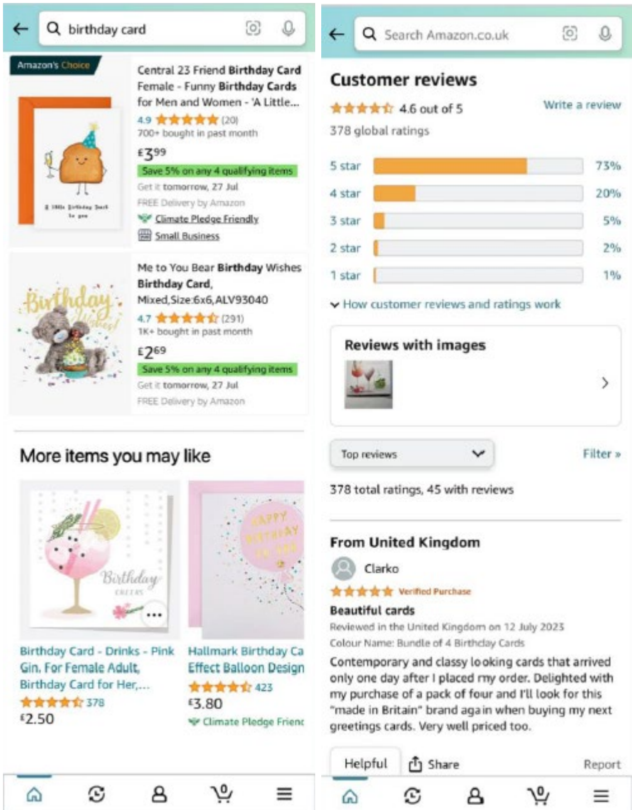

Figure 1: User Interface A

Taking the Amazon app as a prototype, three graphical UIs were created to simulate purchasing a birthday card. UI A (low explainability) resembled existing UIs of the Amazon.uk app. Figure 1 shows UI A. UI B (medium explainability) incorporated an explanatory interface detailing why a product was recommended, a filter of reviews based on chosen star ratings and extracted key terms of reviews. Figure 2 shows UI B. UI C (high explainability) was built upon UI B with added features that displayed the accuracy of a recommendation and word cloud. Figure 3 shows UI C.

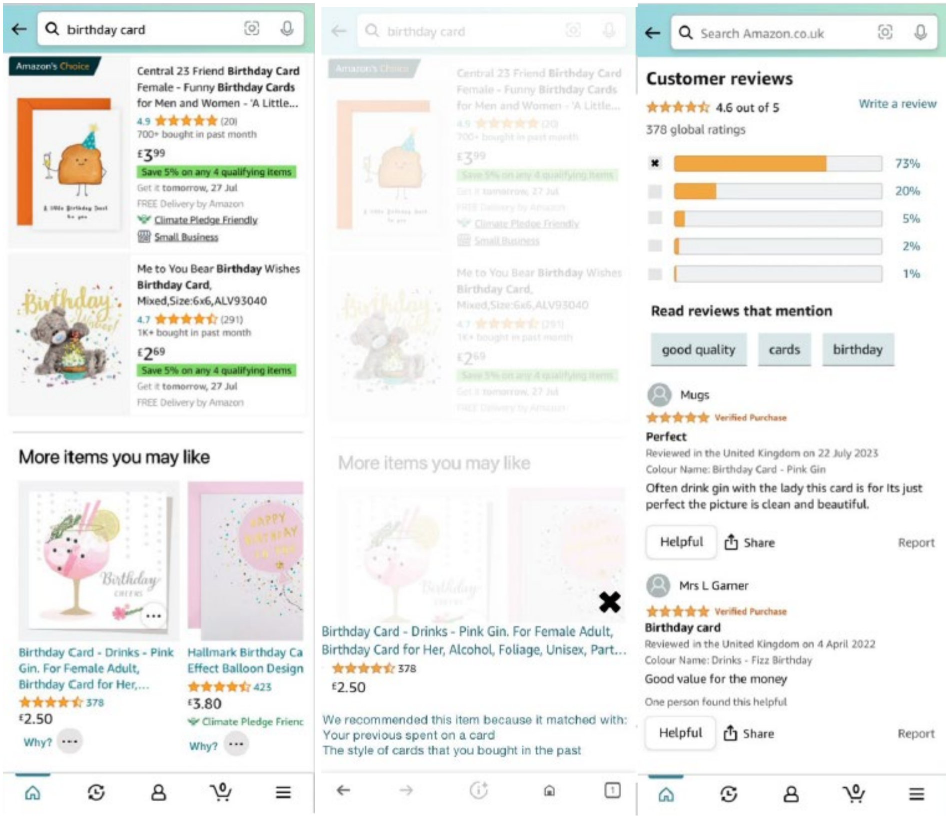

Figure 2: User Interface B

A total of 15 participants with prior online shopping experience participated in a 30-minute study. Participants interacted with all three UIs in counterbalanced order. After interaction with each UI, participants were requested to provide feedback on perceived transparency, trust, usefulness and acceptance. Standardised questionnaires were used to assess users perceived usefulness and acceptance (Van der Laan et al. (1997), transparency (Hellmann et al., 2022) and trust (Gulati et al., 2019). After interacting with all UIs, participants ranked the UIs and were interviewed. The following questions were asked during the interview:

- Please explain your reason for your ranking and your impression of the three interfaces.
- How do you think the UI you rank highest affects your perceived usefulness? Please explain which elements or features make it easier for you to understand and accept the recommended results.
- Does the interface you rank highest enhance your trust in the recommender system compared to another interface? If yes, please explain what gave you confidence in the the system in the reliability and accuracy of the system.
- Is it important to you to understand the reasons why a particular item was recommended? To which extent would you be willing to invest time (by viewing the explanation behind the recommendation) in order to understand why something was recommended?
- Do you have any other comments or suggestions on the interface?

Ethical approval was sought and granted by the ethics board of the School of Computer Science of the University of Nottingham. A pilot study was also conducted before the commencement of the actual study.

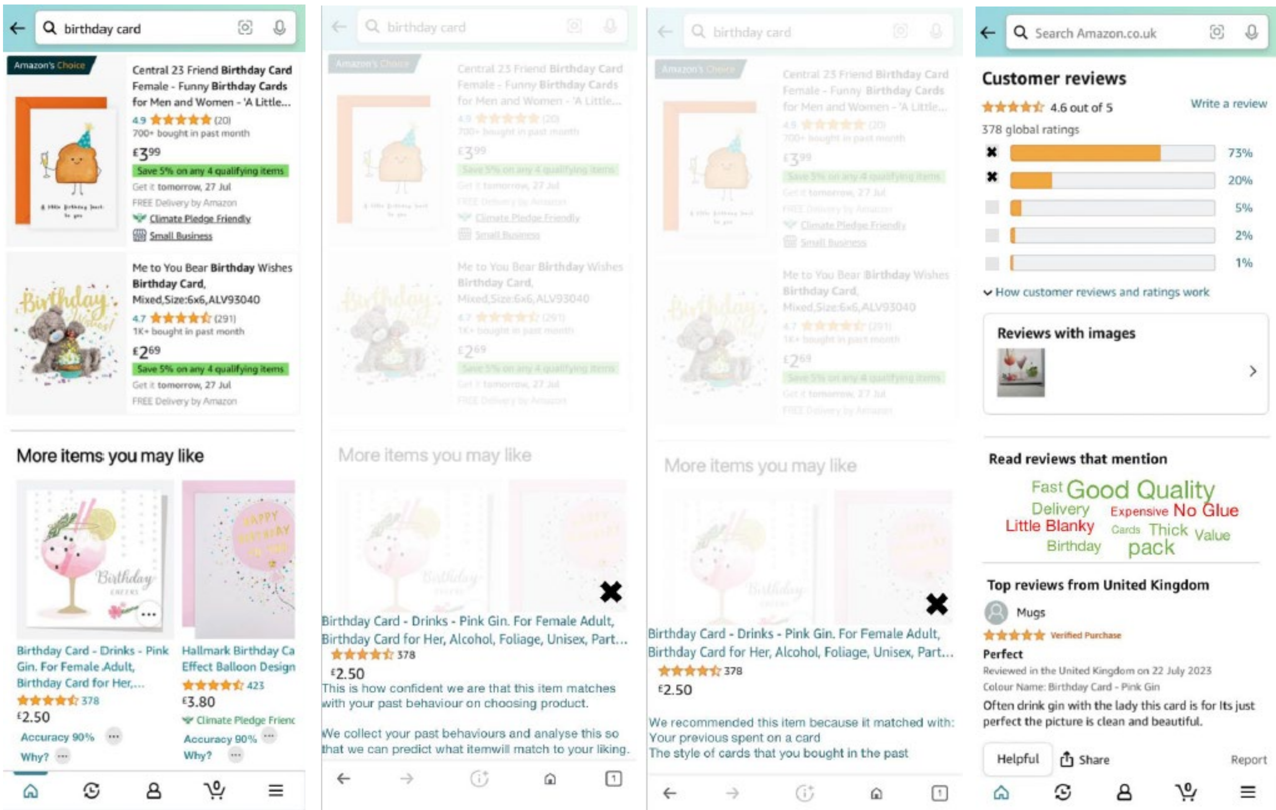

Figure 3: User Interface C

Results

Table 1 below shows the demographic information and online shopping behaviours of participants while Table 2 shows the median values of transparency, trust, acceptance, usefulness and ranking for each UI. Statistical analysis revealed that there were significant differences in perceived transparency between the UIs (χ^2 (2) = 30, p < .001). Pairwise comparison of mean rank scores showed that participants significantly perceived UI C (mean rank = 3) as the most transparent compared to UI A (mean rank = 1) (Dunn =2, p < 0.001) and UI B (mean rank = 2) (Dunn = 1, p = 0.019). The participants also perceived a significant difference between UI A and UI B (Dunn = 1, p = 0.019).

Table 1: Demographics and online shopping behaviours of participants

Age	26.5 ± 9
Frequency of using online shopping sites	Less than once a month (n=1), Less than once a week (n=9), (n=5)
Frequency of browsing and exploring recommended products when shopping online	Rarely (n=3), Sometimes (n=9), Often (n=3)

There were significant differences in perceived trust between the UIs (χ^2 (2) = 20.25, p < .001). Pairwise comparison of mean rank scores showed that participants significantly perceived UI A (mean rank = 1.1) as the least trusted compared to UI B (mean rank = 2.27) (Dunn = 1.17, p=0.04) and UI C (mean rank = 2.63) (Dunn = 1.53, $p < 0.001$). There was no significant difference in trust between UI B and UI C. There were also significant differences in perceived acceptance (χ^2 (2) = 6.377, p=0.041) and usefulness (χ^2 (2) =6.453, p=0.04). However, subsequent pairwise comparisons for perceived acceptance and usefulness could not identify significant differences among all conditions due to the small number of participants. Analysis of the ranking of the UIs also showed a significant difference in the ranking (χ^2 (2) = 7.153, p=0.028). Pairwise comparison of mean rank scores showed that participants significantly ranked UI A (mean rank = 1) worse than UI B (mean rank = 2.4) (Dunn = 1.4, $p<0.001$) and UI C (mean rank = 2.6) (Dunn = 1.6, $p<0.001$). However, there was no significant difference in ranking between UI B and UI C.

Table 2: Median values of transparency, trust, acceptance and usefulness for each UI

	UI A	**UI B**	**UI C**
Transparency	8.96	15.59	17.73
Trust	10.32	13.66	14.65
Acceptance	0	0.5	0.5
Usefulness	0.2	0.6	0.4
Ranking	1	2	3

The interview revealed the ranking of the reason behind the UIs among participants. Participants who ranked UI B highest (n=5) stated that the features provided in this UI ("why" button, filterable user reviews and keywords) were sufficient and additional information was unnecessary. A participant's comment reflected this view: *"I consider the keyword function very important. With keywords, users only need to click on the aspects of the product they want to know about, saving a lot of time and improving purchasing efficiency. If users look for too long, they might get tired and do not want to continue, affecting the purchase. As for accuracy and the word cloud, I don't like them to be overly complicated, but I can accept them since I can simply ignore the word cloud."* On the other hand, participants who ranked UI C highest (n=10) believed that the additional features ("accuracy" button and word cloud) provided them with more confidence and addressed their concerns about the recommendations' reliability and accuracy. This was reflected from one of participants' comment: "*The percentage accuracy relative to the recommended reasons significantly affects whether I trust the system's suggestions. The word cloud, rather than keywords, better reflects the system's understanding of my shopping tendencies, which enhances my trust in the recommendations. '* The interview also revealed that some participants viewed excessive information to enhance explainability negatively and impacted on the overall usefulness of UI as reflected by one of the participants comment: "*Comparing B and C, the excessive feature design makes me find it cumbersome and unpleasant*."

The interview also revealed that 87% of the participants viewed explainability as important, especially when they came across product recommendations, they were unfamiliar with, or when the products they wished to buy were expensive. In these circumstances, the participants were prepared to dedicate some time, and thus willing to use the various features of explainability in the UIs. This shows that the participants changed their behaviour based on their perception of risk that is associated with their purchase. This was reflected by comments from two participants*: "... I find that when purchasing unfamiliar products, understanding the reasons behind recommendations becomes more necessary. It also encourages me to explore the recommended rationale in more detail. Otherwise, I might not invest time in understanding the reasons..."* and *"If the item I want to buy is of high value, then I would be willing to invest more time. If it were something more common, I would not spend a lot of time comparing."*

Discussion

The findings of this study show that an UI with low levels of explainability is significantly perceived by users as untrustworthy by users and does not create trust and transparency. This is in line with the view of Sinha and Swaringen (2002) that increasing explainability helps increase user trust. Explanations assist users in filtering and comprehending recommendation outcomes, facilitating swift decision-making (Tintarev and Masthoff, 2007). When users understand the rationale and formulation process behind recommendations, their reliance on the system increases (Tintarev and Masthoff, 2007; Berkovsky et al., 2017). In other words, convincing explanations contribute to persuasion if users are assured of the appropriateness of recommendations. Qualitative analysis of interview results further reveals the agreement among participants that recommendations coupled with explanations are crucial. Explanations behind the rationale of recommendations offer transparency into the workings of the recommendation system. When users understand the reasons behind the recommendations, they are more likely to trust them. Explanations assist users in comprehending the processes of the recommend recommendation system, understanding its strengths and weaknesses, in line with findings from Herlocker et al.(2004) and Muramatsu and Pratt (2001). They concurred that providing explanations for the decisions or suggestions of AI-based systems yields multiple advantages in user perceptions. Particularly within recommendation systems, providing explanations contributes positively to perceptions of transparency and effectiveness. This study finding is in alignment with the conclusion of Pu and Chen (2006) and underscores that providing explanations can enhance user transparency.

However, our study also reveals that, except for transparency, higher explainability significantly improved perceived trust, acceptance, and usefulness only up to a certain point. Beyond this point, providing additional features to support explainability would not result in significant improvement, and, in the case of usefulness, it can impact negatively as users view additional features as superfluous. This is reflected in our findings, showing the non-significant differences in perceived acceptance and usefulness between UI B and UI C and that participants who chose UI B as the best UI argued that it provided sufficient information (in contrast to UI C, which provided excessive information). This study posits that for some users, UI C offers them an overwhelming array of functionalities and information, potentially leading to information overload and a diminished user experience. As demonstrated in the study by Caro-Martínez et al. (2021), the objectives of explanations, user expectations, available knowledge, and modes of expression should be considered and balanced in order to produce an effective explanation. Therefore, a balancing act is required to ensure that explainability does not lead to: 1) information overload, sparing users the need to invest excessive time and effort in comprehending the rationale behind recommended products (Chazette and Schneider, 2020), and 2) the possibility of misusing the system (Kizilcec, 2016).

This study is not without limitations. Firstly, the chosen recommendation products are birthday cards, which are inexpensive and as such the cost of making low-quality decision is low. Existing research highlights the importance of minimising decision effort; however, in high-risk choice tasks (such as buying a car or a house), ensuring the most optimum choice becomes crucial. The cost of making low-quality decisions in such cases can be considerably high. In this paper, our study design revolves around low-risk and low cost. Therefore, the applicability of the study findings might only be extended or applicable to low-risk products with low costs. Second, the sample size for our study is quite small and has been shown to affect our ability to establish where the significant differences in acceptance and usefulness lie between UIs.

Conclusion

Integrating the explainability and transparency of recommended products into the graphical user interface of online shopping recommendation systems affects trust, perceived usefulness, acceptance, and transparency. Both excessive and insufficient information in recommendation systems influence trust, transparency, usefulness, and acceptance. When the information presented about recommended products is too limited, trust is not established due to low transparency, while excessive information can lead to lower usefulness. Higher levels of explainability are positively correlated and trust and transparency, with users favour and accepting recommendations they perceive as transparent.

Acknowledgment

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreements No 952071 (DIGITbrain).

References

Bansah, A.K., & Darko, A. D. (2022). Perceived convenience, usefulness, effectiveness and user acceptance of information technology: evaluating students' experiences of a Learning Management System. Technology, pedagogy and education, 31(4), pp. 431–449.

Berkovsky, S., Taib, R., & Conway, D. (2017). How to Recommend?, In Proceedings Of the 22nd International Conference On Intelligent User Interfaces, pp. 287–300.

Bridge, D., Göker, M. H., McGinty, L., & Smith, B. (2005). Case-based recommender systems. Knowledge Engineering Review, 20(3), pp. 315–320.

Caro-Martínez, M., Jiménez-Díaz, G., & Recio-García, J.A. (2021). Conceptual modelling of explainable recommender systems: An ontological formalization to guide their design and development. The Journal of Artificial Intelligence Research, 71, pp. 557–589.

Chazette, L., & Schneider, K. (2020). Explainability as a non-functional requirement: challenges and recommendations. Requirements Eng 25, pp. 493–514.

Cosley, D., Lam, S. K., Albert, I, Konstan, J. A., & Riedl, J. (2003). Is seeing believing?: how recommender system interfaces affect users' opinion. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 585–592.

Gulati, S., Sousa, S., & Lamas, D. (2019). Design, development and evaluation of a human-computer trust scale. Behaviour & Information Technology, 38(10), pp. 1004–1015.

Hellmann, M., Hernandez-Bocanegra, D. C., & Ziegler, J. (2022). Development of an instrument for measuring users' perception of transparency in recommender systems. In Proceedings of the International Conference on Intelligent User Interfaces, pp. 156 – 165.

Herlocker, J., Konstan, J., & Riedl, J. (2000). Explaining collaborative filtering recommendations. In Proceedings of the 2000 ACM conference on Computer supported cooperative work, pp. 241–250.

Hernandez-Bocanegra, D.C., & Ziegler, J. (2023). Explaining Recommendations through Conversations: Dialog Model and the Effects of Interface Type and Degree of Interactivity. ACM Transactions on Interactive Intelligent Systems, 13(2), pp. 1 – 47.

Hu, X., Zhu, W., & Li, Q. (2014). HCRS: A hybrid clothes recommender system based on user ratings and product features. In Proceedings of the International Conference on Management of e-Commerce and e-Government, pp. 270-274.

Johnson, H., & Johnson, P. (1993) Explanation facilities and interactive systems. In Proceedings of the 1st International Conference on Intelligent User Interfaces, pp. 159-166.

Kizilcec, R. (2016). How much information? Effects of transparency on trust in an algorithmic interface. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 2390–2395. NewYork, Association for Computing Machinery.

Lim, Y.F., Haw, S. C., Ng, K. W., & Anaam, E. A. (2023). Hybrid-based Recommender System for Online Shopping: A Review. Journal of Engineering Technology and Applied Physics, 5(1), pp. 12–34.

Muramatsu, J., & Pratt, W. (2001). Transparent queries: Investigating users' mental models of search engines. SIGIR forum, pp. 217–224.

Ozok, A.A., Fan, Q., & Norcio, A.F. (2010). Design guidelines for effective recommender system interfaces based on a usability criteria conceptual model: results from a college student population. Behaviour & Information Technology, 29(1), pp. 57–83.

Pu, P., & Chen, L. (2006). Trust building with explanation interfaces. In Proceedings of the International Conference on Intelligent User Interfaces, pp. 93–100.

Sinha, R., & Swearingen, K. (2002). The role of transparency in recommender systems. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. pp. 830–831.

Tintarev, N., & Masthoff, J. (2007). Effective explanations of recommendations. In Proceedings of ACM Conference on Recommender Systems, pp. 153–156.

Van Der Laan, J.D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. Transportation Research. Part C, Emerging technologies, 5(1), pp. 1–10.

Viappiani, P., Pu, P., & Faltings, B. (2007). Conversational recommenders with adaptive suggestions. In Proceedings of the 2007 ACM conference on Recommender systems, pp. 89–96.

Zhang, Q., Lu, J., & Jin, Y. (2021). Artificial intelligence in recommender systems. Complex & intelligent systems, 7(1), pp. 439–457.