On-demand Personalized Explanation for Transparent Recommendation

Mouadh Guesmi
University of Duisburg-Essen
Duisburg, Germany
mouadh.guesmi@stud.uni.de

Mohamed Amine Chatti
University of Duisburg-Essen
Duisburg, Germany
mohamed.chatti@uni-due.de

Laura Vorgerd
University of Duisburg-Essen
Duisburg, Germany
Laura.vorgerd@stud.uni-due.de

Shoeb Joarder
University of Duisburg-Essen
Duisburg, Germany
shoeb.joarder@stud.uni-due.de

Shadi Zumor
University of Duisburg-Essen
Duisburg, Germany
shadi.zumor@stud.uni-due.de

Yiqi Sun
University of Duisburg-Essen
Duisburg, Germany
yiqi.sun@stud.uni-due.de

Fangzheng Ji
University of Duisburg-Essen
Duisburg, Germany
fangzheng.Ji@stud.uni-due.de

Arham Muslim
National University of Sciences and Technology
Islamabad, Pakistan
arham.muslim@seecs.edu.pk

ABSTRACT

The literature on explainable recommendations is already rich. In this paper, we aim to shed light on an aspect that remains underexplored in this area of research, namely providing personalized explanations. To address this gap, we developed a transparent Recommendation and Interest Modeling Application (RIMA) that provides on-demand personalized explanations with varying levels of detail to meet the demands of different types of end-users. The results of a preliminary qualitative user study demonstrated potential benefits in terms of user satisfaction with the explainable recommender system. Our work would contribute to the literature on explainable recommendation by exploring the potential of on-demand personalized explanations, and contribute to the practice by offering suggestions for the design and appropriate use of personalized explanation interfaces in recommender systems.

CCS CONCEPTS

• **Information systems** → **Personalization**; **Recommender systems**.

KEYWORDS

Transparency, Recommendation Explanations, Personalized Explanations, User Modeling

ACM Reference Format:
Mouadh Guesmi, Mohamed Amine Chatti, Laura Vorgerd, Shoeb Joarder, Shadi Zumor, Yiqi Sun, Fangzheng Ji, and Arham Muslim. 2021. On-demand Personalized Explanation for Transparent Recommendation. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21 Adjunct), June 21–25, 2021, Utrecht, Netherlands.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3450614.3464479

UMAP '21 Adjunct, June 21–25, 2021, Utrecht, Netherlands

ACM ISBN 978-1-4503-8367-7/21/06…$15.00
https://doi.org/10.1145/3450614.3464479

1 INTRODUCTION

In recent years, various attempts have been made to address the black-box issue of recommender systems (RS) by providing explanations that enable users to understand the recommendations [22]. Recent research has surveyed the explainable recommendation domain and proposed several classifications dimensions [6, 17, 21, 22]. For instance, Guesmi et al. [6] classified explainable RS based on four dimensions, namely the explanation aim (transparency, effectiveness, efficiency, scrutability, persuasiveness, trust, satisfaction), explanation focus (input: user model, process: algorithm, output: recommended items), explanation type (collaborative-based, content-based, social, hybrid) and explanation display (textual, visual).

Another crucial design choice in explainable recommendation relates to the level of explanation detail that should be provided to the end-user [15]. Users may not be interested in all the information that the explanation can produce [14]. Different users demand different levels of explanation information and explanations may cause negative effects if an explanation is difficult to understand [23]. Thus, it is important to provide explanations with enough details to allow users to build accurate mental models of how the RS operates without overwhelming them. In terms of design choice (i.e., explanation focus, type, display, or level of detail), explainable RS have traditionally followed a one-size-fits-all model, whereby the same explanation is provided to each user, without taking into consideration an individual user's context, i.e., abilities, goals, needs, or preferences. In the explainable recommendations field, research regarding personalized explanation has emerged only recently, showing that personal characteristics may have an impact on the perception of explanations, and that there is potential for the development of personalized explanation. Researchers have focused on investigating what specific characteristics may play a role in a user's interaction with an explainable RS [8, 13]. However, concrete solutions that devise mechanisms to provide users with personalized explanations at a design choice level are scarce in the literature on explainable recommendations.

In this paper, we aim at a shift from a one-size-fits-all to a personalized approach to explainable recommendations. To do so, we developed a transparent Recommendation and Interest Modeling Application (RIMA) that provides on-demand personalized explanations with varying levels of detail, which we argue are more effective as explanations to meet the needs and preferences of different users.

The remainder of this paper is organized as follows. We first outline the background for this research (Section 2). We then present the implementation of the personalized explanations in RIMA (Section 3) followed by a preliminary evaluation of the application (Section 4). Finally, we summarize the work and outline future research plans (Section 5).

2 RELATED WORK

In the following, we discuss related work on explainable recommendations that provides explanation with varying level of details and attempts to personalize the explanations to meet the demands of different users.

2.1 Explanation with varying level of details

In this work, the level of detail refers to the amount of information exposed in an explanation. In the field of explainable AI (XAI) in general, Mohseni et al. [15] argue that different user groups will have other goals in mind while using such systems. For example, while machine learning experts might prefer highly-detailed visual explanations of deep models to help them optimize and diagnose algorithms, lay-users do not expect fully detailed explanations for every query from a personalized agent. Instead, systems with lay-users as target groups aim to enhance the user experience with the system through improving their understanding and trust. In the same direction, Miller [14] argue that providing the exact algorithm which generated the specific recommendation is not necessarily the best explanation. People tend not to judge the quality of explanations around their generation process, but instead around their usefulness. Besides the goals of the users, another vital aspect that will influence their understanding of explanations are their cognitive capabilities [23]. Only when users have enough time to process the information and enough ability to figure out the meaning of the information, a higher level of detail in the explanation will lead to a better understanding. But as soon as the amount of information is beyond the users' comprehension, the explanation could lead to information overload and bring confusion. Without the understanding of how the system works, users may perceive the system as not transparent enough, which could, in turn, reduce the users' trust in the system [5, 23].

A critical question in the research of explainable recommendation is whether the relationship between the level of detail and transparency is a linear one. To answer this question, we need first to discriminate between objective transparency and user-perceived transparency. Objective transparency means that the RS reveals the underlying algorithm of the recommendations. However, the algorithm might be too complex to be described in a human-interpretable manner. Therefore, it might be more appropriate to provide "justifications" instead of "explanations", which are often superficial and more user-oriented. On the other hand, user-perceived transparency is thus based on the users' subjective opinion about how good the system is capable of explaining its recommendations [5].

In general, it can be assumed that a higher level of explanation detail increases the system's objective transparency but is also associated with a risk of reducing the user-perceived transparency, and that this risk depends on the user's characteristics. Recent studies on explainable recommendation showed that personal characteristics have an effect on the perception of explanations and that it is important to take personal characteristics into account when designing explanations [8, 13]. Drawing on these findings, Millecamp et al. [13] suggest that (1) users should be able to choose whether or not they wish to see explanations and (2) explanation components should be flexible enough to present varying level of details depending on users' preferences. Following these design guidelines, the authors developed a music RS that not only allows users to choose whether or not to see the explanations by using a "Why?" button but also to select the level of detail by clicking on a "More/Hide" button. However, providing explanations with varying levels of details remains rare in the literature on explainable recommendations.

2.2 Personalized explanation

In recent years, the literature on AI has emphasized the need for explanations that are tailored to individuals, i.e., personalized explanations. For example, Arya et al. [1] stressed that one explanation does not fit all, as different AI stakeholders present different requirements for explanations and may desire different kinds of explanations (e.g., feature-based, instance-based, language-based). The authors presented an AI toolkit, which contains eight state-of-the-art explainability algorithms that can explain an AI model in different ways to a diverse set of users. Jung and Nardelli [7] pointed out that XAI is challenging since explanations must be tailored (personalized) to individual users with varying backgrounds and proposed an algorithm that allows constructing personalized explanations that are optimal in an information-theoretic sense. Assuming that, based on varying backgrounds like training, domain knowledge and demographic characteristics, individuals have different understandings and hence mental models about the learning algorithm, Kuhl et al. [10] investigated how personalized explanations of learning algorithms affect employees' compliance behavior and task performance. On a conceptual level, Schneider and Handali [18] derived a conceptualization of personalized explanation in machine learning (ML) based on a framework covering desiderata of personalized explanations, dimensions that can be personalized, what and how information can be obtained from individuals and how this information can be utilized to customize explanations. In terms of personalization automation, i.e., "who does the personalization?" [4], the authors distinguished between automatic personalization by the system providing explanations (system-driven personalized explanation) and manual personalization which is done by the explainee, actively setting the explanation parameters, e.g., choosing the number of features to visualize (system-driven personalized explanation). Below, we use this classification to categorize related work on personalized explanation in the recommendation field.

2.2.1 System-driven personalized explanation. Kouki et al. [8, 9] conducted a user study to analyze the effects of different personalized explanation styles, their number, and format as well as personality characteristics on user preferences for explanations. Gedikli et al. [5] presented and discussed the results of a user study where recommendation systems were provided with different types of explanation. The study revealed that the content-based tag cloud explanations were effective and well accepted by the majority of users. They were particularly helpful to increase user satisfaction as well as the user-perceived level of transparency thanks to its personalized variant. However, they found that personalization was detrimental to effectiveness. Similarly, Tintarev and Masthoff [20] investigated the impact of personalizing simple feature-based explanations on effectiveness and satisfaction. They also reported that their personalization method hindered effectiveness, but on the other hand increased the satisfaction with the explanations. Musto et al. [16] presented a framework for generating personalized natural language explanations of the suggestions produced by a graph-based recommendation model based on the information available in the Linked Open Data (LOD) cloud. Their user study results revealed that their strategy outperformed both a non-personalized explanation baseline and a popularity-based one. McInerney et al. [12] presented a method (Bart) that combines bandits and recommendation explanations. This method is able to jointly learn which explanations each user responds to (personalized explanation), and learn the best content to recommend for each user (personalized recommendation). The conducted experiments revealed that personalizing explanations and recommendations provides a significant increase in estimated user engagement. Lu et al. [11] presented a multi-task learning framework that simultaneously learns to perform rating prediction and generate personalized recommendation explanation. They employed a matrix factorization model for rating prediction, and a sequence-to-sequence learning model for explanation generation by generating personalized reviews for a given recommendation-user pair as they consider user-generated reviews as explanations of the ratings given by users. Donkers et al. [3] proposed a conceptual framework that automates the process of extracting arguments from reviews to come up with personalized argumentative item-level explanations for recommendations. Inspired by how people explain word-of-mouth recommendations, Chang et al. [2] designed a process, combining crowd-sourcing and computation, that generates personalized natural language explanations.

In all the approaches outlined above, the most common personalized explanation technique is content-based. That is, the explanation design choice (e.g., explanation focus, type, display, level of detail) was always kept fixed and only the explanation's content was personalized to each user's data and personality. As Schneider and Handali [18] noted, explanations for RS are often inherently personalized due to the nature of the task. For example, users' reviews, tags, or preferred item features, serve as input for the recommendation algorithm as well as explainee data used for explanations. In general, the explanations are personalized by marking certain parts of the recommended item (e.g., item features) which are relevant to the explainee. Thus, content-based personalized explanations are common in the literature on explainable recommendations. By contrast, personalized explanations that focus on tailoring a certain explanation design choice, such as explanation focus, type, display, or level of detail are under-researched. To the best of our knowledge, only the work presented in [19] takes the personalized explanation to the design choice level. The authors proposed a hybrid method of personalized explanation of recommendations, which combines basic explanation styles to provide the appropriate type of personalized explanation to each user. Based on this method, each user will be given an explanation adapted to what most impressed her (i.e., explanation style which she prefers).

2.2.2 User-driven personalized explanation. Compared to system-driven personalized explanation, research that follows a user-driven personalized explanation approach that provides different explanation displays, styles, or levels of detail and then hands over control to the user to actively set the explanation parameters, e.g., choose the level of details that she wants to see, is under-investigated. Only the work presented in [13] follows a user-driven personalized explanation by providing on-demand explanations with varying levels of details.

3 RIMA

We developed the transparent Recommendation and Interest Modeling Application (RIMA) with the goal of explaining the recommendations as well as the underlying interest models. RIMA follows a user-driven, on-demand personalized explanation approach by providing explanations with different levels of detail and empowering users to steer the explanation process the way they see fit. The application provides on-demand explanations, that is, the users can decide whether or not to see the explanation and they can also choose which level of explanation detail they want to see. In this work, we focus on recommending tweets and Twitter users and leveraging explanatory visualizations to provide insights into the recommendation process (see Figure 1).

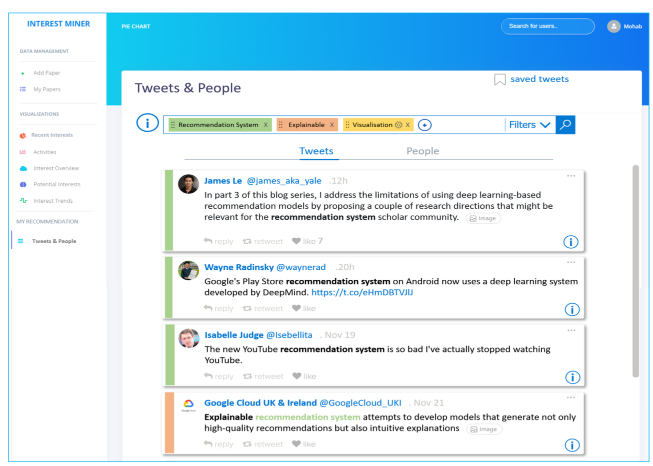


Figure 1: Recommendation Interface in RIMA

3.1 Explaining the interest model

3.1.1 Interest model generation. The aim of opening and exposing the interest model in RIMA is to let users become aware of the underlying interest model used for recommendation. These interest models are generated from users' publications and tweets. The application uses Semantic Scholar and Twitter IDs provided by users to gather their publications and tweets. It applies unsupervised keyphrase extraction algorithms on the collected publications and tweets to generate keyphrase-based interests. In order to address semantic issues, Wikipedia is leveraged as a knowledge base to map the keyphrases to Wikipedia pages and generate Wikipedia-based interests.

3.1.2 Personalized explanation of the interest model. The aim of explaining the interest model in RIMA is to foster user's awareness of the raw data (publications and tweets) and the derived data (interest model) that the RS uses as an input to generate recommendations, in order to increase transparency and promote understandability of the recommendation. Moreover, this may let users become aware of system errors and consequently help them give feedback and correction in order to improve future recommendations.

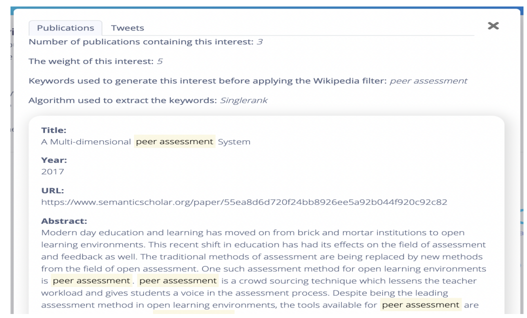
The application provides an on-demand personalized explanation of the interest model with three different levels of detail. The user can hover over an interest in the word cloud to see its source (i.e. publications or tweets) as a *basic explanation* (Figure 2a). When the user clicks on an interest in the word cloud, she will get more information through a pop-up window highlighting the occurrence of the selected interest in the tweets or title/abstract of publications, which represents the *intermediate explanation* (Figure 2b). The next level of detail is provided in the *advanced explanation* which follows an explanation by example approach to show in detail the logic of the algorithm used to infer the interest model (Figure 2c).

3.2 Explaining the recommendation

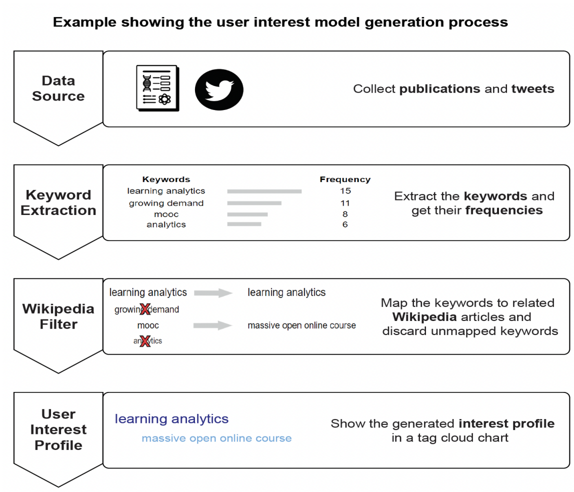
3.2.1 Recommendation generation. The aim of this part of the application is to provide tweet recommendations based on the generated interest model. For obtaining the candidate tweets, we use the Twitter API to fetch tweets that contain one or more user interests that are used as input for the recommendation. We then apply an unsupervised keyphrase extraction algorithm on the fetched tweets to extract keywords from the tweet text. In order to compare the similarity between the user interests and the candidate tweets, we use word embedding techniques to generate vector representations of the interest model and the tweets. After getting the two embedding representations (i.e., interest model embedding and tweet embedding), we calculate the cosine similarity between them in order to obtain a semantic similarity score. Tweets with a semantic similarity score above a threshold of 40 % will then be displayed to the user.

3.2.2 Personalized explanation of the recommendation. The application provides an on-demand personalized explanation of the recommendations with three different levels of detail. The *basic explanation* (Figure 3a) aims at explaining "why" a specific tweet was recommended in an abstract manner. The search box is initially populated with the user's top five interests, ordered by their

(a) Basic explanation



(b) Intermediate explanation



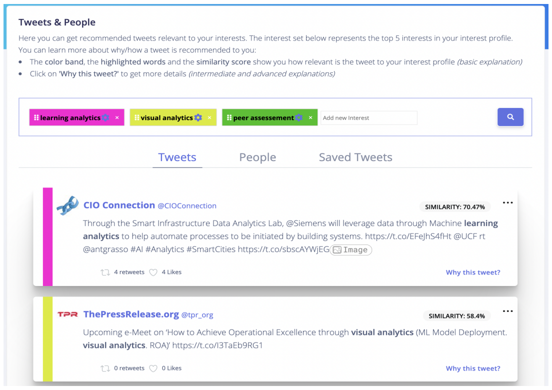
(c) Advanced explanation

Figure 2: Explaining the interest model with three levels of details

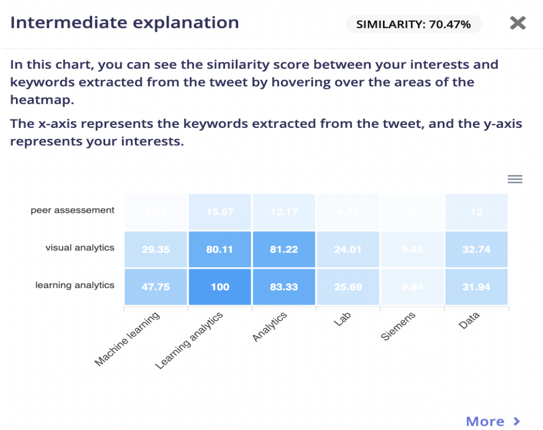
weights as generated by the system. Users can also add new interests in the search box or remove existing ones. The system will use these interests as input for the recommendation process. the basic explanation is achieved using a color band to map the tweet to the related interest(s). Also, the interest will be highlighted in the text of the tweet to show that this tweet contains this specific word (interest). In addition to these two visual elements, we display the similarity score on the top right corner of the tweet to show the level of similarity between the user interests and the recommended tweet. The answer to the "why" question in this level is a visual representation of "because the tweet text contains your interest X and this tweet is Y% similar to your interest profile".

For more details, the user can choose the *intermediate explanation* level (Figure 3b) by clicking on "Why this tweet?" on the bottom right of the tweet. Similar to the basic level, the intermediate level also aims at answering the "why" question, but with more details. We used a heatmap chart to show the semantic similarity between the user interest profile and the keywords extracted from the text of the tweet. The x-axis represents the keywords extracted from the tweet and the y-axis represents the user's interests used in the recommendation. The cells show the computed semantic similarity scores between each interest and keyword. At this level, the answer to the "why" question is a visual representation of the similarities between all user interests used as input for the recommendation and all the keywords extracted from the tweet.

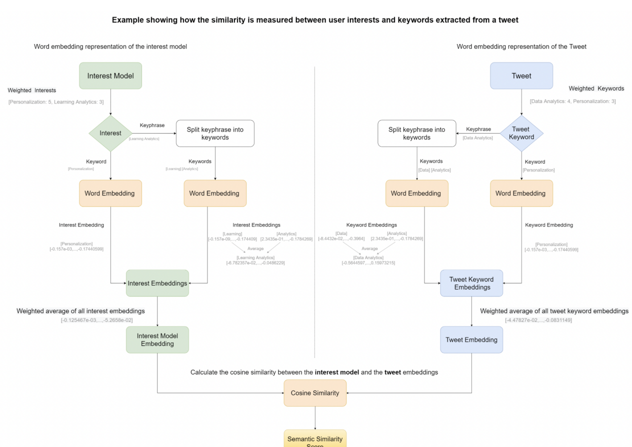
To move to the *advanced explanation* level (Figure 3c), the user has to click on the "more" button on the bottom right of the intermediate explanation window. The aim of the advanced explanation is to explain "how" the recommendation algorithm works. This is achieved by following an explanation by example approach to show in detail the logic of the algorithm used to semantically compare the keywords extracted from the recommended tweet and the user interests.



(a) Basic explanation



(b) Intermediate explanation



(c) Advanced explanation

Figure 3: Explaining the tweet recommendation with three levels of details

4 EVALUATION

We conducted a preliminary qualitative user study to evaluate the users' perception of the on-demand personalized explanations provided in the RIMA application. The user study was targeted at researchers or people who have at least one scientific publication (e.g. journal article, conference paper, book). All participants were recruited via e-mail, word-of-mouth, and groups in social media networks (e.g. Facebook, LinkedIn).

To participate in the study, participants had to fulfill three requirements: (1) they had to have at least one scientific publication, (2) they had to have a Semantic Scholar ID, and (3) they had to complete the study on a computer or laptop. The user study was conducted completely online using a video-conferencing tool. Participants interacted with the application by performing a set of pre-defined tasks. First, the participants were asked to use the RIMA application to create their interest models based on their Semantic Scholar and Twitter IDs and to see the visualizations corresponding to their interest models. Then, the participants were presented with the three visualizations representing the basic, intermediate, and advanced explanations of their generated interest models. Thereafter, they were asked to explore the recommended tweets and try to understand the provided recommendations based on the basic, intermediate, and advanced explanations. At the end of the study, the participants were asked to answer a questionnaire. As this was a qualitative study, the questionnaire had mostly open-ended questions enabling participants to provide free feedback and describe their ideas or suggestions.

A total of 23 participants completed the study. To achieve a diverse sample, the study included participants from different countries, educational levels, and study backgrounds. Out of the 23 participants, 14 (60.9%) reported to live in Germany, where nine (39.1%) were international users from eight different countries. All participants had sufficient English language skills to participate in the study. The highest level of education reported by most participants was master's degree (61.3%). They came from different disciplines such as psycholgy, medicine, biology and computer science. Out of the 23 participants, 12 participants (52.1%) were considered as Twitter users who reported to use Twitter at least 1 hour a week, where 10 (43.4%) reported to never use Twitter in a typical week. Regarding the domain knowledge, 12 (52.1%) participants reported that they have knowledge about recommender system and machine learning.

In the first part of the questionnaire, participants could provide direct feedback to the different explanations provided in the application. Next, for both the interest model and the recommendations, participants chose which explanation level (i.e., basic, intermediate, or advanced) they liked the most, and they could optionally justify their choice. Regarding the explanations of the interest model, ten participants (43.48 %) stated that they prefer the advanced explanation because it provides the most information about the generation of their interest model. These participants seemed to be fascinated by recommender systems and liked getting a deeper insight into the underlying logic. For example, one participant wrote "I find recommender systems interesting and I enjoyed learning about the way they operate". In contrast to the common assumption that providing too many details in explanations could result in information overload [23], one participant stated that "the information is detailed but not too complicated to understand". However, two participants reported that they could not understand why their scientific interests were compared with Wikipedia articles. Nine participants (39.13%) preferred the basic explanation, and found that it is a simple and understandable feature that provides a quick insight into where the interest keywords were extracted. They also liked that this explanation provides interactivity by providing "direct feedback to mouse hovering". However, two participants found this explanation to be superficial and limited. The intermediate explanation was preferred by four participants (17.39%). Participants found it helpful to see from which publication the specific interest keyword was extracted. Similarly, one participant stated that the level of detail in this explanation is suitable to understand how the interest model was generated without "having to know all the steps that are taking place in the backend". However, one participant believed that providing the name of the used algorithm may not be helpful for users that do not have the technical knowledge.

For the explanations of the recommendations, ten participants (43.48%) stated that they prefer the intermediate explanation. They liked that the heatmap is interactive and provides colors that help to quickly identify the similarities between their interests and the recommended tweet. While one participant found that the details in this explanation are presented in a simple manner, other participants stated that it might be overwhelming and hard to understand. One participant suggested making the chart bigger and using rounded percentages (e.g., 43% instead of 42.56%). The basic explanation was preferred by seven participants (30.43%). These participants perceived the explanation to be easy to understand and most time-efficient, as it helps them to determine the quality of the tweet very quickly without "having to click anywhere to see the explanation". They also found that the level of detail in this explanation is sufficient and that they don't need more detailed information. However, several participants suggested using colors instead of bold text to highlight the respective interest keyword in the tweet. Six participants (26.09%) preferred the advanced explanation. Participants found that this explanation most explains the inner logic of the system and that it is a great way of showing the user how the system came to the recommendations. One participant believed that this explanation would only be needed once since it explains the system via example and does not differ between the tweets. However, some participants reported that they could not understand the explanation and felt overwhelmed by the provided details. For example, one participant wrote "It is much information and I lost track during viewing the figure". Moreover, several participants suggested optimizing the chart by making it bigger and more compact by showing only the most important steps.

To further investigate how users perceive on-demand personalized explanations, we additionally asked participants about their opinion on this feature in general. The majority of the participants liked the idea of offering on-demand explanations with varying levels of detail to meet the needs and preferences of different user groups. They stated that different users might be interested in different amounts of information, depending on their knowledge, expertise, or curiosity. Besides individual differences, they believed that the need for information also depends on the situation, as one participant wrote "When something goes wrong, I might be interested in more detailed explanations to fully understand the mistakes". Moreover, participants believed that "providing all information in a single step would be too much", and that on-demand explanations give users the opportunity to discover as much as they want. Only one participant suggested providing just one explanation instead of "offering so many on the plate".

In sum, the results of the user study show that (1) the participants had different opinions regarding what level of detail they prefer to see. This confirms that potential individual user differences and their goals influence their preferences towards the explanation level, as also reported in other studies (e.g., [8, 13]) and (2) providing on-demand personalized explanation with varying level of details has a positive impact on user satisfaction with the explainable RS.

5 CONCLUSION AND FUTURE WORK

In this paper, we addressed an aspect that remains under-research in the literature on explainable recommendation, namely providing personalized explanations. To this end, we proposed a transparent Recommendation and Interest Modeling Application (RIMA) that provides on-demand personalized explanations with varying levels of detail to meet the demands of different types of end-users. The results of a preliminary qualitative user study demonstrated the potential benefits of the RIMA approach in helping users with different backgrounds understand how the recommender system works without overwhelming them. This work represents a first step to inform the design and appropriate use of personalized explanation interfaces in recommender systems. In future work, we will follow a human-centered design (HCD) approach to develop other possible explanatory visualizations at the three levels of detail. In addition to recommending tweets, we are planning to recommend publications and researchers and develop different visualizations to explain these recommendations. Furthermore, a more extensive quantitative and qualitative user study will be conducted to investigate the relationship between the users' characteristics and the level of detail of the explanations, and the effects of these two variables on the perception of and interaction with the explainable recommender system.

REFERENCES

[1] Vijay Arya, Rachel KE Bellamy, Pin-Yu Chen, Amit Dhurandhar, Michael Hind, Samuel C Hoffman, Stephanie Houde, Q Vera Liao, Ronny Luss, Aleksandra Mojsilović, et al. 2019. One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques. *arXiv preprint arXiv:1909.03012* (2019).

[2] Shuo Chang, F Maxwell Harper, and Loren Gilbert Terveen. 2016. Crowd-based personalized natural language explanations for recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*. 175–182.

[3] Tim Donkers, Benedikt Loepp, and Jürgen Ziegler. 2018. Explaining Recommendations by Means of User Reviews.. In *IUI Workshops*.

[4] Haiyan Fan and Marshall Scott Poole. 2006. What is personalization? Perspectives on the design and implementation of personalization in information systems. *Journal of Organizational Computing and Electronic Commerce* 16, 3-4 (2006), 179–202.

[5] Fatih Gedikli, Dietmar Jannach, and Mouzhi Ge. 2014. How should I explain? A comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies* 72, 4 (2014), 367–382.

[6] Mouadh Guesmi, mohamed Amine Chatti, and Arham Muslim. 2020. A Review of Explanatory Visualizations in Recommender Systems. In *Companion Proceedings 10th International Conference on Learning Analytics and Knowledge (LAK20)*. 480–491.

[7] Alexander Jung and Pedro HJ Nardelli. 2020. An information-theoretic approach to personalized explainable machine learning. *IEEE Signal Processing Letters* 27 (2020), 825–829.

[8] Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. 2019. Personalized explanations for hybrid recommender systems. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 379–390.

[9] Pigi Kouki, James Schaffer, Jay Pujara, John O'Donovan, and Lise Getoor. 2020. Generating and Understanding Personalized Explanations in Hybrid Recommender Systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 10, 4 (2020), 1–40.

[10] NIklas Kuhl, Jodie Lobana, and Christian Meske. 2020. Do you comply with AI?–Personalized explanations of learning algorithms and their impact on employees' compliance behavior. *arXiv preprint arXiv:2002.08777* (2020).

[11] Yichao Lu, Ruihai Dong, and Barry Smyth. 2018. Why I like it: multi-task learning for recommendation and explanation. In *Proceedings of the 12th ACM Conference on Recommender Systems*. 4–12.

[12] James McInerney, Benjamin Lacker, Samantha Hansen, Karl Higley, Hugues Bouchard, Alois Gruson, and Rishabh Mehrotra. 2018. Explore, exploit, and explain: personalizing explainable recommendations with bandits. In *Proceedings of the 12th ACM conference on recommender systems*. 31–39.

[13] Martijn Millecamp, Nyi Nyi Htun, Cristina Conati, and Katrien Verbert. 2019. To explain or not to explain: the effects of personal characteristics when explaining music recommendations. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 397–407.

[14] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267 (2019), 1–38.

[15] Sina Mohseni, Niloofar Zarei, and Eric D Ragan. 2018. A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. *arXiv* (2018), arXiv–1811.

[16] Cataldo Musto, Fedelucio Narducci, Pasquale Lops, Marco De Gemmis, and Giovanni Semeraro. 2016. ExpLOD: a framework for explaining recommendations based on the linked open data cloud. In *Proceedings of the 10th ACM Conference on Recommender Systems*. 151–154.

[17] Ingrid Nunes and Dietmar Jannach. 2017. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction* 27, 3-5 (2017), 393–444.

[18] Johanes Schneider and Joshua Handali. 2019. Personalized explanation in machine learning: A conceptualization. *arXiv preprint arXiv:1901.00770* (2019).

[19] Martin Svrcek, Michal Kompan, and Maria Bielikova. 2019. Towards understandable personalized recommendations: Hybrid explanations. *Computer Science and Information Systems* 16, 1 (2019), 179–203.

[20] Nava Tintarev and Judith Masthoff. 2012. Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction* 22, 4-5 (2012), 399–439.

[21] Nava Tintarev and Judith Masthoff. 2015. Explaining recommendations: Design and evaluation. In *Recommender systems handbook*. Springer, 353–382.

[22] Yongfeng Zhang and Xu Chen. 2018. Explainable recommendation: A survey and new perspectives. *arXiv preprint arXiv:1804.11192* (2018).

[23] Ruijing Zhao, Izak Benbasat, and Hasan Cavusoglu. 2019. Do users always want to know more? Investigating the relationship between system transparency and users' trust in advice-giving systems. In *Proceedings of the 27th European Conference on Information Systems*.