

Personalising Explainable Recommendations: Literature and Conceptualisation

Mohammad Naiseh¹, Nan Jiang¹, Jianbing Ma², and Raian Ali³

¹ Faculty of Science and Technology, Bournemouth University, United Kingdom

² Chengdu University of Information Technology, China

³ Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar

{mnaiseh,njiang}@bournemouth.ac.uk

jianbingma@qq.com

raali2@hbku.edu.qa

Abstract. Explanations in intelligent systems aim to enhance a users' understandability of their reasoning process and the resulted decisions and recommendations. Explanations typically increase trust, user acceptance and retention. The need for explanations is on the rise due to the increasing public concerns about AI and the emergence of new laws, such as the General Data Protection Regulation (GDPR) in Europe. However, users are different in their needs for explanations, and such needs can depend on their dynamic context. Explanations suffer the risk of being seen as information overload, and this makes personalisation more needed. In this paper, we review literature around personalising explanations in intelligent systems. We synthesise a conceptualisation that puts together various aspects being considered important for the personalisation needs and implementation. Moreover, we identify several challenges which would need more research, including the frequency of explanation and their evolution in tandem with the ongoing user experience.

Keywords: Explanations · Personalisation · Human-Computer Interaction · Intelligent Systems

1 Introduction

Information systems that have intelligent or knowledge components have been widely used including knowledge-based systems, decision support systems, intelligent agents, and recommender systems. With the increase in data volume, velocity and types, the adoption of solutions where intelligent agents and end-user interact and work closely have been increased in various application domains. The services provided to end-users range from presenting recommendations to the more interactive and engaging forms chatbots and social robots. The end-users' trust is one of the main requirement for the success and acceptability of such services in real-world scenarios [19]. Generating understandable explanations has been considered as a fundamental demand to increase trust in recommendations made by the system [9, 49, 27, 31].

In the literature, there are explanation models proposed and employed in the field of recommendation systems [45, 15, 43, 8], machine learning [47], decision support systems [11, 6] and robots [24]. For example in the field of machine learning, Sokol et al. [47] developed a voice-enabled device to answer user’s questions about the automated loan application decisions to help the end-users to understand the system rationale and identify potential errors and biases. Moreover, explanations help to address an openness culture around artificial intelligence applications and encourage the adoption of good practices around accountability [29, 32], ethics [4] and compliance with regulations such as the General Data Protection Regulation in Europe (GDPR) [20].

The provided models for explanations in the literature were proposed to give enough information and compliance with regulations and social responsibility. The elicitation of users’ needs from these explanations remains limited, and the literature around how users want these explanations to be designed, timed and communicated has only recently started to become a pressing topic. Ribera et al. [44] suggested considering three groups of users, based on their goals, background and roles in the system. Various factors in the personalisation are yet to be researched and considered, and this includes the cognitive styles of users, their prior beliefs and personality and cultural characteristics such as agreeableness and uncertainty avoidance.

In this paper, we survey the current approaches to personalising explanations in the literature of intelligent systems. We elicit and synthesise the main factors which are considered necessary for the design of personalised explanations and discuss research gaps for future research. The findings of the survey suggest that only a limited number of works have been conducted so far and that the number of publications started to increase in the last three years. Areas of interest include the need to evolve explanations together with user experiences and their dynamic trust level and usage context.

2 Literature Review Process

To review the literature around personalised explanations in intelligent systems, we used several search engines; Google Scholar, IEEEExplore, Association for Information Systems (AIS), ScienceDirect (Elsevier), Springer and the ACM digital library.

We started with a start-set of papers using the following keywords ‘personalized explanation’ and ‘personalised explanation’ for inclusivity of British and American English. As the number of papers retrieved was relatively low, the snowballing method was applied to ensure the best possible coverage of the literature. We followed the guidelines provided by Wohlin [53] for conducting systematic literature reviews using snowballing approach. Then, we iterated the backward and forward snowballing until no new papers were found. We did not restrict the starting date, and we found papers implementing personalisation for explanations from 1996. We stopped the search for papers at the end of November 2019.

To decide whether to include a paper in the study, we needed to define the scope. The paper and its literature review method are mainly focused on *explainable intelligent systems*, by which we mean systems that inform the user about their reasoning and decision-making process including aspects about the dataset and training used and their certainty and chance for errors. *Personalisation* in the context of our study refers to “a toolbox of technologies and application features used in the design of an end-user experience“ [17]. The included papers needed to match our definitions above and also discuss *personalised explanations in intelligent systems*, referring to how explanations can be tailored to a specific user or group of users.

The number of papers retrieved by the keyword search and the snowballing was 206 papers. However, reading the paper and matching them to our criteria and scope above, we have a total of 48 papers. We excluded papers which talked of personalised explanations as a general requirement but did not focus on it as the main aspect. We also excluded papers which discussed personalisation of explanations as a general usability requirement without particular consideration of the nature of intelligent systems.

Through a first read of the resulted 48 papers, we observed two main directions for the research in the area. The first focuses on the *needs* for personalised explanations from the users perspective and their requirements of such services. The second focuses on the *implementation* aspect and the process of designing explanation systems. This category of papers expresses partly of fully implemented personalisation technique in explainable systems. Fig. 1 illustrates the number of papers in each category over the years. The figure denotes that papers that emphasise the needs for personalised explanations started in 2017. Generally, the number of papers in the needs category increased in the last three years. The reason for that could be attributed to public demand and new rights of people such as the right to automated decision making, including profiling in the European General Data Privacy Regulations (GDPR). On the other hand, we can see that the research on the implementation aspect started early in 1996. Even though the number of studies (48 studies) is relatively low, the increasing number of studies in the last three years shows that research is getting more interest in the academic community.

3 Personalised Explanations: Needs-focused research

The complexity of intelligent systems and their wide adoption in real-world and daily life applications like healthcare made them more visible and familiar but at the same time more questionable. Driven by theories of the ethical responsibility of technology development, informed consent and informed decision making, regulations started to emphasise and even demand the right of citizens to be explained how systems work [20]. Researchers argued that the current approaches are limited in personalising explanations to end-users and their needs, which motivates the researchers to match different end-users with different explanations types. For instance, Tomsett et al. [52] propose a model that define six different

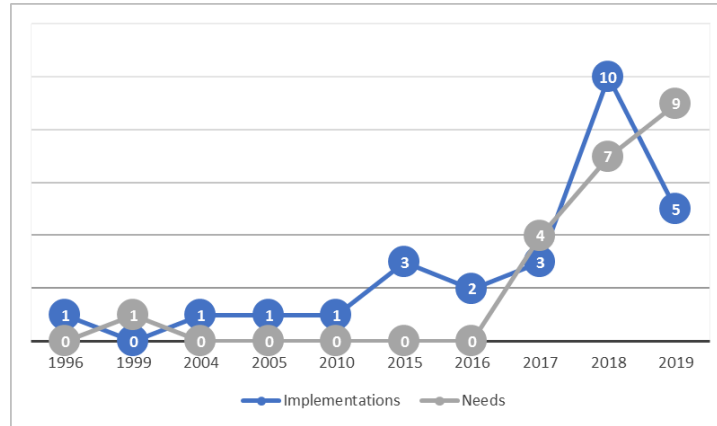


Fig. 1. The number of relevant papers in both categories (Needs and Implementations) over the years.

user roles in the machine learning system and argued that designers of the system should consider providing different explanations that match the needs of each role in the system. Also, Avi et al. [46] presented three categories of users who are different in their demands of the nature and level of explanations; Regular user, Expert user, and External entity. These results are supported by Millecamp et al. [36] who investigated how the end-user characteristics, including their personal profile and role in the system, affect the design of the explanations and the interaction with the system. We synthesised four categories of needs which can drive the personalisation of explanations and we explain them in the rest of this section.

User’s motivations and goals. Users might pursue different goals throughout their interaction with the system. Goals such as curiosity, verify the output, learn from the system and improve future interaction require adaptive and personalise explainable systems to meet these motivations [7, 21]. For instance, Gregor et al. [21] argued that novice users use the explanations for learning where experts need explanations for verification. Hind et al. [23] highlight that explanation should be provided to end-users based on their motivations and present four groups of users with different motivations. Those groups were i) End-users decision-makers, ii) Users affected by the decisions, iii) Regulatory bodies and iv) AI system developers.

Cognitive load. Personalisation can also consider the different ways people process information and use their thoughts, perception and memory to make a judgement. The cognitive load can also relate to the individual personality cognitive resources and learning style. For example, Feng et al. [18] point out that expert users process more information provided by the explanation interface to understand the output of the system compared to novice users. The user’s level

of knowledge in the domain can also be a parameter that affects the cognitive load required to process explanations.

Cost of the decision. Personalising the explanation of a recommendation made by an intelligent system would also need to consider how expensive following that recommendation is for the user and whether it is critical (e.g. recommending changing the password v.s purchasing new security device) as argued by Kleinerman et al. [26] and Bunt et al. [5]. This estimation of costs and the acceptance of the recommendation and the demand for explanation and its personalisation could differ between users depending on personal and cultural factors, such as openness to new experience and uncertainty avoidance. We need user models for personalised explanations shall reflect such diversity.

Compliance with regulations. Krebs et al. [28] and Tomsett et al. [52] discuss the personalisation problem through the lens of compliance with the European GDPR (General Data Protection Regulations) where the users have the right to automated decision making, and this included being explained how they are being profiled and based on what data and processes.

4 Personalised Explanations: Implementation-focused research

In this section, we report on the literature that provided more practical processes and approaches to implementing personalisation techniques. This category considered either partly or fully such implementation through approaches like user modelling with the aim of adapting to the different individuals and groups. We centred our analysis of the literature in this area on the six-dimension model introduced by Fa et al. [17]. The adoption of this model helps our analysis to highlight different design facets of personalisation. The dimensions are *Recipient, Entity, Channel, Mode, Tactics, and Unit of Analysis*. In the following subsection, we elaborate how each of these dimensions was tackled in the literature and synthesise a conceptualisation of its facets.

4.1 Recipient.

This dimension refers to users that receive and consume the explanation. More understanding about the user who is receiving the explanation and their previous interactions with the system is essential in order to time and convey the right explanation. In this regard, various factors were deemed important in the implementation of personalised explanations, and we detail them in the rest of this section.

User preferences. It refers to accommodating users' preferences around the subject and filtering the recommendations and their explanations. In terms of information preferences, Chen et al. [9] provided personalisation approach based on analysing users' reviews to understand users' preferences in a shopping recommender system. For example, the explanation communicated to user A emphasised the preference of the cuff of the T-shirt, whereas, user B explanation

indicated the preference about the T-shirt collar. Similarly, Chang et al. [8] presented an approach based on analysing users' reviews to find what are the movies features that the users prefer to be given in the explanation.

User personality. Kouki et al. [27] show how the end-user personality characteristics affect the choice of the explanations design elements where participants with low neuroticism prefer different explanation style compared to participants with high neuroticism. This means that emotional triggers, trust propensity and persistence in the explanation are main design facets. Cai et al. [7] state how different example-based explanations styles would be useful for users who under-trust the system more than users who over-trust the system. While trust is a dynamic phenomenon in its intrinsic nature [16], some users are more likely to be trusting than others due to personality differences. To date, little is known about the effects of user personality on the acceptance and usefulness of explanations.

User profile. Besides personality, that relates to the users mental and emotional states and typical attitude. The literature also considered user profile, which is more static and factual and relating to interaction with the system. For example, Kaptein et al. [25] compared the preference of explanation styles between adults and children. Similarly, Roitman et al. [45] generated explanations personalised to the patients' medical record in a health recommender system. More personalised explanations have been studied in the literature and considered information like gender [26], domain knowledge [3, 15] and user roles [3].

User-controlled personalisation. This style is defined in [22, 30, 12] where users are given control around how personalisation should be conducted. The approach considers the design of variable explanation system where users are responsible for controlling and configuring their own explanation. This can overlap with the other approaches. For example, users can configure what can be used from their profile and previous interaction to tailor a recommendation and whether they like algorithms like collaborative filtering to work and on what metrics of similarity index.

4.2 Entity.

Fa et al. [17] define personalised entity as the object, entity or substance to be personalised and as being the instantiation of the action of personalisation. We reviewed the literature to conceptualise those entities in the content of explanations, and we discuss the obtained results in the following.

Information. It refers to the information conveyed through the explanation and supposed to be consumed by the recipients. Carolis et al. [11] demonstrate how doctors in decision support systems require different explanation information compared to nurses and patients. From the doctors' perspective, the explanation is not only about how the system comes up with the results, but also the cost of following the recommendation or overriding it. This means that the information content would be customised to the role, responsibility and liability of the recipients. Our literature review showed that implementation papers are driven by personalising information mainly. 27 papers out of 27 papers in the category of implementation-focused research tackled the personalisation of the

information in the explanations. This entity could be personalised via different strategies, and we discuss this later in the Personalised Tactics section.

Interface. It refers to the interface and interaction method utilised in order to present and convey the explanation to the recipients [9, 27, 12]. Different users may require different interface design in terms of complexity, interactivity, layout and multimedia used. There are also factors relating to the domain and its mission and sensitivity, which may affect the design choice. For example, patients in health recommender systems would rather prefer simple and direct interface design like text, user-friendly components, and easy to differentiate layout and colours. However, the domain can also impose some challenges on the simplicity requirements. Professionals like doctors and pharmacists interfaces might unavoidably need to be more complex to include graphs and advanced dialogues and visualisations. Learning is part of the personalisation of interfaces and its evolution. Diaz-Agudo et al. [12] build a Cased-Based Reasoning system to personalise the explanation interface to fit the explanation goal and the type of the user based on previous cases and feedback. Three papers out of 24 implementation-focused papers related to the interface design.

In this section, we also outline open research challenges that may be considered in future research in term of entity dimension. The challenges are mainly focused on determining the frequency, which is the number of occurrences of explanations. In other words, it refers to the need overtime for the intelligent system to explain itself and how this can change. The main aim of using frequency in the personalisation process is to avoid information redundancy and overload and to fit the dynamics of users needs and the context. There are methods and implementations in the literature which can be potentially used for personalising the frequency. One example is the approach described by Sokol et al.[47], where the process of personalising frequency is left for the users to specify. On the other hand, Huang et al. [24] introduced an approach in which the frequency of the explanations is a system task. This approach is based on explaining the critical states and actions of the robots, rather than explaining all the robot actions.

4.3 Channel.

This dimension refers to different methods and communication facilities through which users can access personalised content. Examples include online interfaces, printed documents, email, voice, non-verbal cues and haptic feedback. Our literature review noted a lack of research around this facet, and it is assumed that the choice is mainly relevant to the availability of resources within the task context where recipients and their characteristics and goals are not considered.

4.4 Personalisation mode.

This dimension refers to the type of interaction and dialogue between the system and the user to accomplish the personalisation. The *user mode* gives the individual users a choice to opt-in and specify their preferences and choose their

explanations elements. Diaz-Agudo et al. [12] provide a configurable explanation interface which allows the user to select between different visualisation charts, change colours and sizes of the text. User mode can also enable users to request information when they need it. Chiyah Garcia et al. [19] provided an approach for on-demand queries for explanations from the intelligent system using Natural Language Generation techniques. *Designer mode* refers to the anticipatory or adaptive logic provided by the designers on how explanations made by the system should be derived and issued. An *anticipatory* personalisation is based on rules about the users' profile and their characteristics such as preference, demographics, needs and cognitive. Bofeng et al. [3] classified the explanation content to five different groups for users based on their level of knowledge. The research matched the formed user models with the predefined explanation based on questions asked by the system in an expert system for earthquake prediction. Similarly, Quijano-Sanchez et al. [43] provide personalisation approach in a group recommender system based on a decision tree that steers the generation process of the personalised explanation. This approach combines knowledge extracted from Social Media, the knowledge generated by the user and group recommendations, and a number of additional factors like tie strength, satisfaction and personality to create word variances in the explanation. *Adaptive techniques* are more dynamic than anticipatory, where the system models the user behaviour based on its previous interaction with the system. Suzuki et al. [50] provide an approach to personalise the explanations using recurrent neural networks model that uses the users' reviews as a training set for the generated explanations. We categorised the existing work based on the personalisation mode types. The results are shown in Table 1.

4.5 Personalisation Tactics.

Fa et al. [17] define personalisation tactics as the different technological measures and strategies available for the designer to manipulate and enhance the effect of personalisation. From the papers we reviewed, we synthesise seven tactics that could be considered when designing personalised explanations and discuss them in the following. The choice of the strategy is not exclusive as the designer could have different strategies in the explainable interface (e.g. [3, 12, 11]).

Complexity-based personalisation. This tactic reflects the adaptation of the explanation based on end-users' ability to utilise the explanation concerning the level of complexity provided [19, 3]. The complexity of the explanation could have a different number of lines, number of chunks, number of new concepts, rules, reasons, or level of details. Identifying what complexity factors of explanation to personalise is essential for balancing between the response time, information overload and the users' information needs.

Content-based personalisation. Information should be structured in a way that can be proceeded to meet the variable needs and context, and this concerns the indexing, tagging and filtering of the information in tandem with the users' roles and their tasks and other usage characteristics. Chen et al and Lim et al. showed that users have different information needs [9, 33], or users may request

multiple information content from the system [13]. Stumpf et al. [49] showed the need to provide multiple information content to help the user to make informed decisions.

Table 1. Categorisation of the existing work on personalised explanation using personalisation mode types.

Reference	Personalisation Mode			Approach
	User	Anticipatory	Adaptive	
[42]	x	-	-	Configure the explanation
[25]	-	-	-	Not Available
[30]	x	-	-	Debug the explanation
[3, 2]	-	x	x	Fuzzy user model
[51]	-	-	x	Hybrid Explanation Method with regard to user preferences
[45]	-	x	-	Prioritization
[47, 19]	x	-	-	Dialogue
[15]	-	-	x	An algorithm that exploits users' reviews and ratings
[43]	-	x	-	Decision Tree
[27, 10]	-	-	x	Collaborative Filtering
[9]	-	-	x	Collaborative Filtering and exploiting users' reviews
[50]	-	-	x	RNN that exploit users' reviews
[8]	-	-	x	An algorithm based on modelling users' interaction with the recommender system
[12]	x	-	x	Case-Based Reasoning and manual configuration
[14]	-	-	x	An algorithm based on analysing users' reviews
[41]	-	-	x	A framework which takes the user profile and the recommendations to generate personalised explanations
[1]	-	-	x	An algorithm based on users' goal
[38]	-	x	-	An algorithm based on Hierarchical Task Network
[34]	-	-	x	A novel multi-task learning framework
[40]	-	-	x	A methodology based on exploiting users' reviews
[11]	x	-	-	Questions for building user models
[35]	-	-	x	Multi-armed bandit algorithm

Order-based personalisation. It refers to the order of the information content presented in the explanation and the phasing of explanation so that it meets the evolving users experience. This strategy appeared twice in our relevant papers [11, 39] e.g. Carlis et al. [11] discussed how the order and the priorities of the presented information might differ between patients, nurses and doctors in medical decision support system.

Evolvable personalisation. The decision-making process underlying explanations derivation and delivery can be designed to learn and evolve during the time based on user feedback and actual interaction. Miller [37] argued that the intelligent system should consider the information that has already explained to end-users in order to evolve the explanation during the time. Milliez et al. [38] develop an algorithm based on this idea to update the users' knowledge model so that the system can adapt the explanation to this level of knowledge. Similarly, Bofeng et al. [2] provide an evolvable approach based on an adaptive interview, which asks users questions to update and re-evaluate their knowledge during the usage time.

Style-based personalisation. It indicates the orientation, level, granularity and framing adopted by the intelligent system explains the action and the underlying goal of the explanation. Our analysis showed that the explanation style is inherently related to the domain and the explainable algorithm. One example can be taken from the field of robots where Kaptein et al. [25] found that goal-based explanation, which provides information about the desired outcome from the decision was preferable by adults. Children's preference, on the other hand, was more towards belief-based explanation, which explain the behaviour of the intelligent agent based on the reasons that let the agent choose one action over the others. Machine learning is also another field where the explanation style is studied. For example, Dodge et al. [13] showed how different explanation styles affect users' perception of the fairness of the machine learning decisions, particularly, the difference between the local explanations and the global explanations. Moreover, Kouki et al. [27] used this tactic in recommender systems to personalise the explanation in a different format (text and graphics) and explanation styles (item-based, user-based and social-based).

Presentation-based personalisation. It refers to the method used to convey explanations, including whether and how users can interact with the communication medium. Preferences towards presentation methods could differ between users based on their goals, level of knowledge and familiarity with the method. Feng et al. [18] indicated the importance of personalising explanations to end-user by studying the effect of the presentation method on expert and novice users. Results from their study produced a more accurate and realistic evaluation for machine learning explanations methods. Kouki et al. [27] used explanation presentation method as a control variable to find which visualisation method is more persuasive for the end-users over other methods.

Format-based personalisation. It refers to change the language, framing and layout, including colours and font sizes to reflect the importance of certain parts of the explanation. Diaz-Agudo et al. [12] provide an approach to enable end-users to choose their explanation template in terms of colours, size and charts.

4.6 Unit of analysis.

This dimension refers to the view of the user that the personalisation design takes. The explanation can be designed to deal with categories of users, such as children or adults, experts or novices. In a different setting, it can be designed

toward addressing a unique individual, assumed to be different from all others, e.g. the head of the emergency unit in a hospital. The literature around the unit of analysis dimension is summarised in details in Table 2. The majority of the personalisation techniques reviewed considered individual users as the unit of analysis. This is because most of the literature belonged to recommender systems research community where the user profile construction is the main personalisation practice. The groups profiling and their dynamics have been considered in the fields of machine learning and expert systems, but still the amount of research in that collective and group context is limited.

Table 2. Research brief summary

Data gathering methods	Implicit	[9, 27, 43, 8, 10, 15, 45, 50, 14, 41, 34, 40, 51]
	Explicit	[30, 42, 47, 3, 19, 11, 35]
	Mixed Method	[12, 2]
	Not Available	[25]
Recipient gathered data	User preferences	[9, 27, 8, 43, 15, 50, 47, 14, 41, 11] [34, 40, 51]
	User profile	[10, 45, 25]
	Users Knowledge	[3, 2, 12, 1, 11, 38]
	No Information	[42, 19, 30]
Personalised content	Information	[9, 27, 8, 43, 15, 50, 14, 41, 34, 40, 25, 30] [2, 12, 34, 42, 45, 10, 3, 19, 47, 11, 51]
	Interface	[27, 12, 9]
Personalised tactics	Content	[9, 30, 42, 8, 43, 41, 40] [15, 47, 45, 3, 2, 50, 34, 10, 14, 1, 19, 11, 38]
	Style	[25, 27, 41, 11, 51, 35]
	Order	[11, 39]
	Complexity	[3, 2, 38]
	Evolvable	[2, 38]
	Presentation	[27, 12]
	Format	[12]
Unit of Analysis	Individual	[9, 30, 27, 42, 8, 43, 10, 15, 51] [47, 12, 50, 14, 41, 34, 1, 40, 38]
	Group	[25, 3, 2, 11]

5 Discussion and Future work

In this paper, we reviewed the literature of personalising the explanations in intelligent systems. We synthesised two classifications from different research domains such as Personalisation, User Experience and Explainable Artificial Intelligence [17, 37]. Results from the analysis showed that current literature and approaches of explainable models in intelligent systems lacked user-based research. Such research is important for gathering the lived experience of users and their perception and preferences in the process and the match to their roles in the system [44, 48], user personality [27, 7], domain knowledge [1, 38, 18] and user goals [21, 23]. Moreover, we also reviewed the current approaches in the Implementations category and outlined different algorithms, frameworks and methods for realising personalisation in explainable models. Table 2. presents a categorisation of the existing work in personalising the explanations. We note here that the consideration of users' input and choices was not a main focus in

the implementation strategies, although it was considered important, especially for giving a sense of control to users. The amount of interactivity and user control and the effect on user experience when balancing between intelligent and user-administered personalisation is an open research issue.

Our aim in future research is to focus on the end-users' perspective and their various personas and expectations, and we may need to produce explanations of a particular persona. To derive the goals of users and keep up to date with their dynamic nature, additional data are to be captured from the users, and this will require further explanation, i.e. meta explanation. We also need to provide approaches which enable the detection of needs and goals dynamics during the time to adapt these changes to the end-users. User studies such as diary studies, interviews and observations are needed to determine the nature of these dynamics. Results from these studies will provide us for guidelines to design long term explainable system and what data should be collected from the user to enhance this personalisation.

6 Acknowledgments

This work is partially funded by iQ HealthTech and Bournemouth university PGR development fund.

References

1. Barria-Pineda, J., Akhuseyinoglu, K., Brusilovsky, P.: Explaining need-based educational recommendations using interactive open learner models. In: Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization. pp. 273–277. UMAP'19 Adjunct, ACM, New York, NY, USA (2019). <https://doi.org/10.1145/3314183.3323463>, <http://doi.acm.org/10.1145/3314183.3323463>
2. Bofeng, Z., Na, W., Gengfeng, W., Sheng, L.: Research on a personalized expert system explanation method based on fuzzy user model. In: Fifth World Congress on Intelligent Control and Automation (IEEE Cat. No. 04EX788). vol. 5, pp. 3996–4000. IEEE (2004)
3. Bofeng, Z., Yue, L.: Customized explanation in expert system for earthquake prediction. In: 17th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'05). pp. 5–pp. IEEE (2005)
4. Bostrom, N., Yudkowsky, E.: The ethics of artificial intelligence. *The Cambridge handbook of artificial intelligence* **1**, 316–334 (2014)
5. Bunt, A., Lount, M., Lauzon, C.: Are explanations always important?: a study of deployed, low-cost intelligent interactive systems. In: Proceedings of the 2012 ACM international conference on Intelligent User Interfaces. pp. 169–178. ACM (2012)
6. Bussone, A., Stumpf, S., O'Sullivan, D.: The role of explanations on trust and reliance in clinical decision support systems. In: 2015 International Conference on Healthcare Informatics. pp. 160–169. IEEE (2015)
7. Cai, C.J., Jongejan, J., Holbrook, J.: The effects of example-based explanations in a machine learning interface. In: Proceedings of the 24th International Conference on Intelligent User Interfaces. pp. 258–262. ACM (2019)

8. Chang, S., Harper, F.M., Terveen, L.G.: Crowd-based personalized natural language explanations for recommendations. In: Proceedings of the 10th ACM Conference on Recommender Systems. pp. 175–182. ACM (2016)
9. Chen, X., Chen, H., Xu, H., Zhang, Y., Cao, Y., Qin, Z., Zha, H.: Personalized fashion recommendation with visual explanations based on multimodal attention network: Towards visually explainable recommendation. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 765–774. ACM (2019)
10. Coba, L., Rook, L., Zanker, M., Symeonidis, P.: Decision making strategies differ in the presence of collaborative explanations: Two conjoint studies. In: Proceedings of the 24th International Conference on Intelligent User Interfaces. pp. 291–302. ACM (2019)
11. De Carolis, B., de Rosis, F., Grasso, F., Rossiello, A., Berry, D.C., Gillie, T.: Generating recipient-centered explanations about drug prescription. *Artificial Intelligence in Medicine* **8**(2), 123–145 (1996)
12. Díaz-Agudo, B., Recio-García, J.A., Jiménez-Díaz, G.: Data explanation with cbr. ICCBR 2018 p. 64
13. Dodge, J., Liao, Q.V., Zhang, Y., Bellamy, R.K., Dugan, C.: Explaining models: an empirical study of how explanations impact fairness judgment. In: Proceedings of the 24th International Conference on Intelligent User Interfaces. pp. 275–285. ACM (2019)
14. Dragovic, N., Madrazo Azpiazú, I., Pera, M.S.: From recommendation to curation: When the system becomes your personal docent (IntRS 2018)
15. Dragovic, N., Pera, M.S.: Exploiting reviews to generate personalized and justified recommendations to guide users’ selections. In: The Thirtieth International Flairs Conference (2017)
16. Falcone, R., Castelfranchi, C.: Trust dynamics: How trust is influenced by direct experiences and by trust itself. In: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems, 2004. AAMAS 2004. pp. 740–747. IEEE (2004)
17. Fan, H., Poole, M.: Perspectives on personalization. AMCIS 2003 Proceedings p. 273 (2003)
18. Feng, S., Boyd-Graber, J.: What can ai do for me?: evaluating machine learning interpretations in cooperative play. In: Proceedings of the 24th International Conference on Intelligent User Interfaces. pp. 229–239. ACM (2019)
19. García, F.J.C., Robb, D.A., Liu, X., Laskov, A., Patron, P., Hastie, H.: Explainable autonomy: A study of explanation styles for building clear mental models. In: Proceedings of the 11th International Conference on Natural Language Generation. pp. 99–108 (2018)
20. Goodman, B., Flaxman, S.: Eu regulations on algorithmic decision-making and a “right to explanation”. In: ICML workshop on human interpretability in machine learning (WHI 2016), New York, NY. (2016)
21. Gregor, S., Benbasat, I.: Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS quarterly* pp. 497–530 (1999)
22. Groce, A., Kulesza, T., Zhang, C., Shamasunder, S., Burnett, M., Wong, W.K., Stumpf, S., Das, S., Shinsel, A., Bice, F., et al.: You are the only possible oracle: Effective test selection for end users of interactive machine learning systems. *IEEE Transactions on Software Engineering* **40**(3), 307–323 (2013)
23. Hind, M., Wei, D., Campbell, M., Codella, N.C., Dhurandhar, A., Mojsilović, A., Natesan Ramamurthy, K., Varshney, K.R.: Ted: Teaching ai to explain its decisions.

- In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. pp. 123–129. ACM (2019)
24. Huang, S.H., Bhatia, K., Abbeel, P., Dragan, A.D.: Establishing appropriate trust via critical states. In: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 3929–3936. IEEE (2018)
 25. Kaptein, F., Broekens, J., Hindriks, K., Neerincx, M.: Personalised self-explanation by robots: The role of goals versus beliefs in robot-action explanation for children and adults. In: 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). pp. 676–682. IEEE (2017)
 26. Kleinerman, A., Rosenfeld, A., Kraus, S.: Providing explanations for recommendations in reciprocal environments. In: Proceedings of the 12th ACM Conference on Recommender Systems. pp. 22–30. ACM (2018)
 27. Kouki, P., Schaffer, J., Pujara, J., O’Donovan, J., Getoor, L.: Personalized explanations for hybrid recommender systems. In: Proceedings of the 24th International Conference on Intelligent User Interfaces. pp. 379–390. ACM (2019)
 28. Krebs, L.M., Alvarado Rodriguez, O.L., Dewitte, P., Ausloos, J., Geerts, D., Naudts, L., Verbert, K.: Tell me what you know: Gdpr implications on designing transparency and accountability for news recommender systems. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. p. LBW2610. ACM (2019)
 29. Kroll, J.A., Barocas, S., Felten, E.W., Reidenberg, J.R., Robinson, D.G., Yu, H.: Accountable algorithms. *U. Pa. L. Rev.* **165**, 633 (2016)
 30. Kulesza, T., Burnett, M., Wong, W.K., Stumpf, S.: Principles of explanatory debugging to personalize interactive machine learning. In: Proceedings of the 20th international conference on intelligent user interfaces. pp. 126–137. ACM (2015)
 31. Lamche, B., Adigüzel, U., Wörndl, W.: Interactive explanations in mobile shopping recommender systems. In: Joint Workshop on Interfaces and Human Decision Making in Recommender Systems. p. 14 (2014)
 32. Lepri, B., Oliver, N., Letouzé, E., Pentland, A., Vinck, P.: Fair, transparent, and accountable algorithmic decision-making processes. *Philosophy & Technology* **31**(4), 611–627 (2018)
 33. Lim, B.Y., Dey, A.K.: Assessing demand for intelligibility in context-aware applications. In: Proceedings of the 11th international conference on Ubiquitous computing. pp. 195–204. ACM (2009)
 34. Lu, Y., Dong, R., Smyth, B.: Why i like it: multi-task learning for recommendation and explanation. In: Proceedings of the 12th ACM Conference on Recommender Systems. pp. 4–12. ACM (2018)
 35. McInerney, J., Lacker, B., Hansen, S., Higley, K., Bouchard, H., Gruson, A., Mehrotra, R.: Explore, exploit, and explain: personalizing explainable recommendations with bandits. In: Proceedings of the 12th ACM Conference on Recommender Systems. pp. 31–39. ACM (2018)
 36. Millecamp, M., Htun, N.N., Conati, C., Verbert, K.: To explain or not to explain: the effects of personal characteristics when explaining music recommendations. In: IUI. pp. 397–407 (2019)
 37. Miller, T.: Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* (2018)
 38. Milliez, G., Lallement, R., Fiore, M., Alami, R.: Using human knowledge awareness to adapt collaborative plan generation, explanation and monitoring. In: The Eleventh ACM/IEEE International Conference on Human Robot Interaction. pp. 43–50. IEEE Press (2016)

39. Muhammad, K., Lawlor, A., Rafter, R., Smyth, B.: Great explanations: Opinionated explanations for recommendations. In: International Conference on Case-Based Reasoning. pp. 244–258. Springer (2015)
40. Musto, C., Lops, P., de Gemmis, M., Semeraro, G.: Justifying recommendations through aspect-based sentiment analysis of users reviews. In: Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization. pp. 4–12. ACM (2019)
41. Musto, C., Narducci, F., Lops, P., de Gemmis, M., Semeraro, G.: Linked open data-based explanations for transparent recommender systems. *International Journal of Human-Computer Studies* **121**, 93–107 (2019)
42. Olah, C., Satyanarayan, A., Johnson, I., Carter, S., Schubert, L., Ye, K., Mordvintsev, A.: The building blocks of interpretability. *Distill* **3**(3), e10 (2018)
43. Quijano-Sanchez, L., Sauer, C., Recio-Garcia, J.A., Diaz-Agudo, B.: Make it personal: a social explanation system applied to group recommendations. *Expert Systems with Applications* **76**, 36–48 (2017)
44. Ribera, M., Lapedriza, À.: Can we do better explanations? a proposal of user-centered explainable ai. In: *IUI Workshops* (2019)
45. Roitman, H., Messika, Y., Tsimmerman, Y., Maman, Y.: Increasing patient safety using explanation-driven personalized content recommendation. In: Proceedings of the 1st ACM International Health Informatics Symposium. pp. 430–434. ACM (2010)
46. Rosenfeld, A., Richardson, A.: Explainability in human-agent systems. *Autonomous Agents and Multi-Agent Systems* pp. 1–33 (2019)
47. Sokol, K., Flach, P.A.: Glass-box: Explaining ai decisions with counterfactual statements through conversation with a voice-enabled virtual assistant. In: *IJCAI*. pp. 5868–5870 (2018)
48. Srinivasan, R., Chander, A., Pezeshkpour, P.: Generating user-friendly explanations for loan denials using gans. *arXiv:1906.10244* (2019)
49. Stumpf, S., Rajaram, V., Li, L., Wong, W.K., Burnett, M., Dietterich, T., Sullivan, E., Herlocker, J.: Interacting meaningfully with machine learning systems: Three experiments. *International Journal of Human-Computer Studies* **67**(8), 639–662 (2009)
50. Suzuki, T., Oyama, S., Kurihara, M.: Toward explainable recommendations: Generating review text from multicriteria evaluation data. In: 2018 IEEE International Conference on Big Data (Big Data). pp. 3549–3551. IEEE (2018)
51. Svrcek, M., Kompan, M., Bielikova, M.: Towards understandable personalized recommendations: Hybrid explanations. *Computer Science & Information Systems* **16**(1) (2019)
52. Tomsett, R., Braines, D., Harborne, D., Preece, A., Chakraborty, S.: Interpretable to whom? a role-based model for analyzing interpretable machine learning systems (2018)
53. Wohlin, C.: Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th international conference on evaluation and assessment in software engineering. p. 38. Citeseer (2014)