Explainable Recommendations in Intelligent Systems: Delivery Methods, Modalities and Risks

Mohammad Naiseh[1], Nan Jiang[1], Jianbing Ma[2], and Raian Ali[3]

[1] Faculty of Science and Technology, Bournemouth University, United Kingdom
[2] Chengdu University of Information Technology, China
[3] Information and Computing Technology Division, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar
{mnaiseh,njiang}@bournemouth.ac.uk
mjb@cuit.edu.cn
raali2@hbku.edu.qa

Abstract. With the increase in data volume, velocity and types, intelligent human-agent systems have become popular and adopted in different application domains, including critical and sensitive areas such as health and security. Humans' trust, their consent and receptiveness to recommendations are the main requirement for the success of such services. Recently, the demand on explaining the recommendations to humans has increased both from humans interacting with these systems so that they make an informed decision and, also, owners and systems managers to increase transparency and consequently trust and users' retention. Existing systematic reviews in the area of explainable recommendations focused on the goal of providing explanations, their presentation and informational content. In this paper, we review the literature with a focus on two user experience facets of explanations; delivery methods and modalities. We then focus on the risks of explanation both on user experience and their decision making. Our review revealed that explanations delivery to end-users is mostly designed to be along with the recommendation in a push and pull styles while archiving explanations for later accountability and traceability is still limited. We also found that the emphasis was mainly on the benefits of recommendations while risks and potential concerns, such as over-reliance on machines, is still a new area to explore.

Keywords: Explainable Recommendations, Human Factors in Information Systems, User-Centred Design, Explainable Artificial Intelligence

1 Introduction

The fast development in the fields of artificial intelligence and machine learning introduced more complexity in human-agent systems where humans and the algorithms interact with each other [31] (e.g. recommender systems, social robots

and decision support systems). It is becoming increasingly important to offer explanations on how algorithms decisions and recommendations are made so that humans stay informed and make better decisions whether or not to follow them and to which extent. The need for explanations is reinforced by the demand on openness culture around artificial intelligence applications and the adoption of good practices around accountability [44, 53], ethics [65] and compliance with the new regulations such as the General Data Protection Regulation in Europe (GDPR) [28].

An explanation is an information that communicates the underlying reasons for an event [58]. Explanation in artificial intelligence is a multi-faceted concept embracing elements from transparency, causality, bias, fairness and safety [31]. End-users need explanations for various reasons such as the verification of the output, learning from the system and improving its future operation [71]. Recent studies and surveys in this field explored the user experience facets of explanations such as the explanation goals, content and the different forms of presenting and communicating these explanations including natural language and charts [69, 64, 1]. However, an understanding of the existing research on the delivery methods and modalities is becoming also needed. Recent studies showed that the development of explaining the intelligent human-agent recommendations often faces problems and raises questions that must be addressed [79] (e.g. users ask for more functionalities in the explainable interface to satisfy their needs [23]). Failing in accommodating these facets and coping with the increasing complexity in the explanation interface and content leads to failure in meeting user needs and goals [11, 23]. Moreover, explanations could lead to undesirable effects on end-users and introduce new errors such as over-trust [20], when the end-users fail to recognise the absence of the correct recommendations.

Given the above research challenges and the increasing number of papers in the field of explaining intelligent human-agent recommendations is evidence that user experience facets had been an open research challenge recently. Hence, we conduct a systematic review around two design facets of explanations in intelligent human-agent systems: delivery methods and modalities types. These facets have not been explored in previous surveys and this, together with the increasing demand for usable explanations, motivated us to do this work. Also, we identify and present several risks of explanation both on user experience and their decision making with the purpose of informing the design process and help to detect explanation risks and to mitigate them proactively. The main goals of this study are to (i) identify classes of current explanation delivery methods and explainable interface modalities and their design considerations; (ii) identify potential risks while users are interacting with the explainable interface along with the potential design solutions; (iii) assist researchers in positioning the research challenges and problems to be resolved in this domain appropriately.

The remainder of this paper is structured as follows. Section 2 summarises the methodology and defines research questions. Section 3 outlines the results of the review organised according to each research question defined in Section 2. Section 4 discusses the results and future research challenges.

2 Research Method

We carry out a systematic review to classify, describe, and analyse existing literature around explainability in intelligent human-agent systems. A systematic review is a valuable tool to provide a holistic picture of the research in a particular area. It can also help in providing facets to consider when designing software systems and its results can be seen as a reference model. For example, Hosseini et al. performed a systematic review of crowdsourcing [34] to inform engineers on what to consider in their analysis and design processes in crowdsourcing projects. We follow PRISMA [59] rationale and method and conduct a systematic study for explanations with a focus on two design facets: delivery methods and modalities types. Also, and through the analysis of the literature, we extract several design challenges considering the explanation risks and present them as a road-map of future research for researchers and practitioners in the field. This systematic study will focus on addressing the following questions to get a clear depiction of the concept and the distribution of the research about it:

1. Delivery - What are the methods proposed to deliver the explanations to end-users and their design implications?
2. Modalities - What are the proposed modalities to be used by end-users to provide input to the explanation interface?
3. Reported risks with explanations - What are the main risks while users are interacting with the explanation interface?

Search string and relevant data sources. In our search for literature, we relied on four popular search engines that contain a large number of Journals and conferences of information systems which are: Google Scholar, IEEEXplore, ACM Digital Library, and Science Direct. We started the formation of the search string intending to cover the literature that combines intelligent systems, explainability and HCI. We select (“explanation” OR “Justification” OR “explainable” OR “Explainability”) AND (“Intelligent” OR “Smart”) AND (“System” OR “Agent”) AND (“HCI” OR “User experience” OR “Human-Centred” OR “User-centred”) as the search string. In order to address our research questions in the initial filtering phase, we choose to filter the papers through their title, abstract, and keywords. If there were some doubts about the relation between a paper and our scope, an additional reading through the introduction and the key parts of the paper was required to decide on the relevance. Based on the initial filtering search, we came up with 460 papers. We present our search results in Table 1.

Table 1. Data sources and results from literature search .

Data Source	Total results	Initial Filtering	Content Scanning
IEEEXplore	35	20	7
Google Scholar	443	218	29
ACM Digital Library	552	152	18
ScienceDirect	322	90	12

Content scanning. For each of the papers which we retrieved based on the initial filtering, we conducted a full-text content scanning to assess the relevance of the papers to our research questions ensuring that the paper was within the scope of this systematic study. The number reduced to 66 papers after the content scanning phase. The full set of Inclusion Criteria (IC) and Exclusion Criteria (EC) used in this reduction included:

- Recency (IC-1): Since the aim of the study is to identify the emerging research trends, challenges and gaps, we chose to focus on papers published in the last decade (2009- December 2019).
- Relevance (IC-2): The paper has to relate to one or more of our research questions. The reviewed papers should define explicitly one or more of our user experience facets (delivery methods, modalities and reported risks with the explanations).
- Full Access(IC-3): To include the paper, the content of the paper should be accessible in full-text.
- Duplicated papers (EC-1): We excluded repeated papers which have been published in an extended or complete version and considered the more inclusive version.
- Language and peer review (EC-2): We restricted our selected papers on papers that are written in English and published in recognised peer-reviewed journals and conferences.
- Domain-related (EC-3): The paper must be centred around the intelligent human-agent systems domain. For example, our search results introduced us with papers addressing the explanations from psychology, social science and theoretical computing perspective without direct relation to the user-experience aspect; these papers were excluded.

Data extraction and synthesis process. Considering the aforementioned criteria, 66 papers were selected for the data extraction and synthesis phase. After the content scanning phase, we formed data extraction forms to record the extracted data needed to answer our research question. The data extraction process was performed by the first author. However, an inter-rater reliability test was performed in which the other authors confirmed the first author results by a randomly selected set of papers. Then, we used an iterative process between the research team to formulate, combine and conceptualise the emerged concepts.

3 Results

This section summarises the results of our analysis of the reviewed papers and answers our three research questions. Later in Section 4, we comment on the overall picture of the research in this area and the challenges to address in future work. Also, Table 2 lists our reviewed papers with their corresponding aspects of explanations.

3.1 Delivery Methods

In this section, we answer RQ1 around the different delivery methods of explanations with a particular focus on how the delivery methods inter-relate with other design considerations. Delivery methods are not mutually exclusive, and multiple delivery options can be used in the same interface based on the context, the recipient of the explanation and the nature of the application. Our results revealed four delivery methods which have been studied in our reviewed papers and comment on their motivations and goals in the next sub-sections.

Persistent-specific: Explanations are delivered to the users for along with the recommendation in a straightforward and accessible way and without waiting for the user to request the explanation. The lifetime of the explanation in this method is specific to the user interaction time with the recommendation. In other words, the user is unable to consume the explanation after finishing the task. The main goal is to inform the user decides whether to accept the recommendation. This method used in the literature to foster trust [27], transparency [27], persuasiveness [72], user acceptance [39] and prevent errors and bias [73]. The cost-benefit analysis is challenging design consideration [47, 11], as users may perceive the cost of reading explanations to exceed their benefits [11].

Ad-hoc: The explanation in this category is designed to be delivered to the end-users when it is necessary and needed. This method is used in the literature in two ways:

On-demand: This method enables the users to request the explanation where the explanation is embedded in a separate view, and the users can ask for it. This is meant to reduce information overload in the interface [57, 5] when explanations are not always beneficial or crucial for the performing task [11, 84]. Also, this delivery method could blend well with the persistent-specific method, e.g. when users ask for further details in order to reveal the full set of explanation features [60]. On-demand method is useful where explanations contain a high level of information so users may get distracted and need more time to consume it [27, 76]. Also, it is argued to be more effective to reduce users cognitive effort and avoid overwhelming end-users with unnecessary information [67]. On the other hand, embedding explanations in a separate view argued in the literature that it might not fulfil the goal of presenting the explanation and become an additional burden on user-experience. Eslami et al. [25] and Leon et al. [52] found that users might not benefit from this method as end-users may hardly notice the on-demand button due to factors like their main focus and flow state.

Exploration: The users in this method are able to explore the nature of the explanation and the agent process and increasing the understanding of the reasoning behind the recommendation [9, 83]. This exploration could be: (a) feature-based exploration, where user can investigate how individual feature contributes to the recommendation and explanation output [49], (b) subset-based where input features specified by users are leveraged [46, 77] and (c) global exploration

where the nature of the data and its distribution are exploited [74]. Exploration techniques help users to build useful mental models and provide the user with the ability to discover more knowledge and about the agent in an interactive and engaging way [45]. Examples of such tools help the users in some problems like detecting bias in data [43], combat the filter-bubble effect in social media [38].

Persistent-generic: Explanations are stored as a report for later investigation, and the explanation is persistent without time limit. The report may include more information compared to persistent-specific and ad-hoc methods. For instance, information about the underlying processes of the algorithm decision making on each step of the process and the reasons for selecting each decision point [50]. This is essential in some application domains, such as clinical decision support systems where the explanation is a crucial factor for accountability, traceability and ethics [6, 17]. Most of our reviewed studies did not focus on developing approaches with the ability to access the explanation after finishing the task. Main approaches provided in the literature to apply this method include i) embedding the explanation in the "help page" [22] and ii) providing a dialogue interface to navigate the archive interactively [87].

Autonomous: This method appeared twice in our reviewed studies. The system in this method is responsible for deriving users' needs for an explanation based on the context. In other words, it is about the autonomy of the system to choose the time and the context to deliver the explanation. In contrary to the ad-hoc approach, which is a user-based delivery method, autonomous approaches are a systems-based method. Lim et al. [55] argued that this method could be used to provide privacy-sensitive information when the recommendation could provoke privacy violation so that it acts as a precautionary measure. Understanding the nature of the application and the different users' personas is essential to lunch this approach in human-agent systems. The papers that studied this method appeared in the domains of ubiquitous computing [55] and robots [36]. For instance, Huang et al. [36] develop an approach to explain the intelligent agent behaviour only in the critical situations, e.g. there is no need to explain why the autonomous vehicles slow down when the road is empty. This method was helpful to calibrate user trust and avoid over-trust and under-trust states.

3.2 Modalities types

Explanations in common applications are presented either as text or graphical representations in a static way [64]. However, explanations can be designed as interactive systems where the initial explanation represents a starting point for further user interactions, e.g. asking the user for correct parts of the explanation. Designers use such modalities to streamline user functionalities to explore more details about the underlying algorithm and put the user into control the output. Providing such interactive explainable interfaces can fulfil both persuasion and

over-trust reduction requirements by demonstrating the algorithmic reasoning in a thoroughness and experimental way to the end-users [73]. Research in this area is still limited, and it is unclear how to design interactive explanation interfaces in a way that is tailored and fit to users in standardised or personalised ways. Kulesza et al. [46] mentioned that supporting users with interactive explanation could lead to more complexity, as it needs a level of knowledge in software engineering and machine learning and also burden on the user experience. In this section, we focus on discovering common input modalities in the literature that typical explanation not only conveys information but also might trigger an interactive approach. We highlight these types and their potential usage scenarios.

Control: Users are enabled to play, change, regenerate or elicit some preferences about the agent in order to enhance their understanding of the underlying system [51]. The main principles behind this interaction style include boosting transparency and interpretability of the system processes and giving users control on their output [49, 56]. Studies found that the control functionalities can enhance the user experience as well as enriching mental models. The research in this area focused on providing dynamic explanations more than static approaches when users need to observe the inter-relation between different factors that influence the output - e.g. Tsai et al. [81] presented an approach based on a user-controlled function that include different explanation components, which allows tuning recommendation parameters for exploring social contacts at academic conferences.

Configure: This modality gives end-users the ability to choose what information, presentation, colours, order and size are suitable to reflect the importance, relevance and focus of certain parts of the explanation. This method is rarely studied in the literature as it appeared twice in our selected studies and without elaborating on the design considerations [16].

Dialogue: It indicates the explanations provided to the end-users in an interactive bi-directional style. The user can ask for specific information about the recommendations [87]. This approach is argued to be beneficial for the design of explanation interfaces and balance between the amount of information presented to the end-users and their cognitive efforts to process that information [68]. Our findings show that users have specific information requirements before they are willing to use recommendation such as system capability, the algorithmic reasoning and detailed information about the recommended item. For instance, Eiband et al. [23] revealed that users called for more accurate information about the recommended item rather than relying on the item-based and user-based explanations. These requirements could be fulfilled by using a dialogue interaction, as the user will be assured asking specific questions about the recommendation.

Debug: In this approach, the system presents its explanation to the end-users, where, on the other hand, users are enabled to provide corrections and feedback

about the explanation to the systems in order to improve output in the future recommendations [47, 45, 37]. User debugging can occur by different types of inputs, such as providing ratings to the explanation [24] and correcting parts of the explanations explicitly [45]. Providing the debug modality is argued to increase the algorithm accuracy by putting the Human-In-The-Loop.

3.3 Reported risks

In this section, we present the main risks and side effects while users are receiving the explanations from the human-agent systems. In the following, we compile a list of potential risks and side-effects of explanations which are likely to arise when the design process overlooks the user experience aspect.

Over-trust. In the situations where the cost of adopting the recommendations is high such as diagnosis and medical recommendations, it is risky to follow a recommendation rashly i.e. over-reliance. This effect could be enhanced through explanations [12, 20]. Research on how to reduce over-trust effect suggested different solutions, but it needs more investigation to measure and adjust the relationship between different variable including trust, certainty level, cognitive styles, personality and liability. Existing proposals revolve around comparative explanations [13], argumentation [20], personalised explanation based on user personality [75], uncertainty and error presentation [76]. Research is still needed to investigate how to embed these solutions in the interfaces considering other usability and user experience factors such as the timing, the level of details, the feedback to collect from end-users and the evolution of explanation to reflect it. It is worth noting that over-trust can be seen as a merging property which requires observing throughout the life-cycle of the intelligent human-agent systems. For example, users may over-trust a system due to cognitive anchoring and overconfidence biased, when it proves to be correct in a number of previous occasions.

Under-trust. As explanations could promote over-trust, it also could lead to under-trust issues [73], when the explanation is perceived to have a limited quality or fitness to the user intentions and context. Research of explanation quality discussed that improving the quality of the explanation is not about increasing transparency and recommendation rationale only. Springer et al. [76] showed that increasing the level of explanation details in an intelligent system may not necessarily lead to trust and it can lead to confusing users and harming their experience with the system. Another study showed that users could have an algorithm disillusionment when the algorithm use information derived about users as part of the explanation [25]. Explanations should be designed to simulate natural human interaction patterns so trust can be taken in a way similar to what a user would do in real life [73]. Another research linked under-trust with end-user personality. Millecamp et al. [57] found that explainable recommendations have under-trust issues to users with a high need for cognition e.g.the need to

interpret and understand how the situation is composed. On the other hand, the explanations increased the trust to the users with a low need for cognition.

Suspicious motivations. The motivation behind the explanation could be perceived as an attempt to manipulate the users. For example, marketing companies may try to explain with the purpose of enhancing the chance of purchasing the item recommended rather than informing the decision of the customer. This case is discussed by Chromik et al. [14] as a dark pattern of explainability. They discuss the problem in terms of explanation style, which is the phrasing of explanations and the modality type. The correction of the perception of a motivation to be suspicious could be conveyed to the user through the explanation itself. Eiband et al. [21] found that placebic explanations which do not supply the user with enough information about the algorithm decision-making process invoke a perception of a suspicious motivation behind the explanation.

Information overload. Explanations could cause information overload and overwhelm the end-users. They can become confusing and complicated [47]. Bunt et al. [11] argued that robust system design should help the users to derive its underling reasoning without much need for explanations and must avoid overwhelming users with unnecessary information. More transparency in the explanation affects users ability to detect the errors in the recommendations themselves [67] and increase users response time [62] and this may lead to losing timeliness,e.g. in taking an offer available for a limited time. Hence, approaches for balancing between the soundness and completeness of the explanations need to be developed [45].

Perceived loss of control. Users may perceive a loss of control when the system presents static explanations rather than dynamic and interactive explanations allowing them to query and investigate further. Holliday et al. [33] studied this effect when they examined the perceived control as a factor in two conditions (with and without explanations). They concluded that users in the absence of explanations showed more control-exerting behaviour of intelligent systems. Andreou et al. [2] and Eiband et al. [23] also found that static explanations are going to be seen incomplete for some users and sometimes misleading. This calls for personalised and more dynamic interfaces for explanations.

Refusal. Refusing the explanations may happen when users feel that putting cognitive efforts to read the explanations does not lead to better recommendations or better understanding. Moreover, users cab be typically focused on completing their tasks more than reading the explanations and improving their mental models [11]. The conflict between the explanations and prior beliefs, cultural backgrounds, the nature of the application, level of knowledge and interests could be other reasons for refusing the explanations. For example, Eiband et al. [23] found that some users were more interested to know information about

the specific recommendation rather than the recommendation process itself in the everyday intelligent systems (e.g. social media), whereas, this kind of explanations could be critical for other users in other application domains. This calls again for user-centred approaches to meet users' explainability needs that take users personality and contextual variables into account.

Table 2. The categorisation of the reviewed papers.

Delivery Methods	Persistent-specific	[8][21][15][6][18][17] [49][50][86][7][85][37] [60][35][4][70]
	Ad-hoc	[74][43][80][66][49] [40][5][60][52][4][68][83][30]
	Persistent-geniric	[8][22][50]
	Autonomous	[55][36]
Modalities types	Control	[74][38][43][80][66][9] [29][82][57][56][82][51][81]
	Configure	[78][16]
	Dialogue	[3][87][63][68]
	Debug	[10][46][78][54][49][24] [37][45]
Reported risks	Over-trust	[48][12][20][13][75] [76]
	Under-trust	[76][42][25][73][13][57]
	Refusal	[23][25][11][32]
	perceived loss of control	[2][82][33][23]
	Information overload	[47][62][41][55][76][67]
	Suspicious motivations	[14][21]

4 Discussion and Research Challenges

Our systematic review study investigated two design facets of explainable recommendations and we reported on the results and synthesised main dimensions and facets and focus areas in each of these facets in the previous section. In this section, we reflect on the status of the research in them and present a set of research challenges as open issues for future research.

Delivery methods. The popular methods in the literature focused on delivering the explanations to end-user while performing the task and while looking for recommendations. We still lack studies and the long term retrieval of such explanations, e.g. through a digital archive, and the effect of that on the accountability and traceability of these systems and users trust and adoption of them. Such approaches are important with the increasing adoption of intelligent human-agent systems in sensitive areas such as security and health. Also, it remains a challenge to design autonomous delivery which is able to consider the context and the situation when and where the users need explanations from the system. A cost-benefit analysis would then need to integrate explanations

well with the good practice of user experience. Personal factors are various and they can affect that autonomous-based delivery method (e.g. users with a low level of curiosity need less frequent and simple explanations). Techniques such as UI adaptation [26] can be used to adapt the delivery method based on users' personal factors. The perceived cost of the decision is another factor (e.g. recommending changing password v.s buying security device). Privacy can also determine how recommendations and their explanation are derived and delivered. For example, some recommendations can be based on simple demographic information about the user while others utilise usage and real-time data of the user.

Modalities. Explanations can be required to provide functionalities to allow users to navigate through them and query them in an interactive style rather than being only passive recipients of static information content. For instance, Bostandjiev et al. [9] studied whether adding additional interaction functionalities (e.g. supporting a "what-if" scenario) affects the user experience and they concluded that it increases the recommendation accuracy and enhances user experience. The integration of specific modalities could lead to different experiences during the interaction [46] i.e. some modalities cannot be utilised by end users without a high-level of understanding of the agent algorithm. Hence, modalities should be used with the consideration of the automatic usability evaluation (AUE). Also, user-friendliness and intelligent modalities would need to learn the explanation that best fit users goals and needs. For instance, simple feedback such as "explain more", "redundant explanation" or "different explanation" can support users who wish to involve with the explanations and improve the explanations in future interactions. In a previous paper [61], we reported on the results related to input modalities meant for tailoring the explanations for a specific user or group of users i.e. personalisation.

Reported risks. The researchers in our systematic review reported several challenges for HCI researchers and practitioners to develop explainability solutions and avoid the potential risks. Explaining recommendations can offer benefits for users trust and acceptance. Additionally, the emphasis on the benefits and overlooking the side effects can lead to less critical consequences in low-cost recommendation services, e.g., movies. In high-cost recommendations, e.g. prescription recommendations, users may over-trust, or under-trust the advice provided by the system and this may lead to critical consequences. Hence, the design of the explanations needs to consider the potential risks of presenting the explanations as a first-class issue. Also, the research needs to design explanations to evolve during the time considering what has been explained before to work for long-term interaction with the end-users and consider techniques from learning (e.g. constructive feedback [19]) that could mitigate these risks. We would need to develop evaluation metrics and questionnaires that cover the user-centred aspects of explanations and evaluate error-proneness and potential risks.

5 Conclusion and Future work

Driven by a growing need for, and interest in, intelligent human-agent systems, this paper presented a systematic review to clarify, map and analyse the relevant literature in the last ten years. The findings present the results regarding two main explanation design facets which are the delivery methods and the modalities. Also, we reflected on our systematic review and presented several challenges considering the risks while users are receiving the explanations. We elaborated on the status of the field and where research is lacking to aid future research in the area. We made the argument that explanations should be engineered using user-centred approaches and be evolved and adapted iteratively as their acceptance and trust are not only reliant on the information content and correctness but rather require consideration of a wider set of factors around users and their usage context and experience.

Acknowledgments

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

References

1. Al-Taie, M.Z., Kadry, S.: Visualization of explanations in recommender systems. Journal of Advanced Management Science Vol **2**(2), 140–144 (2014)
2. Andreou, A., Venkatadri, G., Goga, O., Gummadi, K., Loiseau, P., Mislove, A.: Investigating ad transparency mechanisms in social media: A case study of facebook's explanations (2018)
3. Arioua, A., Buche, P., Croitoru, M.: Explanatory dialogues with argumentative faculties over inconsistent knowledge bases. Expert Systems with Applications **80**, 244–262 (2017)
4. Bader, R., Woerndl, W., Karitnig, A., Leitner, G.: Designing an explanation interface for proactive recommendations in automotive scenarios. In: International Conference on User Modeling, Adaptation, and Personalization. pp. 92–104. Springer (2011)
5. 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. ACM (2019)
6. Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., Shadbolt, N.: 'it's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. p. 377. ACM (2018)
7. Biran, O., McKeown, K.R.: Human-centric justification of machine learning predictions. In: IJCAI. pp. 1461–1467 (2017)
8. Blake, J.N., Kerr, D.V., Gammack, J.G.: Streamlining patient consultations for sleep disorders with a knowledge-based cdss. Information Systems **56**, 109–119 (2016)

9. Bostandjiev, S., O'Donovan, J., Höllerer, T.: Tasteweights: a visual interactive hybrid recommender system. In: Proceedings of the sixth ACM conference on Recommender systems. pp. 35–42. ACM (2012)
10. Brooks, M., Amershi, S., Lee, B., Drucker, S.M., Kapoor, A., Simard, P.: Featureinsight: Visual support for error-driven feature ideation in text classification. In: 2015 IEEE Conference on Visual Analytics Science and Technology (VAST). pp. 105–112. IEEE (2015)
11. 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)
12. 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)
13. 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)
14. Chromik, M., Eiband, M., Völkel, S.T., Buschek, D.: Dark patterns of explainability, transparency, and user control for intelligent systems. In: IUI Workshops (2019)
15. Coba, L., Zanker, M., Rook, L., Symeonidis, P.: Exploring users' perception of collaborative explanation styles. In: 2018 IEEE 20th conference on business informatics (CBI). vol. 1, pp. 70–78. IEEE (2018)
16. Díaz-Agudo, B., Recio-Garcia, J.A., Jimenez-Díaz, G.: Data explanation with cbr. ICCBR 2018 p. 64
17. 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)
18. Dominguez, V., Messina, P., Donoso-Guzmán, I., Parra, D.: The effect of explanations and algorithmic accuracy on visual recommender systems of artistic images. In: Proceedings of the 24th International Conference on Intelligent User Interfaces. pp. 408–416. ACM (2019)
19. Du Toit, E.: Constructive feedback as a learning tool to enhance students' self-regulation and performance in higher education. Perspectives in Education **30**(2), 32–40 (2012)
20. Ehrlich, K., Kirk, S.E., Patterson, J., Rasmussen, J.C., Ross, S.I., Gruen, D.M.: Taking advice from intelligent systems: the double-edged sword of explanations. In: Proceedings of the 16th international conference on Intelligent user interfaces. pp. 125–134. ACM (2011)
21. Eiband, M., Buschek, D., Kremer, A., Hussmann, H.: The impact of placebic explanations on trust in intelligent systems. In: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. p. LBW0243. ACM (2019)
22. Eiband, M., Schneider, H., Buschek, D.: Normative vs. pragmatic: Two perspectives on the design of explanations in intelligent systems. In: IUI Workshops (2018)
23. Eiband, M., Völkel, S.T., Buschek, D., Cook, S., Hussmann, H.: When people and algorithms meet: user-reported problems in intelligent everyday applications. In: Proceedings of the 24th International Conference on Intelligent User Interfaces. pp. 96–106. ACM (2019)
24. Elahi, M., Ge, M., Ricci, F., Fernández-Tobías, I., Berkovsky, S., David, M.: Interaction design in a mobile food recommender system. In: CEUR Workshop Proceedings. CEUR-WS (2015)

25. Eslami, M., Krishna Kumaran, S.R., Sandvig, C., Karahalios, K.: Communicating algorithmic process in online behavioral advertising. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. p. 432. ACM (2018)
26. Galindo, J.A., Dupuy-Chessa, S., Mandran, N., Céret, E.: Using user emotions to trigger ui adaptation. In: 2018 12th International Conference on Research Challenges in Information Science (RCIS). pp. 1–11. IEEE (2018)
27. Gedikli, F., Jannach, D., Ge, M.: How should i explain? a comparison of different explanation types for recommender systems. International Journal of Human-Computer Studies **72**(4), 367–382 (2014)
28. 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)
29. Gretarsson, B., O'Donovan, J., Bostandjiev, S., Hall, C., Höllerer, T.: Smallworlds: visualizing social recommendations. In: Computer Graphics Forum. vol. 29, pp. 833–842. Wiley Online Library (2010)
30. Gutiérrez, F., Charleer, S., De Croon, R., Htun, N.N., Goetschalckx, G., Verbert, K.: Explaining and exploring job recommendations: a user-driven approach for interacting with knowledge-based job recommender systems. In: Proceedings of the 13th ACM Conference on Recommender Systems. pp. 60–68 (2019)
31. Hagras, H.: Toward human-understandable, explainable ai. Computer **51**(9), 28–36 (2018)
32. ter Hoeve, M., Heruer, M., Odijk, D., Schuth, A., de Rijke, M.: Do news consumers want explanations for personalized news rankings. In: FATREC Workshop on Responsible Recommendation Proceedings (2017)
33. Holliday, D., Wilson, S., Stumpf, S.: The effect of explanations on perceived control and behaviors in intelligent systems. In: CHI'13 Extended Abstracts on Human Factors in Computing Systems. pp. 181–186. ACM (2013)
34. Hosseini, M., Shahri, A., Phalp, K., Taylor, J., Ali, R.: Crowdsourcing: A taxonomy and systematic mapping study. Computer Science Review **17**, 43–69 (2015)
35. Hu, J., Zhang, Z., Liu, J., Shi, C., Yu, P.S., Wang, B.: Recexp: A semantic recommender system with explanation based on heterogeneous information network. In: Proceedings of the 10th ACM Conference on Recommender Systems. pp. 401–402. ACM (2016)
36. 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)
37. Hussein, T., Neuhaus, S.: Explanation of spreading activation based recommendations. In: Proceedings of the 1st International Workshop on Semantic Models for Adaptive Interactive Systems, SEMAIS. vol. 10, pp. 24–28. Citeseer (2010)
38. Kang, B., Tintarev, N., Höllerer, T., ODonovan, J.: What am i not seeing? an interactive approach to social content discovery in microblogs. In: International Conference on Social Informatics. pp. 279–294. Springer (2016)
39. Karga, S., Satratzemi, M.: Using explanations for recommender systems in learning design settings to enhance teachers' acceptance and perceived experience. Education and Information Technologies pp. 1–22 (2019)
40. Katarya, R., Jain, I., Hasija, H.: An interactive interface for instilling trust and providing diverse recommendations. In: 2014 International Conference on Computer and Communication Technology (ICCCT). pp. 17–22. IEEE (2014)
41. 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)

42. Knijnenburg, B.P., Kobsa, A.: Making decisions about privacy: information disclosure in context-aware recommender systems. ACM Transactions on Interactive Intelligent Systems (TiiS) **3**(3), 20 (2013)
43. Krause, J., Perer, A., Bertini, E.: A user study on the effect of aggregating explanations for interpreting machine learning models. In: ACM KDD Workshop on Interactive Data Exploration and Analytics (2018)
44. 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)
45. 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)
46. Kulesza, T., Stumpf, S., Burnett, M., Kwan, I.: Tell me more?: the effects of mental model soundness on personalizing an intelligent agent. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1–10. ACM (2012)
47. Kulesza, T., Stumpf, S., Burnett, M., Yang, S., Kwan, I., Wong, W.K.: Too much, too little, or just right? ways explanations impact end users' mental models. In: 2013 IEEE Symposium on Visual Languages and Human Centric Computing. pp. 3–10. IEEE (2013)
48. Lai, V., Tan, C.: On human predictions with explanations and predictions of machine learning models: A case study on deception detection pp. 29–38 (2019)
49. Lamche, B., Adıgüzel, 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)
50. Langley, P., Meadows, B., Sridharan, M., Choi, D.: Explainable agency for intelligent autonomous systems. In: Twenty-Ninth IAAI Conference (2017)
51. Le Bras, P., Robb, D.A., Methven, T.S., Padilla, S., Chantler, M.J.: Improving user confidence in concept maps: Exploring data driven explanations. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. p. 404. ACM (2018)
52. Leon, P.G., Cranshaw, J., Cranor, L.F., Graves, J., Hastak, M., Xu, G.: What do online behavioral advertising disclosures communicate to users?(cmu-cylab-12-008) (2012)
53. 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)
54. Li, T., Convertino, G., Tayi, R.K., Kazerooni, S.: What data should i protect?: recommender and planning support for data security analysts. In: IUI. pp. 286–297 (2019)
55. 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)
56. Loepp, B., Herrmanny, K., Ziegler, J.: Blended recommending: Integrating interactive information filtering and algorithmic recommender techniques. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. pp. 975–984. ACM (2015)
57. 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)
58. Miller, T.: Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence (2018)

59. Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G.: Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. Annals of internal medicine **151**(4), 264–269 (2009)
60. 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)
61. Naiseh, M., Jiang, N., Ma, J., Ali, R.: Personalising explainable recommendations: Literature and conceptualisation. In: WorldCist'20 - 8th World Conference on Information Systems and Technologies,. Springer (2020)
62. Narayanan, M., Chen, E., He, J., Kim, B., Gershman, S., Doshi-Velez, F.: How do humans understand explanations from machine learning systems? an evaluation of the human-interpretability of explanation (2018)
63. Nguyen, T.N., Ricci, F.: A chat-based group recommender system for tourism. In: Information and Communication Technologies in Tourism 2017, pp. 17–30. Springer (2017)
64. Nunes, I., Jannach, D.: A systematic review and taxonomy of explanations in decision support and recommender systems. User Modeling and User-Adapted Interaction **27**(3-5), 393–444 (2017)
65. Paraschakis, D.: Towards an ethical recommendation framework. In: 2017 11th International Conference on Research Challenges in Information Science (RCIS). pp. 211–220. IEEE (2017)
66. Parra, D., Brusilovsky, P., Trattner, C.: See what you want to see: visual user-driven approach for hybrid recommendation. In: Proceedings of the 19th international conference on Intelligent User Interfaces. pp. 235–240. ACM (2014)
67. Poursabzi-Sangdeh, F., Goldstein, D.G., Hofman, J.M., Vaughan, J.W., Wallach, H.: Manipulating and measuring model interpretability (2018)
68. Ramachandran, D., Fanty, M., Provine, R., Yeh, P., Jarrold, W., Ratnaparkhi, A., Douglas, B.: A tv program discovery dialog system using recommendations. In: Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue. pp. 435–437 (2015)
69. Rosenfeld, A., Richardson, A.: Explainability in human–agent systems. Autonomous Agents and Multi-Agent Systems **33**(6), 673–705 (2019)
70. Ruiz-Iniesta, A., Melgar, L., Baldominos, A., Quintana, D.: Improving childrens' experience on a mobile edtech platform through a recommender system. Mobile Information Systems **2018** (2018)
71. Samek, W., Wiegand, T., Müller, K.R.: Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models (2017)
72. Sato, M., Ahsan, B., Nagatani, K., Sonoda, T., Zhang, Q., Ohkuma, T.: Explaining recommendations using contexts. In: 23rd International Conference on Intelligent User Interfaces. pp. 659–664. ACM (2018)
73. Schäfer, H., Hors-Fraile, S., Karumur, R.P., Calero Valdez, A., Said, A., Torkamaan, H., Ulmer, T., Trattner, C.: Towards health (aware) recommender systems. In: Proceedings of the 2017 international conference on digital health. pp. 157–161. ACM (2017)
74. Schaffer, J., Giridhar, P., Jones, D., Höllerer, T., Abdelzaher, T., O'donovan, J.: Getting the message?: A study of explanation interfaces for microblog data analysis. In: Proceedings of the 20th International Conference on Intelligent User Interfaces. pp. 345–356. ACM (2015)
75. Schaffer, J., O'Donovan, J., Michaelis, J., Raglin, A., Höllerer, T.: I can do better than your ai: expertise and explanations. In: IUI. pp. 240–251 (2019)

76. Springer, A., Whittaker, S.: Progressive disclosure: empirically motivated approaches to designing effective transparency pp. 107–120 (2019)
77. 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)
78. Stumpf, S., Skrebe, S., Aymer, G., Hobson, J.: Explaining smart heating systems to discourage fiddling with optimized behavior. In: CEUR Workshop Proceedings. vol. 2068 (2018)
79. Svrcek, M., Kompan, M., Bielikova, M.: Towards understandable personalized recommendations: Hybrid explanations. Computer Science & Information Systems **16**(1) (2019)
80. Tamagnini, P., Krause, J., Dasgupta, A., Bertini, E.: Interpreting black-box classifiers using instance-level visual explanations. In: Proceedings of the 2nd Workshop on Human-In-the-Loop Data Analytics. p. 6. ACM (2017)
81. Tsai, C.H., Brusilovsky, P.: Providing control and transparency in a social recommender system for academic conferences. In: Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization. pp. 313–317. ACM (2017)
82. Tsai, C.H., Brusilovsky, P.: Explaining recommendations in an interactive hybrid social recommender. In: Proceedings of the 24th International Conference on Intelligent User Interfaces. pp. 391–396. ACM (2019)
83. Verbert, K., Parra, D., Brusilovsky, P., Duval, E.: Visualizing recommendations to support exploration, transparency and controllability. In: Proceedings of the 2013 international conference on Intelligent user interfaces. pp. 351–362. ACM (2013)
84. Wiebe, M., Geiskkovitch, D.Y., Bunt, A.: Exploring user attitudes towards different approaches to command recommendation in feature-rich software. In: Proceedings of the 21st International Conference on Intelligent User Interfaces. pp. 43–47. ACM (2016)
85. Zanker, M., Ninaus, D.: Knowledgeable explanations for recommender systems. In: 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. vol. 1, pp. 657–660. IEEE (2010)
86. Zanker, M., Schoberegger, M.: An empirical study on the persuasiveness of fact-based explanations for recommender systems. In: Joint Workshop on Interfaces and Human Decision Making in Recommender Systems. vol. 1253, pp. 33–36 (2014)
87. Zhao, G., Fu, H., Song, R., Sakai, T., Chen, Z., Xie, X., Qian, X.: Personalized reason generation for explainable song recommendation. ACM Transactions on Intelligent Systems and Technology (TIST) **10**(4), 41 (2019)