

The Impact of Explanation Characteristics on the Effects of Recommender Systems Explanations

KATHRIN WARDATZKY, University of Zurich, Switzerland

OANA INEL, University of Zurich, Switzerland

LUCA ROSSETTO, Dublin City University, Ireland

ATHINA KYRIAKOU, University of Zurich, Switzerland

ABRAHAM BERNSTEIN, University of Zurich, Switzerland

There seems to be a consensus that adding explanations to AI systems, such as recommender systems, can have positive effects, such as increasing a user's trust, the transparency of the system, or the efficiency of a decision-making process. It is to this date unclear, though, how an explanation needs to be designed to evoke these effects. To address this gap, we conducted a systematic literature survey of 180 articles. The corpus was created by first querying six literature databases with predefined keyword searches and sorting the retrieved articles by applying inclusion and exclusion criteria. We then extracted and categorized the explanation characteristics, such as their modalities and content, as well as the effects measured in user evaluations. Finally, we combine both dimensions to analyze the impact of the characteristics on the effects. The findings suggest that it is not possible to make general statements about the effects of explanations. Instead, the measured effects vary in different contexts, and the results depend on whether the evaluated explanation was compared to another explanation or to a system without explanation. Our survey identifies and highlights research gaps and promising areas for future work, such as the effects of non-local explanations on the user.

CCS Concepts: • **Information systems** → **Recommender systems**; • **Human-centered computing** → **User studies**; • **General and reference** → **Surveys and overviews**.

Additional Key Words and Phrases: Explainable AI, Recommender Systems, Explanation Characteristics, Explanation Effects

ACM Reference Format:

Kathrin Wardatzky, Oana Inel, Luca Rossetto, Athina Kyriakou, and Abraham Bernstein. 2026. The Impact of Explanation Characteristics on the Effects of Recommender Systems Explanations. *ACM Comput. Surv.* 1, 1 (January 2026), 36 pages. <https://doi.org/10.1145/3856207>

1 Introduction

Recommender systems are an integral part of the everyday life of many people. They not only influence how we inform ourselves about the latest news [76, 159], the media we consume for entertainment [25, 29], or what products we buy [199], they can also impact decisions affecting our health [59], what jobs we apply for [51], or social interactions [151, 206]. Recent years have seen an increased effort to help users understand recommendations through explanations [194]. Aside from a better understanding, adding explanations to recommender systems can have multiple effects on a user. They can increase the trust of the user, enable users to exercise control over the recommendations, or increase the

Authors' Contact Information: Kathrin Wardatzky, wardatzky@ifi.uzh.ch, University of Zurich, Zurich, Switzerland; Oana Inel, inel@ifi.uzh.ch, University of Zurich, Zurich, Switzerland; Luca Rossetto, luca.rossetto@dcu.ie, Dublin City University, Dublin, Ireland; Athina Kyriakou, kyriakou@ifi.uzh.ch, University of Zurich, Zurich, Switzerland; Abraham Bernstein, bernstein@ifi.uzh.ch, University of Zurich, Zurich, Switzerland.



This work is licensed under a Creative Commons Attribution 4.0 International License.

© 2026 Copyright held by the owner/author(s).

Manuscript submitted to ACM

Manuscript submitted to ACM

effectiveness of the recommendations [143, 181]. These effects can correlate with each other [14, 181]. For instance, a better understanding of why the user received a recommendation might lead to an increase in trust or a more effective and efficient decision-making process. We focus our investigation on recommender systems instead of general explainable AI (XAI), as, in contrast to other XAI application areas, recommender systems aim to predict a user's preference, which can vary depending on the context. Restricting the scope of the survey to recommender systems applications results in a clearer focus on the evaluated explanation types and effects.

Given the broad application areas of recommender systems and the diverse user group that many of them target, the effects that explanations have on each user might vary. There is no one-size-fits-all explanation that works best for all users and in all application areas [11]. Tran et al. [186] have shown that for movie recommendations, users often only want to see an explanation when they receive an unexpected recommendation, but in application domains where there is more at stake for the user if they do not understand the recommendations (e.g., the medical domain), the need for an explanation might be higher. The worst thing that might happen to a user when following a movie recommendation without understanding it is wasting time on a movie they do not enjoy and can turn off at any time. Following, e.g., medical recommendations without understanding them might affect the user's health, which is why they might be more interested in engaging with explanations and scrutinizing them more closely. Similarly, the optimal explanation to achieve a desired effect might differ between user types. Furthermore, it is challenging to design explanations that have a specific effect on a user. Balog et al. [14] illustrate this by instructing the crowd to generate explanations that have a specific effect on the recipient of movie recommendations, and found that these explanations often have another effect. This raises the question of how individual components of an explanation might impact the effect that the explanation has on a user. *What is the difference in the explanations for which an effect was found and those for which the same effect was not found?* We analyzed 180 articles in a systematic literature survey and investigated three dimensions related to recommender systems explanations and their effects on users to answer this overarching question. We first look at the characteristics of the evaluated explanations and describe their modalities and content. We then describe the evaluated effects and how they were measured. Ultimately, we bring these two dimensions together to analyze how the measured effects differ for explanations with varying characteristics. In particular, we answer the following research questions:

RQ 1: What kind of explanations are commonly evaluated in recommender systems in terms of content and modalities?

RQ 2: What effects were measured in the explanation evaluation?

RQ 3: What characteristics of an explanation are likely to have a specific effect on a user?

To analyze the differences between explanations, we first determined their *characteristics*. We extracted the available descriptions of the explanations and categorized this information depending on how they are displayed to the user, if they are interactive or static, what aspect of the recommendations process they focus on, and their information type. Then, we investigated the *effects* that the explanations were evaluated on. We extracted the measured effect along with the question or task with which the effect was evaluated and proposed a fine-grained taxonomy of the effects in order to capture the differences in explanations. Finally, we connect both dimensions by analyzing what explanation effects were measured for which explanation characteristics to see *which characteristic is likely to evoke an effect*. Such a study is usually done with a meta-analysis. A meta-analysis, however, requires the results to be obtained by following the same methodology, which unfortunately does not apply to the studies in our corpus. Instead, we extract the results of the explanation evaluation and analyze the differences in terms of explanation modalities and content between explanations that were able to evoke a specific effect and explanations that were not.

In summary, the contributions of this survey are threefold:

- a large-scale overview of explanation modalities and content for recommender systems,
- a fine-grained taxonomy about explanation effects that were typically measured in user studies, and
- a systematic analysis of the connection between the explanation characteristics and the effect that was measured.

Although some surveys have looked at the explanation modalities of recommendation systems [143, 215], their content [143], the effects they can have [39, 160, 181, 194], and the evaluation methodology [39, 160], the connection of these dimensions has, to the best of our knowledge, not been systematically analyzed in a literature survey.

The remainder of this paper is structured as follows: Section 2 provides an overview of the existing work related to this topic to set the scene before we present the methodology of how we collected and extracted the data for this survey paper in Section 3. Our results are covered in three sections: Section 4 describes the explanation characteristics that we identified in our corpus, Section 5 presents the effects that were measured in the user studies before we combine both in Section 6, where we analyze the impact of explanation characteristics on the measured effects. Our survey ends with a discussion of the results and limitations in Section 7 and conclusions in Section 8.

2 Related Work

A plethora of work about XAI, its evaluation, and the effects of adding explanations to recommender systems has been published. In this section, we present related work covering the evaluation of XAI approaches, explanation characteristics, and the application of the aforementioned aspects to the recommender systems domain.

2.1 Explainable AI Evaluation

The evaluation of XAI approaches is commonly seen as an open research challenge, with a lack of established standards and guidelines [215, 219]. To understand and interpret the impact of explanation characteristics on the evaluated results in the right context, it is, however, crucial to discuss the different evaluation approaches that generate these results. Zhou et al. [219] started by describing the existing explainability approaches and the properties of machine learning explanations (i.e., interpretability-based, such as clarity, parsimony, and broadness, and fidelity-based, such as completeness and soundness). Then, the authors used the taxonomy of evaluation types proposed by Doshi-Velez et al. [56], namely application-, human-, and functionality-grounded evaluations, to identify for each of them possible quantitative metrics and the properties of explainability they account for. For human-grounded evaluation, the authors conclude that there are no agreed-upon criteria in the literature regarding the use of subjective measures.

The research community has been working toward solving this issue by developing taxonomies and design guidelines. Chromik et al. [41] performed a structured literature review to develop a taxonomy for human subjects evaluations in explainable AI. The taxonomy focuses on three dimensions: the task (i.e., including the explanation goals), the study design setup, and the participants. The goal of the taxonomy is to guide researchers and practitioners when designing and conducting user evaluation of explanations. The survey, however, does not explicitly focus on recommender systems. Pu et al. [152] address the issue by defining a taxonomy of user criteria for the evaluation of recommender systems and providing a set of design guidelines for effective recommender systems. Regarding explanations, the guidelines focus on the design of explanation interfaces and explanation content with the aims of building users' trust in the system, increasing users' satisfaction, sense of control, intention to save effort, and intention to return, effects that we also survey, to a much larger extent. Mohseni et al. [129] present a categorization of XAI design goals for various user groups and evaluation methods from 42 papers. The recommender systems approaches included in the review use textual explanations. The goal of such explanations is to increase user trust and reliance. More recently, Nauta

et al. [137] reviewed 312 papers published between 2014 and 2020 in the area of XAI and identified 12 conceptual properties, called *12-Co properties*, that explanations should be quantitatively evaluated for. The review shows that around 33% of the reviewed papers evaluate exclusively with anecdotal evidence, 22% of them evaluate the proposed method in user studies, while the majority, namely 58%, use quantitative metrics, such as the 12-Co properties. On a similar note, Donoso-Guzmán et al. [55] surveyed 29 articles, extracted properties of the presented explanations, and information on how these properties were evaluated. They extended the explanation properties presented in Miller [126] and the user-centered evaluation framework of Knijnenburg et al. [96] and mapped the explanation properties to both. Neither of these surveys provide any insights regarding evaluations through user studies and how many of the identified explanation properties are applicable to recommender systems. Building on top of the taxonomies developed by Mohseni et al. [129] and Nauta et al. [137], Kim et al. [92] surveyed the human-centered evaluation approaches applied in 73 articles with a focus on decision-support systems. They investigated the in-context quality of the evaluated explanations, the contribution to human-AI interaction, and human-AI performance, and analyzed the findings when a system with an explanation was compared to a system without an explanation. They conclude that more research into the contextual variables that impact the findings of different evaluation approaches is needed.

2.2 Explanation Characteristics

One of the contextual variables that might impact the findings of different evaluation approaches refers to the characteristics of the evaluated explanations. Zhang et al. [215] provide a comprehensive overview by focusing on information sources and display styles, explainable recommendation methods, and application areas for explainable recommender systems. Furthermore, the survey also focuses on the evaluation methods for explainable recommendations, but it does not go into detail regarding the measurement of explanation effects. Nunes et al. [143] also cover the characteristics of explanations in terms of display style and information type but focus their systematic literature review on explanations for advice-giving systems, in general. Alongside the characteristics of explanations, they cover three more aspects: generation approaches for explanations, explanation goals and evaluation methodologies, and explanation types supporting explanation goals. Papadimitriou et al. [147] categorized the explanations for traditional and social recommender systems into three categories, namely based on previous items chosen by the user, based on similar users, and based on features and hybrid approaches of these. Then, by analyzing the studies conducted by Bilgic et al. [27], Symeonidis et al. [174], Vig et al. [192], they concluded that hybrid explanations are the most effective explanation styles. While there is an overlap in describing the display style of the explanations, all three reviews present different approaches to describing their content, with varying levels of detail.

2.3 Explanations in Recommender Systems

After reviewing related work on XAI in general, we now discuss relevant work related to recommender systems. One of the first reviews of explanations for recommender systems was conducted by Tintarev et al. [181] and looked at understanding what a good explanation entails. As a result, they identified seven aims of explanations for recommender systems, namely trust, transparency, satisfaction, scrutability, persuasiveness, effectiveness, and efficiency. These explanation goals are still widely used to measure the quality of explanations.

The reviews closest to our survey are Chen et al. [39], Rong et al. [160], and Chatti et al. [34]. Chen et al. [39] surveyed more than 100 papers in the area of explainable recommender systems and identify taxonomies for evaluation perspectives or effects and evaluation methods. Compared to our survey, however, they only acknowledge four explanation effects, namely effectiveness, persuasiveness, scrutability, and transparency, and summarize the types of

evaluation approaches (i.e., case studies, quantitative metrics, crowdsourcing, and online experiments) that target these particular effects. In the human-centered XAI domain, Rong et al. [160] analyzed papers focusing on user evaluation of model explanations. While they observed that many applications are in the recommender systems domain, they do not provide a clear overview of explainability models, explanation types, and experimental methodology and design per domain. Furthermore, they only provide a list of studies that report on whether they found or not found explanation effects when the explanation is compared to a “no explanation” or placebo explanation as a baseline. Lastly, Chatti et al. [34] reviewed 33 explainable recommendation tools that produce visual explanations. While their focus lies on the visual components of the tools, they also outline whether the explanations of each tool achieved a certain aim or not. They did not explore how the individual components of an explanation might impact this result, though.

The previous paragraphs show that multiple survey papers have covered explanation characteristics in recommender systems and their effects on a user. However, combining both dimensions to see what part of an explanation impacts the effect has not been thoroughly investigated. Instead, existing works point out the need for further research on the contextual variables that impact the measured effects of explanations. In a prior survey [200], we already looked at the participant sampling and their user characteristics as a potential factor that needs to be considered when interpreting the results. While the number of papers that analyzed explanation effects on varying user characteristics is small, we were able to conclude with a set of recommendations to increase the focus on these aspects. In this survey, *we investigate the connection between explanations and their effects* as a contextual factor and consider not only explanations that were compared with a recommender system without explanations, but also comparisons of two explanations.

3 Methodology

This section describes the paper’s methodology by explaining the data collection, selection, and annotation process.

3.1 Data collection and selection

We conducted a systematic literature survey of peer-reviewed articles published between 2017 and 14.07.2025 (the day of the query) to answer the research questions outlined in Section 1. We opted to start the collection with the year 2017 as this year seems to have been the starting point of a steadily increasing number of publications in the explainability field [16, 194]. To collect the papers, we queried ACM Digital Library,¹ IEEEXplore,² Taylor & Francis,³ Web of Science,⁴ Wiley Online,⁵ and SemanticScholar.⁶ The six library databases were selected to cover major computer science venues and interdisciplinary outlets. The following search term was used:

(explain* OR explanation* OR interpretab* OR intelligib* OR justification OR transparen*) AND
(recommender OR recommendation OR personalization OR personalized)

We constructed the search term to contain words frequently and interchangeably used to refer to both explainable AI and recommender systems or recommendations. The first part of the search term includes terms related to the explanation part. Given that there are currently no commonly agreed-upon definitions of AI explanations [69, 123],

¹<https://dl.acm.org/>

²<https://ieeexplore.ieee.org>

³<https://www.tandfonline.com/>

⁴<https://www.webofscience.com>

⁵<https://onlinelibrary.wiley.com/>

⁶<https://www.semanticscholar.org/>

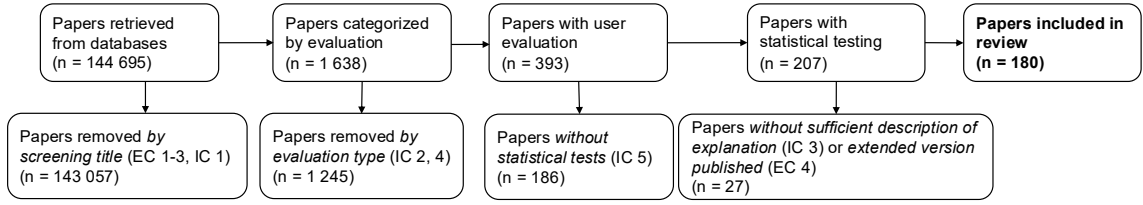


Fig. 1. Paper selection process.

we included additional terms aiming to capture potentially relevant articles using a different terminology, along with variants of *explainability* and *interpretability*. Some articles distinguish between *justifications* and *explanations* while others use the term *explanation* for a *justification*. Therefore, we opted to include *justifications* in the search. The variations of *transparency* aim to include those articles that opt for a different wording.

This initial search returned a total of 144 695 articles. Figure 1 depicts the paper selection process that was conducted by two authors. In the first round, we manually filtered out all papers that, based on their titles and abstracts, were clearly not related to explainable AI or recommender systems. Second, we further filtered using the exclusion and inclusion criteria (EC and IC) summarized in Table 1. Hereby, we excluded all publications that were not written in English and for which we could not access the full text through the licenses of our institution (EC-1 and EC-2). Furthermore, we excluded papers with results published in another paper matching our inclusion and exclusion criteria (EC-3). Duplicated papers that appeared in multiple databases were only considered once (EC-4).

The remaining papers had to fulfill all of the following inclusion criteria to be selected. We included papers that propose an explainable or interpretable recommender system or an explanation generation method (IC-1), and the explanations are evaluated in a recommender systems context (IC-2). To ensure that the explanations are the focus of the papers, we only included publications that provide examples or detailed descriptions of the explanations (IC-3). This left 1 638 articles on explainable or interpretable recommender systems. The remaining inclusion criteria refer to the evaluation method. As we analyze the impact of explanation effects on users, we only included papers that evaluated the explanations in a user study (IC-4) and the analysis had statistical testing (IC-5).

Table 1. Exclusion and inclusion criteria for the paper selection process.

Exclusion Criteria (EC)	
EC-1	The paper was already retrieved from another database
EC-2	The paper is not written in English
EC-3	We have no access to the full paper
EC-4	The content of the paper was published in an extended paper that matches the inclusion criteria
Inclusion Criteria (IC)	
IC-1	The paper proposes an explainable or interpretable recommender system or an explanation generation method
IC-2	The paper presents explanations that are evaluated in a recommendation scenario
IC-3	The paper provides an example or detailed description of the explanation
IC-4	The explanations are evaluated on users
IC-5	The results were tested for statistical significance

The selection process narrowed the corpus to 180 papers containing a total of 226 user studies (some papers reported on multiple studies), from which we extracted the information reported on the user study results. Next (Section 3.2), we elaborate on the annotation process of the papers in our corpus.

3.2 Data annotation

We extracted general information about the explanation that was evaluated, as well as the evaluation methodology. The data annotation process was conducted by three authors. We randomly selected five articles, and each of the three authors annotated them independently. Then, we discussed the differences in annotations. From these discussions, we derived annotation guidelines that were revised as more edge cases appeared. Whenever an update happened, the already annotated articles were revised, and the data was updated. The corpus spans articles published in 79 unique venues, the most common ones being ACM International Conference on User Modeling, Adaptation and Personalization (UMAP, 23 articles), ACM Recommender Systems Conference (RecSys, 14), ACM Conference on Intelligent User Interfaces (IUI, 13), ACM CHI conference on Human Factors in Computing Systems (CHI, 10), and User Modeling and User-Adapted Interaction (UMUAI, 9). The articles in our corpus cover 21 different application domains, with movie recommendations as the most popular (35 articles), followed by education (30), point-of-interest (20), health-related (20), and e-commerce recommendations (20) (for a detailed overview, see the supplemental material⁷).

To analyze the first research question (explanation characteristics), we focused on the description of the explanations that were evaluated in the 226 user studies of our corpus. We followed the taxonomy of Nunes et al. [143] and categorized the content and appearance of the extracted explanations based on the given descriptions and examples.

To answer the second research question (explanation effects), we annotated the findings in the results sections of each paper. More precisely, we extracted the dependent and independent variables of each finding, which evaluated an explanation along with the outcome of the evaluation (i.e., whether or not an effect was found). The dependent variables or measured explanation effects were then categorized in the next step. We noticed that the effects did not consistently follow the same definitions and were frequently named differently. Therefore, we referred to the evaluation task or question used for the evaluation, wherever possible (i.e., when they were made available by the authors of the papers). We categorized the explanation effects by the seven explanation goals defined by Tintarev et al. [181]: effectiveness, efficiency, transparency, persuasiveness, trust, satisfaction, and scrutability. By looking at the task design of the user studies, we noticed that the papers stating to evaluate the scrutability of the system were doing so by assessing the perceived system control. We, therefore, added it to an *interface quality* category along with other related aspects, such as the perceived information quality.

The final research question combined the annotations described in the previous paragraphs.

4 Explanation Characteristics

In this section, we provide an overview of the types of explanations typically generated and provided for recommender systems and their characteristics. We extracted 665 explanations from the articles in our corpus. This means that, on average, 2.9 explanations were evaluated in each of the 226 user studies in our corpus. The explanations of our corpus cover a total of 21 application domains, with the majority explaining recommendations from the movie and point-of-interest (POI) domains (144 and 122 explanations, respectively). For every explanation, we extracted and categorized the information provided about their modality and content, which we describe in the following sub-sections.

⁷<https://doi.org/10.5281/zenodo.17788874>

Table 2. Overview of the explanation modalities and whether they are interactive (✓) or not (×).

Modality	Interactivity	Articles	# Expl.
audio	×	[38, 47, 150, 203, 216]	25
hybrid	×	[5, 10, 23, 24, 31, 32, 44, 67, 68, 73, 80, 85, 88, 122, 145, 146, 168, 180, 187, 188, 203, 204, 221]	43
	✓	[3, 10, 18, 33, 49, 52, 61, 66, 68, 82, 85, 111, 116, 121, 140, 141, 156, 157, 163, 168, 186–188]	44
tabular	×	[28, 187, 188]	3
	✓	[81, 155]	2
textual	×	[2, 7, 9, 10, 13–15, 20–24, 30, 32, 35–38, 40, 53, 54, 60–64, 70–72, 79, 84, 86–90, 93–95, 97–99, 101–103, 108–110, 112–115, 118, 119, 122, 128, 130–133, 136, 138, 142, 144, 145, 149, 153, 161, 162, 164–167, 169, 170, 172, 173, 175–177, 179, 180, 185, 188, 191, 195, 197, 198, 202, 204, 207–209, 212, 213, 216, 218]	367
	✓	[6, 12, 17–19, 49, 53, 61, 65, 68, 77, 78, 81, 82, 100, 102, 104, 105, 135, 138, 140, 148, 154, 155, 158, 165, 172, 178, 186, 190, 193, 198, 205, 214, 217, 220]	63
visual	×	[8, 26, 31, 43, 45, 53, 62, 63, 80, 87, 88, 91, 98, 99, 107, 120, 125, 139, 153, 180, 182, 188, 202, 204, 210, 211]	72
	✓	[1, 4, 18, 28, 33, 52, 57, 81, 91, 99, 102, 107, 111, 121, 124, 128, 134, 139, 171, 188, 189, 214]	30

4.1 Explanation modalities and interactivity

We were able to extract information about explanation modalities and whether they provide options for the user to interact with the explanations for 658 out of the 665 explanations.

Modalities. First, we look at the aspects related to how the explanations are displayed to the user. We categorized the explanations into five modalities: text, visual, tabular, audio, or a hybrid of those. We also examine whether and how the user can interact with the explanations. Table 2 provides an overview of the modalities and interactivity of the explanations in our corpus. We observe that static textual explanations are predominant in our corpus, making up 55% (367 out of 665 explanations) of all explanations. The second most common explanation type is the static visual explanation, with 72 occurrences. Popular visual types are bar charts or histograms showing the importance of individual features to a recommendation or depicting the overlap between the item features and the user preferences for these features. Other visual explanations that show the overlap of the user history with the recommendations include Venn diagrams (e.g., in Kouki et al. [99]) or contour plots (seen in Liang et al. [107]). Most hybrid explanations (64 out of 87) are combinations of text and visual elements. The text either provides additional information to the visual element(s) (as in Tran et al. [186]) or explains them (e.g., Coba et al. [44]).

Interactivity. Explanations with interactive elements are less common across all modalities than static explanations. Measured against the total number of explanations in each modality, hybrid explanations are the most interactive, while text-based explanations are the most static. 60 of the 143 interactive explanations in our corpus are interactions to show/hide explanations (e.g., [6, 53, 116, 121, 134, 186]) or provide more detail in a drill-down approach (e.g., [33, 68, 81]). Some drill-down approaches [33, 68] guide the user from a local explanation of an individual recommendation to a global explanation of how the entire recommender system works. Other interaction types include providing feedback on the explanations [66, 154, 205] or adapting the weights of input features to see how the recommendations change [91, 107, 111, 214, 217]. In conversational recommendation settings, users can usually freely respond to the received explanations and use this possibility to interact with the explanations [3, 141, 190, 193].

Table 3. Overview of the articles covering the different content dimensions and the number of explanations. We omit explanations for which we could not determine the type and focus.

	Output	#	Input	#	Process	#	Multiple	#
Local	[2, 3, 6–8, 12–15, 17–21, 23, 24, 26, 30–32, 35, 40, 44–48, 52–54, 57, 60–67, 70, 72, 73, 78–82, 84–87, 89–91, 93–95, 97–99, 101, 103–107, 109, 112–115, 118–121, 124, 125, 128, 130–136, 138, 140–142, 145, 146, 148, 150, 153–158, 161–163, 165–169, 172, 175, 176, 178–180, 185–191, 193, 197, 198, 202–205, 207–213, 216, 218, 220, 221]	486	[43, 67, 68, 88, 139, 171, 173]	25	[86, 88]	2	[12, 33, 49, 87, 121, 203]	16
Global	[67, 116, 154, 164]	5	[1, 67, 77, 91, 111, 166, 182, 214]	11	[68, 100, 195, 217]	5	[33, 195, 203]	4
Pairwise	[20, 21, 144, 150, 204, 205]	15					[144]	4
Top-n	[4–6, 9, 10, 38, 71, 74, 102, 107, 109, 131, 138, 140, 170, 173, 176, 195, 202, 214]	55						
Multiple	[22, 36, 37, 108, 131, 149, 154, 177]	15					[110, 111, 122]	12

4.2 Explanation content

The second aspect related to the explanations is their content. We first look at three different aspects of the content: the part of the recommender system the explanation focuses on, the type of explanation, and the information type. Table 3 provides an overview of the distribution of types and foci of the evaluated explanations in our corpus. We classified the information that the explanations provide based on the taxonomy developed by Nunes et al. [143].

Explanation focus. Explanations can focus on three aspects of the recommender system: input, output, or the recommendation process. Explanations focusing on input describe the data or user model that was used to generate them. Explanations focusing on the process describe the workings of the recommender system. As suggested, output explanations focus on the recommendations themselves. We identified the explanation focus for 660 of the 665 explanations. In our corpus, explanations focusing on the output are the most frequent, with 580 explanations, followed by the input with 36 explanations, and the process with 8 explanations. The remaining 36 explanations combine multiple foci.

Explanation type. The explanation type is connected to the explanation focus and describes a different notion of what is being explained. Typically, the literature on explainable AI distinguishes between local and global explanations [69]. While local explanations focus on the system’s output and explain one individual recommendation, global explanations focus on the input or the process and explain the user model or the system’s workings, respectively. In short, a local explanation explains why a user received a specific recommendation, and a global explanation answers the question of how the recommendations were generated. We were able to identify the explanation type of 655 of the 665 explanations. Local explanations dominate our corpus, with 529 explanations, while global explanations occur 25 times. Explanations that focus on the output can be further classified into pairwise explanations that compare two items with each other and highlight aspects in which each item satisfies different needs (19 occurrences), a ranked list of top-k items ranked by the computed likelihood of the user being interested in the items (55 times), or a combination of multiple types. Table 3 shows the distribution of explanation types across the different explanation foci for the explanation for which we were able to identify both characteristics. The dominance of local explanations that focus on the output is obvious.

Information type. For the classification of the information type of the explanations, we follow the taxonomy of Nunes et al. [143] with 17 labels grouped into four categories about the information that the explanations in recommender and decision support systems convey. Given that we exclusively analyze recommender systems explanations, we

adapted the wording of the labels and definitions to match the commonly used vocabulary in the recommender systems community. We refer to the *Explanation Characteristics* sheet in the supplemental material for the exact categorization of the explanations and to the original paper for the definition of the information types.

The most common information type in our corpus is *preference match* with 125 explanations, followed by *decisive features* and *knowledge about the community* with 111 and 93 explanations, respectively. The least common information types are *self-reflective statistics*, *relationship between knowledge objects*, and *decision method side-outcomes*, each occurring in less than 1% of all explanations. Table 4 shows the distribution of the information types for the three most common modalities and all explanations. For all explanations, the most common information types are *preference match* with 16.76%, *decisive features* with 14.75%, and *knowledge about the community* with 12.47% of the explanations. Looking at the individual modalities, we see different focal points of information. A total of 42.86% of all visual explanations contain either *knowledge about the community* or *decisive features*, which is also the most common information type for hybrid explanations with 18.97%. The textual explanations are generally more distributed across the information types, the most common being *preference match* with 15.87%, *background data* with 12.73%, and *decisive features* with 12.53%.

Table 4. Distribution of the information types across each modality as a percentage of the total number of explanations for this modality (e.g., 18.10% of all hybrid explanations contain the information type “preference match”). The most common information types for each explanation modality are highlighted in **bold** with the follow-up being *italicized*.

Information Type		% of all hybrid	% of all textual	% of all visual	% of all modalities
Background & complementary information	Background data	4.31	<i>12.73</i>	0.84	9.52
	Knowledge about peers	0.86	10.44	7.56	8.04
	Knowledge about similar items	1.72	6.89	3.36	5.23
	Knowledge about the community	14.66	9.39	21.85	12.47
	Relationship between knowledge objects	1.72	0.42	1.68	0.80
Decision inference process	Decision method side-outcomes	0.86	0.84	0.84	0.94
	Inference and domain knowledge	12.93	9.19	5.04	8.98
	Inference trace	6.90	5.22	7.56	5.76
	Self-reflective statistics	2.59	0.00	0.00	0.40
Items & their features	Decisive features	18.97	12.53	<i>21.01</i>	14.75
	Feature-based domination	0.86	4.59	2.52	3.62
	Pros and Cons	6.90	4.38	1.68	4.96
User preferences & input	Decisive input values	4.31	5.43	3.36	4.83
	Feature importance analysis	2.59	0.63	3.36	1.34
	Preference match	<i>18.10</i>	15.87	16.81	16.76
	Suitability estimate	1.72	1.46	2.52	1.61

4.3 Section summary

In this section, we described the modalities and content of the explanations of our corpus. The majority of explanations are static text-based explanations that explain one individual recommendation. Global explanations covering the general recommendation process are rare. Commonly, the explanations of our corpus provide information about how the recommendation matches the preferences and constraints of the user, the item features that are key for the recommendation, or the preferences of other users that support the recommendation.

5 Explanation Effects

After investigating the explanation characteristics, the next step is to look at the effects that were evaluated by the articles in our corpus. This allows us to investigate the overarching question of *what makes an explanation likely to evoke a specific effect*. In this section, we present a taxonomy of explanation effects that allows for a more precise understanding of the evaluated explanation capabilities. To develop the taxonomy, we extracted the measured effects from the findings reported in the articles. We consulted the task description or the questions with which the effects were measured to gain a more fine-grained notion of them. If task descriptions or questions were not stated in the articles, we extracted as much detail as provided from the context given in the article. The extracted 500 fine-grained effects were then clustered in multiple iterations until all effects were assigned to a category.

We divided the effects into effects on the user, effects on the system perception, and effects that impact both the user and the system perception. In each of these categories, we describe the different facets each effect entails, present the number of times it was measured, found, or not found, and report on evaluated correlations and interaction effects. Table 5 provides an overview of the effects that we present in the following subsections and their descriptions, the number of times each effect was found or not found, as well as the associated references to the articles.

5.1 Effects on the User

The first set of effects in our taxonomy describes the impact that explanations can have on the user. It consists of trust, persuasiveness, effectiveness, and satisfaction. Following, we take a closer look at how often each of these effects was found (✓) or not found (×).

5.1.1 Trust.

Notions and measures of trust. Defining trust and how to measure it is a complex topic [145]. Definitions of trust in the literature vary between context and application area [117]. Furthermore, trust is not static but evolves and changes over time [83, 183]. Tintarev et al. [181] define trust in the context of recommender systems explanations as a means to “increase users confidence in the system”. This definition aligns with the cognitive side of the trust taxonomy developed by Madsen et al. [117]. In their taxonomy, the authors describe trust as a dual concept that consists of cognition- and affect-based components. More precisely, cognition trust is based on the user’s perceptions of the system’s characteristics. The result of cognition trust is high confidence of the user in the system, while affect trust components are based on the user’s emotional responses to the system, such as the personal attachment [117].

Another notion of trust that we observed is behavioral trust, where the participants show trust through their behavior. Overall, we identified 448 findings for trust, which we sorted into these four categories.

Affective Trust The affective trust was exclusively evaluated as the perceived benevolence of the system. We found evaluations by nine articles [23, 24, 35, 72, 101, 130, 141, 145, 172] which did not find an effect in the majority of the conditions (found ✓: 13, not found ×: 35).

Behavioral Trust Behavioral trust was mainly evaluated by asking the participants if they intend to return to the system and use it in the future [1, 22–24, 26, 46, 48, 52, 72, 85, 104, 106, 108, 116, 125, 138, 145, 146, 148, 149, 153, 172, 205]. The effect was evaluated in 92 conditions (✓: 42, ×: 51).

Cognitive Trust Cognitive trust was evaluated in 270 conditions, which makes it the trust category with the highest number of evaluations. We identified four notions of cognitive trust. Nine articles evaluated how the participants perceive the competence of the system [1, 23, 24, 35, 101, 130, 145, 172, 214] (✓: 12, ×: 33) and the perceived

Table 5. Overview of the extracted effects and their descriptions. The number and references that follow a ✓ indicate how often and by which papers an effect was found, × indicates how often and by which papers the effect was evaluated but not found. Note that some papers ran multiple evaluations. Hence, the number of findings and articles are not the same.

Effect	Description	Findings
Effects on Users	Affective	Trust based on a user's emotional responses to the system, e.g., personal attachment [117]. ✓(13): [23, 35, 72, 145, 172] × (35): [23, 24, 35, 72, 101, 130, 141, 145]
	Behavioral	Trust shown by users' behavior, e.g., their intention to return and use the application. ✓(42): [1, 24, 26, 46, 48, 52, 72, 85, 106, 116, 125, 138, 149, 153] × (51): [22–24, 32, 46, 48, 52, 72, 104, 108, 116, 145, 146, 148, 153, 172, 205]
	Cognitive	Measures trust by the increased user confidence in the system [181]. ✓(129): [1, 5, 7, 15, 20, 21, 23, 24, 30, 31, 35, 36, 38, 49, 52, 57, 71, 72, 81, 94, 95, 101, 104, 110, 116, 122, 128, 131, 133, 145, 149, 155, 164, 189, 190, 198, 210, 211] × (141): [10, 15, 20, 22–24, 31, 35, 38, 48, 52, 53, 71, 72, 80, 81, 85, 94, 97, 101, 116, 122, 130, 131, 133, 141, 145, 146, 163, 172, 198, 204, 214, 221]
	Overreliance	Measures if the users rely on the agent even for bad recs. ✓(6): [31] × (3): [31]
	Trusting Beliefs	Combined trust ratings of multiple notions, often sys. competence, benevolence, and integrity. ✓(15): [33, 48, 67, 68, 145, 158] × (13): [48, 67, 145, 158, 205]
	Explanation	Persuasiveness of the expl. ✓(105): [20, 21, 35, 60, 72, 81, 93, 99, 112, 131, 133, 161, 162, 167, 175, 191] × (90): [20, 35, 60, 67, 80, 97, 99, 131, 133, 145, 161, 162, 167, 175, 191]
	Recommendation	Persuasiveness of the rec. ✓(1): [204]
	Likelihood to Follow Rec.	The impact of the expl. on the likelihood that a user will follow the rec. ✓(70): [2, 3, 15, 17, 36, 38, 40, 58, 63, 70, 97, 114, 122, 141, 150, 153, 164, 198, 212, 216, 221] × (53): [2, 3, 15, 18, 38, 122, 141, 148, 149, 153, 163, 216, 221]
	Recommender	How effective the recs or the system are perceived by the users. ✓(27): [28, 35, 46, 67, 71, 81, 130, 193, 197, 213] × (33): [35, 46, 67, 71, 79–81, 189]
	Explanation	How effective the expl. are in assisting the user to make better decisions. ✓(98): [8, 20, 21, 36, 49, 68, 74, 77, 114, 128, 132, 144, 150, 156, 167, 177, 197, 210, 216] × (33): [8, 18, 20, 21, 97, 144, 156, 157, 167, 216]
Effects on System Perception	Decision	The users' (perceived) decision quality. ✓(77): [3, 26, 28, 30–32, 70, 86, 87, 90, 105, 132, 141, 144, 146, 148, 170, 179, 180, 188, 190] × (69): [3, 26, 30, 31, 70, 86, 87, 90, 118, 132, 139, 141, 144, 146, 148, 163, 169, 180, 188, 204]
	Recommender	Impact of expl. on the satisfaction with the rec. system and its output. ✓(102): [7, 8, 19, 21, 36, 40, 47, 49, 52, 53, 61, 90, 94, 95, 110, 114, 116, 130, 131, 138, 148, 149, 153, 165, 179, 184–187, 190, 197, 198, 210, 216, 221] × (106): [8, 18, 52, 53, 61, 80, 94, 100, 110, 116, 120, 131, 148, 153, 165, 185–187, 189, 197, 198, 216]
	Decision	Expl. impact on satisfaction with their decision. × (6): [118, 140, 170]
	Efficiency	Whether the expl. help the users to make faster decisions [181]. ✓(66): [28, 35, 36, 67, 68, 85, 113, 114, 153, 185, 193, 198, 204, 207, 210] × (60): [32, 35, 67, 79, 80, 97, 114, 153, 163, 185, 204, 207]
	Control	Impact of expl. on the (perceived) control over the recs. ✓(23): [5, 10, 26, 32, 35, 62, 68, 90, 97, 104, 110, 116, 188] × (44): [5, 10, 35, 37, 67, 97, 102, 107, 110, 116, 148, 188]
	Info Quality	Impact of expl. on the perceived quality of information provided in the application. ✓(67): [7, 26, 32, 36, 58, 61, 72, 80, 85, 103, 106, 107, 112, 116, 125, 131, 133, 138, 139, 148, 172, 213, 221] × (26): [5, 26, 61, 72, 80, 116, 131, 133, 139, 148, 172, 213]
	Rec. Quality	The impact of expl. on the perceived rec quality. ✓(115): [1, 4, 5, 9, 10, 15, 18, 36, 40, 45, 49, 52, 63, 66, 71, 87, 91, 94, 102, 104, 109, 116, 124, 138, 140, 142, 148, 166, 168, 171, 190, 203, 205, 207, 209, 211, 213] × (190): [5, 9, 10, 15, 19, 45, 52, 71, 87, 94, 100–102, 107, 108, 116, 125, 140–142, 148, 154, 166, 171, 189, 190, 203, 207, 213]
	Other	Expl. impact on the sys. complexity, ease of use, effort, sys. acceptance, and sys. usefulness. ✓(72): [5, 6, 10, 28, 31, 36, 57, 61, 94, 102, 106, 116, 118, 138, 149, 153, 158, 190, 196, 198, 210, 221] × (81): [1, 10, 28, 31, 32, 61, 94, 102, 104, 105, 116, 118, 138, 140, 146, 148, 153, 169, 170, 188, 190, 198, 221]
	Recommendations	Perceived transparency of the recs. ✓(16): [4, 36, 97, 111, 145, 148, 172, 190, 218] × (12): [9, 80, 97, 111, 131, 145, 148, 155]
	System	Perceived transparency of the rec process or the overall sys. ✓(17): [6, 21, 24, 79, 81, 104, 116, 188] × (14): [23, 24, 79, 81, 116, 188]
Effects on Users & System Perception	Understanding	(Perceived) understanding of the users about why they are getting the recs or how they were generated. ✓(136): [5, 7, 8, 10, 12, 20, 30, 35, 38, 46, 48, 49, 52, 53, 64, 67, 68, 71, 85, 90, 93, 94, 105, 112, 114, 122, 131, 133, 139, 150, 155, 164, 166, 172, 179, 182, 189, 190, 197, 198, 208, 211] × (156): [1, 8–10, 12, 19, 20, 30, 35, 46, 48, 64, 67, 71, 80, 94, 102, 107, 115, 122, 131, 139, 150, 154, 155, 163, 166, 198, 202, 204, 205, 208]
	Engagement	Impact of expl. on the engagement of the users with the system. ✓(92): [1, 10, 18, 26, 37, 53, 57, 63, 81, 91, 95, 107, 119, 124, 125, 134, 158, 176, 188, 207] × (41): [1, 10, 26, 107, 119, 134, 188, 207]
	Information Retention	Impact of expl. on the information retention of the user ✓(4): [72] × (2): [72]
	Rec. Specific	Specific measurements to the rec scenario. ✓(53): [2, 148, 153, 185] × (78): [141, 148, 153, 185]

integrity of the system [23, 24, 35, 71, 101, 130, 141, 145, 214] ($\checkmark$: 9, $\times$: 38). Furthermore, we found evaluations of the participants' trust in the received recommendations [10, 15, 31, 52, 53, 57, 97, 149, 189, 198] ($\checkmark$: 11, $\times$: 20) and their trust in the overall system revised [5, 7, 20, 21, 23, 24, 30, 36, 48, 49, 72, 80, 81, 85, 94, 95, 104, 110, 116, 128, 131, 133, 141, 145, 146, 155, 163, 164, 190, 204, 210, 221] ($\checkmark$: 79, $\times$: 46). Three articles evaluated a combination of trust in the recommendation and the system [38, 122, 211] ($\checkmark$: 18, $\times$: 4).

Overreliance A user's overreliance on the recommendation agent was measured 9 times by one article Bućinca et al. [31] ($\checkmark$: 6, $\times$: 3).

In 28 instances, the authors combined trust ratings from multiple notions, commonly consisting of perceived system competence, benevolence, and integrity, to a unified trusting belief score [33, 48, 67, 68, 145, 158, 205] ($\checkmark$: 15, $\times$: 13).

Correlations and interaction effects. We identified 62 interaction and mediating effects and ten correlations of trust with other variables. Chatti et al. [33], Guesmi et al. [68], Göbel et al. [72], and Hernandez-Bocanegra and Ziegler [82] evaluated interaction effects between trust and user characteristics of their participants (need for cognition, visualization familiarity, personal innovativeness, technical expertise, propensity to trust others, and domain knowledge), but found no interactions. However, Zhang et al. [214] found that trust is mediated by the perceived personalization, while Mauro et al. [122] and Chen et al. [38] found an interaction between trust and the perceived interpretability of the system. Dominguez et al. [53] and Guo et al. [71] found that the perceived understandability of the explanations is mediated by trust, which aligns with the positive correlation between these two variables found by Shin et al. [164]. Nelekar et al. [141] found positive correlations of the perceived system competence, benevolence, integrity, and overall trust with the participant's propensity to trust others. Hernandez-Bocanegra et al. [80] found correlations between the participants' decision-making style and social awareness, and their explicit trust in the system and recommendations, and confirmed the correlation between social awareness and trust in a different study [79].

5.1.2 Persuasiveness.

Notions and measures of persuasiveness. Tintarev et al. [181] define persuasiveness as “convince users to try or buy” the recommended items. We found that the evaluation of persuasiveness was either measured through the likelihood that participants follow the recommendations, or they evaluate the (perceived) persuasiveness of the explanations or recommendations. Overall, we identified 319 evaluations of persuasiveness and sorted these into three categories. Two categories specify what element of the system was evaluated (explanation and recommendation persuasiveness), while the last one (likelihood to follow a recommendation) measures persuasiveness through the (intended) user action after seeing the explanation.

Explanation Persuasiveness The (perceived) persuasiveness of the explanations was evaluated in 195 instances [20, 21, 35, 60, 67, 72, 80, 81, 93, 97, 99, 112, 131, 133, 145, 161, 162, 167, 175, 191] ($\checkmark$: 105, $\times$: 90).

Recommendation Persuasiveness The effect explanations have on the persuasiveness of the recommendations was only evaluated once by Wibowo et al. [204] ($\checkmark$: 1).

Likelihood to Follow Recommendation The impact of explanations on the likelihood of the users to follow recommendations or adapt their behavior according to the recommendations was evaluated in 123 instances [2, 3, 15, 17, 18, 36, 38, 40, 58, 63, 70, 97, 114, 122, 141, 148–150, 153, 163, 164, 198, 212, 216, 221] ($\checkmark$: 70, $\times$: 53).

Correlations and mediating effects. Neither of the six articles [33, 68, 72, 122, 133, 167] investigating the interaction or mediation of different user characteristics with persuasiveness found an interaction between persuasiveness and

need for cognition, visualization familiarity, personal innovativeness, technical expertise, propensity to trust others, domain knowledge, or technical expertise. However, Mauro et al. [122] found an interaction between persuasiveness and transparency, Göbel et al. [72] between persuasiveness and information uncertainty and the time spent interacting with the system, and Silva et al. [167] an interaction and correlation of persuasiveness with the participants' familiarity with the items. Kouki et al. [99] found positive correlations between persuasiveness and the personality type of their participants (calm and anxious).

5.1.3 Effectiveness.

Notions and measures of effectiveness. Tintarev et al. [181] define the effectiveness of an explanation as support to “help users make good decisions”. Overall, we identified 337 measurements of effectiveness, which we sorted into three different groups. Every group measures the effectiveness of a different element, namely the recommender, the explanations, and the decision-making process.

Recommender Effectiveness The perceived effectiveness of the recommender was evaluated in 60 instances [28, 67, 79–81, 189, 193, 197, 213] (✓: 27, ×: 33).

Explanation Effectiveness The (perceived) impact of explanations on the effectiveness of the explanations was evaluated in 131 instances [8, 18, 20, 21, 36, 49, 68, 74, 77, 97, 114, 128, 132, 144, 150, 156, 157, 167, 177, 197, 210, 216] (✓: 98, ×: 33).

Decision-Making Effectiveness The impact of explanations on the effectiveness of the users' decision-making process was evaluated 146 times [3, 26, 28, 30–32, 70, 86, 87, 90, 105, 118, 132, 139, 141, 144, 146, 148, 163, 169, 170, 179, 180, 188, 190, 204] (✓: 77, ×: 69).

Correlations and mediating effects. Six articles investigate correlations or mediating effects for effectiveness. Chatti et al. [33], Guesmi et al. [68], and Hernandez-Bocanegra and Ziegler [82] investigate mediating effects of effectiveness with user characteristics (need for cognition, visualization familiarity, technical expertise, personal innovativeness, propensity to trust others, and domain knowledge). While Chatti et al. [33] found an interaction between effectiveness and participants' personal innovativeness, no interactions were found for the other ones. Silva et al. [167] further found an interaction with the participants' familiarity with the items, Lee et al. [104] with the framing of the recommendation source, Hernandez-Bocanegra and Ziegler [82] with trusting beliefs, and the perceived explanation quality. Takami et al. [177] found an interaction of effectiveness with the participants' engagement with the system. Hernandez-Bocanegra et al. [80] found correlations between the participants' perceived effectiveness and their decision-making style and social awareness, while Silva et al. [167] found positive correlations with satisfaction and persuasiveness.

5.1.4 Satisfaction.

Notions and measures of satisfaction. Tintarev et al. [181] define satisfaction as “increase the ease of usability or enjoyment.” Overall, satisfaction was evaluated in 214 instances, which we sorted into satisfaction with the recommendation or system and satisfaction with the users' decision.

Recommender Satisfaction The impact of explanations on the users' satisfaction with the recommendation or system was evaluated 208 times [7, 8, 18, 19, 21, 36, 40, 47, 49, 52, 53, 61, 80, 90, 94, 95, 100, 110, 114, 116, 120, 130, 131, 138, 148, 149, 153, 165, 179, 184–187, 189, 190, 197, 198, 210, 216, 221] (✓102, ×: 106).

Decision Satisfaction The impact of explanations on the users' satisfaction with their decision was evaluated in 6 instances [118, 140, 170] (×: 6).

Correlations and mediating effects. Chatti et al. [33] and Guesmi et al. [68] analyzed potential interactions of their participants' satisfaction and did not find interactions with visualization familiarity, personal innovativeness, technical expertise, the propensity to trust others, or domain knowledge. In addition to the interactions with user characteristics, Dominguez et al. [53] found that participants' satisfaction is mediated by the understandability of the system and the time spent in the application. Regarding correlations, Tran et al. [185] found positive correlations between the satisfaction level of participants and their perceived fairness and consensus within a group. Starke [168] found a positive correlation between satisfaction with the system and the number of recommendations that a participant followed. Silva et al. [167] found a correlation between satisfaction and the participants' familiarity with the items, and Ahmad et al. [7] found correlations of satisfaction with trust, transparency, and effectiveness.

5.2 Effects on the system perception

The effects of explanations can not only impact the user but also the system that the user is interacting with. We identified two groups of explanation effects in this category: the impact of explanations on a system's efficiency and their impact on the application's interface quality. In the following, we present the findings for these effects.

5.2.1 Efficiency.

Notions and measures of efficiency. Tintarev et al. [181] define efficiency as “help users make decisions faster”. Thus, the explanation improves the speed with which a user reaches their goal. The efficiency measurements in our corpus were generally not further specified. Out of the 126 instances [28, 32, 35, 36, 67, 68, 79, 80, 85, 97, 113, 114, 153, 163, 185, 193, 198, 204, 207, 210] evaluating efficiency (✓: 66, ×: 60), one instance evaluated the impact of explanations on the perceived recommendation efficiency [80] (×: 1) and the perceived efficiency of the explanation was evaluated in four instances [67, 68] (✓: 2, ×: 2). For better readability, we analyze all efficiency measurements as one category.

Correlations and mediating effects. Chatti et al. [33] and Guesmi et al. [68] both investigated possible mediating effects between efficiency and user characteristics. Guesmi et al. [68] could not find any effects between efficiency and need for cognition, personal innovativeness, propensity to trust others, domain knowledge, and technical expertise. While Chatti et al. [33] confirmed these findings for the need for cognition, they did find an interaction effect between efficiency and the technical expertise and personal innovativeness of their participants. Additionally, they investigated potential interactions of efficiency with the participants' visualization familiarity but did not find an effect. Hernandez-Bocanegra et al. [80] analyzed possible correlations between efficiency and the participants' decision-making style and social awareness. While they found a correlation between efficiency and a rational decision-making style, they did not find an effect on intuitive decision-making or social awareness.

5.2.2 Interface quality.

Notions and measures of interface quality. The impact of explanations on the (perceived) quality of the interface entails three categories: the degree to which explanations enable a user to exert control over the recommendations, the quality of the provided information, and the quality of the recommendations. We further identified a set of other measures related to how explanations impact the interactions with the system and its usability. Overall, we identified 619 measurements, which makes this the largest category.

Control The 67 measurements of (perceived) control [5, 10, 26, 32, 35, 37, 62, 67, 68, 90, 97, 102, 104, 107, 110, 116, 148, 188] ($\checkmark$: 23, $\times$: 44) were often equated with the scrutability goal of Tintarev et al. [181] which is defined as “allow users to tell the system it is wrong”.

Information Quality The impact of explanations on the perceived system information quality was measured 93 times ($\checkmark$: 67, $\times$: 26). The quality of the information was either measured as the overall information quality [26, 106, 139] ($\checkmark$: 15, $\times$: 3) or further specified to novelty [112, 131, 133, 213] ($\checkmark$: 12, $\times$: 14), sufficiency [5, 7, 32, 58, 61, 85, 116, 125, 138, 139, 148] ($\checkmark$: 18, $\times$: 6), perceived credibility [72] ($\checkmark$: 2, $\times$: 1), perceived information uncertainty [72] ($\checkmark$: 3) or perceived information usefulness [36, 80, 85, 103, 106, 107, 172, 221] ($\checkmark$: 20, $\times$: 2).

Recommendation Quality The impact of explanations on the perceived recommendation quality was measured 306 times in total, out of which it was found in 115 instances. It was either evaluated as general recommendation quality [10, 15, 19, 49, 100, 102, 104, 108, 168, 189] ($\checkmark$: 7, $\times$: 24), novelty [5, 36, 40, 49, 102, 107, 116, 125, 171, 207] ($\checkmark$: 25, $\times$: 19), perceived recommender performance or accuracy [5, 10, 18, 36, 40, 63, 66, 87, 91, 94, 101, 116, 138, 140, 141, 189, 190, 205, 209, 211] ($\checkmark$: 31, $\times$: 26), recommendations diversity [10, 52, 102, 116, 154, 171, 189] ($\checkmark$: 4, $\times$: 27), perceived serendipity [4, 5, 213] ($\checkmark$: 7, $\times$: 4), fairness [1, 9, 108, 166, 203] ($\checkmark$: 17, $\times$: 50), the preference for a specific recommender type [45, 109] ($\checkmark$: 8, $\times$: 1), and the alignment of the users’ expectation with the recommender system behavior [71, 91, 142, 148, 190] ($\checkmark$: 16, $\times$: 39).

Other A total of 153 measurements were grouped into this category. The effects circle around the perceived system complexity [31, 190] ($\checkmark$: 8, $\times$: 9), its ease of use [6, 10, 28, 32, 57, 116, 138, 149, 188, 190, 196, 210, 221] ($\checkmark$: 11, $\times$: 12), the perceived effort to navigate the system [1, 6, 105, 118, 148, 169, 170, 188, 190, 198] ($\checkmark$: 5, $\times$: 23), and how participants perceived their own performance [190] ($\checkmark$: 2, $\times$: 3), system acceptance [138, 146] ($\checkmark$: 2, $\times$: 3), perceived adequacy of the interactions [61, 116, 140] ($\checkmark$: 4, $\times$: 7), negative perception of the system (i.e., irritation or perceived uselessness of system) [6, 10, 28, 94, 188, 190, 221] ($\checkmark$: 6, $\times$: 10), perceived system usefulness [10, 36, 106, 116, 149, 153, 196, 221] ($\checkmark$: 26, $\times$: 12), a combined usability score Yang et al. [210] ($\checkmark$: 1), and perceived user experience (not further specified) [5, 102, 158, 221] ($\checkmark$: 7, $\times$: 3).

Correlations and mediating effects. Chatti et al. [33] and Guesmi et al. [68] investigated mediating and interaction effects for interface quality aspects on user characteristics. They did not find interactions between the perceived control over the recommender system and the need for cognition, visualization familiarity, domain knowledge, or technical expertise, but found interactions with personal innovativeness and propensity to trust others. In terms of the perceived quality of the system, Kunkel et al. [101] found that the perceived explanation quality mediates the perceived recommendation quality. They also investigated mediating effects of social presence and the perceived recommendation quality with the perceived system competence and the perceived recommendation quality with the perceived system integrity, but did not find an effect. Liao et al. [109] looked into interactions between the perceived recommender system quality and the bandwagon effect, but did not find an interaction effect. In contrast to this, Liao [108] found an interaction between the perceived recommendation quality and the bandwagon perception, and additionally an interaction with homophily. Göbel et al. [72] evaluated interactions between the perceived credibility of the information and need for cognition, the familiarity of the items, and conscientiousness, but did not find an effect. They did, however, find an interaction between credibility and information uncertainty. In terms of correlations, Feldman-Maggor et al. [61] found positive correlations with trust and information quality and interface adequacy.

5.3 Effects on user and system perception

Aside from effects that either impact the user or the system perception, the impact of explanations on the (perceived) transparency affects both the user and the system perception. Additionally, we identified two categories of effects that impact users and system perception: engagement and RS-specific effects. While the transparency measurements are approximately equal in the number of times they were found or not found, the majority of the engagement measurements found an effect, and the majority of the effects specific to the recommender scenario (RS-specific) were not found. Next, we report the detailed results for these categories (bundling the discussion of engagement and RS-specific effects into one sub-section).

5.3.1 Transparency.

Notions and measures of transparency. Tintarev et al. [181] define transparency in the context of recommender systems explanations as “*explain how the system works*”, implying that explanations increase the user’s understanding of how the recommender system works. Along with the impact of explanations on a user’s understanding of the system, we also found measurements of the perceived transparency of the recommendations and the recommendation system or process. We extracted a total of 351 findings from the articles in our corpus that fit into these transparency categories.

Transparency of Recommendations The perceived transparency of the recommendations was measured 25 times [4, 9, 36, 80, 111, 131, 145, 148, 155, 172, 190, 218] (✓: 14, ×: 11).

Transparency of the System The perceived transparency of the system or the recommendation process was evaluated in 34 instances [6, 21, 23, 24, 79, 81, 97, 104, 116, 188] (✓: 19, ×: 15).

Understanding The most frequent effect is the (perceived) understanding of the users which was evaluated 292 measurements times [1, 5, 7–10, 12, 19, 20, 30, 35, 38, 46, 48, 49, 52, 53, 64, 67, 68, 71, 80, 85, 90, 93, 94, 102, 105, 107, 112, 114, 115, 122, 131, 133, 139, 150, 154, 155, 163, 164, 166, 172, 179, 182, 189, 190, 197, 198, 202, 204, 205, 208, 211] (✓: 136, ×: 156). With 260 evaluations, the majority of these measured the users’ perceived understanding (✓: 122, ×: 138). Only 32 evaluations used a method to measure the objective understanding of a user [12, 64, 154, 155, 172, 182, 202] (✓: 14, ×: 18).

Correlations and interaction effects. We identified four findings in which transparency correlates with a different variable and 82 findings in which transparency was either the interaction effect or it was found through an interaction effect. Starke et al. [169] and Hernandez-Bocanegra et al. [80] both found that the perceived transparency of the users correlates with their social awareness, and Ahmad et al. [7] found correlations between transparency and satisfaction. The majority of the interaction effects that were evaluated involve user characteristics. No interaction effects were found between transparency and need for cognition [33, 68], visualization familiarity [33, 82], technical expertise [33], trust propensity [68], and decision-making style Chen et al. [38]. Guesmi et al. [68] found a significant interaction between transparency and personal innovativeness, which contrasts the findings of Chatti et al. [33], where no such interaction was found. Similar contrasts were identified for domain knowledge, where an interaction effect was found by Loepp et al. [111] but not found by Guesmi et al. [68]. Interactions of transparency and purchase intention [109, 122] (✓: 3, ×: 1), and trust [38, 71, 82, 109, 122, 220] (✓: 16, ×: 1) were also found in some instances but not in others. Interaction effects were found between transparency and the perceived explanation quality [82, 205], perceived recommendation quality [109, 111], persuasiveness [122], and trust in technology [111].

5.3.2 Other effects.

Notions and measures of other effects. This category entails the impact of explanations on the engagement of the users and the effects that are specific to a recommendation scenario. The effects specific to a recommendation scenario are related to aspects that are only relevant in certain application areas, such as the rapport built with an agent in a conversational recommender setting or the perceived fairness of the recommendations toward group members in a group recommender system. Furthermore, we added the engagement measurements for which we did not have further information about the effect that was evaluated (i.e., the time spent in the application can measure efficiency, but it can also be an indicator of the persuasiveness, effectiveness, or complexity of explanations).

Engagement Engagement measurements occur 133 times in our corpus [1, 10, 18, 26, 37, 53, 57, 63, 81, 91, 95, 107, 119, 124, 125, 134, 158, 176, 188, 204, 207] (✓: 92, ×: 41).

Recommendation Specific We identified a total of 131 measurements by only five articles [2, 141, 148, 153, 185] that are specific to the recommendation scenario. The majority is related to group recommendations, with 112 measurements [141, 153, 185] (✓: 51, ×: 61). For conversational recommendations, we identified a total of 19 measurements [2, 141, 148] (✓: 2, ×: 17).

Information Retention Measures on the impact of explanations on the participants' information retention occurred six times [72] (✓: 4, ×: 2).

Mediated effects and correlations. Barria-Pineda et al. [18] looked into mediating and interaction effects involving this category and found an interaction effect between the explanation format and engagement with the explanation.

5.4 Measurement of effects

As we can see from the majority of the findings of the effects listed in this section, it is not possible to make general statements about the effect of explanations in recommender systems without providing additional context. Understanding how the effects were measured adds such a contextual dimension to the results of the evaluation, which helps interpret their impact. We refrain from an in-depth analysis of the evaluation methodology to keep the focus of this paper on the explanations. Instead, we point at Rong et al. [160] for a comprehensive survey of user evaluation in XAI evaluation and Wardatzky et al. [200] for recommendations to improve current evaluation practices. However, we highlight some observations that differ from their findings or have not been investigated in their study.

We found two general approaches to how the effects of explanations were measured: (1) through objective measures such as clicks on recommendations or the quality of a participant's decision, and (2) through the participant's self-assessed, subjective assessment of the explanation. In total, our corpus spans 2 176 findings that measure the self-reported perception of the users and only 523 objective measurements.

The second point to take into consideration when interpreting the results of the user studies is that out of the 123 user studies that provided information on this aspect, 107 studies are short-term evaluations where the participants interact with the system for the first time. The 16 long-term studies ranged from a duration between 1 and 6 weeks to an entire teaching term and were often conducted as A/B tests.

5.5 Section summary

This section presents the effects measured when evaluating recommender systems' explanations. We developed a taxonomy of the effects and grouped them into three categories: effects on the user, effects on the system perception, and effects on the user and system perception. We further analyzed how often the effects were found or not found. Overall, we cannot derive general statements about the effects of explanations in recommender systems without further

Table 6. Overview of the number of times an effect was measured for each modality in cases where an explanation was compared to a system without an explanation and the articles that evaluated the combination. ✓ denotes that an effect was found, × denotes that the effect was not found. We colored the cells that show effects that were evaluated by at least two articles and where the effect was always found (blue) or not found (red). Yellow cells denote explanations that were evaluated at least 20 times on an effect with a small difference between the number of found and not found effects.

Displ.	Interactivity	Result	Trust	#	Persua-siveness	#	Effective-ness	#	Satis-faction	#	Effi-ciency	#	Interface Quality	#	Trans-parency	#	Other	#
hybrid	interactive	✓	[116]	1			[9]	2					[52, 66, 116]	4	[52, 116]	4		
		×	[52, 116]	7			[156, 157]	3	[18, 52, 116, 186]	7			[52, 116]	8				
hybrid	static	✓	[5, 31, 122, 145]	18	[122, 221]	3	[31, 146]	3	[221]	1			[5, 31, 221]	18	[5, 122, 145]	9		
		×	[31, 145, 146, 221]	8	[145, 221]	2	[31, 146]	3					[5, 31, 146, 203, 221]	14				
textual	interactive	✓	[104, 158, 172, 190]	6	[17]	2	[105, 148, 190]	6	[19, 148, 165, 190]	9	[198]	1	[6, 104, 148, 158, 172, 190]	17	[6, 12, 104, 105, 148, 172, 190, 198]	17	[158]	1
		×	[104, 148, 158, 172, 205]	13	[148]	2	[148]	1	[18, 100, 148, 165, 186, 198]	9			[19, 100, 104, 105, 148, 190, 198]	25	[12, 19, 148, 205]	6	[148]	6
	static	✓	[30, 38, 72, 122, 149, 153, 198]	21	[2, 40, 72, 97, 122, 153, 175, 198]	23	[30, 70, 90, 130, 132, 144, 197]	36	[40, 53, 90, 130, 149, 153, 197, 198]	12	[153, 198]	8	[9, 40, 72, 90, 97, 118, 149, 153]	25	[30, 53, 64, 90, 97, 122, 197, 198]	18	[53, 72, 153]	16
		×	[22, 48, 53, 72, 122, 130, 145]	25	[122, 145, 149, 175]	4	[70, 87, 90, 97, 118, 132, 144]	32	[22, 118, 197]	4			[9, 72, 87, 118, 198, 213]	19	[9, 48, 64, 145, 202]	10	[72, 119]	3
visual	interactive	✓	[1, 57, 189]	5									[1, 57, 139]	3	[52, 111, 139, 189]	4	[1, 57, 124]	3
		×	[52]	2			[139, 189]	2	[18, 52]	2			[1, 52, 107, 139, 189]	7	[1, 111]	3	[1]	2
	static	✓	[26, 31, 48, 125, 153, 211]	12	[153]	1	[26, 31, 87]	5	[53, 153]	2	[153]	1	[26, 31, 125, 153, 211]	9	[48, 53, 211]	3	[26, 53, 125, 153]	5
		×	[31, 48, 53]	7			[31, 87]	3	[120]	4			[31, 87, 125]	6	[202]	2		

context, as most effects were found in certain conditions and not found in others. We have a closer look at one of these context factors, the explanation itself, in Section 6, but we also recommend looking into the impact of the evaluation methodology on the reported results in future work. Finally, we briefly addressed the main issues that we found with the applied evaluation approaches in our corpus, namely a lack of objective measurements of the effects and the dominance of one-off short-term evaluations.

6 Effects of Explanation Characteristics

In this section, we present the results from analyzing the explanation effects from Section 5 in combination with the explanation characteristics presented in Section 4. The results are based on the findings in which a recommender system with an explanation was compared to a system without an explanation or an explanation was compared to another explanation. To trace down the effects of the explanation characteristics, we only consider findings in which one individual explanation was evaluated and disregard findings in which explanations with multiple characteristics were grouped together. Given the distribution of the data, we focus our analysis on the textual, visual, and hybrid display in the analysis of the effects of modalities, as with 5 tabular and 25 audio explanations, the number of explanations is too low to identify significant effects.

In the following, we will highlight the results that indicate an impact of the explanation characteristics on the effects and the findings where future work should have a closer look at the differences between the evaluated explanations.

6.1 Explanation compared to no explanation

We first investigate explanation modalities and information types that are likely to evoke an effect when a recommender system with an explanation is compared to a system without explanation. We extracted 586 results for this analysis.

6.1.1 Explanation modalities. We first look at how the measured effects differ for the explanation modalities. Table 6 summarizes the results. Overall, when distributing the findings across their modalities and effects, the number of findings for each modality-effect combination is too low to draw general conclusions. We can see, however, some indications for possible impacts of the modalities on the measured effects and identify gaps for future work. Interactive visual and hybrid explanations do not seem to have an effect on user satisfaction. In both cases, the effect was evaluated by multiple articles, but was never found. The same applies to visual interactive explanations and effectiveness. Both interactive and static hybrid explanations seem to have a positive impact on transparency, as all evaluations found an effect so far. There is also an indication that static text-based explanations have a positive effect on efficiency and static visual explanations on the effects grouped into the other category. As seen in Table 6, the impact of interactive explanations on persuasiveness and the general impact of explanation characteristics on efficiency have not been investigated much when comparing systems with and without an explanation. Our corpus contains two articles (with two evaluations each) that evaluated interactive text explanations on persuasiveness; one found an effect, and the other did not. Efficiency was only evaluated by two articles overall, Quijano-Sanchez et al. [153] with seven explanations overall, and Wang et al. [198] with three explanations. An effect was always found for all evaluated explanations. No indications can be found for the impact of any explanation modality on the trust and interface quality effects. The difference between the number of found and not-found effects of trust and interface quality for static text-based explanations, and on the interface quality for static hybrid explanations, is quite small. This could be a good starting point to explore the detailed differences between the explanations for which the effect was found and the explanations for which it was not found.

6.1.2 Explanation content. Regarding explanation type and focus, our corpus contains too few non-local explanations that do not focus on the output of the recommender to assess their ability to evoke an effect. For local explanations focusing on the output, the only indication we can find is a positive effect on efficiency with ten measurements from two articles [153, 198] that all found an effect. The remaining effects were both found and not found in similar proportions.

The results indicate that the information type of an explanation might impact the measured effect when being compared to a system without explanation. There seems to be a positive impact of multiple information types (decisive features, knowledge about the community, and inference trace) on transparency. Explanations containing knowledge about peers or decisive input values seem to positively impact satisfaction. Background data and feature-based domination might also impact satisfaction, but only one article evaluated this. Explanations with the *suitability estimate* information type always had an impact on the evaluated effect, but only two articles used this information type in their explanations. Table 7 reveals further gaps that can be filled by future research. Efficiency and persuasiveness were not evaluated very frequently in a setting where an explanation is compared to a system without explanations in general, but specifically for explanations that contain information types in the *items and their features* and *user preferences and input* categories; the results are very sparse. The difference between the number of times effects of interface quality were found and not found when *decisive features*, *inference and domain knowledge*, or *background data* explanations were evaluated is small. The same applies to effectiveness and explanations containing *background data*, *feature-based domination*, or *preference match* and trust and *decisive features* information types. We recommend starting to investigate the detailed differences between these explanations to gain further insights. Further starting points for future research that are indicated by our findings can be the impact of explanations belonging to: (1) feature-based domination type on satisfaction, (2) knowledge about similar items on persuasiveness, (3) inference trace on transparency and trust, and (4) background data on satisfaction.

Table 7. Overview of how many times an effect was measured for each information type when an explanation was compared to a system without explanations. ✓ denotes an effect was found, × denotes an effect was not found. For readability, we omit the results for effects and information types that were only evaluated by one article. We highlight the combinations of information type and effect that were evaluated by at least two papers and always found (blue) or not found (red) an effect. Yellow cells denote explanations that were evaluated at least 20 times on an effect with a small difference between the number of found and not found results.

Information Type	Result	Trust	#	Persua- siveness	#	Effectiveness	#	Satisfaction	#	Interface Qual.	#	Effi- ciency	#	Transparency	#	Other	#	
Background & compl. info	Background data	✓	[5, 172, 190]	3		[70, 132, 144, 190]	14	[190]	1	[5, 9, 172, 190, 203]	13			[172, 190]	5			
		×	[172]	2		[132, 144]	10			[9, 190, 203]	10			[9]	3			
	Knowledge about peers	✓	[1, 153]	9	[153, 221]	9	[130]	2	[130, 153, 221]	5	[1, 153, 221]	11	[153]	6		[1, 153]	13	
		×	[130, 221]	4	[221]	1				[1, 221]	3			[1, 202]	5	[1, 119]	4	
	Knowledge about similar items	✓			[40, 97]	5			[40, 53, 165]	6	[40, 97]	6		[53, 97]	3	[53]	1	
	×	[22, 53]	2			[97, 157]	3		[22, 165]	2	[213]	2		[202]	1			
	Knowledge about the community	✓	[38, 190]	3		[90, 132, 144, 190]	8	[90, 165, 190]	3	[90, 190, 203]	5			[90, 148, 190]	4			
		×	[148]	1	[148]	1	[70, 90, 144, 148]	11	[148, 165, 186]	4	[148, 190, 203]	12				[148]	2	
Decision inference process	Decision method side-outcomes	✓	[31]	3	[17]	2	[31]	1		[31]	4							
		×	[31]	3			[31]	3		[31]	2							
	Inference & domain knowledge	✓	[116, 122, 145]	5	[40]	2	[105]	2	[40]	2	[40, 52, 116, 203]	11		[12, 52, 105, 116, 122]	10			
		×	[22, 52, 116, 145]	13	[145]	1	[157]	2	[22, 52, 116, 186]	7	[52, 105, 116, 203]	16			[12]	1		
	Inference trace	✓	[5, 122]	5	[175]	1				[5]	9			[5, 122]	2			
		×			[175]	1	[156]	1	[100]	1	[5, 100]	6						
Items & their features		✓	[26, 31, 38, 48, 49, 72, 104, 116, 122, 158]	24	[72, 122]	9	[3, 26, 31, 49, 130, 146, 197]	11	[49, 130, 197]	6	[26, 31, 49, 72, 104, 116, 118, 158]	20		[48, 49, 104, 116, 122, 197]	15	[72, 158]	5	
	Decisive features	×	[31, 48, 72, 104, 116, 122, 130, 146, 158]	25	[122]	1	[31, 118, 146]	6	[116, 118, 120, 197]	8	[31, 72, 104, 116, 118, 146, 213]	15				[72]	1	
	Feature-based domination	✓	[26, 149]	3			[105, 132, 144]	18	[149]	1	[149]	2		[105]	2	[26]	1	
		×	[205]	4	[149]	1	[132, 144]	17		[105]	1			[205]	2			
	Pros and Cons	✓					[132, 144]	9										
	×					[132, 144]	10											
User preferences & input	Decisive input values	✓	[1, 5, 172, 190, 198, 211]	8	[198]	1	[87, 148, 190]	3	[148, 190, 198]	4	[1, 5, 148, 172, 190, 203, 211]	15	[198]	3	[5, 64, 111, 172, 190, 198, 211]	17	[1]	1
		×	[148, 172]	3	[148]	1	[87]	3	[198]	1	[1, 5, 87, 148, 190, 198, 203]	26		[1, 64, 111, 148]	7	[1, 148]	6	
	Feature importance analysis	✓	[48, 189]	2						[139]	1			[48, 139, 189]	3			
		×	[48]	6			[87, 139, 189]	4		[87, 139, 189]	6			[48]	1			
	Preference match	✓	[57, 122, 125]	5	[122]	2	[3, 90, 132, 144]	17	[19, 53, 90]	3	[57, 66, 90, 125, 139]	5		[52, 53, 90, 122, 139]	8	[53, 124]	3	
	×	[52, 53, 122, 205]	10	[122]	1	[90, 132, 139, 144, 157]	15	[18, 52, 186]	8	[19, 52, 107, 125, 139, 213]	8		[19, 205]	4	[125]	2		
	Suitability estimate	✓	[49, 153]	7	[153]	7	[49]	1	[49, 153]	5	[49, 153]	6	[153]	5	[49]	1	[153]	10

6.2 Explanation compared to a baseline explanation

Here, we present the results from analyzing the findings when one explanation is compared to another explanation as a baseline. In 1 583 findings, this baseline was generated by the research team to compare it with the evaluated explanations. External baselines from already published work were only used in 199 cases.

6.2.1 Explanation modalities. In terms of modalities (Table 8), we found an indication that static visual explanations could increase persuasiveness, interactive visual explanations could impact trust, and interactive text explanations could impact efficiency. All these combinations were evaluated by multiple articles, and an effect was always found. Interactive text-based explanations and persuasiveness are the only combination of modality and effect in our corpus, where an effect was never found despite being evaluated by multiple articles. A starting point for future research could be further investigating the impact of interactive visual explanations on satisfaction, which was evaluated only once in our corpus, and an effect was found. Table 8 additionally shows that no article in our corpus has evaluated the impact of visual interactive explanations on efficiency and hybrid static explanations on persuasiveness and satisfaction, which could be a focus for future work. The results of static textual explanations on trust and efficiency, and static visual explanations on effectiveness, interface quality, and transparency are close regarding how often an effect was found or not. This opens opportunities to examine the more detailed differences between the evaluated explanations.

Table 8. Overview of how many times an effect was measured for each modality when an explanation was compared to a baseline explanation. ✓ denotes an effect was found, × denotes an effect was not found. We highlight the combinations of information type and effect that were evaluated by at least two papers and always found (blue) or not found (red). Yellow cells denote explanations that were evaluated at least 20 times on an effect with a small difference between the number of found and not found effects.

Displ.	Interact.	Result	Trust	#	Persuasiveness	#	Effectiveness	#	Satisfaction	#	Efficiency	#	Interface Qual.	#	Transparency	#	Other	#
hybrid	interactive	✓	[33, 52, 85]	8	[3]	1	[156]	1	[52, 61, 187]	5	[85]	2	[5, 10, 52, 61, 85]	16	[10, 52, 85]	4	[10]	17
		×	[10, 85, 141, 163]	16	[3, 141, 163]	16	[3, 141, 156, 157, 163, 188]	10	[61, 187]	4	[163]	1	[10, 61, 141, 188]	30	[10, 163, 188]	6	[10, 141, 188]	22
	static	✓	[23, 85, 145]	5			[180, 188]	8			[204]	1	[80]	1	[145]	1	[10]	4
		×	[23, 145, 204]	8			[31, 169, 180, 204]	10			[204]	1	[10, 169]	6	[23, 204]	2	[10]	6
textual	interactive	✓	[138, 172]	2			[77, 148, 193]	4	[138, 148, 165, 186]	8	[193, 198]	2	[18, 138, 148, 198]	8	[172]	3	[148]	1
		×	[81, 148, 172, 205, 214]	12	[18, 148]	2	[81]	2	[53]	1			[148, 154, 172]	7	[81, 148, 154]	24	[148]	2
	static	✓	[7, 15, 20, 21, 23, 24, 30, 35, 38, 46, 71, 94, 95, 101, 131, 133, 145, 153, 164]	90	[15, 20, 21, 35, 63, 70, 99, 112, 131, 133, 153, 161, 162, 164, 167, 191, 198]	94	[20, 21, 30, 35, 46, 71, 74, 86, 167, 170, 179, 197]	87	[7, 21, 94, 95, 131, 138, 153, 179, 185, 197, 198]	26	[35, 113, 153, 185, 198, 207]	38	[7, 10, 15, 35, 71, 94, 102, 112, 131, 133, 138, 153, 166, 198, 207, 218]	54	[7, 20, 21, 24, 30, 35, 46, 71, 94, 112, 131, 133, 145, 164, 166, 179, 197, 208, 218]	64	[2, 10, 63, 95, 153, 185, 207]	62
		×	[10, 15, 20, 23, 24, 35, 38, 46, 71, 94, 101, 131, 133, 145, 153, 198]	108	[2, 15, 20, 35, 99, 131, 153, 161, 162, 167, 191]	68	[20, 30, 35, 46, 71, 86]	38	[53, 94, 131, 153, 170, 185, 197]	41	[35, 153, 185, 207]	37	[10, 15, 35, 71, 94, 101, 102, 131, 153, 166, 170, 198]	76	[10, 20, 23, 30, 35, 46, 71, 94, 102, 131, 166, 198, 202, 208]	84	[10, 153, 185, 207]	46
	interactive	✓	[52, 128]	3			[128]	1	[52]	1			[18, 91, 102, 107, 139]	7	[52]	1	[91, 107]	3
		×			[18]	1	[139]	1					[52, 102, 107, 139]	15	[102, 107, 111, 139]	5	[107]	1
visual	static	✓	[48, 153, 210]	5	[63, 153]	2	[8, 26, 180, 188, 210]	13	[8, 153, 210]	6	[153, 204, 210]	4	[26, 45, 62, 80, 153, 188, 210]	13	[8, 48, 182, 188]	11	[26, 63, 153, 188]	6
		×	[31, 48, 204]	8			[8, 26, 31, 180, 188, 204]	10	[8, 120]	4	[204]	4	[26, 31, 45, 188]	10	[8, 188, 202, 204]	11	[188]	2

6.2.2 Explanation content. We cannot derive any observation regarding the effects of local output explanations, as effects were found and not found in similar proportions. The biggest difference between the number of evaluations in which an effect was found vs. not found is for effectiveness, which was found 130 times out of 211 instances. For pairwise explanations that focus on the output, we found an indication that they have a positive impact on the explanation’s persuasiveness. Three articles [20, 21, 150] evaluated it in ten conditions and always found an effect.

As shown in Table 9, the impact of feature-based domination and inference and domain knowledge on persuasiveness was evaluated by multiple articles and always found an effect. The same goes for decisive features and efficiency, and decisive input values and satisfaction. The evaluations of the impact of decisive features on satisfaction and feature importance analysis on trust and effectiveness were both evaluated by multiple papers, and no effects were found. Potential starting points to verify the impact of information types on explanation effects are the (1) impact of decision-method side-outcomes on any effect, (2) impact of decision inference process explanations on effects grouped in the other category, (3) combinations of information type and effect that were only evaluated by one article and an effect was found, and (4) detailed differences of explanations in combinations where there is only a small difference between the number of times an effect was found and not found (yellow cells in Table 9), as well as (5) all information type-effect combinations that have not been evaluated at all, such as the impact of inference trace on satisfaction.

6.2.3 Baselines and information type combinations. The final analysis we conducted is an exploration of how the baseline that the evaluated explanation is compared to impacts the results. Given the dominance of static textual explanations focusing on explaining one recommendation to a user, most of the measured effects (1 309 out of 1 555) are from comparisons between the same explanation modalities and types. The remaining comparisons of explanations with different modalities are sparsely distributed across the effects, so we could not derive any indications for a possible impact on the effect. We, therefore, focus the analysis on the information type categories of the explanations. Up

Table 9. Overview of often an effect was measured for each information type in cases where an explanation was compared to a baseline explanation. ✓ denotes an effect was found, × denotes an effect was not found. We highlight the combinations of information type and effect that were evaluated by at least two papers and always found (blue) or not found (red). Yellow cells denote explanations that were evaluated at least 20 times on an effect, and the difference between found and not-found effects is small.

Information Type	Result	Trust	#	Persuasiven.	#	Effectiveness	#	Satisfaction	#	Efficiency	#	Interface Qual.	#	Transparency	#	Other	#	
Background & compl. info	Background data	✓	[15, 23, 24, 35, 138, 172]	24	[15, 35, 70, 99, 150]	25	[35, 77, 114, 150, 180, 193]	18	[114, 138]	2	[35, 114, 193]	10	[15, 35, 138, 203]	12	[24, 35, 150, 172]	12	[2]	1
		×	[15, 23, 24, 35, 81, 141, 172]	77	[2, 15, 35, 99, 141]	43	[35, 81, 141, 180]	26	[138]	1	[35, 114]	18	[15, 35, 141, 172, 203]	27	[23, 35, 81, 150]	28	[141]	4
	Knowl. about peers	✓	[153]	9	[153]	9			[153, 185]	11	[153, 185]	22	[153]	9			[153, 185]	38
		×	[153]	4	[153]	4			[153, 185]	26	[153, 185]	17	[153]	4	[202]	12	[153, 185]	40
	Knowl. about similar items	✓	[23, 24, 114, 131]	16	[112, 131]	3	[114]	1	[114, 131]	2			[102, 131, 166]	9	[24, 112, 114, 131, 166]	7		
		×	[23, 24, 131]	8	[131, 161]	11			[131]	11	[114]	1	[102, 131, 166]	32	[23, 102, 131, 166, 202]	27		
	Knowl. about the community	✓	[23, 38, 85, 138]	7	[38, 63, 70]	3	[20, 180]	7	[61, 138]	3	[85, 207]	6	[45, 61, 80, 85, 112, 138, 203, 207]	35	[85]	2	[63, 207]	7
		×	[23, 38, 85]	6	[38]	1	[169]	6	[61, 138]	4	[207]	2	[45, 61, 169, 203]	11	[23]	1	[207]	4
	Relationship betw. knowl. objects	✓	[164]	1	[164]	1									[164]	1		
		×	[163]	1	[163]	1	[163]	1			[163]	1	[28]	1	[163]	1		
Decision method side-outcomes	✓	[33]	2			[28]	2			[28]	1	[28]	1					
	Inference & domain knowl.	✓	[7, 30, 52, 110, 122, 145]	13	[63, 99, 162]	11	[30, 86, 114, 170]	5	[7, 52, 110, 114]	5	[114]	1	[7, 52, 110, 203]	6	[7, 30, 52, 122, 145]	6	[63]	3
		×	[108, 122, 145]	5			[30, 86, 157, 169]	9	[110, 170]	2	[114]	1	[52, 108, 110, 154, 169, 170, 203]	17	[30, 154]	10		
	Inference trace	✓	[46, 122, 128]	3			[46, 128]	2					[62, 166]	5	[46, 122, 166]	5		
×		[163]	1	[163]	1	[163]	1			[163]	1	[166]	7	[122, 163, 166]	9			
Items & their features	Decisive features	✓	[20, 48, 85, 122, 133, 164]	14	[3, 20, 93, 122, 133, 150, 164, 167]	14	[20, 26, 150, 167, 170, 180, 188]	18			[85, 113]	5	[10, 26, 85, 133, 166, 188]	21	[20, 48, 85, 93, 133, 150, 164, 166, 182, 188, 218]	25	[10, 188]	10
		×	[10, 31, 38, 48, 85, 133]	9	[3, 20]	4	[3, 26, 31, 180, 188]	12	[120, 170]	3			[10, 26, 31, 154, 166, 170, 188]	30	[10, 115, 122, 150, 154, 166, 188]	29	[10, 188]	6
	Feature-based domination	✓	[20, 128]	13	[20, 21]	14	[20, 21, 128, 150]	26							[20, 21, 150]	12	[26]	1
		×	[205]	2			[20]	1							[20, 150]	3		
Pros & Cons	✓	[36, 85, 138]	5	[36]	2	[36, 74]	2	[36, 61, 138]	4	[36, 85, 204]	6	[36, 61, 85, 138]	24	[36, 85]	4			
	×	[85, 204]	5			[204]	4	[61, 138]	3	[204]	5	[61, 166]	7	[166, 204]	8			
User preferences & input	Decisive input values	✓	[33, 172]	3	[162, 191, 198]	8	[148]	1	[148, 198]	2	[198, 207]	3	[148, 198, 207]	7	[166, 172]	4	[148, 207]	6
		×	[148, 172, 198]	6	[148, 162, 191]	3					[207]	1	[148, 166, 172, 198]	9	[148, 166, 198]	5	[148]	2
	Feat. importance analysis	✓											[10, 91, 139]	5	[10]	1	[10, 91]	19
		×	[10, 163, 214]	8	[163]	1	[139, 163]	2			[163]	1	[10, 139]	21	[10, 139, 163]	6	[10]	8
	Preference match	✓	[7, 20, 71, 94, 95, 114, 122, 138, 210]	12	[3, 20, 21, 93, 114, 122, 161]	22	[8, 20, 21, 71, 114, 188, 210]	43	[7, 8, 47, 94, 114, 138, 186, 210]	16	[113, 114, 210]	6	[7, 18, 71, 80, 94, 107, 138, 139, 210]	19	[7, 8, 20, 21, 71, 93, 114]	18	[10, 95, 107, 188]	7
		×	[20, 71, 94, 141, 205]	30	[3, 18, 20, 141, 161, 167]	25	[3, 8, 20, 71, 139, 141, 157, 188]	20	[8, 94, 138, 187, 217]	7	[114]	2	[10, 71, 94, 107, 139, 141, 188]	15	[8, 20, 71, 94, 107, 122, 139, 188]	23	[10, 107, 141, 188]	16
	Suitability est.	✓	[46, 153]	10	[153]	8	[46]	2	[94, 153, 187]	5	[153]	8	[153]	8	[46, 94]	2	[153]	13
		×	[46, 94, 153]	6	[153]	4	[46]	1	[153, 187]	3	[153]	4	[94, 153]	6	[46]	2	[153]	7

until now, we have looked at each information type individually. For this analysis, we additionally aim to find out if a combination of multiple information types in one explanation impacts the measured effects.

As seen in Table 10, the number of findings for each effect is low, and we can only derive potential indications for the impact of information types on the effects. First, we look at the comparisons of explanations where both the evaluated and the baseline explanations belong to only one information type category. Explanations containing information on items and their features were compared to three such baselines. Bellini et al. [20] found that explanations containing information about *items and their features* are more likely to improve transparency and trust than explanations containing *background and complementary information*. For the same information types, Bellini et al. [20] found that explanations of *items and their features* were better in increasing trust. Kouki et al. [99] found that compared to an explanation containing information on the *decision inference process*, the evaluated explanation is worse for persuasiveness. For

Table 10. Explanation effects of comparing different information types when taking the combination of the evaluated with another information type into consideration. The combination of information types is denoted by +, $\uparrow$ denotes the evaluated information type was better than the baseline, $\downarrow$ that the baseline was better, and = that no significant differences between them were found. We omit effects and explanation type combinations without an indication of a potential impact. The full table can be found in the supplementary material. Blue cells indicate that one information type was exclusively found to be better in evoking an effect than the other by at least two articles.

Evaluated Information Type	Baseline Information Type	Result	Trust	#	Persuasiveness	#	Effectiveness	#	Satisfaction	#	Efficiency	#	Interface Quality	#	Transparency	#	Other	#
Background & complementary information	Decision inference process	$\uparrow$			[99]	1							[203]	1				
		=			[99]	2							[28, 203]	10				
		$\downarrow$			[99]	4	[28]	2				[28]	1					
	User preferences & input	$\uparrow$										[207]	2	[207]	4		[207]	2
		=	[141, 148]	4	[141, 148, 161, 167, 191]	9	[141]	1	[187]	2	[207]	2	[141, 148]	4	[148]	1	[141, 148, 207]	8
		$\downarrow$	[20]	3	[20, 191]	6	[20, 148]	7	[148, 187]	2	[207]	1	[148, 207]	4	[20]	3	[148, 207]	3
	+ Items & their features	=			[38]	1	[180]	2	[61]	1			[61]	1				
		$\downarrow$					[180]	3					[61]	1				
	+ User preferences & input	$\uparrow$	[153]	1	[153]	1					[153]	1	[153]	1			[153]	2
		=	[141, 153]	5	[141, 153]	6	[141]	1			[153]	2	[141, 153]	3			[141, 153]	8
		$\downarrow$	[153]	3	[153]	3			[138, 153]	3	[153]	3	[80, 138, 153]	7			[153]	2
Decision inference process	User preferences & input	$\uparrow$	[33]	1	[162]	1												
		=	[108]	1	[162]	1							[108]	3	[46]	1		
		$\downarrow$	[33, 46]	3	[162]	1	[46]	1	[186]	1								
	+ Items & their features	=							[170]	1			[154, 170]	2	[154]	7		
		$\downarrow$					[170]	1										
	+ User preferences & input	$\uparrow$	[52]	2					[52]	1					[52]	1		
		=											[52]	1				
		$\downarrow$	[46]	1			[46]	1							[46]	1		
	Background & compl. info	$\uparrow$	[20]	6	[20]	5	[20, 74]	10							[20]	6		
		=			[20]	1	[169, 180]	5					[169]	2				
		$\downarrow$			[99]	1	[20, 180]	6										
Items & their features	Decision inference process	=	[31]	2			[31]	1					[31, 154]	3	[154]	9		
		$\downarrow$			[99]	2												
		$\uparrow$	[48]	3									[10]	7	[48]	1	[10]	8
	User preferences & input	=	[10, 20, 48, 205, 214]	19	[20]	7	[20]	12					[10]	17	[10, 20]	16	[10]	6
		$\downarrow$	[20]	9	[20]	11	[20]	24					[10]	2	[10, 20]	5	[10]	7
	+ Background & compl. info	$\uparrow$					[180]	1										
		$\downarrow$	[164]	1	[150, 164]	2	[150]	2							[150]	2	[150, 164]	3
	+ Decision inference process	=											[154]	1	[154]	7		
		$\downarrow$	[128]	1			[128]	1										
	+ User preferences & input	$\uparrow$			[167]	1	[167]	1										
		$\downarrow$	[205]	2	[93]	1									[93]	1		
	+ Items & their features	=	[205]	2														
		$\downarrow$			[21]	2	[21]	1							[21]	1		

explanations containing *background and complementary information*, Bogina et al. [28] found that they are worse for efficiency and effectiveness than explanations with *decision inference* information. Tran et al. [186] also found that *decision inference process* information is worse for satisfaction when compared to user preferences and input information.

The second aspect we take a look at is whether extending the information type of the evaluated explanation with a different information type (i.e., a comparison between a *decision inference process* explanation and a baseline with *decision inference process* and *user preferences & input* information) has an impact on the effects. Penha et al. [150] and Shin et al. [164] found that adding *background and complementary information* to an explanation that contains information on *items and their features* increases persuasiveness compared to an explanation containing only *items and their features*. Shin et al. [164] also found it increases trust. When adding *decision inference process* information instead, Mishra et al. [128] found that it might increase effectiveness and trust. There are indications that adding *user preferences*

and input information to background and complementary information increases satisfaction. Adding information on items and their features to user preferences and their input might improve effectiveness. For this combination, Bellini et al. [21] also found additional positive effects for persuasiveness and transparency. We also found a few instances in which adding more information types does not have a beneficial effect on explanation effects. Adding user preference and input information to decision inference process information negatively impacts satisfaction [52], and adding background and complementary information to items and their features decreases effectiveness [167].

Despite these indications, the majority of explanation pairs that were evaluated by multiple articles were better at evoking an effect in some instances and worse in others.

6.3 Section summary

In this section, we analyzed the impact of explanation characteristics on the measured effects. We separated the analysis between explanations that were evaluated against a system without explanation and explanations that were compared to a baseline explanation. While our data sample was too small to show significant results, we were still able to provide starting points for further research, particularly for the information type of the explanations. Most of the baseline explanations are of the same modality and content category as the evaluated explanations, making it challenging to identify the differences that might impact the measured effect. We discuss these results in Section 7.

7 Discussion and Limitations

After extensively describing our results, we discuss their implications and answer the research questions.

7.1 Explanation characteristics

First, we answer RQ 1: *What kind of explanations are commonly evaluated in recommender systems regarding content and modalities?* To do so, in Section 4, we described the characteristics of the explanations typically evaluated for recommender systems in terms of modality and content. With respect to modality, we found a *clear majority of static, textual explanations*. While there have been individual comparisons between text-based explanations and explanations of other modalities (e.g., text compared to audio [216] or text compared to visual explanations [50]), to the best of our knowledge, there has not been a systematic evaluation on which modality works better in which context (i.e., application domain, explanation goal, target user group, interface design). We can, therefore, only hypothesize about the reasons for this dominance. Visual explanations are a powerful tool to communicating complex information [34]. Most explanation approaches in our corpus focus on explaining a single recommendation through user history, similar users, relevant features, etc. This information is easy to convey in a short textual statement. Other reasons could be the increased popularity of generative AI for generating text in recent years and the newly opened research avenues in the area of XAI (see Kim et al. [93], Lubos et al. [114], or Ashaduzzaman et al. [12] for examples from our corpus).

The content mainly focuses on explaining the output of the recommender system and is usually a *local explanation that explains one recommendation*. Hence, we *know very little about the effect of non-local explanations or explanations that explain different aspects of the recommendation process*.

The information types of explanations are more distributed across the categories. Preference match explanations are the most popular information type, followed by decisive features and knowledge about the community. The distribution of explanations across all information type categories indicates that the focus of evaluation in the articles of our corpus is on *experimenting with different content types and not so much with different modalities*. This is further backed up by

the fact that only 17% of the evaluated combinations of explanations compared different modalities, in contrast to 38% that compared different information type categories.

As a consequence, we suggest *future research to investigate the effects of non-local explanation types and explanations that focus on aspects of the recommender system other than the output*. Miller [127] recently summarized some issues with current XAI approaches and called for a paradigm shift in the community to move towards hypothesis-driven decision support based on the cognitive decision-making process of humans. He proposes a conceptual framework that keeps the decision maker in control over which option to explore while the machine provides feedback on the selected option without justifying why the machine is correct. It would be interesting to see if and how this framework could be applied to current recommendation settings and if such an approach that requires the users to engage with their decision-making process cognitively would be accepted by users.

7.2 Explanation effects

Our second research question centers around the effects that were measured in the explanation evaluation. We identified 23 notions of eight effects that we sorted into three categories, depending on whether they impact the user directly, the system perception, or both. Most of the effects in our categorization impact the user. Still, most of the findings in our corpus are about the impact of explanations on the (perceived) quality of the interface and, thus, on the perception of the system. When looking at the measured results, all effects with a considerable number of measurements were both found and not found, which confirms that *context matters when interpreting the reported results and that there is no one-size-fits-all explanation*.

One major aspect to consider when interpreting the measured effects in Section 6 is the method with which the explanations were evaluated. Except for engagement and efficiency metrics, the *majority of the measured effects reflect user perception*. For effects that inherently measure user perception, such as satisfaction, this is to be expected. Still, for effects such as effectiveness or transparency, we expected to see a more balanced distribution, if not a slight majority, of the number of objective measurements compared to the self-reported perceived measurements. Research further indicates that a user's perceived understanding of a system or model does not necessarily match the actual comprehension [42, 75], which is why *it is important to make a distinction between objective vs. perceived effects*.

We generally noticed that the questions or tasks with which the effects were measured frequently varied. The *same concept can be evaluated with different questions*. To provide an intuition for this issue, for example, Zhang et al. [216] asked their participants to rate to what extent the explanation helps them to understand why the item is recommended, Chatti et al. [33] asked about the perceived understanding of what the recommendations are based on, and Afridi [4] asked to rate the extent to which the explanations convey transparency of the recommendations. This raises the issue of whether *the community should start developing and adopting validated survey instruments*. We provide the extracted questions in the supplemental material, but a detailed analysis of this aspect was out of the scope of this survey. It is, nonetheless, an important aspect to keep in mind when interpreting the presented results.

We further identified that most findings are based on participants' first impressions, as the number of studies in which participants interact with the systems repeatedly, over a longer period, is low. Literature suggests that some effects, such as trust, evolve and change over time [83, 183]. *Due to the lack of long-term evaluation, it is mostly unknown if and how the effects of explanations evolve for long-term users of recommender systems*.

Lastly, we noticed only *very few evaluations of the potential negative effects of adding explanations to recommender systems*, such as an overreliance of the user on the recommendations or information overload created through the

explanations. Including potential weaknesses of explanations in the evaluation procedure could contribute to a better understanding of the capabilities of explanations in recommender systems.

As a consequence, it seems as if *the field could greatly profit from contextualized studies that investigate the positive and negative effects over longer periods of time using well-defined measurement instruments*.⁸

7.3 Impact of explanation characteristics on measured effects

The third research question explores how the explanation characteristics from Section 4 influence the effects described in Section 6. While we cannot answer this question conclusively, as there are only a few measurements for each characteristic-effect combination, our findings indicate that the impact of explanations' characteristics might depend on the baseline used in the comparison. We found *different indications for characteristics that impact the effects when the explanation was compared against a system without an explanation and to a baseline explanation*. Interactive and static hybrid explanations, e.g., might positively influence transparency compared to a recommender system without explanations. When the baseline is another explanation, we did not find such an indication for transparency. Instead, static visual explanations might positively impact persuasiveness. The results also indicate that *the characteristics of the baseline play a role in what results are found*. We could not draw any general conclusions such as "adding a second information type to the explanation leads to an improvement" from our results. We found, for example, indications that adding another information type to a decision inference process explanation can lead to a worse result on the evaluated effects compared to the simple explanation. When adding another information type to an item and its features explanation, results indicate that it could improve trust and persuasiveness.

Overall, in the results presented in Section 6, the majority of combinations of explanation characteristics and effects were found by some articles and not found by others. Sometimes, the same article found an effect in some instances, and in others, no effect was found. The differences across articles could be caused by discrepancies in how the effect was measured (see Section 5.4). The differences within the same article indicate that factors other than the explanation can impact the measured effects. Investigating what these factors are and the extent to which they influence the results would greatly improve the understanding that we currently have of the explanations' effects on users. This clearly indicates that *the field could greatly profit from the standardization of experimental procedures to be able to compare and combine the results of the various user studies*.

In this survey, we only analyzed comparisons of two explanations when they are compared to a baseline explanation that has a different modality or information type. A significant part of our corpus compares two explanations with the same modality or information type. 83% of all modality comparisons and 62% of the information type comparisons are between explanations with the same characteristic. Investigating the detailed differences of these pairs is left for future work. Consequently, *the field lacks a systematic understanding of the tradeoffs between the various explanation types* — a crucial endeavor requiring further study.

Another aspect that should be investigated in future work is the role of large language models (LLMs) in explanation generation. In our corpus, nine papers used LLMs [7, 12, 77, 93, 105, 144, 167, 173, 198] to generate explanations, all of which were textual. Using LLMs to generate explanations comes with challenges, though, that require thorough evaluation and further research. Most importantly, LLMs tend to hallucinate information that users might not be able to identify as false or not supported [201].

⁸Cf. [96, 152] for proposals to define standardized user-centric instruments in the recommender systems domain.

7.4 Limitations

One of the main limitations of this work is that we are bound to what has been reported in the papers of our corpus. It is more likely that statistically significant results are reported than results where the evaluated effect was not found, which might impact the results shown in Section 6. However, our work highlights what is known in the research community, so we think the reporting bias did not significantly influence our contribution.

When creating the corpus for a survey, a selection bias might limit the scope of the papers. We mitigated this by querying library databases covering multiple disciplines and not selecting papers based on venue or number of citations. While we only analyzed the papers published between 2017 and July 2025, this period showed the largest increase in explainability research [194], demonstrating that the sample of papers we analyzed covers a significant time period.

Other limitations derive from decisions on what data to focus on for our analysis. We decided to evaluate the explanation modalities separately from the content, as combining the modalities and content variables for each explanation increases the sparsity of the results, leaving it unable to draw conclusions. As this survey aims to investigate which explanation characteristics are likely to evoke a certain effect, we believe that these factors did not impact our reported results and left a thorough exploration of the combination of modalities and content for future work.

8 Conclusions and Future Work

In this literature survey, we analyzed 180 articles in which recommender system explanations were evaluated in a user study and extracted information about the explanations and the effects that were measured in the evaluation. We first described the characteristics of the 665 explanations in terms of modality and content. Out of the five modalities, 55% of all explanations are text-based, and 77% are static. This signifies a lack of knowledge about non-text-based and interactive explanations. Content-wise, the majority of explanations explain one individual recommendation (80%) and 44% of the explanations either contain information about how the recommendation matches a user's preference, the item features relevant to the recommendation, knowledge about the community, or knowledge about a user's peers. We then presented a taxonomy of the effects that were measured in the evaluations, which resulted in eight effect categories that we grouped depending on whether they impact the user, their perception of the system, or both. When analyzing how often each effect was found in the evaluations and how often it was evaluated but not found, we concluded that it is not possible to make general statements about the effects of explanations in recommender systems without further context. We investigated the explanation characteristics as context variables for the explanation effects with the goal of seeing how their modalities and content impact the measured effects. While we could not identify significant impacts of explanation characteristics due to data sparsity, we identified promising gaps for future research.

The main gaps that we identified in our survey that future work should focus on are: (1) standardized methods to measure the explanation effects, (2) effects of explanations that are not local and focus on other aspects of the recommender system than the output, and only when the first two points have progressed, (3) further investigate the differences of explanations for which an effect was found and the explanations for which the same effect was not found, and (4) generally adopting standardized procedures as much as possible. Addressing these four challenges will greatly improve the research community's understanding of the role that the explanation design plays when measuring the effects of explanations in recommender systems.

References

- [1] Solmaz Abdi et al. 2020. Complementing educational recommender systems with open learner models. In *Proc. ACM LAK*. ACM, 360–365. <https://doi.org/10.1145/3375462.3375520>

- [2] Amal Abdulrahman et al. 2019. Belief-based Agent Explanations to Encourage Behaviour Change. In *Proc. ACM IVA*. ACM, 176–178. <https://doi.org/10.1145/3308532.3329444>
- [3] Amal Abdulrahman et al. 2022. Changing users' health behaviour intentions through an embodied conversational agent delivering explanations based on users' beliefs and goals. *Behaviour & Information Technology* 42, 9 (2022), 1338–1356. <https://doi.org/10.1080/0144929x.2022.2073269>
- [4] Ahmad Hassan Afridi. 2018. Visualizing Serendipitous Recommendations in User Controlled Recommender System for Learning. *Procedia Computer Science* 141 (2018), 496–502. <https://doi.org/10.1016/j.procs.2018.10.136>
- [5] Ahmad Hassan Afridi. 2019. Transparency for Beyond-Accuracy Experiences. *Procedia Computer Science* 151 (2019), 335–344. <https://doi.org/10.1016/j.procs.2019.04.047>
- [6] Ahmad Hassan Afridi et al. 2020. Triggers and connection-making for serendipity via user interface in recommender systems. *Personal and Ubiquitous Computing* 25, 1 (2020), 77–92. <https://doi.org/10.1007/s00779-020-01371-w>
- [7] Juan Ahmad et al. 2025. Tell Me the Good Stuff: User Preferences in Movie Recommendation Explanations. In *Adjunct Proc. ACM UMAP*. ACM, 103–108. <https://doi.org/10.1145/3708319.3733676>
- [8] Mohammed Alshammari et al. 2019. Mining Semantic Knowledge Graphs to Add Explainability to Black Box Recommender Systems. *IEEE Access* 7 (2019), 110563–110579. <https://doi.org/10.1109/ACCESS.2019.2934633>
- [9] Gabrielle Alves et al. 2024. User Perception of Fairness-Calibrated Recommendations. In *Proc. ACM UMAP*. ACM, 78–88. <https://doi.org/10.1145/3627043.3659558>
- [10] Ivana Andjelkovic et al. 2019. Moodplay: Interactive music recommendation based on Artists' mood similarity. *Int. J. Hum.-Comput. Stud.* 121 (2019), 142–159. <https://doi.org/10.1016/j.ijhcs.2018.04.004>
- [11] Vijay Arya et al. 2019. One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques. [arXiv:1909.03012](http://arxiv.org/abs/1909.03012) <http://arxiv.org/abs/1909.03012>
- [12] Md Ashaduzzaman et al. 2025. Leveraging Generative AI to Improve Comprehensibility in Social Recommender Systems. In *Adjunct Proc. ACM UMAP*. ACM, 192–201. <https://doi.org/10.1145/3708319.3733696>
- [13] Z.K.A. Baizal et al. 2020. Computational model for generating interactions in conversational recommender system based on product functional requirements. *Data & Knowledge Engineering* 128 (2020), 101813. <https://doi.org/10.1016/j.datak.2020.101813>
- [14] Krisztian Balog et al. 2020. Measuring Recommendation Explanation Quality. In *Proc. ACM SIGIR*. ACM, 329–338. <https://doi.org/10.1145/3397271.3401032>
- [15] Catalin-Mihai Barbu et al. 2019. The influence of trust cues on the trustworthiness of online reviews for recommendations. In *Proc. ACM SAC*. ACM, 1687–1689. <https://doi.org/10.1145/3297280.3297603>
- [16] Alejandro Barredo Arrieta et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion* 58 (2020), 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- [17] Jordan Barria Pineda et al. 2019. Making educational recommendations transparent through a fine-grained open learner model. In *Proceedings of Workshop on Intelligent User Interfaces for Algorithmic Transparency in Emerging Technologies at the 24th ACM Conference on Intelligent User Interfaces, IUI 2019, Los Angeles, USA, March 20, 2019 (CEUR Workshop Proceedings, Vol. 2327)*. CEUR-WS.org, 5 pages.
- [18] Jordan Barria-Pineda et al. 2020. Personalized Remedial Recommendations for SQL Programming Practice System. In *Adjunct Proc. ACM UMAP*. ACM, 135–142. <https://doi.org/10.1145/3386392.3399312>
- [19] Jordan Barria-Pineda et al. 2023. Adaptive Navigational Support and Explainable Recommendations in a Personalized Programming Practice System. In *Proc. ACM HT*. ACM, 1–9. <https://doi.org/10.1145/3603163.3609054>
- [20] Vito Bellini et al. 2018. Knowledge-aware Autoencoders for Explainable Recommender Systems. In *Proc. Workshop DLRS*. ACM, 24–31. <https://doi.org/10.1145/3270323.3270327>
- [21] Vito Bellini et al. 2019. Semantics-Aware Autoencoder. *IEEE Access* 7 (2019), 1166122 – 1166137. <https://doi.org/10.1109/ACCESS.2019.2953308>
- [22] Daniel Ben Zaken et al. 2022. Generating Recommendations with Post-Hoc Explanations for Citizen Science. In *Proc. ACM UMAP*. ACM, 69–78. <https://doi.org/10.1145/3503252.3531290>
- [23] Shlomo Berkovsky et al. 2017. How to Recommend? User Trust Factors in Movie Recommender Systems. In *Proc. ACM IUI*. ACM, 287–300. <https://doi.org/10.1145/3025171.3025209>
- [24] Shlomo Berkovsky et al. 2018. A Cross-Cultural Analysis of Trust in Recommender Systems. In *Proc. ACM UMAP*. ACM, 285–289. <https://doi.org/10.1145/3209219.3209251>
- [25] Moumita Bhattacharya et al. 2022. Augmenting Netflix Search with In-Session Adapted Recommendations. In *Proc. 16th RecSys*. ACM, 542–545. <https://doi.org/10.1145/3523227.3547407>
- [26] Émilie Bigras et al. 2019. Recommendation Agent Adoption: How Recommendation Presentation Influences Employees' Perceptions, Behaviors, and Decision Quality. *Applied Sciences* 9, 20 (2019), 4244. <https://doi.org/10.3390/app9204244>
- [27] Mustafa Bilgic et al. 2005. Explaining recommendations: Satisfaction vs. promotion. In *Beyond personalization workshop, IUI*, Vol. 5. Citeseer, 8 pages.
- [28] Veronika Bogina et al. 2020. Visualizing Program Genres' Temporal-Based Similarity in Linear TV Recommendations. In *Proc. ACM AVI*. ACM. <https://doi.org/10.1145/3399715.3399813>
- [29] Théo Bontempelli et al. 2022. Flow Moods: Recommending Music by Moods on Deezer. In *Proc. 16th RecSys*. ACM, 452–455. <https://doi.org/10.1145/3523227.3547378>

- [30] Zana Bućinca et al. 2020. Proxy tasks and subjective measures can be misleading in evaluating explainable AI systems. In *Proc. ACM IUI*. ACM, 454–464. <https://doi.org/10.1145/3377325.3377498>
- [31] Zana Bućinca et al. 2021. To Trust or to Think. *Proc. ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–21. <https://doi.org/10.1145/3449287>
- [32] Berk Buzcu et al. 2024. Towards interactive explanation-based nutrition virtual coaching systems. *Autonomous Agents and Multi-Agent Systems* 38, 1 (2024). <https://doi.org/10.1007/s10458-023-09634-5>
- [33] Mohamed Amine Chatti et al. 2022. Is More Always Better? The Effects of Personal Characteristics and Level of Detail on the Perception of Explanations in a Recommender System. In *Proc. ACM UMAP*. ACM, 254–264. <https://doi.org/10.1145/3503252.3531304>
- [34] Mohamed Amine Chatti et al. 2024. Visualization for Recommendation Explainability: A Survey and New Perspectives. *ACM TiiS* 14, 3, Article 19 (Aug. 2024), 40 pages. <https://doi.org/10.1145/3672276>
- [35] Changdong Chen et al. 2023. When Post Hoc Explanation Knocks: Consumer Responses to Explainable AI Recommendations. *Journal of Interactive Marketing* 59, 3 (2023), 234–250. <https://doi.org/10.1177/10949968231200221>
- [36] Li Chen et al. 2017. Explaining Recommendations Based on Feature Sentiments in Product Reviews. In *Proc. 22nd IUI*. ACM, 17–28. <https://doi.org/10.1145/3025171.3025173>
- [37] Li Chen et al. 2019. User Evaluations on Sentiment-Based Recommendation Explanations. *ACM TiiS* 9, 4, Article 20 (2019), 38 pages. <https://doi.org/10.1145/3282878>
- [38] Qian Qian Chen et al. 2025. Tailoring explanations in conversational recommendations: The impact of decision contexts and user interfaces. *Journal of Retailing and Consumer Services* 85 (2025), 104281. <https://doi.org/10.1016/j.jretconser.2025.104281>
- [39] Xu Chen et al. 2022. Measuring “why” in recommender systems: A comprehensive survey on the evaluation of explainable recommendation. *CoRR* abs/2202.06466 (2022), 9 pages. arXiv:2202.06466 <https://arxiv.org/abs/2202.06466>
- [40] Jaewon Choi et al. 2017. Examining the effects of personalized App recommender systems on purchase intention: A self and social-interaction perspective. *Journal of Electronic Commerce Research* 18, 1 (2017), 73 – 102.
- [41] Michael Chromik et al. 2020. A Taxonomy for Human Subject Evaluation of Black-Box Explanations in XAI. In *Proc. Workshop on Explainable Smart Systems for Algorithmic Transparency in Emerging Technologies co-located with 25th International Conference on Intelligent User Interfaces (IUI 2020)*, Cagliari, Italy, March 17, 2020 (CEUR Workshop Proceedings, Vol. 2582). CEUR-WS.org, 7 pages. <https://ceur-ws.org/Vol-2582/paper9.pdf>
- [42] Michael Chromik et al. 2021. I think i get your point, AI! the illusion of explanatory depth in explainable AI. In *Proc. 26th IUI*. ACM, 307–317. <https://doi.org/10.1145/3397481.3450644>
- [43] Ludovik Coba et al. 2018. Exploring Users’ Perception of Collaborative Explanation Styles. In *2018 IEEE 20th Conference on Business Informatics (CBI)*, Vol. 01. IEEE, 70–78. <https://doi.org/10.1109/CBI.2018.00017>
- [44] Ludovik Coba et al. 2019. Decision making strategies differ in the presence of collaborative explanations. In *Proc. ACM IUI*. ACM, 291–302. <https://doi.org/10.1145/3301275.3302304>
- [45] Ludovik Coba et al. 2019. Personalised novel and explainable matrix factorisation. *Data & Knowledge Engineering* 122 (2019), 142–158. <https://doi.org/10.1016/j.datak.2019.06.003>
- [46] Yiling Dai et al. 2025. What’s more important when developing math recommender systems: accuracy, explainability, or both? *Research and Practice in Technology Enhanced Learning* 21 (2025), 004. <https://doi.org/10.58459/rptel.2026.21004>
- [47] Khalil Damak et al. 2021. Sequence-Based Explainable Hybrid Song Recommendation. *Frontiers in Big Data* 4 (2021). <https://doi.org/10.3389/fdata.2021.693494>
- [48] Regina de Brito Duarte et al. 2023. AI Trust: Can Explainable AI Enhance Warranted Trust? *Human Behavior and Emerging Technologies* 2023 (2023), 1–12. <https://doi.org/10.1155/2023/4637678>
- [49] Luis M. de Campos et al. 2024. An explainable content-based approach for recommender systems: a case study in journal recommendation for paper submission. *User Model. User-Adapt. Interact.* 34, 4 (2024), 1431–1465. <https://doi.org/10.1007/s11257-024-09400-6>
- [50] Simon Delarue et al. 2024. Evaluating graph-based explanations for AI-based recommender systems. arXiv:2407.12357 [cs.AI] <https://arxiv.org/abs/2407.12357>
- [51] Juhi Dhameliya et al. 2019. Job Recommender Systems: A Survey. In *2019 Innovations in Power and Advanced Computing Technologies (i-PACT)*, Vol. 1. IEEE, 1–5. <https://doi.org/10.1109/i-PACT44901.2019.8960231>
- [52] Vicente Dominguez et al. 2018. Towards Explanations for Visual Recommender Systems of Artistic Images.. In *Intr@RecSys (CEUR Workshop Proceedings, Vol. 2225)*. CEUR-WS.org, 69–73.
- [53] Vicente Dominguez et al. 2020. Algorithmic and HCI Aspects for Explaining Recommendations of Artistic Images. *ACM Transactions on Interactive Intelligent Systems* 10, 4 (2020), 1–31. <https://doi.org/10.1145/3369396>
- [54] Tim Donkers et al. 2020. Explaining recommendations by means of aspect-based transparent memories. In *Proc. ACM IUI*. ACM, 166–176. <https://doi.org/10.1145/3377325.3377520>
- [55] Ivania Donoso-Guzmán et al. 2023. Towards a Comprehensive Human-Centred Evaluation Framework for Explainable AI. In *Explainable Artificial Intelligence*, Luca Longo (Ed.). Springer Nature Switzerland, 183–204.
- [56] Finale Doshi-Velez et al. 2017. Towards A Rigorous Science of Interpretable Machine Learning. , 13 pages. arXiv:1702.08608 [stat.ML] <https://arxiv.org/abs/1702.08608>
- [57] Fan Du et al. 2018. Personalizable and Interactive Sequence Recommender System. In *Proc. ACM CHI EA*. ACM, 1–6. <https://doi.org/10.1145/3170427.3188506>

- [58] U.M. Ene et al. 2025. Enhancing Perceived Transparency and Purchase Decisions through Explanations in Personalized Product Recommendations. *advances in multidisciplinary & scientific research journal publication* 13, 2 (2025), 53–66. <https://doi.org/10.22624/aims/math/v13n2p5>
- [59] Maryam Etemadi et al. 2023. A systematic review of healthcare recommender systems: Open issues, challenges, and techniques. *Expert Systems with Applications* 213 (2023), 118823. <https://doi.org/10.1016/j.eswa.2022.118823>
- [60] Somayeh Fatahi et al. 2023. Investigating the effectiveness of persuasive justification messages in fair music recommender systems for users with different personality traits. In *Proc. ACM UMAP*. ACM, 66–77. <https://doi.org/10.1145/3565472.3592958>
- [61] Yael Feldman-Maggor et al. 2025. The Impact of Explainable AI on Teachers' Trust and Acceptance of AI EdTech Recommendations: The Power of Domain-specific Explanations. *International Journal of Artificial Intelligence in Education* (2025). <https://doi.org/10.1007/s40593-025-00486-6>
- [62] Jibril Frej et al. 2024. Finding Paths for Explainable MOOC Recommendation: A Learner Perspective. In *Proc. ACM LAK*. ACM, 426–437. <https://doi.org/10.1145/3636555.3636898>
- [63] Phyliss Jia Gai et al. 2019. Making Recommendations More Effective Through Framings: Impacts of User- Versus Item-Based Framings on Recommendation Click-Throughs. *Journal of Marketing* 83, 6 (2019), 61–75. <https://doi.org/10.1177/0022242919873901>
- [64] Krzysztof Z. Gajos et al. 2022. Do People Engage Cognitively with AI? Impact of AI Assistance on Incidental Learning. In *27th IUI*. ACM, 794–806. <https://doi.org/10.1145/3490099.3511138>
- [65] Azin Ghazimatin et al. 2020. PRINCE: Provider-side Interpretability with Counterfactual Explanations in Recommender Systems. In *Proc. ACM WSDM*. ACM, 196–204. <https://doi.org/10.1145/3336191.3371824>
- [66] Azin Ghazimatin et al. 2021. ELIXIR: Learning from User Feedback on Explanations to Improve Recommender Models. In *Proc. ACM Web*. ACM, 3850–3860. <https://doi.org/10.1145/3442381.3449848>
- [67] Mouadh Guesmi et al. 2021. Input or Output: Effects of Explanation Focus on the Perception of Explainable Recommendation with Varying Level of Details. In *IntRS@ RecSys (CEUR Workshop Proceedings, Vol. 2948)*. CEUR-WS.org, 55–72. <https://ceur-ws.org/Vol-2948/paper4.pdf>
- [68] Mouadh Guesmi et al. 2022. Explaining User Models with Different Levels of Detail for Transparent Recommendation: A User Study. In *Adjunct Proc. ACM UMAP*. ACM, 175–183. <https://doi.org/10.1145/3511047.3537685>
- [69] Riccardo Guidotti et al. 2018. A Survey of Methods for Explaining Black Box Models. *ACM Comput. Surv.* 51, 5, Article 93 (2018), 42 pages. <https://doi.org/10.1145/3236009>
- [70] Junius Gunaratne et al. 2018. The Persuasive Power of Algorithmic and Crowdsourced Advice. *Journal of Management Information Systems* 35, 4 (2018), 1092–1120. <https://doi.org/10.1080/07421222.2018.1523534>
- [71] Lijie Guo et al. 2022. The Effect of Recommendation Source and Justification on Professional Development Recommendations for High School Teachers. In *Proc. ACM HT*. ACM, 175–185. <https://doi.org/10.1145/3511095.3531280>
- [72] Kyra Göbel et al. 2022. Explanatory machine learning for justified trust in human-AI collaboration: Experiments on file deletion recommendations. *Frontiers in Artificial Intelligence* 5 (2022). <https://doi.org/10.3389/frai.2022.919534>
- [73] Sophia Hadash et al. 2022. Improving understandability of feature contributions in model-agnostic explainable AI tools. In *Proc. ACM CHI*. ACM, Article 487, 9 pages. <https://doi.org/10.1145/3491102.3517650>
- [74] Andi Uswatun Hasanah et al. 2024. User-Centric Evaluation of Novelty and Explanation Aspects of Recommender Systems in an Indonesia E-commerce Platform Based on Perceived Usefulness. In *2024 2nd Int. Conf. ICoseIT*. IEEE, 70–75. <https://doi.org/10.1109/icoseit60086.2024.10497523>
- [75] Peter Hase et al. 2020. Evaluating Explainable AI: Which Algorithmic Explanations Help Users Predict Model Behavior?. In *Proc. 58th Annual Meeting of the Association for Computational Linguistics*. ACL, 5540–5552. <https://doi.org/10.18653/v1/2020.acl-main.491>
- [76] Lucien Heitz et al. 2023. Deliberative Diversity for News Recommendations: Operationalization and Experimental User Study. In *Proc. 17th RecSys*. ACM, 813–819. <https://doi.org/10.1145/3604915.3608834>
- [77] Rully Agus Hendrawan et al. 2024. Explanations in Open User Models for Personalized Information Exploration. In *Adjunct Proc. ACM UMAP*. ACM, 256–263. <https://doi.org/10.1145/3631700.3665188>
- [78] Stella Heras et al. 2020. Recommending Learning Objects with Arguments and Explanations. *Applied Sciences* 10, 10 (2020), 3341. <https://doi.org/10.3390/app10103341>
- [79] Diana C. Hernandez-Bocanegra et al. 2020. Effects of Argumentative Explanation Types on the Perception of Review-Based Recommendations. In *Adjunct Proc. ACM UMAP*. ACM, 219–225. <https://doi.org/10.1145/3386392.3399302>
- [80] Diana C. Hernandez-Bocanegra et al. 2020. Explaining Review-Based Recommendations: Effects of Profile Transparency, Presentation Style and User Characteristics. *i-com* 19, 3 (2020), 181–200. <https://doi.org/10.1515/icom-2020-0021>
- [81] Diana C. Hernandez-Bocanegra et al. 2021. Effects of Interactivity and Presentation on Review-Based Explanations for Recommendations. In *Human-Computer Interaction – INTERACT 2021*. Springer International Publishing, 597–618. https://doi.org/10.1007/978-3-030-85616-8_35
- [82] Diana C. Hernandez-Bocanegra and Jürgen Ziegler. 2023. Explaining Recommendations through Conversations: Dialog Model and the Effects of Interface Type and Degree of Interactivity. *ACM TiS* 13, 2 (2023), 1–47. <https://doi.org/10.1145/3579541>
- [83] Daniel Holliday et al. 2016. User Trust in Intelligent Systems: A Journey Over Time. In *Proc. ACM IUI*. ACM, 164–168. <https://doi.org/10.1145/2856767.2856811>
- [84] Minsung Hong et al. 2019. Improving Explainability of Recommendation System by Multi-sided Tensor Factorization. *Cybernetics and Systems* 50, 2 (2019), 97–117. <https://doi.org/10.1080/01969722.2019.1565119>
- [85] Zhongli Filippo Hu et al. 2023. Service-based Presentation of Multimodal Information for the Justification of Recommender Systems Results. In *Proc. ACM UMAP*. ACM, 46–53. <https://doi.org/10.1145/3565472.3592962>

- [86] Paul Hur et al. 2022. Using Machine Learning Explainability Methods to Personalize Interventions for Students. *International Educational Data Mining Society* (2022).
- [87] Maia Jacobs et al. 2021. How machine-learning recommendations influence clinician treatment selections: the example of antidepressant selection. *Translational Psychiatry* 11, 1 (2021), 108. <https://doi.org/10.1038/s41398-021-01224-x>
- [88] Michael Jugovac et al. 2018. Investigating the Decision-Making Behavior of Maximizers and Satisficers in the Presence of Recommendations. In *Proc. ACM UMAP*. ACM, 279–283. <https://doi.org/10.1145/3209219.3209252>
- [89] Marius Kaminskas et al. 2017. Item-based explanations for user-based recommendations. In *Proceedings of 9th Int. Conf. Information, Process, and Knowledge Management*. IARIA Press, 65–70.
- [90] Soultana Karga et al. 2019. Using explanations for recommender systems in learning design settings to enhance teachers' acceptance and perceived experience. *Educ Inf Technol* 24, 3 (2019), 2953–2974. <https://doi.org/10.1007/s10639-019-09909-z>
- [91] Byung Hyung Kim et al. 2020. Improved Explanatory Efficacy on Human Affect and Workload Through Interactive Process in Artificial Intelligence. *IEEE Access* 8 (2020), 189013–189024. <https://doi.org/10.1109/access.2020.3032056>
- [92] Jenia Kim et al. 2024. Human-centered evaluation of explainable AI applications: a systematic review. *Frontiers in Artificial Intelligence* 7 (2024), 22 pages. <https://doi.org/10.3389/frai.2024.1456486>
- [93] Jieyong Kim et al. 2025. Review-driven Personalized Preference Reasoning with Large Language Models for Recommendation. In *Proc. 48th International SIGIR*. ACM, 1697–1706. <https://doi.org/10.1145/3726302.3730055>
- [94] Akiva Kleinerman et al. 2018. Providing explanations for recommendations in reciprocal environments. In *Proc. ACM RecSys*. ACM, 22–30. <https://doi.org/10.1145/3240323.3240362>
- [95] Akiva Kleinerman et al. 2020. Supporting users in finding successful matches in reciprocal recommender systems. *User Model. User-Adapt. Interact.* 31, 3 (2020), 541–589. <https://doi.org/10.1007/s11257-020-09279-z>
- [96] Bart P. Knijnenburg et al. 2012. Explaining the user experience of recommender systems. *User Model. User-Adapt. Interact.* 22, 4 (2012), 441–504. <https://doi.org/10.1007/s11257-011-9118-4>
- [97] Denis Kotkov et al. 2024. On the Negative Perception of Cross-domain Recommendations and Explanations. In *Proc. ACM SIGIR*. ACM, 2102–2113. <https://doi.org/10.1145/3626772.3657735>
- [98] Pigi Kouki et al. 2017. User Preferences for Hybrid Explanations. In *Proc. 11th ACM RecSys*. ACM, 84–88. <https://doi.org/10.1145/3109859.3109915>
- [99] Pigi Kouki et al. 2020. Generating and Understanding Personalized Explanations in Hybrid Recommender Systems. *ACM TiS* 10, 4 (2020), 1–40. <https://doi.org/10.1145/3365843>
- [100] Thorsten Krause et al. 2025. The Effect of Gender De-biased Recommendations — A User Study on Gender-specific Preferences. In *Proc. ACM CHI*. ACM, 1–16. <https://doi.org/10.1145/3706598.3713155>
- [101] Johannes Kunkel et al. 2019. Let Me Explain. In *Proc. ACM CHI*. ACM, 1–12. <https://doi.org/10.1145/3290605.3300717>
- [102] Johannes Kunkel et al. 2023. A comparative study of item space visualizations for recommender systems. *Int. J. Hum.-Comput. Stud.* 172 (2023), 102987. <https://doi.org/10.1016/j.ijhcs.2022.102987>
- [103] Trung-Hoang Le et al. 2021. Explainable Recommendation with Comparative Constraints on Product Aspects. In *Proc. ACM WSDM*. ACM, 967–975. <https://doi.org/10.1145/3437963.3441754>
- [104] Benjamin Charles Germain Lee et al. 2023. LIMEADE: From AI Explanations to Advice Taking. *ACM TiS* 13, 4 (2023), 1–29. <https://doi.org/10.1145/3589345>
- [105] Yoonjoo Lee et al. 2024. PaperWeaver: Enriching Topical Paper Alerts by Contextualizing Recommended Papers with User-collected Papers. In *Proc. ACM CHI*. ACM, 1–19. <https://doi.org/10.1145/3613904.3642196>
- [106] Shuyang Li et al. 2022. Self-Supervised Bot Play for Transcript-Free Conversational Recommendation with Rationales. In *Proc. ACM RecSys*. ACM, 327–337. <https://doi.org/10.1145/3523227.3546783>
- [107] Yu Liang et al. 2021. Interactive Music Genre Exploration with Visualization and Mood Control. In *Proc. ACM IUI*. ACM, 175–185. <https://doi.org/10.1145/3397481.3450700>
- [108] Mengqi Liao. 2023. Understanding the Effects of Personalized Recommender Systems on Political News Perceptions: A Comparison of Content-Based, Collaborative, and Editorial Choice-Based News Recommender System. *Journal of Broadcasting & Electronic Media* 67, 3 (2023), 294–322. <https://doi.org/10.1080/08838151.2023.2206662>
- [109] Mengqi Liao et al. 2021. When E-Commerce Personalization Systems Show and Tell: Investigating the Relative Persuasive Appeal of Content-Based versus Collaborative Filtering. *Journal of Advertising* 51, 2 (2021), 256–267. <https://doi.org/10.1080/00913367.2021.1887013>
- [110] Mengqi Liao et al. 2022. User Trust in Recommendation Systems: A comparison of Content-Based, Collaborative and Demographic Filtering. In *Proc. ACM CHI*. ACM, Article 486, 14 pages. <https://doi.org/10.1145/3491102.3501936>
- [111] Benedikt Loepp et al. 2019. Interactive recommending with Tag-Enhanced Matrix Factorization (TagMF). *Int. J. Hum.-Comput. Stud.* 121 (2019), 21–41. <https://doi.org/10.1016/j.ijhcs.2018.05.002>
- [112] Hongyu Lu et al. 2023. User Perception of Recommendation Explanation: Are Your Explanations What Users Need? *ACM Transactions on Information Systems* 41, 2 (2023), 1–31. <https://doi.org/10.1145/3565480>
- [113] Sebastian Lubos et al. 2023. Concentrating on the impact: Consequence-based explanations in recommender systems. *arXiv preprint arXiv:2308.16708* (2023).

- [114] Sebastian Lubos et al. 2024. LLM-generated Explanations for Recommender Systems. In *Adjunct Proc. ACM UMAP*. ACM, 276–285. <https://doi.org/10.1145/3631700.3665185>
- [115] Shi-Jun Luo et al. 2023. Automatic Extraction of Effective Relations in Knowledge Graph for a Recommendation Explanation System. In *Proc. ACM SAC*. ACM, 1754–1761. <https://doi.org/10.1145/3555776.3577732>
- [116] Boxuan Ma et al. 2021. CourseQ: the impact of visual and interactive course recommendation in university environments. *Research and Practice in Technology Enhanced Learning* 16, 1 (2021), 18. <https://doi.org/10.1186/s41039-021-00167-7>
- [117] Maria Madsen et al. 2000. Measuring human-computer trust. In *11th australasian conference on information systems*, Vol. 53. Citeseer, 6–8.
- [118] Ayoub El Majjodi et al. 2022. Nudging Towards Health? Examining the Merits of Nutrition Labels and Personalization in a Recipe Recommender System. In *Proc. ACM UMAP*. ACM, 48–56. <https://doi.org/10.1145/3503252.3531312>
- [119] Rwitajit Majumdar et al. 2023. Learning with Explainable AI-Recommendations at School: Extracting Patterns of Self-Directed Learning from Learning Logs. In *2023 IEEE ICALT*. IEEE, 245–249. <https://doi.org/10.1109/icalt58122.2023.00078>
- [120] Avleen Malhi et al. 2020. Explainable Agents for Less Bias in Human-Agent Decision Making. In *LNCS*. Springer International Publishing, 129–146. https://doi.org/10.1007/978-3-030-51924-7_8
- [121] Millecamp Martijn et al. 2022. “Knowing me, knowing you”: personalized explanations for a music recommender system. *User Model. User-Adapt. Interact.* 32, 1-2 (2022), 215–252. <https://doi.org/10.1007/s11257-021-09304-9>
- [122] Noemi Mauro et al. 2022. Justification of recommender systems results: a service-based approach. *User Model. User-Adapt. Interact.* 33, 3 (2022), 643–685. <https://doi.org/10.1007/s11257-022-09345-8>
- [123] Christian Meske et al. 2022. Explainable Artificial Intelligence: Objectives, Stakeholders, and Future Research Opportunities. *Information Systems Management* 39, 1 (2022), 53–63. <https://doi.org/10.1080/10580530.2020.1849465>
- [124] Martijn Millecamp et al. 2019. To explain or not to explain: the effects of personal characteristics when explaining music recommendations. In *Proc. ACM IUI*. ACM, 397–407. <https://doi.org/10.1145/3301275.3302313>
- [125] Martijn Millecamp et al. 2020. What’s in a User? Towards Personalising Transparency for Music Recommender Interfaces. In *Proc. ACM UMAP*. ACM, 173–182. <https://doi.org/10.1145/3340631.3394844>
- [126] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial intelligence* 267 (2019), 1–38.
- [127] Tim Miller. 2023. Explainable AI is Dead, Long Live Explainable AI! Hypothesis-driven Decision Support using Evaluative AI. In *Proc. ACM FAccT*. ACM, 333–342. <https://doi.org/10.1145/3593013.3594001>
- [128] Aditi Mishra et al. 2022. Why? Why not? When? Visual Explanations of Agent Behaviour in Reinforcement Learning. In *2022 IEEE 15th Pacific Visualization Symposium (PacificVis)*. IEEE, 111–120. <https://doi.org/10.1109/PacificVis53943.2022.00020>
- [129] Sina Mohseni et al. 2021. A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM TiS* 11, 3-4 (2021), 1–45.
- [130] Geoff Musick et al. 2023. Knowing Unknown Teammates: Exploring Anonymity and Explanations in a Teammate Information-Sharing Recommender System. *Proc. ACM on Human-Computer Interaction* 7, CSCW2 (2023), 1–34. <https://doi.org/10.1145/3610075>
- [131] Cataldo Musto et al. 2019. Linked open data-based explanations for transparent recommender systems. *Int. J. Hum.-Comput. Stud.* 121 (2019), 93–107. <https://doi.org/10.1016/j.ijhcs.2018.03.003>
- [132] Cataldo Musto et al. 2021. Exploring the Effects of Natural Language Justifications in Food Recommender Systems. In *Proc. 29th ACM UMAP*. ACM, 147–157. <https://doi.org/10.1145/3450613.3456827>
- [133] Cataldo Musto et al. 2023. Harnessing distributional semantics to build context-aware justifications for recommender systems. *User Model. User-Adapt. Interact.* 34, 3 (2023), 659–690. <https://doi.org/10.1007/s11257-023-09382-x>
- [134] Mohammad Naiseh et al. 2021. Nudging through Friction: An Approach for Calibrating Trust in Explainable AI. In *Proc. Int. Conf. BESC*. IEEE, 1–5. <https://doi.org/10.1109/BESC53957.2021.9635271>
- [135] Shabnam Najafian et al. 2021. Exploring User Concerns about Disclosing Location and Emotion Information in Group Recommendations. In *Proc. ACM HT*. ACM, 155–164. <https://doi.org/10.1145/3465336.3475104>
- [136] Shabnam Najafian et al. 2021. Factors Influencing Privacy Concern for Explanations of Group Recommendation. In *Proc. ACM UMAP*. ACM, 14–23. <https://doi.org/10.1145/3450613.3456845>
- [137] Meike Nauta et al. 2023. From Anecdotal Evidence to Quantitative Evaluation Methods: A Systematic Review on Evaluating Explainable AI. *ACM Comput. Surv.* 55, 13s, Article 295 (2023), 42 pages. <https://doi.org/10.1145/3583558>
- [138] Sidra Naveed et al. 2018. Argumentation-Based Explanations in Recommender Systems. In *Adjunct Proc. ACM UMAP*. ACM, 293–298. <https://doi.org/10.1145/3213586.3225240>
- [139] Sidra Naveed et al. 2020. Featuristic: An interactive hybrid system for generating explainable recommendations-beyond system accuracy.. In *IntRS@ RecSys (CEUR Workshop Proceedings, Vol. 2682)*. CEUR-WS.org, 14–25.
- [140] Sidra Naveed et al. 2020. On the Use of Feature-based Collaborative Explanations: An Empirical Comparison of Explanation Styles. In *Adjunct Proc. ACM UMAP*. ACM, 226–232. <https://doi.org/10.1145/3386392.3399303>
- [141] Shreeya Nelekar et al. 2021. Effectiveness of embodied conversational agents for managing academic stress at an Indian University (ARU) during COVID-19. *British Journal of Educational Technology* 53, 3 (2021), 491–511. <https://doi.org/10.1111/bjet.13174>
- [142] Fakhroddin Noorbehhahani et al. 2018. The Impact of Demographic Factors on Persuasion Strategies in Personalized Recommender System. In *2018 8th Int. Conf. Computer and Knowledge Engineering (ICCKE)*. IEEE, 104–109. <https://doi.org/10.1109/ICCKE.2018.8566550>

- [143] Ingrid Nunes et al. 2017. A systematic review and taxonomy of explanations in decision support and recommender systems. *User Model. User-Adapt. Interact.* 27, 3 (2017), 393–444.
- [144] Ayano Okoso et al. 2025. Impact of Tone-Aware Explanations in Recommender Systems. *ACM Transactions on Recommender Systems* 3, 4 (2025), 1–34. <https://doi.org/10.1145/3718101>
- [145] Jeroen Ooge et al. 2022. Explaining Recommendations in E-Learning: Effects on Adolescents' Trust. In *Proc. 27th IUI*. ACM, 93–105. <https://doi.org/10.1145/3490099.3511140>
- [146] Cecilia Panigutti et al. 2022. Understanding the impact of explanations on advice-taking: a user study for AI-based clinical Decision Support Systems. In *Proc. ACM CHI*. ACM, Article 568, 9 pages. <https://doi.org/10.1145/3491102.3502104>
- [147] Alexis Papadimitriou et al. 2012. A generalized taxonomy of explanations styles for traditional and social recommender systems. *Data Mining and Knowledge Discovery* 24 (2012), 555–583.
- [148] Florian Pecune et al. 2019. A Model of Social Explanations for a Conversational Movie Recommendation System. In *Proc. ACM HAI*. ACM, 135–143. <https://doi.org/10.1145/3349537.3351899>
- [149] Florian Pecune et al. 2022. Designing Persuasive Food Conversational Recommender Systems With Nudging and Socially-Aware Conversational Strategies. *Frontiers in Robotics and AI* 8 (2022), 22 pages. <https://doi.org/10.3389/frobt.2021.733835>
- [150] Gustavo Penha et al. 2022. Pairwise Review-Based Explanations for Voice Product Search. In *Proc. ACM CHIIR*. ACM, 300–304. <https://doi.org/10.1145/3498366.3505828>
- [151] Luiz Pizzato et al. 2013. Recommending people to people: the nature of reciprocal recommenders with a case study in online dating. *User Model. User-Adapt. Interact.* 23, 5 (2013), 447–488. <https://doi.org/10.1007/s11257-012-9125-0>
- [152] Pearl Pu et al. 2012. Evaluating recommender systems from the user's perspective: survey of the state of the art. *User Model. User-Adapt. Interact.* 22 (2012), 317–355.
- [153] Lara Quijano-Sanchez et al. 2017. Make it personal: A social explanation system applied to group recommendations. *Expert Systems with Applications* 76 (2017), 36–48. <https://doi.org/10.1016/j.eswa.2017.01.045>
- [154] Marissa Radensky et al. 2022. Exploring the Role of Local and Global Explanations in Recommender Systems. In *Proc. ACM CHI EA*. ACM, Article 290. <https://doi.org/10.1145/3491101.3519795>
- [155] Antonio Rago et al. 2020. Argumentation as a Framework for Interactive Explanations for Recommendations. In *Proc. Int. Conf. KR*. International Joint Conferences on Artificial Intelligence Organization, 805–815. <https://doi.org/10.24963/kr.2020/83>
- [156] Arpit Rana et al. 2018. Explanations that are Intrinsic to Recommendations. In *Proc. ACM UMAP*. ACM, 187–195. <https://doi.org/10.1145/3209219.3209230>
- [157] Arpit Rana et al. 2022. Extended recommendation-by-explanation. *User Model. User-Adapt. Interact.* 32, 1-2 (2022), 91–131. <https://doi.org/10.1007/s11257-021-09317-4>
- [158] Neha Rani et al. 2023. Explanation for User Trust in Context-Aware Recommender Systems for Search-As-Learning. In *2023 IEEE ICALT*. IEEE, 47–49. <https://doi.org/10.1109/icalt58122.2023.00019>
- [159] Shaina Raza et al. 2022. News recommender system: a review of recent progress, challenges, and opportunities. *Artificial Intelligence Review* 55, 1 (2022), 749–800. <https://doi.org/10.1007/s10462-021-10043-x>
- [160] Yao Rong et al. 2024. Towards Human-Centered Explainable AI: A Survey of User Studies for Model Explanations. *IEEE Trans. Pattern Anal. Mach. Intell.* 46, 4 (2024), 2104–2122. <https://doi.org/10.1109/TPAMI.2023.3331846>
- [161] Masahiro Sato et al. 2018. Explaining Recommendations Using Contexts. In *Proc. ACM IUI*. ACM, 659–664. <https://doi.org/10.1145/3172944.3173012>
- [162] Masahiro Sato et al. 2019. Action-Triggering Recommenders: Uplift Optimization and Persuasive Explanation. In *Proc. ICDMW*. IEEE, 1060–1069. <https://doi.org/10.1109/ICDMW.2019.00155>
- [163] Roan Schellingerhout et al. 2024. Creating healthy friction: determining stakeholder requirements of job recommendation explanations. In *Proc. 4th Workshop on Recommender Systems for Human Resources, HR 2024, co-located with ACM RecSys (RecSys 2024), Bari, Italy (arXiv preprint arXiv:2409.15971)*.
- [164] Donghee Shin et al. 2022. In Platforms We Trust? Unlocking the Black-Box of News Algorithms through Interpretable AI. *Journal of Broadcasting & Electronic Media* 66, 2 (2022), 235–256. <https://doi.org/10.1080/08838151.2022.2057984>
- [165] Dorin Shmaryahu et al. 2020. Post-hoc Explanations for Complex Model Recommendations using Simple Methods.. In *Intr@ RecSys (CEUR Workshop Proceedings, Vol. 2682)*. CEUR-WS.org, 26–36. <https://ceur-ws.org/Vol-2682/paper3.pdf>
- [166] Avital Shulner-Tal et al. 2022. Fairness, explainability and in-between: understanding the impact of different explanation methods on non-expert users' perceptions of fairness toward an algorithmic system. *Ethics and Information Technology* 24, 1 (2022), 2. <https://doi.org/10.1007/s10676-022-09623-4>
- [167] Ítallo Silva et al. 2024. Leveraging ChatGPT for Automated Human-centered Explanations in Recommender Systems. In *Proc. 29th IUI*. ACM, 597–608. <https://doi.org/10.1145/3640543.3645171>
- [168] Alain Starke. 2019. The effectiveness of advice solicitation and social peers in an energy recommender system, In *Intr@ RecSys. CEUR Workshop Proceedings* 2450, 65 – 71.
- [169] Alain Starke et al. 2021. Promoting Energy-Efficient Behavior by Depicting Social Norms in a Recommender Interface. *ACM Transactions on Interactive Intelligent Systems* 11, 3-4 (2021), 1–32. <https://doi.org/10.1145/3460005>

- [170] Alain Starke et al. 2021. “Serving Each User”: Supporting Different Eating Goals Through a Multi-List Recommender Interface. In *Proc. ACM RecSys*. ACM, 124–132. <https://doi.org/10.1145/3460231.3474232>
- [171] Emily Sullivan et al. 2019. Reading News with a Purpose. In *Adjunct Proc. UMAP*. ACM, 241–245. <https://doi.org/10.1145/3314183.3323456>
- [172] Yuan Sun et al. 2022. Exploring the Effects of Interactive Dialogue in Improving User Control for Explainable Online Symptom Checkers. In *Proc. ACM CHI EA*. ACM, Article 417. <https://doi.org/10.1145/3491101.3519668>
- [173] Zhu Sun et al. 2024. Revisiting Bundle Recommendation for Intent-aware Product Bundling. *ACM Transactions on Recommender Systems* 2, 3 (2024), 1–34. <https://doi.org/10.1145/3652865>
- [174] Panagiotis Symeonidis et al. 2009. MoviExplain: a recommender system with explanations. In *Proc. 3rd RecSys*. ACM, 317–320. <https://doi.org/10.1145/1639714.1639777>
- [175] Kyosuke Takami et al. 2022. Educational Explainable Recommender Usage and its Effectiveness in High School Summer Vacation Assignment. In *LAK22: 12th International Learning Analytics and Knowledge Conference*. ACM, 458–464. <https://doi.org/10.1145/3506860.3506882>
- [176] Kyosuke Takami et al. 2023. Personality-based tailored explainable recommendation for trustworthy smart learning system in the age of artificial intelligence. *Smart Learning Environments* 10, 1 (2023). <https://doi.org/10.1186/s40561-023-00282-6>
- [177] Kyosuke Takami et al. 2024. Evaluating the Effectiveness of Bayesian Knowledge Tracing Model-Based Explainable Recommender. *International Journal of Distance Education Technologies* 22, 1 (2024), 1–23. <https://doi.org/10.4018/ijdet.337600>
- [178] Kensuke Takii et al. 2024. Explainable eBook recommendation for extensive reading in K-12 EFL learning. *Research and Practice in Technology Enhanced Learning* 20 (2024), 027. <https://doi.org/10.58459/rptel.2025.20027>
- [179] Yiyi Tao et al. 2019. The FacT. In *Proc. ACM SIGIR*. ACM, 295–304. <https://doi.org/10.1145/3331184.3331244>
- [180] Serge Thill et al. 2018. Driver adherence to recommendations from support systems improves if the systems explain why they are given: A simulator study. *Transportation Research Part F: Traffic Psychology and Behaviour* 56 (2018), 420–435. <https://doi.org/10.1016/j.trf.2018.05.009>
- [181] Nava Tintarev et al. 2007. A Survey of Explanations in Recommender Systems. In *IEEE ICDE Workshop*. IEEE, 801–810. <https://doi.org/10.1109/ICDEW.2007.4401070>
- [182] Nava Tintarev et al. 2018. Knowing the unknown. In *Proc. ACM SAC*. ACM, 1396–1399. <https://doi.org/10.1145/3167132.3167419>
- [183] Suzanne Tolmeijer et al. 2021. Second Chance for a First Impression? Trust Development in Intelligent System Interaction. In *Proc. ACM UMAP*. ACM, 77–87. <https://doi.org/10.1145/3450613.3456817>
- [184] Helma Torkamaan et al. 2019. Rating-based Preference Elicitation for Recommendation of Stress Intervention. In *Proc. ACM UMAP*. ACM, 46–50. <https://doi.org/10.1145/3320435.3324990>
- [185] Thi Ngoc Trang Tran et al. 2020. Social Choice-based Explanations: An Approach to Enhancing Fairness and Consensus Aspects. *JUCS - Journal of Universal Computer Science* 26, 3 (2020), 402–431. <https://doi.org/10.3897/jucs.2020.021>
- [186] Thi Ngoc Trang Tran et al. 2021. Do Users Appreciate Explanations of Recommendations? An Analysis in the Movie Domain. In *Proc. ACM RecSys*. ACM, 645–650. <https://doi.org/10.1145/3460231.3478859>
- [187] Thi Ngoc Trang Tran et al. 2023. User Needs for Explanations of Recommendations: In-depth Analyses of the Role of Item Domain and Personal Characteristics. In *Proc. ACM UMAP*. ACM, 54–65. <https://doi.org/10.1145/3565472.3592950>
- [188] Chun-Hua Tsai et al. 2019. Evaluating Visual Explanations for Similarity-Based Recommendations. In *Proc. ACM UMAP*. ACM, 22–30. <https://doi.org/10.1145/3320435.3320465>
- [189] Chun-Hua Tsai et al. 2020. The effects of controllability and explainability in a social recommender system. *User Model. User-Adapt. Interact.* 31, 3 (2020), 591–627. <https://doi.org/10.1007/s11257-020-09281-5>
- [190] Chun-Hua Tsai et al. 2021. Exploring and Promoting Diagnostic Transparency and Explainability in Online Symptom Checkers. In *Proc. ACM CHI*. ACM. <https://doi.org/10.1145/3411764.3445101>
- [191] Kosetsu Tsukuda et al. 2020. Explainable Recommendation for Repeat Consumption. In *Proc. ACM RecSys*. ACM, 462–467. <https://doi.org/10.1145/3383313.3412230>
- [192] Jesse Vig et al. 2009. Tagsplanations: explaining recommendations using tags. In *Proc. 14th IUI*. ACM, 47–56. <https://doi.org/10.1145/1502650.1502661>
- [193] Sergey Volokhin et al. 2022. Generating and Validating Contextually Relevant Justifications for Conversational Recommendation. In *ACM SIGIR Conference on Human Information Interaction and Retrieval*. ACM, 284–289. <https://doi.org/10.1145/3498366.3505789>
- [194] Alexandra Vultureanu-Albiş et al. 2022. A survey on effects of adding explanations to recommender systems. *Concurrency and Computation: Practice and Experience* 34, 20 (2022), 15 pages. <https://doi.org/10.1002/cpe.6834>
- [195] Clarice Wang et al. 2023. When Biased Humans Meet Debaised AI: A Case Study in College Major Recommendation. *ACM TiS* 13, 3 (2023), 1–28. <https://doi.org/10.1145/3611313>
- [196] GaoShan Wang et al. 2020. Research on the influence of interpretability of artificial intelligence recommendation system on users' behavior intention. In *Proc. ACM EITCE*. ACM, 762–766. <https://doi.org/10.1145/3443467.3443850>
- [197] Nan Wang et al. 2018. Explainable Recommendation via Multi-Task Learning in Opinionated Text Data. In *Proc. SIGIR*. ACM, 165–174. <https://doi.org/10.1145/3209978.3210010>
- [198] Xinru Wang et al. 2025. Less or More: Towards Glanceable Explanations for LLM Recommendations Using Ultra-Small Devices. In *Proc. ACM IUI*. ACM, 938–951. <https://doi.org/10.1145/3708359.3712074>
- [199] Zongyi Wang et al. 2023. An Industrial Framework for Personalized Serendipitous Recommendation in E-commerce. In *Proc. 17th RecSys*. ACM, 1015–1018. <https://doi.org/10.1145/3604915.3610234>

- [200] Kathrin Wardatzky et al. 2025. Whom do Explanations Serve? A Systematic Literature Survey of User Characteristics in Explainable Recommender Systems Evaluation. *ACM RecSys* 3, 4, Article 49 (2025), 35 pages. <https://doi.org/10.1145/3716394>
- [201] Kathrin Wardatzky et al. 2026. The Utility of LLMs in Recommender Systems Explanation Evaluation. In *Proc. 20th RecSys* (Minneapolis, MN, USA). ACM, New York, NY, USA, To appear. <https://doi.org/10.1145/3773078.3831835>
- [202] Cedric Waterschoot et al. 2025. With Friends Like These, Who Needs Explanations? Evaluating User Understanding of Group Recommendations. In *Proc. ACM UMAP*. ACM, 253–262. <https://doi.org/10.1145/3699682.3728345>
- [203] Helena Weith et al. 2023. Information provision measures for voice agent product recommendations— The effect of process explanations and process visualizations on fairness perceptions. *Electronic Markets* 33, 1 (2023). <https://doi.org/10.1007/s12525-023-00668-x>
- [204] Agung Toto Wibowo et al. 2018. Understanding how to Explain Package Recommendations in the Clothes Domain.. In *IntRS@ RecSys (CEUR Workshop Proceedings, Vol. 2225)*. CEUR-WS.org, 74–78. <https://ceur-ws.org/Vol-2225/paper11.pdf>
- [205] Darcia Wilkinson et al. 2021. Why or Why Not? The Effect of Justification Styles on Chatbot Recommendations. *ACM Transactions on Information Systems* 39, 4 (2021), 1–21. <https://doi.org/10.1145/3441715>
- [206] Peng Xia et al. 2016. Design of reciprocal recommendation systems for online dating. *Social Network Analysis and Mining* 6, 1 (2016), 32. <https://doi.org/10.1007/s13278-016-0340-2>
- [207] Xuhai Xu et al. 2020. Understanding User Behavior For Document Recommendation. In *Proc. ACM Web*. ACM, 3012–3018. <https://doi.org/10.1145/3366423.3380071>
- [208] Aobo Yang et al. 2021. Explanation as a Defense of Recommendation. In *Proc. ACM WSDM*. ACM, 1029–1037. <https://doi.org/10.1145/3437963.3441726>
- [209] Aobo Yang et al. 2022. Comparative Explanations of Recommendations. In *Proc. ACM Web*. ACM, 3113–3123. <https://doi.org/10.1145/3485447.3512031>
- [210] Po-Kai Yang et al. 2022. Touching the Explanations: Explaining Movie Recommendation Scores in Mobile Augmented Reality. In *2022 IEEE AIVR*. IEEE, 157–162. <https://doi.org/10.1109/aivr56993.2022.00032>
- [211] Zekun Yang et al. 2021. Interpretable video tag recommendation with multimedia deep learning framework. *Internet Research* 32, 2 (2021), 518–535. <https://doi.org/10.1108/intr-08-2020-0471>
- [212] Jing Yao et al. 2025. Neural Recommendation Reasoning with Logic Rules. *ACM Transactions on Information Systems* 43, 5 (2025), 1–28. <https://doi.org/10.1145/3742856>
- [213] Run Yu et al. 2021. Orienting Students to Course Recommendations Using Three Types of Explanation. In *Adjunct Proc. ACM UMAP*. ACM, 238–245. <https://doi.org/10.1145/3450614.3464483>
- [214] Jingjing Zhang et al. 2017. Exploring Explanation Effects on Consumers’ Trust in Online Recommender Agents. *International Journal of Human–Computer Interaction* 34, 5 (2017), 421–432. <https://doi.org/10.1080/10447318.2017.1357904>
- [215] Yongfeng Zhang et al. 2020. Explainable recommendation: A survey and new perspectives. *Foundations and Trends® in Information Retrieval* 14, 1 (2020), 1–101.
- [216] Zhirun Zhang et al. 2022. Effects of Feature-Based Explanation and Its Output Modality on User Satisfaction With Service Recommender Systems. *Frontiers in Big Data* 5 (2022), 17 pages. <https://doi.org/10.3389/fdata.2022.897381>
- [217] Yong Zheng et al. 2021. The role of transparency in multi-stakeholder educational recommendations. *User Model. User-Adapt. Interact.* 31, 3 (2021), 513–540. <https://doi.org/10.1007/s11257-021-09291-x>
- [218] Jinfeng Zhou et al. 2021. CRFR: Improving Conversational Recommender Systems via Flexible Fragments Reasoning on Knowledge Graphs. In *Proc. 2021 EMNLP. ACL*, 4324–4334. <https://doi.org/10.18653/v1/2021.emnlp-main.355>
- [219] Jianlong Zhou et al. 2021. Evaluating the quality of machine learning explanations: A survey on methods and metrics. *Electronics* 10, 5 (2021), 593.
- [220] Jihua Zhu et al. 2022. Joint Reason Generation and Rating Prediction for Explainable Recommendation. *IEEE TKDE* (2022), 1–1. <https://doi.org/10.1109/tkde.2022.3146178>
- [221] Robert Zimmermann et al. 2022. Enhancing brick-and-mortar store shopping experience with an augmented reality shopping assistant application using personalized recommendations and explainable artificial intelligence. *Journal of Research in Interactive Marketing* 17, 2 (2022), 273–298. <https://doi.org/10.1108/jrim-09-2021-0237>