Comparative Explanations for Recommendation: Research Directions

Meysam Varasteh[1,*], Elizabeth McKinnie[2], Amanda Aird[2], Daniel Acuña[1] and Robin Burke[2]

[1]Department of Computer Science, University of Colorado, Boulder, USA
[2]Department of Information Science, University of Colorado, Boulder, USA

Abstract

Explanations have a long history in recommender systems. Researchers have studied the different roles explanations can play, the value of explanations for users, and different techniques for generating explanations for a given output. To date, we have rarely seen recommender systems make use of comparative explanations, a technique that social scientists emphasize as important in human explanatory behavior. We believe that comparative explanation could be a very powerful tool to augment explanations that recommender systems currently provide and to offer new types of transparency. In this paper, we provide a taxonomy of different types of comparative explanations for recommender systems, emphasizing in particular the potential value of comparative explanations for recommender system providers. We suggest directions for future research to realize this potential rather than providing solutions.

Keywords

recommender systems, explanation, multistakeholder recommendation

1. Introduction

In an extensive survey of the social science literature, the psychologist Tim Miller concluded that when people ask "Why P?" questions, they are typically asking "Why P rather than Q?," where Q is often implicit in the context [1]. An explanation that answers such a question Miller terms as *constrastive*. That term has taken on a specific meaning in fair machine learning and explainable AI (XAI), so we will use the synonym *comparative* to mean a general strategy of using comparison as a means of creating explanations. We believe that comparative explanations can provide greater insight into recommender system operation and help engender user trust. In this paper, we examine the little-studied realm of comparative explanations for recommender systems and provide a taxonomy of different types of such explanations. We also propose comparative explanations as a useful component of interfaces for providers, stakeholders who are rarely considered in designing recommender system interfaces.

Following Miller, we will define comparative explanation in recommendation as comparing two different (actual or possible) outcomes from a recommender system with the aim of providing greater transparency into the system's operation. Note that this definition excludes comparing a recommended item to an item already rated by the consumer. We examine comparative explanations by considering two aspects of recommendation explanations:

- **Audience**: To whom is the explanation being delivered? Recommender systems are best understood as multistakeholder applications [2] and explanations delivered to different stakeholder audiences will require different approaches. There are the three main stakeholders of any recommender system [2]: consumer, provider, and system. We concentrate on consumer and provider, and note that very little attention has been paid to provider-side explanations in recommender systems research.

IntRS'24: Joint Workshop on Interfaces and Human Decision Making for Recommender Systems, October 18, 2024, Bari (Italy)
meysam.varasteh@colorado.edu (M. Varasteh); elizabeth.mckinnie@colorado.edu (E. McKinnie); amanda.aird@colorado.edu (A. Aird); daniel.acuna@colorado.edu (D. Acuña); robin.burke@colorado.edu (R. Burke)
0009-0003-0346-4951 (M. Varasteh); 0009-0002-8721-5700 (E. McKinnie); 0009-0002-0348-5843 (A. Aird); 0000-0002-7765-1595 (D. Acuña); 0000-0001-5766-6434 (R. Burke)

- **Scope**: Comparative explanations will differ in form depending on whether we are comparing recommendations of individual items or the general behavior of the recommender system over time or over some set of items.

Because this paper is interested in comparative explanation and because we are taking a multistakeholder approach, there are a wide variety of different scopes to be considered, most of which have not seen any research attention. For our purposes in this paper, we consider scope as a two-level construct. First, there is the question of what aspect of recommendation output is being compared. Possible answers include:

- The presence of an **item** in a recommendation list,
- The **rank** of an item in a recommendation list, or
- A **pattern** of recommendation that spans multiple items in a list or multiple lists.

The next aspect of scope for comparative explanation is the question of the source of the comparison. We can compare:

- Recommendations given to a single recommendation **consumer**,
- Recommendations given to two different consumers: **cross-consumer** comparison, or
- Recommendations of items from a given **provider**.[1]

Finally, we have the question of audience. Not all types of explanation are appropriate for all audiences. There may be privacy or business reasons why not all types of explanations should be available to all audiences.[2] Table 1 illustrates the space of comparative explanations using a letter code to uniquely identify each possibility. We use dashes to indicate those explanations that are unlikely to be used in real systems because of potential violations of users' privacy or confidentiality expectations.

For example, consider an item-oriented cross-consumer explanation being delivered to an individual, the missing `iccC` cell in the table. Such an explanation would answer a question like "Why is User A being recommended Item X and I am not?". There are several reasons to doubt whether such an explanation could or should be part of an explainer's output. First, this question assumes that the user knows what is being recommended to someone else, and in addition, any answer to this question would by necessity divulge information about User A's profile. Similar considerations exist about divulging user information to providers or cross-provider information, which might be considered confidential business data in an e-commerce setting. In this paper, we focus on explanation types that would not create such risks to confidentiality.

There are nine types of comparative explanations that we consider. Examples of the questions such explanations would respond to are shown here:

- **Consumer-oriented questions**
  - Why is Item A being recommended to me instead of Item B? (icC)
  - Why is Item A being ranked ahead of Item B in my recommendations? (rcC)
  - Why are items of type A being recommended to me more than items of type B? (tcC)
- **Provider-oriented questions**
  - Why is my Item A being recommended to users of type X and not users of type Y? (iccP)
  - Why is my Item A being recommended and my Item B is not being recommended? (ipP)
  - Why is my Item A being ranked higher to users of type X than users of type Y? (rccP)
  - Why is my Item A being ranked higher than my Item B in recommendation lists? (rpP)

[1] It is possible to have comparative explanations across providers, but these would likely violate providers' privacy and so would be accessible only to system stakeholders.

[2] As noted above, we are setting aside the System stakeholder for future work. Providing explanations to this set of users amounts to building a dashboard for examining all aspects of recommendation outputs. This is a worthy task, but we are more interested in users who would not be experts in the design and operation of the system.

		Source of Comparison	Audience	
			Consumer (C)	Provider (P)
Scope	Item (i)	Consumer (c)	icC	—
		Cross-Consumer (cc)	—	iccP
		Provider (p)	—	ipP
	Rank (r)	Consumer	rcC	—
		Cross-Consumer	—	rccP
		Provider	—	rpP
	Pattern (t)	Consumer	tcC	—
		Cross-Consumer	—	tccP
		Provider	—	tpP

Table 1
Different types of comparative explanations. Blank entries indicate explanations that would be likely to violate privacy or confidentiality.

- Why are my items of type A being recommended to users of type X and not users of type Y? (tccP)
- Why are my items of type A being recommended more often than my items of type B? (tpP)

In the rest of this paper, we will examine in detail hypothetical scenarios that illustrate each type of comparative explanation, discuss the benefit of such explanations, and consider what would be required to generate appropriate explanations in response to these hypothetical questions. Our scenarios are non-exhaustive and created to illustrate our comparative explanation taxonomy.

2. Related Work

This paper brings together ideas from explanation in recommender systems, multistakeholder recommendation, and comparative / constrastive explanation from explainable AI.

2.1. Explanation in Recommender Systems

Explanation has a long history in recommender systems research, starting with the seminal work of Herlocker et al. [3] and summarized more recently in the survey in [4]. This research trajectory has concentrated entirely on consumers as the explanation audience and with very limited exceptions, has concentrated on explanations for individual recommended items. As Tintarev and Masthoff discuss in [4], research has explored additional types of explanations meant to place recommendations in context or provide users with support in understanding their own goals and consumption behavior. Comparative explanations could serve these roles as well.

The comparative explanations in recommendation explored in [5, 6, 7] fall under a slightly different framework than that explored here. These works aim to justify a single recommendation by comparing it to items the user has already rated. In these cases, comparison serves to provide a known anchor point against which a recommendation is evaluated and draws from sentences extracted from reviews [5] or recipes [6, 7] as the underlying explanatory text. Our vision of comparative explanation involves comparing two possible recommender system outputs in line with Miller's concept, not comparing user input with user output.

2.2. Contrastive Explanation in XAI

An explanation can be considered an answer to the question, "Why P?" where "P" represents the explicit event that occurred and needs to be explained. However, studies in social science and philosophy show that "why" questions are often more complex than this straightforward approach [8, 9, 10]. These studies indicate that such questions go beyond the occurrence of event P and expect the explanation to address more than just a single event.

This raises an important question in the field of Explainable AI (XAI): "What constitutes a good explanation?" To address this, Miller in [1] proposes three key criteria for effective explanations. First, explanations should be contrastive, meaning they should explain why a particular input yields a specific output rather than an alternative output. Second, explanations should be selective, presenting only the relevant information and avoiding the inclusion of all possible causes to reduce cognitive load for both the explainer and the explainee. Lastly, explanations should be social, recognizing that they are a form of communication between the explainer and the explainee.

Contrastive explanation has been studied in image classification and other areas of AI. For image classification, a contrastive explanation could be framed as follows: a classifier predicts label Y for input X because, if X were slightly modified to X_c, the classifier would instead predict label Y_c where X can be important object pixels in an image [11, 12, 13]. [11] proposes a contrastive explanation method for explaining image black-box classifiers. This method identifies the minimal set of pixels that must be present in an image to justify its classification and the minimal set of pixels that must be absent to distinguish it from a similar input close to the original. Note that the contrastive aspect is between the actual image and a hypothetical alternative that would be classified differently.

In text generation settings, where the output is a sequence of words rather than a single label, the explanation might be framed as the LLM outputs a reply to a given prompt because if the prompt was slightly modified, the LLM would have given a different response [14, 15, 16, 17].

Jacovi et al. [18] introduce a comparative explanation method for model interpretability by generating a contrastive latent representation. This method projects the latent representation of the input space into a new space that distinguishes between two different decisions made by the model. They use an interventionist approach [19] to determine the causality of a factor by intervening on it, thus generating a counterfactual. In all of these contrastive explanation studies, although the questions do not directly compare two types of events, they explain the event in a comparative manner for the purpose of justification.

2.3. Provider Perspectives

Although the multistakeholder perspective on recommender systems has achieved recognition in recent research [2], there is little research specifically on how recommender systems should interface with item providers. Recommender systems designers have at times considered providers' perspectives as irrelevant to their efforts to find appropriate content for users, leading them to concentrate on ways to defeat provider manipulation of ranking systems [20]. The general problem of recommender system transparency for providers has been approached in qualitative work seeking provider perspectives. For example, content creators and dating app users identified transparency as an important part of fairness in [21], creating transparency metrics that expose matching mechanisms to users or discussing content revoking reasoning and "what factors into a successful post". Music artists mostly agreed that they wanted more transparency in [22] and [23], with one artist specifically stating a desire for the system to describe what artists should do to be recommended more often [23]. YouTube creators also discussed transparency, particularly how difficult it is to understand the relationship between creative choices and the impact of their videos, and wanted to know how the algorithm operates [24]. An example scenario that the creators generated during one of the workshops was that the algorithm can explain that your video is not doing well due to how long it is [24].

3. Consumer-Oriented Comparative Explanations

As noted above, explanation in recommender systems has historically been directed towards consumers. Providing consumer-oriented explanations can build trust and satisfaction among users by helping them understand why certain recommendations are being made [25, 4]. This section focuses on comparative explanations for consumers examining different kinds of comparative explanations by providing hypothetical examples of questions and explanations. As we note above, the set of comparisons

that we expect users to seek will be focused on their own recommendation results. Other types of comparisons may violate privacy or confidentiality.

In demonstrating these explanation types, we will rely on hypothetical scenarios from two very different domains: music streaming recommendation (an application oriented towards a general audience) and the recommendation of scientific literature (an application targeted towards specialists). Also note that music streaming is more asymmetrical (with consumers less likely to have provider roles as well) whereas the consumers of recommendations about scientific papers are likely to be individuals who also publish such papers.

3.1. Comparing Individual Items (icC)

Question template: Why is Item A being recommended to me instead of Item B?

The first type of comparative explanation for a single consumer involves comparing two different individual items when one item was explicitly recommended and the other item was not. In Miller's terminology, the first item is referred to as the *fact*, and the second as the *foil* [1].

Consider the following hypothetical example: Sarah is a frequent user of the *Tunester* platform, which she uses to listen to music. The platform uses collaborative filtering to recommend music to users by finding similar users (neighbors) and recommending songs that those neighbors like. Each time it recommends 10 songs in descending order based on the predicted score for each track. Sarah received a recommendation list from the platform, and the first song on the list "Alice Abroad" was from her favorite band, Wolf Law. Since she is a fan of this artist, she is familiar with other songs by them, especially "Because, Because" which she really likes. She becomes curious about why the recommendation system recommended "Alice Abroad" to her instead of "Because, Because." In answer to this question, the comparative explainer might say, "85% of your close neighbors listened to 'Alice Abroad' while only 10% of your neighbors listened to 'Because, Because'. That's why 'Alice Abroad' was recommended." This explanation makes it clear that a collaborative recommender is in use and that Sarah's recommendations are a function of what her peers on the platform are listening to.

This kind of explanation would be useful in any environment where the user is likely to have extensive knowledge of potential items for recommendation, such as popular culture or media, because it depends on the consumer having specific knowledge of items they might expect to be recommended to them. In other settings, it might be less useful; for example, for restaurant recommendations in an unfamiliar city, the consumer might be unlikely to have an alternative 'foil' restaurant about which they are seeking an explanation.

To generate such an explanation, the comparative explainer needs access to key steps in the recommendation computation: the selection of peers, the characteristics of these peers' profiles, and the extrapolation of recommendations. Since the user is supplying the two entities to be contrasted, the explainer can focus on the difference between how these entities were treated in the original recommendation calculation, or it could run the recommendation calculation again, with these two entities as targets, and extract the differences.

3.2. Comparing Item Ranking (rcC)

Question template: Why is Item A being ranked ahead of Item B in my recommendations?

Items are typically recommended to users in a ranked list, with each item's rank indicating its importance and priority relative to the others. Users expect that the first item is the most relevant and best aligned with their preferences. However, the ranking process is often not transparent, leaving users uncertain as to how rankings are derived.

We can examine this explanation style with reference to Sarah and the *Tunester* platform. Recently, Sarah noticed that in a list of recommendations, "All the Things You Are" (the first recommendation) and "Bolivar Blues" (the seventh recommendation) are both in the Jazz genre and by the same artist. Curious about the ranking, she asks why "All the Things You Are" is prioritized over "Bolivar Blues".

The comparative explainer responds by saying "Our recommender likes to promote artists' newer work. 'All the Things You Are' is recommended over 'Bolivar Blues' because it is a recently-released track."

We see that the explanation in this case focuses attention on a particular feature that the recommender system takes into account in ranking, one that the user might not be aware of. From Sarah's point of view, these tracks are very similar; the explanation provides an opportunity for the recommender to indicate what distinctions it is making. This type of recommendation could also be useful in contexts where the user might not be aware of some of the key distinctions between items. For example, consider an e-commerce setting, where a consumer is purchasing an electronics product, such as a laptop dock. These products come in many different configurations and capabilities, so inquiring about the recommender's ranking may help the user understand the differences between the products.

As in our first example, the explainer needs access to the recommendation computation. To get the type of explanation we have envisioned in this case, we can imagine a causal model linking the songs' features to their weight in the item representation and to the recommendation calculation. The release date would be a key difference between the casual reasoning in each case since the songs are similar in other ways, and could be used as the basis for generating our (admittedly hypothetical) explanation.

3.3. Comparing Patterns of Items (tcC)

Question template: Why are items of type A being recommended to me more than items of type B?

A pattern of recommendation can span multiple items in a list or across multiple lists. There are many different types of patterns that users might ask about. Questions that are related to item features will have explanations similar in nature to those discussed above about comparative ranking.

A different type of pattern is one that takes place over time, when the recommender is making certain recommendations at one time and different recommendations at another. Such temporal changes in recommendation patterns can be categorized as either contextual shifts or preference drifts. Contextual shift refers to changes in the context or environment in which recommendations are made. This context can include observable factors such as the time of day, location, the user's current mood, or even external factors that might not be observable to users. Preference drift, on the other hand, describes the gradual evolution in a user's preferences over time. Unlike contextual shifts, which are temporary and context-dependent, preference drift reflects long-term changes in what a user likes or needs [26]. While users might not always be aware of these underlying changes, they may notice the changed nature of the content or products recommended to them. Providing explanations can help users recognize and understand these changes, such as why certain recommendations may have shifted or diversified over time.

Consider the following hypothetical example: Sanjay is a final-year PhD student in mathematics; the title of his dissertation is "Unconstrained Optimization: a New Faster Method for Solving Nonlinear Equations." During his PhD studies, he has published several papers. He is a regular user of *Moogle Scholar*, a website that allows users to subscribe to a daily newsletter of recommended scientific papers according to their interests. He has found that he generally receives paper recommendations that are highly relevant to his research area. However, recently, he has noticed a shift in the recommendations coming from the system, with a focus on optimization methods for machine learning papers – an area in which he has neither published nor has substantial knowledge. Sanjay is curious about why this change in recommendation patterns has occurred and asks: "Why are papers on machine learning-oriented optimization methods now being recommended more frequently than those on mathematical optimization that I used to get?" The comparative explainer responds: "Sometimes we recommend papers based on patterns of citations coming from your work. There has been a recent increase in citations of your work from researchers in the machine learning topic area and we are recommending papers by some of those researchers to you."

Without this comparative explanation, Sanjay might not understand the reasons for the changing patterns in the recommendations he see. With the insights the explanation provides, he may want to learn more about the researchers who have picked up on his work and what they are doing. Without the

transparency provided by the explanation, he might have been tempted to ignore the recommendations of these papers. We envision that temporally-oriented comparative explanations could be very helpful in explaining trends in any area. The recommender has access to the constantly changing item catalog and the evolution of user behavior and typical users will not. We will also see the value of these types of explanations for providers in the next section.

Generating such temporal explanations is not simple, however. First, it assumes that the explainer either has access to or can re-generate a trace of the recommender system's behavior sufficient to derive an explanation for the phenomenon of interest. We cannot assume that the explainer is tracking every kind of change in anticipation of some user possibly asking about it. There will be too many possible changes and too many users and most such computation would be wasted. With this capability in place, the system would still need the ability to identify the trend to which the user is referring. In this specific case, the system would need to be able to make enough sense of the question to detect that the user is asking about a change in the topic area of recommended papers. After that, the types of casual inference that we have referred to above would be needed to isolate the effect of changes in citation patterns which cause the changes in recommendations that the user sees.

4. Provider-Oriented Comparative Explanations

A multistakeholder approach to recommendation requires that we consider the needs of stakeholders other than recommendation consumers, especially item providers [2]. As noted above, providers note acutely the lack of transparency that they perceive in their interactions with recommender systems. Yet, research on provider-oriented explanations in recommender systems is practically non-existent. There is a lot to address in this gap. For the purposes of this paper, we believe that providers can benefit from explanations that help them understand to whom their products are recommended (and to whom they are not). We note that in the music context, artists (as providers) have expressed a strong desire for more transparency [22, 23] in recommender systems. One artist specifically noted a desire for the system to explain what actions they could take to be recommended more frequently [23]. We noted above that comparative explanations are more informative for consumers and align with how humans naturally make decisions. For slightly different reasons, this approach is equally valuable for providers. Comparative questions, such as why two items from the same provider have different levels of recommendation activity, can reveal potential issues like algorithmic bias or intrinsic quality differences. This understanding can lead to improved marketing strategies, enabling providers to more effectively reach their desired audiences.

4.1. Single Provider

4.1.1. Comparing Individual Items (ipP)

Question template: Why is my Item A being recommended and my Item B is not being recommended?

Ideally, providers want all their items to have a fair chance of being recommended and purchased or otherwise being received by interested consumers or audiences. There is of course natural variation in the appeal of products, but it is very hard for a provider, especially one whose work is intrinsically tied to a recommendation-oriented platform (think YouTube or TikTok), to understand the interplay between the item's properties and the relatively inscrutable behavior of the recommender system. This lack of transparency in existing algorithms often leads providers to try to develop their own *folk theories* from others' experiences or ad-hoc experimentation [27, 28]. Such theories can be far afield from actual algorithm behavior. Comparative explanations of the type that we are advocating for in this paper can help close this gap. Analogous to the consumer-oriented comparative explanations discussed above, we can imagine explanations through which providers can compare the treatment of different items at the hands of the recommender system.

Consider the following hypothetical example: Maria is a PhD student working on the evaluation of large language models. In the final years of her PhD, she writes a conference paper based on her

results and then follows it up with a more comprehensive journal article. While she is developing her application materials in advance of a job search, she consults the provider-side interface of *Moogle Scholar* which gives her information about her profile relative to the recommendations it generates. She notices that the conference paper is being recommended more often than the journal article, even though the journal article was published in a top journal and the conference was a relatively small one. Curious about the reasons behind this difference, she asks the system: "Why is my conference paper A being recommended more often than my journal article B?" The system responds: "One factor in our recommendation algorithm is how often a paper is cited and by whom. Publication A has a higher citation count and higher quality score (citations * centrality of citing author) than Publication B, which outweighs Publication B's higher venue centrality score and so Publication A is more likely to be recommended in contexts where both items might be relevant."

We see that this explanation provides a valuable bit of transparency to the researcher who can use it to understand the difference in reception of her work and might cause her to investigate how her work has been cited. We imagine such explanations to be highly useful in other recommendation applications. The YouTube creators interviewed in [24] explicitly noted their inability to understand why some of their videos were viewed heavily and other, similar, videos received hardly any attention. `ipP` explanations could go a long way towards helping such individuals have a more productive relationship with the recommendation platform.

Our example here assumes a capability not present in today's recommendation platforms: a provider-side interface that helps a provider understand when and in what contexts their items have been recommended. We believe that robust interfaces of this type are necessary for providers to be incorporated as first-class users of recommender systems; this is corollary of adopting a multistakeholder view of recommendation in the first place. In this example and others in this section, we proceed under the assumption that such an interface exists and comparative explanation can become one of its features.

The types of casual explanation mechanisms that we have discussed earlier in this paper could come into play in generating explanations of this type. There is a key difference, however. Providers would not have access to individual recommendation lists to be able to ask questions about how individual lists were generated. So, provider-focused explanations are by necessity explaining a pattern that has developed over multiple recommendation interactions: item A was recommended over time, item B was not. This creates complexities similar to those discussed in relation to the pattern-oriented explanations of type `tcC`. The recommender would need to be able to revisit a history of recommendation decisions and the underlying processes that generated them.

4.1.2. Comparing Item Ranking (rpP)

Question template: Why is my Item A being ranked higher than my Item B in recommendation lists?

As described earlier, items are usually recommended to users in a ranked list, where items higher in the list are expected to be the most relevant and the most aligned with user preferences. While providers likely do not have access to individual user's recommendation lists, they may be told average recommendation ranks across many users' recommendation lists of their own items, and may wonder why there is a discrepancy in rank between two of their items (on average).

Consider the following hypothetical example: Saylor Twift is an up-and-coming singer-songwriter who just released a new EP. Her new album has five songs on it, and she and her manager thought the first song, "Orange" – a poppy, upbeat, fun song – would perform better than "Mosaic", a slow, acoustic ballad. Twift and her team have been advertising the whole album but have been focusing on the first song. Twift's album is on *Tunester*, the hypothetical music streaming platform we have seen before. The artist interface for *Tunester* shows that in recommendation lists, "Mosaic" is ranked, on average, higher than "Orange" on these daily playlists. Twift wants to know why this is the case. The comparative explainer responds by saying, "Our recommender considers the music tastes of listeners when recommending songs. 'Mosaic' is recommended over 'Orange' on average because listeners in the Pop music category who have you and artists like you in their listening history tend to listen more

to slow acoustic songs with guitar than uptempo songs with synthesizer."

In this case, the comparative explainer reveals a glimpse of how the recommender algorithm works – clearly, some assigned categorization of the songs are considered and playlists depend on engagement with artists. We can also see that the recommender characterizes music listeners and uses that to make decisions. Here, comparative explanation is particularly advantageous over simple factual explanation – asking "why does my song not have a higher ranking" may be difficult to explain by itself, since ranking is inherently rivalrous. The simple answer is that other things are ranked higher, but the comparative explanation can help the provider narrow in on what features of their items the recommendation algorithm is picking up on, which can help influence decisions on newer productions or marketing. Armed with this new knowledge, maybe Twift decides to focus more on the acoustic songs featured on her new EP. This type of explanation would be informative to any provider with multiple items being recommended and wanting greater transparency in recommender system operation.

As with the prior provider-oriented explanation, the explainer needs access to the history of recommendation slates delivered to users and the ability to generalize over multiple executions of the recommendation algorithm. A casual approach would be useful as we have seen, to identify the key features distinguishing between the treatment of one item versus another. Our hypothetical explanation assumes the system performs certain kinds of categorization ("slow" vs "uptempo", "artists like you") and these can be included in the explanation.

4.1.3. Comparing Patterns of Items (tpP)

Question template: Why are my items of type A being recommended more often than my items of type B?

Although comparing individual items or individual items' rankings may be useful for a provider, a provider might notice larger discrepancies in the recommendations of items according to specific types of their items or across time. Instead of considering a single recommendation list, the provider would ask about the recommendations of their items across multiple users.

Consider the following hypothetical example: Jo is a tenured professor of information science of 10 years. They have 600+ citations, and although much of their early work was in ethical AI and policy, recently they' have been working more on research into ethical AI curriculum. Their papers are all available on *Moogle Scholar*. Recently, Jo noticed that their ethical AI and policy papers have been recommended significantly more to users. Jo wants to know why their older papers on ethical AI and policy are being recommended more often than their more recent papers on ethical AI curricula and pedagogy. The comparative explainer responds by saying, "We recommend papers that are highly relevant to our subscribers' research interests. There are fewer scholars working on ethical AI curriculum and fewer opportunities to recommend these papers compared to the number of AI and policy researchers."

Here the explainer is focusing on the recommendation opportunities, based on the popularity of the research area. Jo might not have a good sense for the relative popularity of different research fields, especially if they are not currently doing research in this area. Similar to the provider-side explanations above, this kind of insight could be useful to any provider seeking to understand larger scale dynamics of which the recommender system is aware by its position.

The explainer would need the same historical perspective that the other provider-oriented explanation types require. It would also need to understand what Jo is referring to in describing the classes of papers they wish to contrast ("ethical AI and policy" vs "ethical AI curriculum"). A causal model of how recommendations are delivered over time could highlight the difference between papers with many potential opportunities for recommendation and those with fewer.

4.2. Cross-Consumers

Providers often aim for equitable distribution or recommendations of their products across different consumer groups (e.g. geographic location, gender, age, race/ethnicity, or stated preferences), but

achieving this can be challenging due to various factors including consumer behavior, marketing strategies, or potential biases. Therefore, providers may be particularly interested in understanding the rationale behind differing levels of engagement or utility among these consumer groups. To the best of our knowledge, this type of explanation has not been studied yet in recommender systems.

4.2.1. Comparing Individual Items (iccP)

Question template: Why is my Item A being recommended to users of type X and not users of type Y?

In previous examples, the key comparison occurred between two items, item rankings, or groups of items. For the cross-consumer source, however, the comparison is instead between groups of users. A provider may want to know why a specific item of theirs is recommended to one group of users and not another.

We'll return to the example of Jo, a tenured information science professor. They have recently published a new paper, "Ethical AI Curriculum in U.S. Public High Schools: A Literature Review". When looking at their *Moogle Scholar* dashboard, which shows analytics of recommendations on their papers, Jo notices that this paper is only being recommended to professors and not to grad students. Jo wants to know why this is the case, as they have recently been talking to several grad students interested in this area who might benefit from reading such a paper. The comparative explainer responds by saying, "We recommend papers that are similar in topic to papers that users have published and depend on a minimum of five papers to establish a user's topic profile. Grad students have fewer published papers on average than professors and may not have established profiles in the system."

The explanation reveals more about how the underlying recommender system works. Clearly, if one doesn't have any published papers, their recommendations may not be the same as someone with published papers, even if they share common interests. This may not be desirable behavior, but at least Jo now knows why their paper is not being recommended in the way that they expect. The added benefit of comparative explanation in this context is that it can serve two purposes at once, by allowing the provider to discover why a certain group is not being recommended to and to discover why a certain group is being recommended to. Each of these may be useful in different contexts. For example, a queer Instagram content creator might want to know why their content is being served to transphobic people, which they do not want, and why their content is not being served to LGBTQIA+ people, which they do want. Asking only one of these questions may not reveal the full picture.

On top of the capabilities already discussed for provider-oriented explanations, this type of explanation requires information about different user categories: "professor" vs "graduate student". Such categorization is assumed in the ability for Jo to even formulate their question, so this capacity is similar to what we have articulated for the other explanation types.

4.2.2. Comparing Item Ranking (rccP)

Question template: Why is my Item A being recommended more highly to users of type X than users of type Y?

There are many factors that may cause recommendations to vary across demographic groups. There may be shared group preferences; consider teenage music listeners versus elderly ones. There may be cultural factors, such as the language in which lyrics are sung, and there may be third parties differentially marketing items to one group over another.

Consider again our musical artist Saylor Twift. From *Tunester*'s artist dashboard, she discovered that "Mosaic" was recommended to European listeners at a higher average rank compared to U.S. listeners, despite equal advertising efforts in both regions. Since a higher ranking increases the likelihood of consumption, she wants to understand why the "Mosaic" track is ranked higher for European listeners than for listeners in the U.S. The comparative explainer replies: "Our recommender considers the demographics of listeners when recommending songs. 'Mosaic' is recommended more highly in the

European market on average because listeners in the Pop music category who have you and artists like you in their listening history tend to be older than their U.S. counterparts and older listeners tend to prefer acoustic music."

It is easy to imagine explanations of this type being of use to any type of creator interested in the audience for their work and wanting to understand how the recommender is (or is not) helping to reach those audiences or markets. Analysis of this type might highlight undesired bias in a recommender system. For example, consider a company that discovers that its job listings for engineers were being recommended much more often to male applicants than female applicants. If the job recommender exhibited the type of gender bias seen in [29], the hiring manager might get an explanation saying that "We recommended your engineering positions to male applicants because they were more similar in background to prior successful applicants for similar positions than the female applicants were." This would be a sign of unacceptable gender bias in the recommender and one would expect the manager to demand that the platform change how its algorithm operates.

4.2.3. Comparing Patterns of Items (tccP)

Question template: Why are my items of type A being recommended to users of type X and not users of type Y?

Besides asking why one specific item is recommended differently between groups of users, a provider may wonder why a group of their items is recommended differently. All of the benefits of a comparative explanation for a single item and comparing across consumers holds in this case as well – namely, knowing that your content is reaching your intended audience, or not, is extremely useful. The only difference here is that by looking at a pattern, group, or trend of items, a provider can potentially understand how more of their content is being recommended (or not) by the recommendation algorithm, which could lead to deeper understanding of how particular content is affected.

Let's return to the example of Saylor Twift, a singer-songwriter who just released a new EP. The artist interface for *Tunester* shows that the songs from the EP overall are recommended at much higher rates to listeners in the Japan than to listeners than in the U.S., even though Twift is American and is based in the U.S. Twift wants to know why this group of songs – her EP – is recommended more to Japanese listeners and not to U.S. listeners. The comparative explainer responds by saying, "We recommend items based on trends that we anticipate when there are new releases. The new EP was just released in Japan and so is getting more recommendations due to our new release expectations while in the U.S., it has been out for three months and is outside the new release window."

This explanation is again similar to those we have already seen in terms of its demands on the explainer. This explanation in particular requires that the explainer have access to what might be considered business rules ("release window") about promotional activity that the platform engages in.

5. Conclusions and Future Work

We have presented a taxonomy of different types of comparative explanations for recommender systems, where a comparative explanation is one that compares two different outcomes from the recommender system with the aim of providing greater transparency into the system's operation. We note that such comparisons are recognized as a common feature of explanations in interpersonal communication.

Our taxonomy considers different types of explanations based on the kinds of comparisons that are being made, between which kinds of recommender output, and for which audiences / stakeholders. We note in particular that not all comparative explanations are appropriate for all audiences for privacy or confidentiality reasons. We emphasize the importance of explanations for multiple stakeholders, in particular providers, who are for the most part neglected as users in recommender systems writ large. As far as we know, this is the first work to date to examine the potential of comparative explanations broadly in recommender systems, and to suggest the possibility of and potential impacts of explanations for providers.

This taxonomy is intended as a pointer to future work as the title would suggest, since there is little or no work on the types of explanations that we describe here. Some explanation types require significant advances in how we represent and track system activity within recommender systems, and how we interface with users to solicit their explanation-oriented questions and to produce acceptable answers. We note that comparative explanation, as a task, orients recommender systems explanation squarely in the direction of system transparency. A question like "Why is A ranked higher than B in my list?" cannot be legitimately answered in a post-hoc manner, as a rationalization. It can only be answered by reference to the process by which the ranking is actually generated, prioritizing transparency. Thus, we expect that new techniques may be needed to generate explanations for complex recommendation models, but as we discuss, we believe that causal techniques may offer some solutions.

Although some existing work has surveyed providers on their perspectives of the platforms they use, no studies have interviewed providers with the intent of designing explanations for them. A detailed survey of providers where their perspectives on explanations are gathered, as well as the current information they receive from a platform's interface, could help influence explanation design and help articulate the value of comparative explanation. Such relationships would also be useful in future user studies of a fully developed comparative explanation system.

We opted to leave comparative explanations targeted at system stakeholders, such as platform owners and operators, from our discussion. We expect that systems designers might want all of the types of explanations included here and then some, in order to understand how their systems are treating different kinds of content and different classes of users. Most recommender system designers and operators function within well-resourced organizations, but with the rise of Very Small Online Platforms (VSOPs) [30], the need for explanation interfaces to support recommender systems in these contexts may increase.

Acknowledgments

Author Burke was supported by the National Science Foundation under grant award IIS-2107577. Author McKinnie was supported by the National Science Foundation under grant award IIS-2232555.

References

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

[2] H. Abdollahpouri, G. Adomavicius, R. Burke, I. Guy, D. Jannach, T. Kamishima, J. Krasnodebski, L. Pizzato, Multistakeholder recommendation: Survey and research directions, User Modeling and User-Adapted Interaction 30 (2020). doi:10.1007/s11257-019-09256-1.

[3] J. L. Herlocker, J. A. Konstan, L. G. Terveen, John, T. Riedl, Evaluating collaborative filtering recommender systems, ACM Transactions on Information Systems 22 (2004) 5–53.

[4] N. Tintarev, J. Masthoff, Beyond explaining single item recommendations, in: Recommender Systems Handbook, Springer, 2022, pp. 711–756.

[5] A. Yang, N. Wang, R. Cai, H. Deng, H. Wang, Comparative explanations of recommendations, in: Proceedings of the ACM Web Conference 2022, 2022, pp. 3113–3123.

[6] A. D. Starke, C. Musto, A. Rapp, G. Semeraro, C. Trattner, "tell me why": using natural language justifications in a recipe recommender system to support healthier food choices, User Modeling and User-Adapted Interaction 34 (2024) 407–440.

[7] C. Musto, A. D. Starke, C. Trattner, A. Rapp, G. Semeraro, Exploring the effects of natural language justifications in food recommender systems, in: Proceedings of the 29th ACM conference on user modeling, adaptation and personalization, 2021, pp. 147–157.

[8] P. Lipton, Contrastive explanation, Royal Institute of Philosophy Supplements 27 (1990) 247–266.

[9] D. Temple, Discussion: The contrast theory of why-questions, Philosophy of Science 55 (1988) 141–151.

[10] P. Ylikoski, The idea of contrastive explanandum, in: Rethinking explanation, Springer, 2007, pp. 27–42.
[11] A. Dhurandhar, P.-Y. Chen, R. Luss, C.-C. Tu, P. Ting, K. Shanmugam, P. Das, Explanations based on the missing: Towards contrastive explanations with pertinent negatives, Advances in neural information processing systems 31 (2018).
[12] Y. Wang, X. Wang, "why not other classes?": Towards class-contrastive back-propagation explanations, Advances in Neural Information Processing Systems 35 (2022) 9085–9097.
[13] Y. Lei, Z. Li, Y. Li, J. Zhang, H. Shan, Lico: explainable models with language-image consistency, Advances in Neural Information Processing Systems 36 (2024).
[14] R. Luss, E. Miehling, A. Dhurandhar, Cell your model: Contrastive explanation methods for large language models, ArXiv abs/2406.11785 (2024). URL: https://api.semanticscholar.org/CorpusID:270560761.
[15] S. A. Chemmengath, A. Azad, R. Luss, A. Dhurandhar, Let the cat out of the bag: Contrastive attributed explanations for text, ArXiv abs/2109.07983 (2021). URL: https://api.semanticscholar.org/CorpusID:237532193.
[16] N. Madaan, I. Padhi, N. Panwar, D. Saha, Generate your counterfactuals: Towards controlled counterfactual generation for text, in: AAAI Conference on Artificial Intelligence, 2020. URL: https://api.semanticscholar.org/CorpusID:228063841.
[17] T. S. Wu, M. T. Ribeiro, J. Heer, D. S. Weld, Polyjuice: Generating counterfactuals for explaining, evaluating, and improving models, in: Annual Meeting of the Association for Computational Linguistics, 2021. URL: https://api.semanticscholar.org/CorpusID:235266322.
[18] A. Jacovi, S. Swayamdipta, S. Ravfogel, Y. Elazar, Y. Choi, Y. Goldberg, Contrastive explanations for model interpretability, in: Conference on Empirical Methods in Natural Language Processing, 2021. URL: https://api.semanticscholar.org/CorpusID:232092617.
[19] J. Woodward, Making things happen: A theory of causal explanation, Oxford university press, 2005.
[20] Y. Deldjoo, T. D. Noia, F. A. Merra, A survey on adversarial recommender systems: from attack/defense strategies to generative adversarial networks, ACM Computing Surveys (CSUR) 54 (2021) 1–38.
[21] J. J. Smith, A. Satwani, R. Burke, C. Fiesler, Recommend me? designing fairness metrics with providers, in: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency, FAccT '24, Association for Computing Machinery, New York, NY, USA, 2024, p. 2389–2399. URL: https://doi.org/10.1145/3630106.3659044. doi:10.1145/3630106.3659044.
[22] K. Dinnissen, C. Bauer, Amplifying artists' voices: Item provider perspectives on influence and fairness of music streaming platforms, in: Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization, UMAP '23, Association for Computing Machinery, New York, NY, USA, 2023, p. 238–249. URL: https://doi.org/10.1145/3565472.3592960. doi:10.1145/3565472.3592960.
[23] A. Ferraro, X. Serra, C. Bauer, What is fair? exploring the artists' perspective on the fairness of music streaming platforms, in: C. Ardito, R. Lanzilotti, A. Malizia, H. Petrie, A. Piccinno, G. Desolda, K. Inkpen (Eds.), Human-Computer Interaction – INTERACT 2021, Springer International Publishing, Cham, 2021, pp. 562–584.
[24] Y. Choi, E. Kang, M. Lee, J. Kim, Creator-friendly algorithms: Behaviors, challenges, and design opportunities in algorithmic platforms, in: CHI 2023 - Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Conference on Human Factors in Computing Systems - Proceedings, Association for Computing Machinery, 2023. doi:10.1145/3544548.3581386, publisher Copyright: © 2023 ACM.; 2023 CHI Conference on Human Factors in Computing Systems, CHI 2023 ; Conference date: 23-04-2023 Through 28-04-2023.
[25] N. Tintarev, J. Masthoff, A survey of explanations in recommender systems, in: Data Engineering Workshop, 2007 IEEE 23rd International Conference on, IEEE, 2007, pp. 801–810.
[26] N. Hariri, B. Mobasher, R. Burke, Adapting to user preference changes in interactive recommendation, in: Twenty-Fourth International Joint Conference on Artificial Intelligence, 2015.

[27] M. A. DeVito, Adaptive folk theorization as a path to algorithmic literacy on changing platforms, Proceedings of the ACM on Human-Computer Interaction 5 (2021) 1–38.
[28] Y. Choi, E. J. Kang, M. K. Lee, J. Kim, Creator-friendly algorithms: Behaviors, challenges, and design opportunities in algorithmic platforms, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 2023, pp. 1–22.
[29] J. Dastin, Amazon scraps secret ai recruiting tool that showed bias against women, in: Ethics of data and analytics, Auerbach Publications, 2022, pp. 296–299.
[30] C. Rajendra-Nicolucci, M. Sugarman, E. Zuckerman, The Three-Legged Stool: A Manifesto for a Smaller, Denser Internet, Technical Report, Initiative for Digital Public Infrastructure, 2023. URL: https://publicinfrastructure.org/2023/03/29/the-three-legged-stool/.