User Modeling in Human-Centered AI

Eelco Herder and Judith Masthoff

Abstract Artificial intelligence is the discipline that pursues the understanding, artificial replication and possible enhancement of human intelligence. Among others, AI-based recommender systems, persuasive systems as well as decision support systems are used for making decisions that have direct impact on people's lives. In AI-supported decision making, the initiative may lie on the user's side (such as recommender systems), or be largely automated (such as navigation systems or automotive driving). Many organizations also employ AI systems to make decisions that concern their users, customers, or citizens. But how do these systems learn about the people involved, how are these people represented, and - most importantly - how complete, realistic, reliable and fair is this representation? In this chapter, we discuss various considerations for user modeling in human-centered AI.

1 Introduction

Even though Artificial Intelligence (AI) still sounds like science fiction to many people, we already encounter AI on a daily basis. Social media feeds, music and video streaming services, and web stores are deeply integrated in our daily lives and the way they work is driven by recommender systems [47].

Recommender systems are a particular type of personalized systems that specifically aim to provide lists of items that may be of interest to the user — such as music to listen to, books to read, or products to buy. Other types of personalized systems include persuasive systems — that help users to reach specific goals, such as losing

Eelco Herder
Utrecht University, e-mail: e.herder@uu.nl

Judith Masthoff
Utrecht University e-mail: j.f.m.masthoff@uu.nl

weight, exercising more, or reducing their carbon footprints — and adaptive learning support [25].

What all these personalized systems have in common is that they provide users with individual advice, in order to help them make decisions. In order to do so, these AI-driven adaptive systems need to *know the user* — in other words, to have a model that represents all — or most — relevant aspects of a user with sufficient reliability. The process of creating such a model is called 'user modeling'.

Originally, user modeling was seen as simulating the cognitive processes that humans employ for getting to know other persons, such as stereotyping, making and adjusting assumptions, and inferring assumptions based on observations. Instead, current large-scale recommender systems largely aim to create user models that are optimal for maximizing user consumption or user engagement (i.e. click-rate optimization) [30].

The interaction process between humans and personalized AI systems can be seen as an iterative communication process. A system may ask some initial information about the user, such as demographics. Subsequently, the system provides suggestions (such as recommendations or persuasive advice), to which the user responds — by accepting, ignoring, liking, or disliking the suggestion. The personalization algorithm — or user modeling engine — will further adjust or optimize its assumptions based on the user responses.

Similarly, users will form expectations on the system's performance and the quality of its suggestions, and may change their behavior accordingly. For instance, they may choose to abandon a social media platform that does not provide relevant recommendations, or they may decide to leave it to an agent to automatically select and deliver their weekly groceries, once they are convinced that the agent sufficiently knows their consumption patterns and preferences.

In this chapter, we provide a human-centered perspective on the process of user modeling, an activity that many AI systems engage in and that has direct impact on our lives. We will pay particular attention to the division of initiative and responsibility when it comes to making decisions.

This chapter is structured as follows. In Section 2, we summarize the state-of-the-art in user modeling, explaining which user characteristics can be taken into account, how they are used for inferring user profiles, and how they are deployed for personalization purposes. In Section 3, we focus on challenges and limitations of AI-based personalization, which are related to the uncertainty of derived assumptions, the inherent propensity of systems to reinforce existing habits, and limited understanding of user intentions. The future directions in Section 4 focus on the opportunities that dialogue-based interfaces may offer to engage users in deliberate decision-making. We end the chapter with some concluding thoughts.

2 State-of-the-art in user modeling

In this section, we briefly summarize the different kinds of user characteristics — as observed or inferred by the system — that are used for personalization purposes. We start with an overview on user characteristics that are typically taken into account, followed by a discussion how these characteristics can be automatically inferred. We end the section with implications for commercial

2.1 Traditional user profiles and characteristics

Conceptually, user modeling involves three different stages: the acquisition of user data or usage data, inference of knowledge from the data, and the representation of a user model [43]. In practice, one or more of these stages may overlap. Apart from characteristics pertaining to the users themselves, this representation may also describe the users' (current) context and earlier actions that may indicate interest in particular topics or items.

This model of 'the current state of the world' is subsequently used for deciding upon personalization or adaptation approaches, which are then applied within the system [41]. Currently on the Web, the most common form of adaptation is recommending items that the user may be interested in. For instance, in social media platforms, the system may recommend certain posts or potential friends to connect with or accounts to subscribe to; advertisements on social media platforms are often personalized as well. Music platforms may recommend songs, artists, albums or playlists.

Traditionally, the processes of user modeling and personalization have been regarded as separate activities. Early approaches were inspired by cognition theory and user models were assumed to represent stereotypes [48], or characteristics such as a user's personality, background knowledge, learning style, or current goals [9].

Typical user characteristics considered in traditional user modeling include *demographics and roles* [29], which may be useful to provide personalization and recommendations that are likely to match preferences of — for example — the typical teenager, middle-aged man, student, teacher, nurse or salesperson. These initial assumptions can be further adjusted by taking a user's *competencies* into account, which includes domain knowledge [12] as well as cognitive styles, such as the Five Factor Model of personality [17].

Building upon characteristics that describe the users themselves, user modeling systems may take the user's *interests and tastes* in the domain of interest into account. In news recommendation, this category would include news topics and regions in the world [8], whereas for music and video streaming genres, artists and languages would be more relevant. In an persuasive system that encourages healthy eating, preferences for food types and ingredients is relevant. In collaborative filtering recommender systems, user interests and tastes are typically derived from previously observed

interaction with items, such as watching a video or listening to a song — and used as proxy for a user's future *intentions and goals* [23].

Other categories of user characteristics — used for instance for persuasive technology — include a user's *attitudes*, *opinions*, and *values* [40], and a user's *motivation* and *stage of change* [34]. For personalization of user interfaces, the user's *physical, sensory, and cognitive functioning* is often used [32]. It is beyond the scope of this chapter to enumerate all possible aspects of users and their contexts, but it may have become clear that there are many possible candidate attributes for inferring relevant aspects of a user.

2.2 AI-based inference of user interests

Cognitive, rather holistic models, as discussed above, which represent a user as a whole (e.g. [20]) have largely been replaced by methods that directly aim to translate user actions into recommendations. Particularly, the popular recommendation approach *collaborative filtering* keeps track of items that users interacted with — such as clicking on, liking or commenting on social media posts — and uses this to recommend posts, users, or advertisements that users with similar behavior interacted with [31].

Seen from a broader AI perspective, an adaptive system aims to *understand* the user and their context, and to use this understanding to optimize its behavior. It is important to realize that this understanding may take place on different levels of abstraction [2]:

- A user model may be based on data directly *provided* by the user. For instance, upon account creation, users may provide demographic information, their level of education or knowledge, their goals, or areas of interest. During further interaction with the system, users may, for instance, provide item ratings, report their physical activities, food intake, or how they feel. Arguably, self-reported information is the most reliable type of user information.
- Alternatively, user characteristics can be obtained by *observing* the users' actions. These actions may for instance involve interactions in social media, items purchased in a web store, songs listened to, and exercise performance. Actions such as clicking on an item, spending time watching a video, or sharing or liking a post can be considered as *relevance feedback* [45]. Activity logs may be an objective registration of a user's actions and responses, but in order to be useful, these activities typically need to be interpreted.
- Interpretation of user-provided data and observations of user actions may lead to a *derived* user profile or stereotype: a particular predefined category that forms the best match with the user. Depending on the application, these categories may represent — for instance — a user's political orientation, music taste or dietary habits. Note that, in contrast to direct observations, these interpretations involve *uncertainty* and can be incorrect.

- A more behaviorist approach is to directly *infer* a user's (expected) interest in an item based on correlations between users and items; this is the approach used by collaborative filtering.

2.3 From adaptation to item recommendation

AI systems and AI algorithms are designed and optimized for a particular goal. In the case of personalized and other adaptive systems, the goal is to assist users in fulfilling their needs. However, in the past decades, there has been a shift in focus with respect to the type of user needs that should be considered.

The original (academic) ambition of personalized systems was to support users in meaningful activities, such as online learning, improving habits, or finding relevant information [9]. Optimizing — and possibly changing — user interaction and user behavior is still a thriving branch of research: *persuasive systems* [16] aim to encourage and support users to exhibit, for instance, healthier or more environment-friendly behavior — see [34] for a more extensive overview of personalization in persuasive systems and application areas. Similarly, there is still much work on online learning; see for example the success of personalized language learning applications such as Duolingo. By contrast, the goal of commercial recommender systems is often to support and encourage *existing* habits and behavior.

It does not come as a surprise that commercial companies mainly develop and deploy recommender systems in order to increase sales and profit (which also includes increasing customer loyalty). A strategy for achieving this goal is to address the user's natural tendency to stick to safe routines and safe choices; in music recommendation, for instance, it is known that users tend to have a small set of music tracks from a small set of artists to listen to [10]. The success of such systems is typically evaluated by using a training set to calibrate an algorithm and then to try and predict the remaining user actions in a test set — which usually consists of actions and choices that have already been made without the system's interference. The success is subsequently measured by metrics that represent a minimization of error rates or maximization of click-through rates [30].

These developments contribute to several challenges in AI-based user modeling practices, which is the topic of the next section.

3 Challenges in deploying user models

In this section, we discuss inherent uncertainties and risks when user models are deployed for inferring assumptions about the user and using these assumptions for personalization and automated decision-making. In line with Nissenbaum [39], we will argue that systems need to be careful with inferring, for example, that someone is vegan or Muslim based on the products that a user buys. Further, we discuss

how AI-based user modeling methods inherently are limited to what Kahneman calls system-1 behavior [26] (recognizing and reproducing patterns in an automatic manner), and that actual reasoning about a user's intentions, or fully establishing a user's context — in a system-2 kind of way — is beyond the capabilities of current AI, as argued as well by [7].

3.1 Uncertainties of derived assumptions

Even though the algorithmic approaches used for making sense of individual users may differ fundamentally from human reasoning, in the end, they serve similar goals: deciding what actions to perform, in order to effectively support, convince, and/or satisfy the user and, ultimately, optimize interaction strategies. This may be well compared to a librarian who is expected to recommend books of interest [48], a salesperson trying to sell products, or a coach advising on actions for behavior change. Similar to interactions between human beings, incorrect assumptions or interpretations may lead to unwanted, unpleasant, awkward, or downright embarrassing (or even illegal) situations [39].

Providing a list of recommended items sounds innocent enough, but there is the risk that a user may associate the recommendations with particular user groups or minorities. Already in 2002, this insight was popularized by the article "My TiVo thinks I'm gay" [59], in which a user suspects that his television program choices had put him in some implicit or hidden 'gay' category. Instead of using implicit correlations or clusters, it may be more transparent to use explicit categories and labels. However, users may feel offended, discriminated against, or even insulted when a system explicitly tells them that it assumes them to be gay, Muslim, or vegan.

The European General Data Protection Regulation[1] (GDPR) already prohibits the use of such sensitive (inferred) characteristics, but even non-sensitive characteristics — such as one's supermarket buying behavior[2] may be considered unpleasant or offending. And even with privacy regulations in place, Facebook practices demonstrate that this is a very thin line: in a statement, they explained that they do not use sensitive personal data, but they could use expressions of interest, such as being subscribed to a Pride page[3].

The risks associated with inferring assumptions on the user have been recognized already in the early times of user modeling research, with prominent researchers stressing the importance of *scrutability* [28]: users should be able to control which observations are used, which assumptions are being formed, who has access to this data, and how this affects the system's behavior — and to repair any incorrect or unwanted aspects of it. Even though the full ambition of scrutability has not been reached, stronger regulations and the increasing calls for transparency have led to

[1] https://gdpr.eu/

[2] https://eenvandaag.avrotros.nl/item/ah-cursus-klantprofielen-haaks-op-vn-campagne/

[3] https://www.theguardian.com/technology/2018/may/16/facebook-lets-advertisers-target-users-based-on-sensitive-interests

the adoption of privacy dashboards and explanations in commercial recommender systems [53].

3.2 Reinforcement and amplification of existing habits

As argued by Bengio [7], current (deep learning) AI systems are most successful at 'perception tasks' (system-1 behavior [26]), which involves learning to recognize certain patterns and to use this for reproducing adequate (learned) responses to these patterns. This works very well for tasks such as image recognition or classification, but AI is far less successful for tasks 'that require a deliberate sequence of steps' (system-2 behavior). Translated to recommender systems, this implies that they inherently tend to reinforce a user's habits.

For individual users, this tendency towards reinforcing — and not challenging — existing habits and preferences may be suboptimal. At first sight, it may be convenient and enjoyable to mainly encounter music, series, books, and other items that feel familiar. However, ultimately most users will feel bored, unfulfilled, or otherwise unsatisfied when they hardly engage novel, engaging situations in which they learn or experience something new. Breaking routines may have positive or negative outcomes, but the so-called 'remembering self' tends to appreciate new experiences best [57].

The effect of users mainly encountering known items that confirm and reinforces their world view, as a result of algorithmic decisions, has been given the label 'filter bubble' by Eli Pariser [42] in 2011. As personalized recommendations are by definition different from user to user, it turned out to be hard to provide concrete evidence for this effect. Our current understanding is that 'hard' filter bubbles only occur in very extreme cases for very extreme users [61]. The far majority of users will still encounter sufficient 'mainstream' content: for instance, a study has shown that search results — for users with very different (simulated) profiles — for German politicians and parties consistently led to largely unbiased result sets [44].

Due to its pattern-reproducing nature, AI-based recommender systems are known to amplify existing bias. On the web, where most user modeling takes place, various forms of bias exist: for instance, most activity on the web — which also serves as training data for recommender algorithms — is carried out by a small number of very active users, who large live in western, English-language countries [6]. As a result, AI-based personalized systems have been observed to amplify this bias in historical interactions, which tends to lead to excessive recommendations of majority, mainstream items [52].

3.3 Limited understanding of users and their intentions

Recommender systems are thriving on the Web, for example in the form of product recommendations and personalized social media streams. Arguably, the popularity and ubiquity of this technology have proven its benefits. Simultaneously, as discussed and argued in this very brief state-of-the-art, the technology also has its limitations.

Traditional, cognitive user modeling approaches aimed to describe a user as a human would do. AI-based personalized systems aim to automatically come to a user model, in order to employ this model for a specific purpose.

By necessity, AI-based user models are incomplete, and as machine learning algorithms need to work with limited data, they are designed with underlying assumptions. Therefore, inferring user characteristics or traits is error-prone and caution should be taken in order to avoid incorrect assumptions — particularly when it comes to sensitive issues, such as a user's religion or orientation.

AI-based personalization algorithms have underlying assumptions as well that do not necessarily hold. For instance, collaborative filtering (see Section 2.2) assumes similarity in selection patterns based on users sharing a particular taste. It took Amazon surprisingly long to realize that this may yield for books, music and videos, but not for vacuum cleaners, olive oil or other items that people buy for different reasons [51].

Finally, user modeling algorithms model users based on historical data. As discussed in this section, this may lead to seemingly contradictory effects, such as the filter bubble (extreme extrapolation of individual users) and regression to the mainstream (extreme extrapolation to the majority of users).

3.4 Issues with user modeling for automated decision-making

Recommender systems and persuasive systems are AI-based personalized systems that many people experience on a daily basis. Often, users do not interact directly with the AI technology itself, but they do indirectly interact with it via (web-based) user interfaces in the form of — among others — product recommendations, social media feeds, and personalized search results.

It is important to note that personal (user) data, either directly provided or inferred, is used for many other forms of decision-making that concern the user as well. Social media platforms, for instance, are known to categorize users to infer their 'ad preferences' — in order to optimize the success rate of paid advertisements, to increase platform revenue [22]. There have also been several examples of platforms experimenting with manipulating the emotions of their users, in order to reach an overall impact on the average among the whole user population [15]. Arguably even more concerning, there is ample evidence that social media have exploited personal data for political manipulation, for instance by targeting users with particular controversial topics and opinions [46].

The use of personal data or user models for AI-supported decisions is not limited to the realm of the web: there is a growing concern regarding the use of AI-based (automated) decisions in various other fields, including health-care, judicial and law enforcement [3]. For instance, AI-based algorithms may be used to decide whether a person is eligible for early release from jail.

Tax authorities are also known to make use of automated analysis of large amounts of (personal) data. Despite the advantages in terms of efficiency and overall error minimization, it also raises concerns regarding privacy rights and the risk of errors that may unjustly impact individual persons [50].

From the above examples, it becomes clear that user modeling in human-centered AI has implications that reach far beyond mere suboptimal product recommendations. Stepping away from fully automated decisions, which according to the European GDPR legislation are only allowed with informed consent of the person involved [4], many decisions are actually made by people who use decision support systems. These systems may be designed for various levels of involvement of their users [5].

As discussed in Section 3.2, current AI systems are most successful at 'perception tasks', in other words recognizing and optimizing responses to overall patterns. By contrast, humans are strong in deliberate reasoning, including evaluating moral, ethical, or societal implications. Consequently, algorithmic decisions may be strongly supported by statistical measures and provide overall best decisions, whereas humans base their decisions on experience and judgment — combined with some level of emotion and irrationality. As argued by [27], for these reasons, human judgment may lead to various types of noise that AI algorithms may not be prone to. However, for *individual* cases and decisions that do not fit overall patterns, human judgment may still be preferable. Moreover, people are more likely to accept decisions that are made by humans than automated decisions [27].

Decision support systems are designed to combine the powers of human reasoning and (AI-based) machine reasoning, limiting either side's weaknesses. Similar to 'ordinary' users interacting with web-based platforms, expert decision-makers interact with interfaces that hide the complexity of AI [5]. Expert interfaces may be implemented as dashboards or interactive interfaces for active decision-making [4], but in many cases decisions are based on ranked result lists and feed-based interfaces that are quite similar to the web platforms that regular users work with.

In summary, user modeling was originally conceptualized as a method to understand and support *individual* users. AI-based user modeling may still have this ambition, but it inherently suffers from biases stemming from the data, which may lead to *overall* optimal results, but also to erroneous assumptions and suboptimal results for *individual users* that may not fit the mainstream model. Given the potentially serious consequences [49], user modeling in human-centered AI is an activity that needs close scrutiny.

[4] https://gdpr.eu/article-22-automated-individual-decision-making/

4 Opportunities and future directions

Many AI systems have been designed with purposes that — directly or indirectly — have an impact on individual users or on groups of people within our society. In order to fulfill their tasks, many of these systems employ personal data in one way or another. Inferring high-order data from lower-level observed or provided personal data is essentially user modeling. With the increasing power of AI inferences, this process may lead to unforeseen, surprising and undesirable effects that require adherence to legal and ethical norms, transparency and accountability [39].

In the field of AI-based personalized web systems, transparency is typically offered in terms of *explanations*. These explanations aim to help users understand how a system works and allow them to tell the system when its wrong, increasing their confidence in the provided recommendations [53]. The most transparent type of explanations would explain the actual personalization process, but many systems use explanations that are generated post-hoc [60], designed to be more user-friendly and understandable than purely technical explanations.

A further approach to improve AI transparency and accuracy is to make personalized systems more *interactive*: user input and feedback have been recognized to be essential for better recognizing user characteristics, preferences and goals [19]. Typical types of user interaction that is considered and investigated involves interactive visualizations to provide insight in the system or to justify recommendations or adaptations.

4.1 Putting the human in the loop

In a sense, approaches that rely on user input and feedback are a form of *human-in-the-loop* artificial intelligence [58], a movement that aims to complement AI reasoning with input and feedback stemming from humans. Human-in-the-loop AI recognizes the fact that modern AI functions in a complex world that is largely created and ruled by humans. However, it seems that for AI-supported decision making — particularly when these decisions have serious implications — simple user feedback is not sufficient; among others reasons, explainable recommender systems do allow users to react on those items that are recommended, but not on items or choices that have *not* been recommended and of which the users are not aware [6].

In an early paper on user modeling via stereotyping [48], Elaine Rich drew a parallel between the user modeling system and a librarian, who tried to guess which type of a book on a topic best matches the user's background knowledge, values, interests and goals. The librarian's strategy involved guessing a stereotype and then refining this stereotype with follow-up questions. This metaphor works well in user modeling systems that act solely in the interest of the user, but should be reconsidered in the field of commercial personalized systems, where the system's behavior rather mimics a *salesperson*. In (commercial) multi-stakeholder recommendations [1], several stakeholders need to be satisfied. For instance, the success of a food

delivery platform does not only depend on the satisfaction of the customers, who wish good meals for a fair price; for participating restaurants, main values include increasing revenue and customer base, and the platform itself needs to earn money as well in order to be sustainable.

Assuming that multi-stakeholder AI-based personalized systems take several ethical aspects into account — we refer to [37] for a more in-depth discussion — it can be observed that even fair recommendations are not necessarily solely in the best interest of the user, but are the outcome of balancing the values and interests of various stakeholders. Similar to the offline commercial landscape, online platforms provide a *choice architecture*: users are provided with suggestions for items that they can buy, read, watch, share or rate, while leaving it up to the users whether they wish to do so. Similar to traditional companies, regulations are in order to prohibit mischievous selling techniques, such as the use of micro-targeting and dark patterns [21].

Similar to a person, systems that engage in user modeling and personalization, are not infallible. Contrary to persons, AI systems do not have a 'morale' and cannot be 'punished', at least not in the sense that is associated with humans. This would imply that any user modeling system should be considered as *an agent of the responsible entity* (company, government, other organization, entrepreneur, private person) and that ultimately the responsible entity is liable for damages (within reasonable expectations, as to be determined by legal scholars) and 'moral wrongdoings'.

However, the literature on fair and transparent recommender systems appears to see users as passive, vulnerable human-beings, who need to be protected from the 'big tech algorithms'. This assumption may be in line with how platforms have taught users to passively consume 'the feed' — as discussed in [30] and Section 3.4 — but looking for solutions *on behalf* of the user, without involving the user, seems inherently too limited and possibly counter-productive. In addition — or instead — we believe it will be more productive to *empower* users, among others by stimulating them to engage in active decision-making [4]. In the next section, we will discuss several future directions in order to reach this goal.

4.2 Conversational personalized systems

Many efforts are being made to improve AI-based user modeling and personalization techniques, be it in terms of accuracy, profitability, or fairness. We believe that while technological progress continues, it is the *interface* between AI and humans that needs more attention, particularly in order to maintain a healthy power balance.

In the past few years, the main user interface paradigm for AI-based personalized systems was the web. Despite research efforts to design interfaces that stimulate active decision-making, it appears that users consistently prefer the ease and convenience of feeds and other list-based interfaces [4]. The combined power of voice interfaces and large language models (LLMs) may result in the emergence and acceptance of a new — and simultaneously quite old — *dialogue-based interaction* paradigm.

Smart voice assistants are increasingly penetrating households, with Amazon's Alexa, Apple's Siri and Google Nest currently being the most prominent examples. While most communication with voice assistants involves the assistant performing a simple task — such as providing the weather forecast — following just one command [14], conversational systems based on Large Language Models (LLMs) are increasingly proficient in interactive dialogues with users [13].

Current AI-based personalized systems already rely to a large extent on relevance feedback from individual users and their fellow users [45], with items that are being liked, clicked, selected, read and/or bought most often considered more popular, relevant or meaningful than items that receive less of them. Even though this relevance feedback is extremely useful in aggregated form — for instance in collaborative filtering [31] — they are arguably very limited means for expressing one's opinion, attitudes, or emotions towards a system response, particularly when the system response is limited to a simple list or feed as well. As discussed in [54], there has been some work on dialogues in recommender systems for requirement elicitation, but these dialogues are still very limited in scope.

As argued by Mercier and Sperber [36], a core element of successful human reasoning is communication, having to convince others through argumentation and reacting on their responses. These dialogues force people to engage in more deliberate, slow 'system-2' thinking [26]. There has been research on dialogues in persuasive systems based on argumentation (e.g. [18, 35]) and motivational interviewing (e.g. [38]). However, current dialogues between users and personalized systems via relevance feedback are very limited, due to limited expressiveness on both the system and user side. Similarly, in personalized e-coaches, which often try to mimic the feeling of a dialogue, the input possibilities for users are normally limited. With the advent of LLMs, the type of generative AI, conversations with voice assistants are becoming increasingly natural [11], which opens the possibility for real dialogue-based AI interface paradigms.

4.3 Dialogue initiation and argument flow

Conversational recommender systems are an active field of research, for reasons discussed above [24]. A particular point of attention is the construction of meaningful dialogues and the elicitation of user needs and preferences.

It should be emphasized that natural-language interaction with a conversational system is not a guarantee for meaningful dialogues. For instance, search engines increasingly provide direct answers and information snippets, in addition to the traditional result sets. The observed reduction in user interaction with the result sets suggests that users are prone to accept plausibly formulated results — or recommendations — as the final answer [56]. The same effect is currently observed in user responses to texts generated by LLM-powered conversational chatbots: well-written AI-formulated texts suggest authority, which may lead users to automatically assume that what the chatbot says is accurate and correct [11].

Putting the users into control is a challenge for conversational interfaces [24], as users may feel that they are forced to 'repair' results that should have been optimal. Instead, conversational techniques and interventions need to be in place that make clear that it is the AI interface which helps the user in making decisions and not the other way around.

Persuasive systems already have design elements that aim to stimulate users to consider alternative choices that may be smarter, healthier, or better for society [16]. Nudges are small design elements that present arguably better choices in a more attractive manner, thereby steering users in a gentle way. Even though this may prompt users to consider alternatives, the initiative still largely remains at the system — arguably, the nudge can be considered an add-on 'smart' proposition, but not really a dialogue.

4.4 Truly interactive decision making

We believe that the growing adoption of voice assistants is an opportunity for changing the interaction paradigm from users accepting or rejecting recommendations or other AI-generated propositions to a dialogue in which both parties explore a solution space together. Early experiments on LLM chain-of-thought prompting have shown that dialogues make intermediate steps explicit [55], which provides opportunities to provide arguments as well as doubts on certain assumptions, values or consequences.

A particular focus of attention in these dialogues would be to explicitly address *uncertainties* with respect to inferred user characteristics that the voice assistant may have, to prevent that the inherent risks of user modeling propagate into 'hallucinations' [33].

Once users would be accustomed to natural-language decision-making dialogues — with voice assistants asking questions such as "Am I correct to assume.." or "Do you really want to..." — and appreciate the benefits (and hopefully the ease and naturalness) of this approach, this paradigm may also be integrated in web-based (visual) interfaces — where until now feed-based recommendations were consistently preferred above more interactive decision-making systems [4].

Naturally, the majority of decisions or choices do not require elaborate deliberation and can safely be automated by AI and feeds of recommendations, in a 'system-1' fashion. However, in addition, there is much to gain if means will be provided to engage in dialogue-based human-AI 'system-2' reasoning when either the system or the user thinks this is necessary, appropriate or otherwise useful.

5 Conclusions

AI-supported user modeling is a powerful technology for systems to better understand and support their individual users. User models are created for particular

purposes, with particular assumptions and particular goals that may differ between stakeholders.

In this chapter, we discussed the inherent risks and limitations of user modeling, personalization and recommender systems. Flawed inferences or misconceptions may lead to suboptimal or even unsuitable results. Furthermore, current AI user modeling tends to reinforce existing habits and mainstream patterns, which is further reinforced by the users themselves, due to the feed-based paradigm of many interfaces.

Persuasive and explanatory interfaces aim to break this routine pattern by putting the user in the loop, but this still leaves too much initiative and responsibility on the system-side. In many cases, the personalization process can be seen as a negotiation between different stakeholders, where the outcome needs to satisfy users and providers as well as the platform. Particularly when the stakes are high, it might be useful to approach user modeling as a human activity with AI in the loop, instead of the other way around.

References

1. Himan Abdollahpouri and Robin Burke. Multi-stakeholder recommendation and its connection to multi-sided fairness. *arXiv preprint arXiv:1907.13158*, 2019.
2. Martin Abrams. The origins of personal data and its implications for governance. *Available at SSRN 2510927*, 2014.
3. Theo Araujo, Natali Helberger, Sanne Kruikemeier, and Claes H De Vreese. In ai we trust? perceptions about automated decision-making by artificial intelligence. *AI & society*, 35:611–623, 2020.
4. Claus Atzenbeck, Eelco Herder, and Daniel Roßner. Breaking the routine: spatial hypertext concepts for active decision making in recommender systems. *New Review of Hypermedia and Multimedia*, pages 1–35, 2023.
5. Verena Bader and Stephan Kaiser. Algorithmic decision-making? the user interface and its role for human involvement in decisions supported by artificial intelligence. *Organization*, 26(5):655–672, 2019.
6. Ricardo Baeza-Yates. Bias on the web. *Communications of the ACM*, 61(6):54–61, 2018.
7. Yoshua Bengio, Yann Lecun, and Geoffrey Hinton. Deep learning for ai. *Communications of the ACM*, 64(7):58–65, 2021.
8. Daniel Billsus and Michael J Pazzani. User modeling for adaptive news access. *User modeling and user-adapted interaction*, 10:147–180, 2000.
9. Peter Brusilovsky. Adaptive hypermedia. *User modeling and user-adapted interaction*, 11(1):87–110, 2001.
10. Òscar Celma and Pedro Cano. From hits to niches? or how popular artists can bias music recommendation and discovery. In *Proceedings of the 2nd KDD Workshop on Large-Scale Recommender Systems and the Netflix Prize Competition*, pages 1–8, 2008.
11. Vinton G Cerf. Large language models. *Communications of the ACM*, 66(8):7–7, 2023.
12. Alexandra I Cristea, David Smits, and Paul ME de Bra. Writing mot, reading aha! converting between an authoring and a delivery system for adaptive educational hypermedia. 2005.
13. Yang Deng, Wenqiang Lei, Lizi Liao, and Tat-Seng Chua. Prompting and evaluating large language models for proactive dialogues: Clarification, target-guided, and non-collaboration. *arXiv preprint arXiv:2305.13626*, 2023.

14. Aarthi Easwara Moorthy and Kim-Phuong L Vu. Privacy concerns for use of voice activated personal assistant in the public space. *International Journal of Human-Computer Interaction*, 31(4):307–335, 2015.
15. Catherine Flick. Informed consent and the facebook emotional manipulation study. *Research Ethics*, 12(1):14–28, 2016.
16. Brian J Fogg. Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002(December):2, 2002.
17. Lewis R Goldberg. The structure of phenotypic personality traits. *American psychologist*, 48(1):26, 1993.
18. Floriana Grasso, Alison Cawsey, and Ray Jones. Dialectical argumentation to solve conflicts in advice giving: a case study in the promotion of healthy nutrition. *International Journal of Human-Computer Studies*, 53(6):1077–1115, 2000.
19. Chen He, Denis Parra, and Katrien Verbert. Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications*, 56:9–27, 2016.
20. Dominik Heckmann, Tim Schwartz, Boris Brandherm, Michael Schmitz, and Margeritta von Wilamowitz-Moellendorff. Gumo–the general user model ontology. In *User Modeling 2005: 10th International Conference, UM 2005, Edinburgh, Scotland, UK, July 24-29, 2005. Proceedings 10*, pages 428–432. Springer, 2005.
21. Mireille Hildebrandt. The issue of proxies and choice architectures. why eu law matters for recommender systems. *Frontiers in Artificial Intelligence*, 5:789076, 2022.
22. Paul Hitlin and Lee Rainie. Facebook algorithms and personal data. *Pew Research Center*, 16, 2019.
23. Dietmar Jannach, Lukas Lerche, and Michael Jugovac. Adaptation and evaluation of recommendations for short-term shopping goals. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 211–218, 2015.
24. Dietmar Jannach, Ahtsham Manzoor, Wanling Cai, and Li Chen. A survey on conversational recommender systems. *ACM Computing Surveys (CSUR)*, 54(5):1–36, 2021.
25. Mathias Jesse and Dietmar Jannach. Digital nudging with recommender systems: Survey and future directions. *Computers in Human Behavior Reports*, 3:100052, 2021.
26. Daniel Kahneman. Maps of bounded rationality: A perspective on intuitive judgment and choice. *Nobel prize lecture*, 8(1):351–401, 2002.
27. Daniel Kahneman, Olivier Sibony, and Cass R Sunstein. *Noise: a flaw in human judgment*. Hachette UK, 2021.
28. Judy Kay. Scrutable adaptation: Because we can and must. In *Adaptive Hypermedia and Adaptive Web-Based Systems: 4th International Conference, AH 2006, Dublin, Ireland, June 21-23, 2006. Proceedings 4*, pages 11–19. Springer, 2006.
29. Alfred Kobsa, Jürgen Koenemann, and Wolfgang Pohl. Personalised hypermedia presentation techniques for improving online customer relationships. *The knowledge engineering review*, 16(2):111–155, 2001.
30. Joseph Konstan and Loren Terveen. Human-centered recommender systems: Origins, advances, challenges, and opportunities. *AI Magazine*, 42(3):31–42, 2021.
31. Greg Linden, Brent Smith, and Jeremy York. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, 7(1):76–80, 2003.
32. Claudia Loitsch, Gerhard Weber, Nikolaos Kaklanis, Konstantinos Votis, and Dimitrios Tzovaras. A knowledge-based approach to user interface adaptation from preferences and for special needs. *User Modeling and User-Adapted Interaction*, 27:445–491, 2017.
33. Potsawee Manakul, Adian Liusie, and Mark JF Gales. Selfcheckgpt: Zero-resource black-box hallucination detection for generative large language models. *arXiv preprint arXiv:2303.08896*, 2023.
34. Judith Masthoff and Julita Vassileva. Personalized persuasion for behavior change. pages 205–235. Walter de Gruyter GmbH & Co KG, 2023.
35. Irene Mazzotta, Fiorella De Rosis, and Valeria Carofiglio. Portia: A user-adapted persuasion system in the healthy-eating domain. *IEEE Intelligent systems*, 22(6):42–51, 2007.

36. Hugo Mercier and Dan Sperber. *The enigma of reason*. Harvard University Press, 2017.
37. Silvia Milano, Mariarosaria Taddeo, and Luciano Floridi. Ethical aspects of multi-stakeholder recommendation systems. *The information society*, 37(1):35–45, 2021.
38. Hien Nguyen and Judith Masthoff. Designing persuasive dialogue systems: Using argumentation with care. In *Persuasive Technology: Third International Conference, PERSUASIVE 2008, Oulu, Finland, June 4-6, 2008.*, pages 201–212. Springer, 2008.
39. Helen Nissenbaum. Contextual integrity up and down the data food chain. *Theoretical Inquiries in Law*, 20(1):221–256, 2019.
40. Kate Pangbourne and Judith Masthoff. Personalised messaging for voluntary travel behaviour change: interactions between segmentation and modal messaging. In *28th Annual Universities' Transport Study Group Conference, Bristol, 6th–8th January*, 2016.
41. Alexandros Paramythis, Stephan Weibelzahl, and Judith Masthoff. Layered evaluation of interactive adaptive systems: framework and formative methods. *User Modeling and User-Adapted Interaction*, 20:383–453, 2010.
42. Eli Pariser. *The filter bubble: What the Internet is hiding from you*. penguin UK, 2011.
43. Wolfgang Pohl. Logic-based representation and reasoning for user modeling shell systems. *User Modeling and User-Adapted Interaction*, 9:217–282, 1999.
44. Cornelius Puschmann. Beyond the bubble: Assessing the diversity of political search results. *Digital Journalism*, 7(6):824–843, 2019.
45. Rachael Rafter, Barry Smyth, and Keith Bradley. Inferring relevance feedback from server logs: A case study in online recruitment. In *11th Irish Conference on Artificial Intelligence and Cognitive Science (AICS 2000)*, 2000.
46. Ulrike Reisach. The responsibility of social media in times of societal and political manipulation. *European journal of operational research*, 291(3):906–917, 2021.
47. Francesco Ricci, Lior Rokach, and Bracha Shapira. Recommender systems: Techniques, applications, and challenges. *Recommender Systems Handbook*, pages 1–35, 2021.
48. Elaine Rich. User modeling via stereotypes. *Cognitive science*, 3(4):329–354, 1979.
49. Ulrik BU Roehl. Automated decision-making and good administration: Views from inside the government machinery. *Government Information Quarterly*, 40(4):101864, 2023.
50. Luisa Scarcella. Tax compliance and privacy rights in profiling and automated decision making. *Internet Policy Review*, 8(4), 2019.
51. Brent Smith and Greg Linden. Two decades of recommender systems at Amazon.com. *IEEE Internet Computing*, 21(3):12–18, 2017.
52. Harald Steck. Calibrated recommendations. In *Proceedings of the 12th ACM conference on recommender systems*, pages 154–162, 2018.
53. Nava Tintarev and Judith Masthoff. Designing and evaluating explanations for recommender systems. In *Recommender systems handbook*, pages 479–510. Springer, 2010.
54. Nava Tintarev and Judith Masthoff. Beyond explaining single item recommendations. In *Recommender Systems Handbook*, pages 711–756. Springer, 2020.
55. Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837, 2022.
56. Zhijing Wu, Mark Sanderson, B Barla Cambazoglu, W Bruce Croft, and Falk Scholer. Providing direct answers in search results: A study of user behavior. In *Proceedings of the 29th acm international conference on information & knowledge management*, pages 1635–1644, 2020.
57. Chris AB Zajchowski, Keri A Schwab, and Daniel L Dustin. The experiencing self and the remembering self: Implications for leisure science. *Leisure Sciences*, 39(6):561–568, 2017.
58. Fabio Massimo Zanzotto. Human-in-the-loop artificial intelligence. *Journal of Artificial Intelligence Research*, 64:243–252, 2019.
59. Jeffrey Zaslow. If tivo thinks you are gay, here's how to set it straight. *Wall Street Journal*, 26, 2002.
60. Yongfeng Zhang and Xu Chen. Explainable recommendation: A survey and new perspectives. *arXiv preprint arXiv:1804.11192*, 2018.
61. Frederik Zuiderveen Borgesius, Damian Trilling, Judith Möller, Balázs Bodó, Claes H De Vreese, and Natali Helberger. Should we worry about filter bubbles? *Internet Policy Review. Journal on Internet Regulation*, 5(1), 2016.