RESEARCH PAPER

Information provision measures for voice agent product recommendations— The effect of process explanations and process visualizations on fairness perceptions

Helena Weith[1] · Christian Matt[1]

Abstract

While voice agent product recommendations (VAPR) can be convenient for users, their underlying artificial intelligence (AI) components are subject to recommendation engine opacities and audio-based constraints, which limit users' information level when conducting purchase decisions. As a result, users might feel as if they are being treated unfairly, which can lead to negative consequences for retailers. Drawing from the information processing and stimulus-organism-response theory, we investigate through two experimental between-subjects studies how process explanations and process visualizations—as additional information provision measures—affect users' perceived fairness and behavioral responses to VAPRs. We find that process explanations have a positive effect on fairness perceptions, whereas process visualizations do not. Process explanations based on users' profiles and their purchase behavior show the strongest effects in improving fairness perceptions. We contribute to the literature on fair and explainable AI by extending the rather algorithm-centered perspectives by considering audio-based VAPR constraints and directly linking them to users' perceptions and responses. We inform practitioners how they can use information provision measures to avoid unjustified perceptions of unfairness and adverse behavioral responses.

Keywords Voice agents · Product recommendations · Fairness perceptions · Process explanations · Process visualizations · Behavioral responses

JEL Classification D81 · D83 · L86 · M15

Introduction

The global volume of voice commerce transactions is forecasted to increase from 4.6 billion USD in 2021 to 19.4 billion USD in 2023 (Statista, 2022). This development is also driven by technological advances of voice agents, such as Alexa or Siri, providing product recommendations and disrupting established retailer-customer interactions. Voice agent product recommendations (VAPRs) make use of artificial intelligence (AI) to support the entire voice interaction and transaction procedures between voice agents and users. This comprises the process steps demand recognition, proactive product recommendations, and product selection, as well as subsequent payment and delivery option settlement (Følstad & Kvale, 2018). VAPRs profit from two core AI components: Machine learning (ML) enables personalized product recommendations (Rai, 2020), while natural language processing (NLP) enables audio-based interactions with users and thus allows for hands-free and eyes-off ubiquitous voice commerce transactions (Knote et al., 2021).

However, practical examples and research have shown that AI-based services might also be perceived as unfair (Kordzadeh & Ghasemaghaei, 2021; Feuerriegel et al., 2020), which could further lead to negative economic consequences for retailers (Wu et al., 2022; Kordzadeh & Ghasemaghaei, 2021;

Responsible Editor: Kevin Clyde Desouza

✉ Helena Weith
  helena.weith@unibe.ch

  Christian Matt
  christian.matt@unibe.ch

[1] Institute of Information Systems, University of Bern, Engehaldenstr. 8, 3012 Bern, Switzerland

🌲 Springer

von Zahn et al., 2021). AI-based recommender systems might provide unfair recommendations by offering products to customers at different prices subject to their characteristics or behavior (Wu et al., 2022; Valentino-DeVries et al., 2012) or pre-select and rank recommendations based on reasons which are not in favor for customers, such as their own margins (Wang et al., 2023). Furthermore, VAPRs communicating via audio are especially limited regarding the variety of information they can transfer as they are restricted to one recommendation at a time. Users' limited information levels compared to providers and voice agents could lead to a lack of users' fairness perceptions (Mavlanova et al., 2012). Due to limited information levels, it can be challenging for users to assess AI-based services and it may provoke perceptions of being treated unfairly even though this may be unjustified from an objective standpoint (Dolata et al., 2021; Kordzadeh & Ghasemaghaei, 2021). With regard to VAPRs, the users' information levels are limited through both their core AI components: First, ML-enabled recommendation engines are accompanied by opacities, characterized by the lack of explanations about how a recommendation is developed and why certain products are recommended (Ochmann et al., 2021; Rai, 2020). Second, NLP-enabled audio-based interactions allow only for limited information to be shared, while processing such auditive information is more complex for users' working memory (Natale & Cooke, 2021), thus resulting in audio-based constraints. Thereby users' information levels are further limited (Ocón Palma et al., 2020; Rhee & Choi, 2020; Sweller et al., 2011).

As both AI components, ML and NLP, are the core enablers for VAPRs, it is crucial to investigate measures that can increase users' information level of VAPRs and thus have the potential to strengthen users' fairness perceptions. To increase information levels, previous research suggests two independent additional information provision measures: process explanations outlining how and why a product is recommended and process visualizations representing the audio content visually (Weith & Matt, 2022; Tintarev & Masthoff, 2007a; Dodge et al., 2019). Both information provision measures can help users to absorb more information (Park et al., 2020; Rhee & Choi, 2020; Sweller et al., 2011) and thus increase the users' information level which is required by them to assess an AI-service and can potentially strengthen their fairness perceptions of VAPRs (Favaretto et al., 2019; Kroll et al., 2017; Kuempel, 2016). While process explanations have received previous attention in the context of recommendation engines (Kordzadeh & Ghasemaghaei, 2021; Rai, 2020), we lack an understanding of their application for VAPRs. Furthermore, current research on VAPRs lacks an understanding of the simultaneous use and effects of process explanations and process visualizations as information provision measures.

Through two experimental studies, we investigate the impact of these two information provision measures on fairness perceptions and users' behavioral responses. Study 1, a 2×2 between-subjects experiment with 235 participants, compares the effects of process explanations and process visualizations for VAPRs on their fairness perceptions. We critically acknowledge the AI advances in the context of VAPRs and provide both information provision measures simultaneously in light of the information processing theory. Since study 1 revealed a positive effect of process explanations on perceived fairness, we further differentiated the process explanation contents (functional-based, product-based, user-based) in study 2 through another online between-subjects experiment with 318 participants. By empirically studying fairness perceptions and behavioral responses leveraging the stimulus-organism-response (SOR) theory, we contribute to the AI fairness and explainable AI (XAI) literature. We further respond to calls by research for more representation and empirical evidence of the user-centered perspectives (Kordzadeh & Ghasemaghaei, 2021; Abdul et al., 2018). Furthermore, by drawing on the information processing theory, we extend a hitherto algorithm-centered perspective of recommendation engines by considering audio-based constraints. Practitioners can benefit from our findings, as we show that particular forms of process explanations serve as feasible information provision measures to reduce unjustified perceptions of unfairness for VAPRs. Practitioners learn that they should prioritize their resources on designing voice-based interactions with user-based explanations instead of integrating process visualizations.

Theoretical background

Fairness perceptions of VAPRs

Aspects of fairness relating to AI have received increasing attention within information systems (IS) research and related research fields (Dolata et al., 2021; Kordzadeh & Ghasemaghaei, 2021; Robert et al., 2020a). Fairness is sometimes conceptualized in relation to trust, for example, as a driver for user trust (Angerschmid et al., 2022; Zhou et al., 2021; Shin, 2020), affected by previous trusting beliefs (Dodge et al., 2019) or as a moderator between perceived fairness and behavioral responses (Kordzadeh & Ghasemaghaei, 2021). While fairness is partially conceptualized as part of collective concepts (Shin, 2020), researchers also increasingly focus on fairness as a standalone concept (Dolata et al., 2021; Kordzadeh & Ghasemaghaei, 2021; Lee et al., 2011). The AI fairness literature distinguishes the technical and social fairness perspectives (Dolata et al., 2021). The technical (objective) perspective applies mathematical approaches to mitigate biases and disparities in data and algorithms (Barocas et al., 2021; Feuerriegel et al., 2020). The social (subjective) fairness perspective

applies a rather user-centered approach by researching fairness perceptions, aspects influencing fairness perceptions, and resulting behavioral responses (Dolata et al., 2021). Ultimately, the technical and social perspectives should be considered collectively as socio-technical fairness by retailers providing VAPRs (Dolata et al., 2021). The technical perspective can be addressed by reducing sensitive attributes from data sets (Feuerriegel et al., 2020) or ensuring that exploited training data is representative and unbiased (Barocas & Selbst, 2016). But even if VAPRs are classified as fair from a technical perspective, it does not ensure subjective fairness perceptions (Dolata et al., 2021). While the technical perspective has received considerable attention (Barocas et al., 2021; von Zahn et al., 2021), a more thorough understanding of the social perspective involving user fairness perceptions of VAPRs is required in IS research (Kordzadeh & Ghasemaghaei, 2021; Feuerriegel et al., 2020).

We draw on the concept of systemic fairness, which consists of four distinct but correlated fairness dimensions, namely procedural, distributive, interpersonal, and informational. Procedural fairness refers to the process to achieve an outcome of a service or task. Distributive fairness refers to the outcome of the process in the form of decisions or recommendations. Interpersonal fairness refers to the social side of the process and the outcome and how it impacts recipients. Informational fairness refers to provided information about the process, the outcome, and how decisions are made (Colquitt & Rodell, 2015; Beugré & Baron, 2001, Greenberg, 1993). The second-order systemic fairness summarizes these four fairness dimensions (Colquitt & Rodell, 2015; Lee et al., 2011; Carr, 2007). Initially resulting from organizational fairness theory, the concept of systemic fairness and its four dimensions are frequently applied in the AI fairness and other IS literature (Dolata et al., 2021; Kordzadeh & Ghasemaghaei, 2021; Robert et al., 2020b). Procedural fairness results from perceptions of process steps and procedures of voice agents to recommend a product (Colquitt & Rodell, 2015; Kordzadeh & Ghasemaghaei, 2021). Distributive fairness results from perceptions of products recommended to users (Robert et al., 2020b; Carr, 2007). Interpersonal fairness results from perceptions about how users are treated by voice agents when recommending a product, thus also resulting from non-verbal information (Kim et al., 2020). It is strengthened if users feel treated with respect by voice agents (Robert et al., 2020b). Informational fairness results from perceptions of metainformation, which are "information about the information" (Kiesel et al., 2021). It can be increased if users perceive information from voice agents as truthful and thorough (Weith & Matt, 2022; Colquitt & Rodell, 2015). Users assess their fairness perceptions based on absorbed information (Favaretto et al., 2019). If VAPRs are not perceived as fair, it could have negative consequences for retailers, not only affecting short-term user behavior but also leading to reputational or long-term economic consequences

for retailers (Dolata et al., 2021). Relevant behavioral responses are, amongst others, the intention to purchase the recommended products (Ebrahimi & Hassanein, 2019), the intention to reuse voice agents for product recommendations (Mehta et al., 2016), and the intention to repurchase from retailers (Fang et al., 2014).

Information provision for VAPRs

VAPRs' two AI components hold two major shortcomings, both limiting users' information level. First, the ML-enabled recommendation engines are accompanied by opacities on recommendation development; thus, users lack information about how a recommendation is developed and why a particular product is recommended (Ochmann et al., 2021). It might further be unclear for users which data has been leveraged for recommendations (Burrell, 2016) and potentially cause negative effects on their usage attitude toward AI-based systems (Kordzadeh & Ghasemaghaei, 2021). These opacities limit users' information level (Ochmann et al., 2021; Kordzadeh & Ghasemaghaei, 2021). Second, NLP-enabled audio interactions hold audio-based constraints, thus limiting the information shared with recipients and reducing their agency (Natale & Cooke, 2021). Research on information processing identified that auditive presented information compared to visual one is cognitively more difficult to process for recipients. Thus, eventually, less information can be absorbed by users, which further limits their information level (Ocón Palma et al., 2020; Park et al., 2020; Rhee & Choi, 2020). Both shortcomings are present in the case of VAPRs, thus limiting users' information level. Hence, it can be challenging for users to assess recommendations, and it might provoke perceptions of being treated unfairly (Dolata et al., 2021; Kordzadeh & Ghasemaghaei, 2021).

Research on product recommendations via e-commerce websites demonstrated that recommendation engine opacities can be addressed by process explanations (Jannach et al., 2021; Xiao & Benbasat, 2007). Process explanations can transfer content to recipients, including messages and meanings of the communicated information in form of metainformation (Kiesel et al., 2021). It could also include descriptions of functionalities of an AI-system (Kim et al., 2020). Thus, additional content via process explanations could focus on how a product recommendation is developed and why a particular product is recommended (Weith & Matt, 2022; Dodge et al., 2019; Xiao & Benbasat, 2007). While explanations of AI-based recommendations for recipients with technical expertise have previously received attention, non-technical explanations understandable by wider audiences are required for recipients of product recommendations (Dodge et al., 2019; Abdul et al., 2018). Two overall types of process explanations can be differentiated: model-based

and case-based. Explanations based on the ML model of the recommendation engine and algorithmic specifications are applied to explain *how* recommendation engines work (Zhou et al., 2021). While these model-based process explanations can be identical for every recommendation provided by VAPRs, case-based process explanations are individually adapted to each recommendation. These explanations reply to the *why*, outlining the reasoning for certain products being recommended to users in a specific, prevalent case (Jannach et al., 2021; Zanker & Ninaus, 2010). Both process explanation types for VAPRs have further subtypes. Model-based process explanations can be differentiated into a technical or functional content base. The technical-based process explanations focus on the technical foundations of the recommendation engine and the voice agent, targeting an audience with specific technical expertise about AI. The audience are developers, engineers, and project managers who want to understand the underlying data used and the algorithms applied (Rai, 2020). Compared to technical-based, functional-based process explanations focus on how a VAPR works in general and provide these explanations in an easily understandable manner without being overly technical (Zhao et al., 2019). These explanations should be clear and help users understand why a certain product is recommended and how recommendation engines operate (Dolata et al., 2021; Dodge et al., 2019). Such functional-based process explanations target an audience without specific technical expertise in AI (Binns et al., 2018). The content of the case-based process explanations can be based on products or users. The product-based process explanations outline details about the recommended product and further compare it to alternative products. Hereby differences and superiorities of the recommended product compared to alternatives are highlighted (Yoshikawa et al., 2019; Nunes & Jannach, 2017; Friedrich & Zanker, 2011). Lastly, user-based process explanations relate and modify the product recommendation to the users' profiles and purchase behavior by considering their previous order history as well as the ones of similar users (Gedikli et al., 2014; Friedrich & Zanker, 2011). Therefore, user-based process explanations leverage the principle of collaborative and content-based filtering (Xiao & Benbasat, 2007).

Former research on information processing by humans suggests that audio-based constraints and limited information levels can be addressed by visualizations (Rhee & Choi, 2020), thus altering the medium used to transfer and present information to a recipient. Medium refers to a way how two parties interact with each other and exchange information, for instance, via audio or visuals (Natale & Cooke, 2021; Qiu & Benbasat, 2009). The information processing theory helps to understand human cognitive processes to encode and decode information (Sweller et al., 2011; Atkinson & Shiffrin, 1968). It separates the human memory into three subsections: sensory, working, and long-term. Sensory memory receives information via sensory inputs such as auditive or visual. It transfers them further via cognitive processes to the working and long-term memory. Working memory is limited in processing chunks of information; however, this capacity increases if processors to handle auditive as well as visual information are activated simultaneously. Therefore, providing visualizations stimulates higher capacities of the working memory to process information and increases the amount of information absorbed by recipients (Park et al., 2020; Sweller et al., 2011). Considering VAPRs, visualizations could be provided along all process steps and procedures of a product recommended via the voice agent to the user. They could visually present the product details as well as details about the process. Therefore, a voice agent could be equipped with an additional screen, such as the Amazon Echo Show 8 (Knote et al., 2021), or an app on the smartphone could present all auditive provided information simultaneously visually.

Although process explanations, as an explicit information provision measure, have already received some recognition amongst XAI research, our knowledge is so far limited to recommendation engines (Kordzadeh & Ghasemaghaei, 2021; Rai, 2020; Dodge et al., 2019). We apply process explanations to VAPRs and consider a second, currently less acknowledged information provision measure, process visualizations (Rhee & Choi, 2020; Knote et al., 2021). We hereby enrich our understanding of the simultaneous provision of information provision measures in the context of VAPRs and we shed more light on how they impact users' fairness perceptions and their behavioral responses. For the latter, we draw on the SOR theory, which originates from psychology but is well grounded in IS and AI fairness research (Kordzadeh & Ghasemaghaei, 2021, Mehrabian & Russell, 1974). The SOR theory serves as an appropriate structure as it conceptualizes that various stimuli (e.g., information provision measures) impact the users' affective or cognitive processes represented by the organism. The organism incorporates the users' fairness perceptions, which further affects the users' responses (Xu et al., 2014). Figure 1 conceptualizes the VAPRs' AI components and their resulting information provision shortcomings. It outlines along the SOR theory, the two information provision measures to address the two previous shortcomings and increase users' information level. It further suggests that information provision measures impact the fairness perceptions, which further might impact behavioral responses.
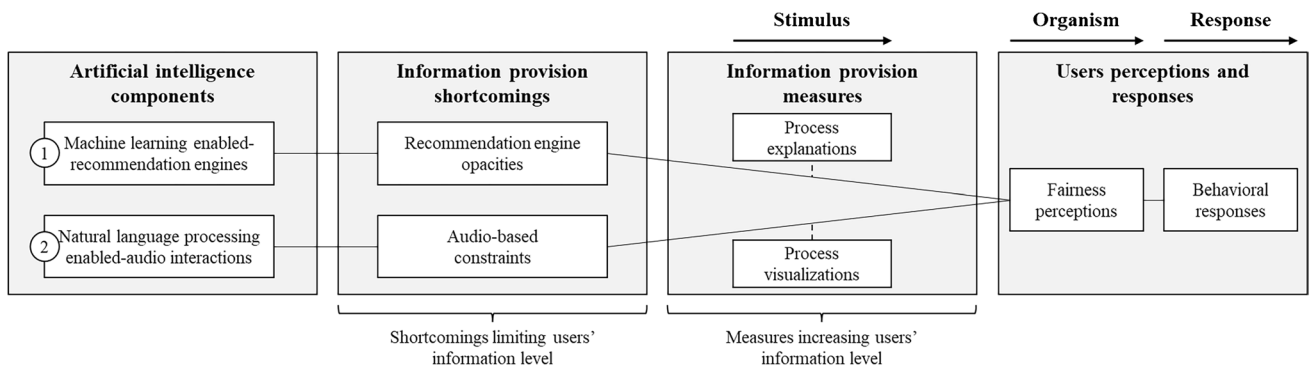
Fig. 1 Conceptualization of shortcomings and information provision measures for VAPRs

Experimental studies and framework

This project aims to identify the effect of information provision measures on users' fairness perceptions and subsequent behavioral responses. For study 1, we implemented process explanations that resembled established practical examples, and provided users with more information on *how* the underlying recommendation engine works as well as *why* certain products are recommended (Ochmann et al., 2021; Friedrich & Zanker, 2011; Rai, 2020). However, acknowledging the leeway that providers have in configuring process explanations, we further differentiated types and content bases for the process explanations in study 2 to shed more light on their differential impact on users' perceived fairness. All four process explanation contents outlined above (technical-based, functional-based, product-based, user-based) increase the users' information level. However, it is unclear how they impact the perceived fairness of VAPRs. We, therefore, in study 2, delve further into three of the four process explanations and seek to identify how they impact perceived fairness. We hereby focus on the functional-, product-, and user-based ones but do not include those that are technical-based, as these are only suitable for

limited audiences with specific technical expertise. Figure 2 provides an overview of the types and content bases for process explanations and highlights the foci of study 1 and study 2.

To identify the effects of the different information provision measures on perceived fairness, we leverage the concept of systemic fairness, which forms a second-order formative construct of the four fairness dimensions procedural, distributive, interpersonal, and informational fairness (Colquitt & Rodell, 2015; Lee et al., 2011; Carr, 2007). As user behavioral responses, we explore three dependent variables. First, the intention to purchase the recommended product to understand the direct economic effect of fair perceived VAPRs (Ebrahimi & Hassanein, 2019). Second, the intention to reuse the voice agent for product recommendations, as it is crucial for retailers to turn one-time users into recurring ones (Mehta et al., 2016). Third, the intention to repurchase from the retailer in general (i.e., not limited to VAPRs as a sales channel), to further assess potential long-term effects on retailers (Fang et al., 2014). Figure 3 provides an overview of the conceptual models for studies 1 and 2, also highlighting the difference between the studies.
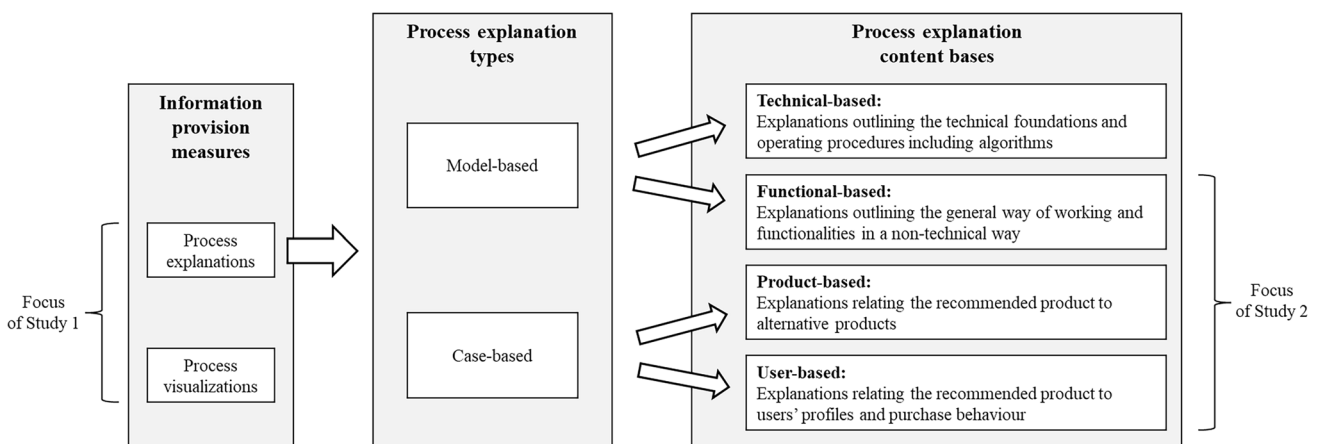

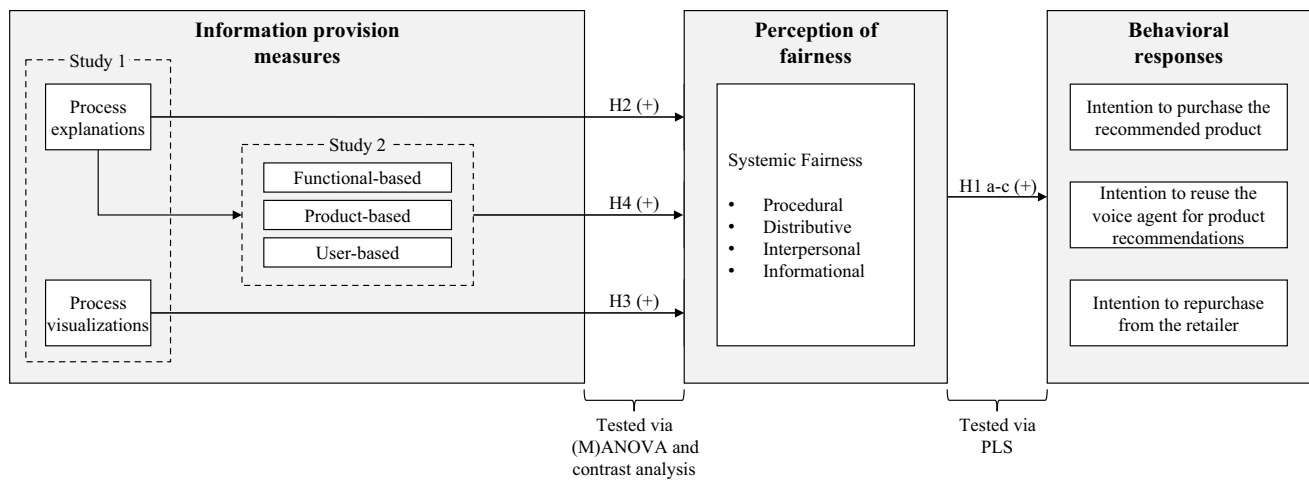**Fig. 2** Process explanation types and content bases

Fig. 3 Conceptual model for studies 1 and 2

Study 1—Process explanations and visualizations

Hypothesis development

Previous research has demonstrated that customer services which are perceived to be unfair cause negative effects by recipients, such as a negative impression of the service quality or reduced service satisfaction (Carr, 2007; Beugré & Baron, 2001), while those being perceived as fair can cause positive behavioral responses, such as higher service reusage (von Zahn et al., 2021; Lee et al., 2011; Carr, 2007). If users are aware of morally discriminating AI-provided recommendations, they have a higher tendency not to accept such recommendations (Ebrahimi & Matt, 2023; Ebrahimi & Hassanein, 2019). Lee et al. (2019) also demonstrated that in the context of food donations, users refused adoption if they perceived algorithms as unfair. On the other hand, if recommendations are developed by algorithms, which are perceived as being fair, this could lead to higher acceptance of algorithms, higher system adoption, and more purchases (Kordzadeh & Ghasemaghaei, 2021). Moreover, algorithm-based services which are perceived as being fair can also benefit from higher acceptance in the long run (Kordzadeh & Ghasemaghaei, 2021; Jussupow et al., 2020). We therefore hypothesize:

H1 a–c: Perceived systemic fairness has a positive effect on the intention to purchase a recommended product (H1 a), reuse the voice agent for product recommendations (H1 b), and repurchase from the retailer (H1 c).

Owing to an increased information level, explanations can increase transparency and thus help to reduce opacities (Zednik, 2021; Tintarev & Masthoff, 2007b). Relevant information, such as individually tailored explanations, are required

to increase users' information level and further to assess a service (Gretzel & Fesenmaier, 2006). Such information allows users to better judge a service or system, including assessing their fairness perceptions (Kordzadeh & Ghasemaghaei, 2021; Lee et al., 2019). Furthermore, explanations are suggested as a driver for fairness perceptions of algorithmic decisions (Binns et al., 2018). Previous research also demonstrated positive effects of explanations on algorithmic transparency and further on users' trust (Ebrahimi & Hassanein, 2019; Springer & Whittaker, 2019), which is often considered in relation to fairness (Angerschmid et al., 2022; Zhou et al., 2021; Shin, 2020). However, while transparency through explanations can increase fairness perceptions, it can also harm users' trust or the user experience if explanations are too long (Springer & Whittaker, 2019) or harm fairness perceptions if users do not agree with the applied reasonings (Lee et al., 2019). Previous research provides findings and propositions also addressing the four underlying fairness dimensions (Kordzadeh & Ghasemaghaei, 2021; Carr, 2007). As process explanations can provide content about why and how a product is recommended, they could thus strengthen perceptions of procedural fairness resulting from the process steps and procedures (Kordzadeh & Ghasemaghaei, 2021). Also, explaining that all users are treated equally could enhance perceptions of distributive fairness (Kordzadeh & Ghasemaghaei, 2021; Carr, 2007). Process explanations might not only affect procedural and distributive fairness, but also interpersonal and informational fairness, as the increased information level also includes additional non-verbal and metainformation. Expedient process explanations could ensure adequate and competent behavior of the voice agent, which might mainly have a positive effect on interpersonal fairness (Robert et al., 2020b; Carr, 2007). As informational fairness amongst others requires thorough explanations, process explanations could thus have a positive

effect on informational fairness (Colquitt & Rodell, 2015; Carr, 2007). Thus, we hypothesize:

H2: Providing process explanations for VAPRs has a positive effect on the perception of systemic fairness.

If visual and auditive information is synchronized, users are more comfortable using a system (Qiu & Benbasat, 2009). Providing visualizations to augment explanations is recommended to increase their effectiveness and users' satisfaction (Lim & Dey, 2009). The lack of visualizations in voice-based interactions makes it more complicated to process information by users (Rhee & Choi, 2020). As users develop their fairness perceptions based on available and absorbed information, an increased information level through visualizations could positively affect fairness perceptions (Kordzadeh & Ghasemaghaei, 2021; Favaretto et al., 2019; Chen & Chou, 2012). Cheng et al. (2019) compared different explanation styles for university admission decisions and identified that broad visual explanations support users' understanding of the algorithm. Different visualization techniques, such as text-based and scatterplots for outcomes and decisions, affected fairness perceptions of the predictors leveraged for the analysis (van Berkel et al.,

2021). Furthermore, visualizations of inputs and outputs in an algorithmic context helped users to better understand the outcome and made them perceive the outcome as fairer (Lee et al., 2019). Previous research indicates how the four underlying fairness dimensions of the second-order systemic fairness might be impacted (Kordzadeh & Ghasemaghaei, 2021; Vimalkumar et al., 2021). Visualizations, including the explanations and product recommendations, help the user to understand how the system works and why products are recommended. Such an increased understanding could support the perception of procedural and distributive fairness (Vimalkumar et al., 2021). Furthermore, an increased information level due to visualizations includes broader non-verbal and meta-information, which constitute the base for perceptions of interpersonal and informational fairness. Therefore, process visualizations could also positively affect the perception of interpersonal and informational fairness. As the second-order systemic fairness aggregates all four fairness dimensions and previous research proposes that visualizations could impact users' fairness perception (Kordzadeh & Ghasemaghaei, 2021), we hypothesize:

H3: Providing process visualizations for VAPRs has a positive effect on the perception of systemic fairness.
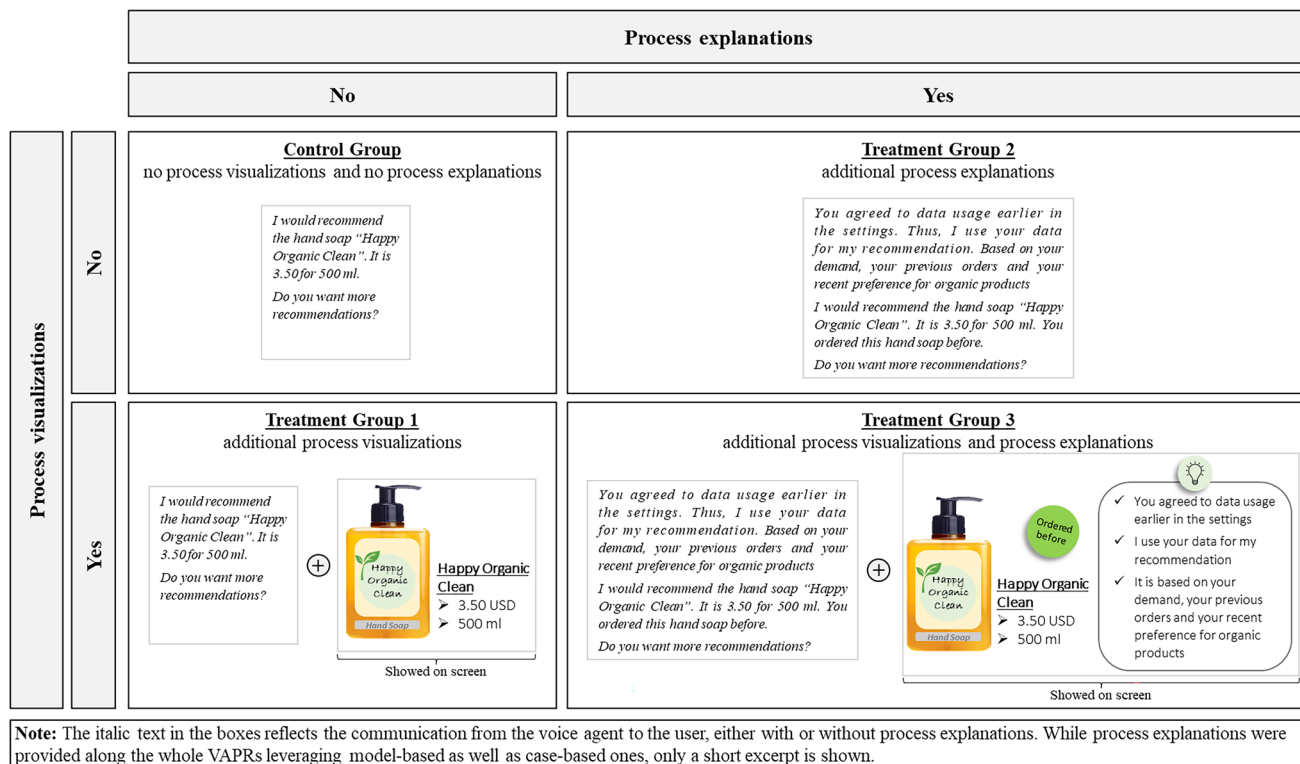

Fig. 4 Overview of control group and treatment groups

Methodology

Experimental design

We test the hypotheses using a 2 (process explanations: yes vs. no) × 2 (process visualization: yes vs. no) between-subjects experiment comparing four groups (Fig. 4). The underlying scenario for the experiment covered a VAPR interaction between a voice agent and a user, inspired by VAPRs of established e-commerce providers. First, the voice agent recognized the users' product needs, then several products were recommended, and finally, the user selected and purchased one. Within the control group, participants only received an audio file presenting the interaction between the user and the voice agent without any additional process explanations. To manipulate process explanations for treatment groups 2 and 3, we provided tangible information explaining how the recommendation engine operates as well as why the products are recommended, thus leveraging both model-based as well as case-based process explanation types. To manipulate process visualizations for treatment groups 1 and 3, we expanded the audio-based interaction by additional visualizations, thus simultaneously presenting the audio content via a screen. Only content provided via audio was presented on the screen, and no additional content was added compared to the control group or treatment group 2. Thus, process visualizations for the control group represented only information about the product, while for treatment group 2, the process visualizations also included the process explanations in visualized form. Figure 4 provides an overview of the groups including an excerpt of the interaction and an example of additional process visualizations.

Compared to the control group, the participants were either prompted with additional process visualizations (treatment group 1), additional process explanations (treatment group 2), or both (treatment group 3). While participants in the control group and treatment group 2 only listened to an audio file, participants in treatment groups 1 and 3 watched a video file with audio. The video demonstrated a voice agent equipped with an additional screen, inspired by real examples such as the Amazon Echo Show 8 (Knote et al., 2021). The voice agent showed process visualizations such as the examples above (Fig. 4), synchronized with auditive communication. We chose hand soap as a product for the experiment for two reasons: first, hand soap is a familiar product to participants as it is used and purchased regularly; second, hand soap is a low-involvement product, as it is bought regularly involving limited effort and a relatively low price (Watkins, 1984). This is in line with the current adoption of voice commerce, which is mainly limited to low-involvement products (Chabria & Someya, 2020). We invented a fictional brand for the voice agent, the retailer, and the hand soaps to avoid brand biases (Lowry et al., 2008).

Measures

We leveraged previously validated and established constructs to measure the four fairness dimensions and behavioral responses and contextualized them slightly. We further added VAPR-specific indicators to the formative fairness constructs (Weith & Matt, 2022) and measured the second-order systemic fairness based on the disjoint two-stage approach and thus on the indicators of the first-order fairness dimensions (Sarstedt et al., 2019). Measures for the intention to purchase were based on Sia et al. (2009). The intention to reuse and the intention to repurchase were based on Benlian et al. (2012). All indicators were measured based on a seven-point Likert scale ranging from "strongly disagree" (1) to "strongly agree" (7). Besides questions focusing on the users' perceived fairness and behavioral responses, we further included questions regarding the participants' e-commerce behavior, their previous usage of voice agents, and further sociodemographic aspects. Before realizing the experiment, we conducted a pre-test with three researchers and four practitioners to validate the design of the experiment as well as the structure and content of the questionnaire. Based on their feedback, we undertook minor changes to the questionnaire. We further discussed the questionnaire with four potential study participants to identify any issues regarding their understanding. We subsequently made minor adjustments regarding the instructions for participants and the process flow of the questionnaire. That included, for example, the opportunity for participants to replay audio or video file at a later stage throughout the questionnaire in case they did not pay enough attention the first time.

Data collection

For study 1, we recruited the participants via the online platform Prolific. We defined three requirements to participate in the study. First, we required the participants to have used e-commerce services before, whereby 96% stated they had used e-commerce at least once a month. Additionally, participants were required to participate in the study not via a mobile phone but rather via laptop or tablet to ensure a proper demonstration of the video in treatment groups 1 or 3. Lastly, we expected the participants to know what voice agents are. Besides communicating these requirements upfront, we verified these and excluded participants who did not meet them. All participants were randomly assigned to one of the four groups. We highlighted that there were no right or wrong answers and that we were solely interested in the participants' individual opinions based on their first impressions, as such instructions can strengthen the participants' commitment with the experiment (Xu et al., 2014). We further introduced each part of the questionnaire by including graphical elements highlighting the participant's

progress and the focus of each subsequent section of the questionnaire. We removed a few samples if they were finished in less than 8 min or more than 40 min (slightly different cut-off points of 7 and 20 min resulted in overall similar results). To assess if participants had noticed the manipulation of the treatment group, we asked two questions about the use case demonstrated. First, we asked if the assigned use case included a video or only consisted of an audio file. Second, we asked if the voice agent provided additional process explanations. Eighty-seven percent of the participants responded correctly to both questions. We excluded the remaining 13%, as failing the manipulation check questions is associated with low attention to the questionnaire. The final sample of study 1 consisted of 235 participants from the UK (61% women; 39% men; 58% with Bachelor's or Master's degree; distribution across age ranges: 11% 15–24 years, 26% 25–34 years, 23% 35–44 years, 22% 45–54 years, 13% 55–64 years, 4% 65–74 years). In total, 93% of the participants had used voice agents before. Of these 93%, 29% used a voice agent less than a month, while 71% used it once or several times a month. 46% of the participants who used a voice agent before, used it once or several times a day.

Analysis

Measurement model testing

We assessed the research model using partial least squares structural equation modeling (PLS-SEM) using SmartPLS 4.0, which allows us to test the measurement model (factors and indicators) and the structural model (relationships between variables and their strength and direction), including second-order constructs (Sarstedt et al., 2019). To assess the second-order formative systemic fairness construct, we applied the disjoint two-stage approach (Sarstedt et al., 2019). To assess the level of collinearity, we calculated the variance inflation factor (VIF), with all values ranging between 1.178 and 2.400 and thus below the critical value of 3.3 (Diamantopoulos & Siguaw, 2006). Next, we analyzed the significance and relevance of the outer weights and loadings. In case the outer weights of the indicators are not significant, the outer loadings should be retained if they exceed a value of 0.5 (Hair et al., 2022). We eventually deleted four initial indicators of the formative constructs. For the disjoint two-stage approach, the latent variable scores

of the four fairness dimensions are integrated as indicators for the second-order construct, followed by applying the same assessment criteria used for the first-order construct (Chin, 2010). We assessed the second-order measurement with satisfying results, as the VIF values for the four fairness dimensions were between 2.032 and 2.852 and thus below any critical values, factor weights were significant ($p < 0.05$), and outer loadings were above 0.5. For the reflective behavioral response constructs, we assessed the internal reliability and convergent and discriminant validity. First, we calculated Cronbach's alpha (CA) to test the internal reliability, whereby all indicators exceeded the threshold of 0.7 (Hair et al., 2022). Second, to assess the convergent validity, we calculated the composite reliability (CR), average variance extracted (AVE), and the factor loadings of the reflective indicators. All indicators surpassed the threshold for composite reliability of 0.7 (Nunally & Bernstein, 1994). Furthermore, the indicators all exceeded the AVE threshold of 0.7 (Hair et al., 2022). Third, we calculated the Fornell-Larcker criterion to test the discriminant validity. The square root of the AVE surpassed the interconstruct correlations for all three factors (Hair et al., 2022).

Hypotheses testing

We divided the research model into two parts to apply the most suitable analysis to each test the respective hypotheses (Xu et al., 2014). To validate the effect of systemic fairness on the three user behavioral responses (H1 a–c), we assessed the structural model using bootstrap resampling methods with 10,000 samples. Table 1 outlines that systemic fairness has a significant positive effect on all three behavioral responses. It has a positive effect on the intention to purchase the recommended product (H1a) ($\beta = 0.669$; $p < 0.05$), the intention to reuse the voice agent for product recommendations (H1b) ($\beta = 0.657$; $p < 0.05$), and the intention to repurchase from the retailer (H1c) ($\beta = 0.630$; $p < 0.05$). Thus, the data supported the structural hypothesis H1 a–c.

To identify the effect of the two information provision measures (process explanations and process visualizations) on fairness perceptions (H2 and H3), we dummy-coded these two as binary variables and conducted a group comparison using a two-way ANOVA. Furthermore, the aim was to identify the potential interaction effects of the two measures (Kim, 2014). Besides the two-way ANOVA to test

Table 1 Path coefficients and R^2 for H1 a–c	Path coefficients (β)	R^2
Systemic fairness (H1a) ➔ Intention to purchase recommended product	0.669*	0.447
Systemic fairness (H1b) ➔ Intention to reuse voice agent for product recommendations	0.657*	0.432
Systemic fairness (H1c) ➔ Intention to repurchase from the retailer	0.630*	0.397

* p < 0.05

Table 2 Descriptive statistics for systemic fairness and its four subdimensions

Groups	Systemic fairness		Procedural fairness		Distributive fairness		Interpersonal fairness		Informational fairness	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Control group	5.919	0.641	5.897	0.694	5.766	0.737	6.296	0.652	5.715	0.802
Treatment group 1	5.852	0.594	5.812	0.700	5.643	0.776	6.312	0.562	5.642	0.691
Treatment group 2	6.032	0.513	5.916	0.654	6.046	0.605	6.302	0.535	5.864	0.596
Treatment group 3	6.080	0.556	5.945	0.658	6.138	0.579	6.315	0.554	5.921	0.735

SD = standard deviation

H2 and H3, we also ran an additional two-way MANOVA with the four fairness dimensions as dependent variables. The MANOVA enabled us to test the effect of more than two groups on several dependent variables (Cole et al., 1993). The aim was therefore to understand how the four fairness dimensions are individually impacted by the two information provision measures (Table 2).

Before running the ANOVA, we deleted four data points as they were classified as univariate extreme outliers based on the z-score (Fidell & Tabachnikc, 2003). The Levene test demonstrated homogeneity of the error variances for systemic fairness ($F(3, 227) = 1.301$, $p = 0.275$), and the Shapiro-Wilk test revealed a normal distribution. Running the two-way ANOVA revealed no statistically significant effect of the interaction of the two variables process explanations and process visualizations on systemic fairness (Table 3). The data showed that providing process explanations has a positive effect on the second-order systemic fairness ($F(1, 227) = 4.934$, $p < 0.001$, $\eta^2 = 0.021$). Thus, the data supported H2. However, it did not reveal any significant effect by providing supporting process-visualizations (Table 3). Thus, H3 was not supported by the data.

Besides the ANOVA to test H2 and H3, to additionally understand how the four fairness dimensions are driving perceptions of systemic fairness, we ran a two-way MANOVA with the four fairness dimensions as dependent variables. Two prerequisites for the MANOVA were partly violated, as the Levene test demonstrated no homogeneity of the error variances for distributive fairness ($F(3, 227)$

$= 2.988$, $p = 0.032$), and the Shapiro-Wilk test revealed issues with normal distribution. However, the MANOVA is robust against small violations of the homogeneity of the error variances and the violation of the normal distribution, especially in the case of big and equal treatment groups (Finch, 2005). Furthermore, visual interpretation of the QQ-plots suggested approximately normally distributed data. Correlations between the four fairness dimensions were below 0.75, suggesting that multicollinearity is not critical (Dattalo, 2013). While the process-visualizations also had no significant effect on any of the four fairness dimensions, process explanations showed a significant effect on distributive fairness perception ($F(1, 227) = 18.869$, $p < 0.001$, $\eta^2 = 0.077$) as well as the informational fairness perception ($F(1, 227) = 5.101$, $p = 0.025$, $\eta^2 = 0.022$). The data showed no effect on the procedural ($F(1, 227) = 0.707$, $p = 0.401$, $\eta^2 = 0.003$) as well as the interpersonal fairness perception ($F(1, 227) = 0.004$, $p = 0.951$, $\eta^2 = 0.000$). Table 4 outlines the results from the two-way MANOVA for the four fairness dimensions.

Table 3 Results of between-subjects effects (ANOVA) for H2 and H3

	DV	F	η^2
Process explanations	Systemic fairness	4.934*	0.021
Process visualizations	Systemic fairness	0.015	0.000
Process explanations x Process visualizations	Systemic fairness	0.552	0.002

DV = dependent variable; F = F-value; *p < 0.05; η^2 = partial eta-square

Table 4 Results of between-subjects effects (MANOVA)

	DV	F	η^2
Process explanations	Procedural fairness	0.707	0.003
	Distributive fairness	18.869*	0.077
	Interpersonal fairness	0.004	0.000
	Informational fairness	5.101*	0.022
Process visualizations	Procedural fairness	0.096	0.000
	Distributive fairness	0.030	0.000
	Interpersonal fairness	0.034	0.000
	Informational fairness	0.007	0.000
Process explanations x Process visualizations	Procedural fairness	0.396	0.002
	Distributive fairness	1.460	0.006
	Interpersonal fairness	0.000	0.000
	Informational fairness	0.470	0.002

DV = dependent variable; F = F-value; *p < 0.05; η^2 = partial eta-square

Study 2—Content bases for process explanations

Hypothesis H2 from study 1 hypothesized a positive effect of process explanations on fairness perceptions, which was supported by the data. However, process explanations can further be differentiated regarding their content base. Thus, in study 2, we differentiated three selected process explanations based on their content (functional-based, product-based, user-based) to identify their impact on fairness perceptions. Large parts of the design were adopted from study 1 and thus only briefly described in the following.

Hypothesis development

Previous research suggested that explanations impact the users' evaluation of a recommendation (Xiao & Benbasat, 2007) and their fairness perceptions (Kordzadeh & Ghasemaghaei, 2021). Providing detailed information about the inner working logic of a recommendation engine can support the users' understanding of the recommendation. However, too much technical focus can also cause negative consequences (Zhao et al., 2019), and users might not perceive it as useful (Dolata et al., 2021). As not every user has the skill set to understand technically complex explanations, explanations focusing on the output and leveraged data in general can strengthen users' fairness perceptions (Springer & Whittaker, 2019). Providing explanations about products by also considering alternative products and outlining the differences and trade-offs between the options has positive effects on building users' trust (Pu & Chen, 2007). Therefore, explanations relating to alternative products could also have a positive effect on fairness perceptions. According to Binns (2019), explanations comparing a result to previous similar results had the highest impact on fairness perceptions. Such an approach comes close to user-based explanations, whereby a current product recommendation is explained in the context of a user's previous order history or the one of similar users. Recommendations matching the users' profiles and purchase behavior help users to have easier

access directly to the relevant information, which helps them to make purchasing decisions (Tiihonen & Felfernig, 2017). Furthermore, providing such process explanations to the user positively affects the users' attitude toward a product compared to non-personalized ones (Rhee & Choi, 2020). Research has demonstrated that personalized recommendations are generally perceived as useful; however, if they are too intrusive, users might also react negatively (Nguyen & Hsub, 2022). Furthermore, process explanations focusing on the recommended product are more personal than the ones focusing on the recommendation engine, especially user-based process explanations, as they consider users' profiles and behaviors the most. Therefore, we hypothesize:

> *H4: Perceived systemic fairness is the lowest for the functional-based process explanations and increases for product-based and user-based ones, with the highest value for the user-based process explanations.*

Methodology

Experimental design and measures

To test the manipulation of the process explanations, we conducted an online between-subjects experiment whereby we leveraged the same use case as for study 1. It covered the same products and same overall interaction structure between the voice agent and the user but included different process explanations. While the process explanations in study 1 were based on all three content bases jointly, we now differentiated them further across the groups (Table 5). Each of the groups' audio files consisted of the interaction between the voice agent and the user and leveraged only one content base for the process explanations. We also included an additional control group without any explanations. We made use of the questionnaire from study 1 as a base and adjusted context-specific aspects such as the attention check questions. We also adjusted the indicators for the formative constructs based on the assessment and results of the measurement model of study 1.

Table 5 Overview of control group and treatment groups

Groups	Control group	Treatment group 1	Treatment group 2	Treatment group 3
Content base for process explanations	No explanations (control group)	Functional-based	Product-based	User-based
Examples	-	"By using algorithms and data analysis techniques, I develop product recommendations." "I compare similar user profiles with each other"	"Compared to many other soaps, "Happy organic clean" is without any additional perfume." "This soap currently has the best rating of all soaps with 4.9 out of 5.0"	"You recently preferred organic products such as soap, washing detergent, and cleaning products." "Other users with a similar profile like yours"

Data collection

For study 2, we used a university mailing list to recruit participants and introduced the experiment as part of a university innovation project on the development of voice agents, to further strengthen participants' commitment. In line with study 1, we required participants to have experience with e-commerce and an understanding of voice agents. Each participant was randomly assigned to one of the four groups. As in study 1, we stated that there were no right or wrong answers and that we were solely interested in their individual perceptions. Each participant could take part in a draw with the possibility of winning one of eight $50 vouchers for an online shop. We removed those participants that did not meet these requirements, or which took less than 8 or more than 40 min to finish the study. As in study 1, we again included an attention check to ensure that the participants paid attention and perceived the manipulation correctly. Therefore, after listening to the group-specific audio file, we asked participants about the type of process explanations that had been provided to them. The question was answered correctly by 93% of the participants, and the remaining 7% of the participants were removed from the data set. The final sample of study 2 comprised 318 participants from Switzerland (70% women; 30% men; 55% with Bachelor's or Master's degree; distribution across age ranges: 69% 15–24 years, 28% 25–34 years, 2% 35–44 years, 1% 45–54 years). In total, 71% of all participants had used a voice agent before. Of these 71%, 70% used a voice agent less than a month, while 30% used it once or several times a month. 12% of the participants who used a voice agent before, used it once or several times a day.

Analysis

Measurement model testing

We applied the same steps as in study 1 to test the research model. After assessing the first level based on the four fairness dimensions, we again applied the disjoint two-stage approach (Sarstedt et al., 2019). On the first level, all VIF values ranged between 1.167 and 2.004 and thus below critical values (Diamantopoulos & Siguaw, 2006). Also, all outer weights were significant. Furthermore, we assessed the second-order measurement again with satisfying results, as the VIF values ranged between 1.956 and 2.234 and thus below any critical values, the weights

were significant ($p < 0.05$), and outer loadings were above 0.5. Also, the assessment of the reflective constructs led to satisfying results, which were further in line with the results from study 1. The values exceeded the threshold of 0.7 for Cronbach's alpha (Hair et al., 2022), the threshold of 0.7 for composite reliability (Nunally & Bernstein, 1994), and the threshold for the AVE of 0.7 (Hair et al., 2022). Lastly, the square root of the AVE was larger than the interconstruct correlations for all three factors.

Hypotheses testing

To ensure the validity, we again tested the impact of systemic fairness on all three behavioral responses. Systemic fairness positively affects the intention to purchase the recommended product ($\beta = 0.597$; $p < 0.05$), the intention to reuse the voice agent for product recommendations ($\beta = 0.588$; $p < 0.05$), and the intention to repurchase from the retailer ($\beta = 0.496$; $p < 0.05$). Thus, also the data of Study 2 supported H1 a-c (Table 6).

To test H4, we applied a contrast analysis, which allows for testing a specific order of variables, i.e., the order of the three process explanations regarding their effect on the perceived systemic fairness (Wiens & Nilsson, 2017). Besides the systemic fairness and thus testing H4, we furthermore conducted an additional contrast analysis for the four individual fairness dimensions to explore how they are impacted by the three process explanation content bases. Table 7 provides a summary of the means for the second-order systemic fairness as well as the four underlying fairness dimensions for each of the four groups.

Based on H4 and the order of independent variables, we coded the contrasts with lambda weights summing up to zero. The weights are −3 for the control group, −1 for functional-based, 1 for product-based, and 3 for user-based process explanations. Thus, they suggest that user-based process explanations have the highest effect on fairness perceptions, followed by product-based and functional-based ones. Before testing H4, we deleted two outliers and assessed for the homogeneity of the error variances. The Levene's test showed that the homogeneity of the error variances was not significant for systemic fairness (F(3, 312) = 0.536, $p = 0.658$), and the Shapiro-Wilk test revealed a normal distribution. The contrast analysis revealed significant differences between the three process explanations for systemic fairness (F(3, 312) = 5.229, $p = 0.002$). The Bonferroni post hoc test for systemic fairness (Table 8) revealed that only the user-based process explanations had a significant effect compared

Table 6 Path coefficients and R^2 for H1a-c

	Path coefficients (β)	R^2
Systemic fairness (H1a) ➔ Intention to purchase recommended product	0.597*	0.357
Systemic fairness (H1b) ➔ Intention to reuse voice agent for product recommendations	0.589*	0.347
Systemic fairness (H1c) ➔ Intention to repurchase from the retailer	0.496*	0.246

*p < 0.05

Table 7 Descriptive statistics for systemic fairness and its four subdimensions

Groups	Systemic fairness Mean	SD	Procedural fairness Mean	SD	Distributive fairness Mean	SD	Interpersonal fairness Mean	SD	Informational fairness Mean	SD
Control group	5.175	0.633	5.293	0.801	4.816	0.892	5.848	0.713	4.744	0.871
Treatment group 1	5.369	0.679	5.397	0.765	5.239	0.752	5.890	0.753	4.950	0.943
Treatment group 2	5.399	0.606	5.355	0.643	5.212	0.772	5.868	0.805	5.160	0.804
Treatment group 3	5.585	0.676	5.497	0.803	5.625	0.752	5.910	0.760	5.307	0.776

SD = standard deviation

to the control group. No other significant differences between the three process explanations were found. Thus, the data partially supported H4.

To better understand how the different process explanations impact the four fairness dimensions, we also ran a contrast analysis for the four fairness dimensions. The Levene's test showed that the homogeneity of the error variances was also not significant for all four fairness dimensions. The process explanations had significant effects on only two of the four fairness dimensions: the distributive fairness ($F(3, 312) = 13.583$, $p < 0.001$) and the informational fairness ($F(3, 312) = 6.595$, $p < 0.001$), which is in line with the results from study 1. There were

no significant differences for the procedural ($F(3, 312) = 1.028$, $p = 0.380$), nor for the interpersonal fairness dimension ($F(3, 312) = 0.097$, $p = 0.961$). The Bonferroni post hoc test for the distributive fairness showed that all three process explanations had a significant effect compared to the control group, with the highest value for the user-based ones (Table 9). Regarding the informational fairness, the product-based and the user-based process explanations were perceived as significantly fairer than the control group, with the higher effect resulting from the user-based ones. Table 9 provides an overview of the results from the Bonferroni test for distributive and informational fairness and highlights the significant effects.

Table 8 Bonferroni results for systemic fairness for H4

Fairness dimensions	Process explanations (I)	Process explanations (J)	Average differences (I-J)
Systemic fairness	Control group	Functional-based	−0.193
		Product-based	−0.223
		User-based	−0.409*
	Functional-based	Product-based	−0.029
		User-based	−0.215
	Product-based	User-based	−0.185

*p < 0.05

Table 9 Bonferroni results for distributive and informational fairness

Fairness dimensions	Process explanations (I)	Process explanations (J)	Average differences (I-J)
Distributive fairness	Control group	Functional-based	−0.423*
		Product-based	−0.395*
		User-based	−0.808*
	Functional-based	Product-based	0.027
		User-based	−0.385*
	Product-based	User-based	−0.412*
Informational fairness	Control group	Functional-based	−0.206
		Product-based	−0.416*
		User-based	−0.563*
	Functional-based	Product-based	−0.209
		User-based	−0.356
	Product-based	User-based	−0.146

*p < 0.05

Discussion and implications

Discussion

Table 10 provides an overview of all hypotheses tested in studies 1 and 2. Both studies revealed a positive effect of systemic fairness on all three behavioral responses (H1 a–c), indicating the relevance of perceived fairness as an important trigger of major e-commerce performance indicators. The two studies revealed the highest effect on purchase intentions and the lowest for repurchase intentions. This could be explained as purchase intention was the most closely related to the experimental use case, while reuse and repurchase intention affect future user behavior which is thus more difficult to forecast for participants (Limayem et al., 2007).

Study 1 revealed no effect of process visualizations on fairness perceptions (H3). Even though visualizations support the information processing by the users' brain (Rhee & Choi, 2020; Sweller et al., 2011) and thus allow for processing more information (Ocón Palma et al., 2020) which are required as a base to make up one's perception of fairness, our study did not find support for this effect. The results could potentially be explained due to the rather short interaction between the user and the voice agent, and the capabilities of the participants to digest information without supporting process visualizations. As the participants, especially in study 1, showed high previous usage experience with voice agents, they might already have become used to processing audio information well. Such previous experience could have impacted the results (Saeed & Abdinnour-Helm, 2008). Both studies consistently demonstrated an effect of explanations on systemic fairness (H2 and H4), which thus strengthens the validity of the research. Considering the perceived fairness based upon the four fairness dimensions throughout the

subsequent analysis, both studies revealed that only two of the four fairness dimensions are impacted by process explanations—distributional as well as informational fairness—while no effect on procedural or interpersonal fairness was identified. Potentially users were under the impression that they were, for example, not capable of assessing procedural fairness of VAPRs, as independent of the provided explanations, the VAPRs inner working model remained a black box (Rai, 2020). Even though we only saw an effect on two of the fairness dimensions, an effect on the second-order systemic fairness was revealed in both studies and, as such, stands above the individual fairness dimensions (Lee et al., 2011). Considering the explanations, the highest positive impact on perceptions of systemic fairness can be achieved by user-based process explanations, which are dedicated to the users' profiles and behaviors. Even though functional- and product-based explanations also showed positive impacts on two of the four fairness dimensions, only user-based explanations demonstrated a significant effect on systemic fairness. The prominent effect of user-based explanations is in line with previous research suggesting that personalizing VAPRs to users' profiles and purchase behaviors generally has positive effects on user satisfactions and responses (Rhee & Choi, 2020; Tiihonen & Felfernig, 2017), as long as users have the time and ability to process and understand the provided explanations (Zhao et al., 2019).

Theoretical implications

By addressing both VAPRs' inherent shortcomings through two information provision measures, we shed light on their simultaneous effect on the users' fairness perceptions. Thus, we outline the role of information contents and the medium to transfer information for fairness perceptions. While previous research adopted

Table 10 Hypothesis overview

Study		Hypothesis	Results
Study 1 + 2	H1	Perceived systemic fairness has a positive effect on …	
		a ... the intention to purchase the recommended product	Supported
		b …the intention to reuse the voice agent for product recommendations	Supported
		c ... the intention to repurchase from the retailer	Supported
Study 1	H2	Providing process explanations for VAPRs has a positive effect on the perception of systemic fairness.	Supported
Study 1	H3	Providing process visualizations for VAPRs has a positive effect on the perception of systemic fairness.	Not supported
Study 2	H4	Perceived systemic fairness is the lowest for the functional-based process explanations and increases for product-based and user-based ones, with the highest value for user-based ones.	Partly supported

the organizational fairness theory to AI (Dolata et al., 2021; Kordzadeh & Ghasemaghaei, 2021), we adopted it to AI-based VAPRs. We hereby expand the social perspective of AI fairness, which so far remains relatively underrepresented compared to the technical perspective (Dolata et al., 2021; Feuerriegel et al., 2020). Furthermore, while previous research on the social fairness perspective has taken a predominantly conceptual perspective, we respond to calls for more empirical evidence on fairness perceptions and subsequent consequences by providing quantified results through two empirical experiments (Kordzadeh & Ghasemaghaei, 2021). We also enrich the application fields of the AI fairness literature, as voice commerce research is currently scarce (von Zahn et al., 2021; Robert et al., 2020b).

Previous research demonstrated a notable focus on the procedural as well as distributive fairness dimensions but had neglected the interpersonal and informational dimensions (Kordzadeh & Ghasemaghaei, 2021; Robert et al., 2020b). Furthermore, the assessment of fairness based on the comprehensive second-order systemic fairness remains scarce. By considering all four fairness dimensions collectively, we provide a comprehensive understanding of fairness perceptions, which is necessary since interpersonal and informational fairness are essential to reflect the social relationships built between humans and computers (Yoo & Gretzel, 2011; Qiu & Benbasat, 2009). The results especially underpinned the importance of informational fairness, as both experiments demonstrated that informational fairness is impacted by process explanations. Moreover, since fairness perceptions are often conceptualized in relation to trust (Angerschmid et al., 2022; Dodge et al., 2019), our research can assist future research advances in trying to achieve a coherent understanding of both concepts.

By comparing different content bases for process explanations, we respond to previous calls for research by focusing on process explanations understandable by audiences without technical expertise (Kordzadeh & Ghasemaghaei, 2021; Dodge et al., 2019; Abdul et al., 2018). By delving into content bases for process explanations, we provide empirical evidence about the superiority of user-based process explanations, which is in line with previous research identifying positive effects of voice agents adjusting their messages to the individual user. However, while previous research focused on aspects such as a positive attitude towards a product or usefulness (Rhee & Choi, 2020; Nguyen & Hsub, 2022), we shed light on the effects on fairness

perceptions, especially crucial in the light of current AI fairness efforts. We furthermore expand an algorithm-centered perspective of recommendation engines' shortcomings by addressing the audio-based constraint through process visualizations motivated by the research stream of information processing theory (Sweller et al., 2011; Atkinson & Shiffrin, 1968). We address the proposition to empirically identify if and how additional visualizations impact fairness perceptions (Kordzadeh & Ghasemaghaei, 2021; Vimalkumar et al., 2021) by identifying that process visualizations have no impact on fairness perceptions.

Lastly, we provide empirical insights through validating how fairness perceptions impact users' behavioral responses. While previous research introduced the SOR theory for fairness of AI, we applied it explicitly to AI-based VAPRs (Kordzadeh & Ghasemaghaei, 2021). Based on the SOR theory and in line with previously identified positive effects of fairness perceptions on users' satisfaction or the tendency to accept algorithms (Carr, 2007), we identified positive effects on behavioral responses across both studies and thus underpinned the economic relevance of fairness perceptions.

Practical implications

Our research provides practitioners with empirical insights for the assessment, development, and design of VAPRs and presents process explanations as a feasible measure to provide information to prevent unjustified fairness perceptions. Since we did not identify any positive effect of additional process visualization for VAPRs on fairness perceptions, firms can focus on the development of VAPRs without additional visualizations. This provides businesses with increased flexibility for the application of VAPRs, as VAPRs do not require devices with screens to be perceived as fair. Practitioners should furthermore focus their resources on designing and providing suitable process explanations to embrace hands-free and eyes-off interactions with VAPRs. Here, practitioners should especially focus on explanations related to the individual users' profiles and purchase behaviors of users, as incorporated by user-based process explanations.

As only two of the four fairness dimensions are impacted by process explanations, practitioners might further wish to identify and apply measures to address the two remaining fairness dimensions. Besides explanations as information provision measures, endeavors to reduce biases and disparities of data and algorithms must be

pursued supplementary by businesses to achieve technical and objective fairness (Dolata et al., 2021; Fuchs et al., 2016). As we identified important behavioral responses resulting from fairness perceptions, these findings can help to convince budget holders to approve the resources required to develop and provide process explanations to ensure users perceive VAPRs as fair.

Besides focusing on the effects of VAPRs for themselves, firms should also bear in mind that perceptions of their VAPRs might also affect fairness perceptions of their competitors' VAPRs, especially in case of negative reputations. While previous research often focused on manipulating the actual product or service recommended or its prices (Weiler et al., 2022; Wu et al., 2022), which are more easily verifiable for the user, variations of content bases for process explanations are less researched and are even more subtle and more difficult to identify by users. However, this might enable them to be abused more easily without being detected. Therefore, our findings help to create awareness about the potential misusages of explanations, as retailers could apply them to cover up objectively unfair recommendations upon biased data and algorithms or provide recommendations leveraging higher margins in their own interest. Therefore, regulatory authorities should be aware of the effects of process explanations when assessing VAPRs and potentially providing fairness labels for them.

Conclusion, limitations, and future research

Despite their advantages for users, VAPRs are at risk of being perceived as unfair due to their AI-inherent shortcomings. Our project adds substantial value to the research field of fairness of AI and XAI by shedding light on the effect of information provision measures on users' perceived fairness of VAPRs. While process visualizations demonstrated no positive effect on fairness perceptions, our studies revealed a positive effect of process explanations on fairness perceptions. Among the different types of process explanations, user-based process explanations demonstrated the strongest positive effect on fairness perceptions. Moreover, by identifying the effect of the users' fairness perceptions

on behavioral responses, we could demonstrate the economic importance of fairness perceptions. The results support practitioners in prioritizing their resources by focusing only on voice-based interactions and providing user-based explanations. Practitioners should not focus on developing visualizations to increase the fairness of VAPRs. This could provide them with less complexity and more flexibility for the application and realization of VAPRs and their integration into related devices.

We acknowledge that our studies are not free of limitations. While we further expanded process explanations in study 2 to obtain more insights on the differential effects of various implementations of process explanation, we did not further consider variations of visualizations. Future research could test whether different types of process visualizations such as only text-based, only graphic-based, or interactive visualizations lead to different effects on fairness perceptions. To strengthen the generalizability, such visualization types could be further adapted to contextual factors, such as different products. Even though, from a practical perspective, the focus solely on audio-based VAPRs might be favorable from perspectives of cost and technological complexity, products of a higher complexity than soap might benefit from visualization as part of a transition phase to only audio-based VAPRs. Furthermore, as users' perception of fairness is driven by societal standards and personal experiences, it can therefore also vary depending on users' technological experiences and competences, which are also influenced over time (Kim & Malhotra, 2005). Therefore, fairness perceptions, as well as behavioral consequences, should be researched in a longitudinal setting to reveal any changes. We defined prerequisites regarding VAPRs, including general knowledge about voice agents, and our participants also demonstrated relatively high previous experience with VAPRs or willingness to use VAPRs in the future. Future research should investigate how users with different backgrounds and intentions perceive VAPRs as fair and how fairness perceptions might differ upon different VAPR applications. Future research should also investigate the potential and risk of retailers abusing process explanations for VAPRs to cover up objectively unfair product recommendations.

Appendix 1

Table 11 Overview of final constructs for studies 1 and 2

Constructs	Items	Origin
Procedural Fairness	The voice agent adapted the activities to the specific user needs.	Colquitt & Rodell (2015)
	The activities were neutral.	
	The voice agent made use of accurate information for its activities.	
	The voice agent's activities were in line with the values of the user.	
	The voice agent ensured a secure process.	Weith & Matt (2022)
Distributional Fairness	The recommendations were appropriate for the data the user provided.	Colquitt & Rodell (2015)
	The recommendations were justified.	
	The recommendations were personalized for the user.	Weith & Matt (2022)
	Alternative products were recommended.	
Interpersonal Fairness	The voice agent treated the user in a polite manner.	Colquitt & Rodell (2015)
	The voice agent treated the user appropriately.	
	The voice agent was competent.	Weith & Matt (2022)
Informational Fairness	The information provided by the voice agent were genuine.	Colquitt & Rodell (2015)
	The explanations provided by the voice agent were reasonable.	
	The voice agent communicated details in time.	
	The voice agent tailored the information to the user's needs.	
	The information provided by the voice agent were truthful.	Weith & Matt (2022)
Intention to purchase	I would consider purchasing from the voice agent Kim.	Sia et al. (2009)
	I would seriously contemplate purchasing from the voice agent Kim.	
	It would be likely that I am going to buy from the voice agent Kim.	
	I would likely make future purchases via the voice agent Kim.	
Intention to reuse	I would intend to continue using the voice agent (Kim) for a product recommendation.	Benlian et al. (2012)
	I predict I would use the voice agent (Kim) for a product recommendation.	
	I would plan to continue using the voice agent (Kim) for a product recommendation.	
	I would continue to consider using the voice agent (Kim) for a product recommendation.	
Intention to repurchase	I would intend to continue purchasing from the online shop (Kim's world).	Benlian et al. (2012)
	I predict I would purchase from the online shop (Kim's world).	
	I would plan to purchase from the online shop (Kim's world).	
	I would continue to consider using the online shop (Kim's world).	

References

Abdul, A., Vermeulen, J., Wang, D., Li, B.Y., & Kankanhalli, M. (2018). Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, Montreal, https://doi.org/10.1145/3173574.3174156

Angerschmid, A., Zhou, J., Theuermann, K., Chen, F., & Holzinger, A. (2022). Fairness and explanation in AI-informed decision making. *Machine Learning & Knowledge Extraction, 4*, 556–579. https://doi.org/10.3390/make4020026

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation, 2*, 89–195. https://doi.org/10.1016/S0079-7421(08)60422-3

Barocas, S., Hardt, M., & Narayanan, A. (2021). *Fairness in machine learning – Limitations and opportunities*. Nips Tutorial, 1, pp. 1–249. Retrieved July 7, 2022, from https://fairmlbook.org/

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review, 104*(671), 671–732. https://doi.org/10.15779/Z38BG31

Benlian, A., Titah, R., & Hess, T. (2012). Differential effects of provider recommendations and consumer reviews in e-commerce transactions: An experimental study. *Journal of Management*

Information Systems, 29(1), 237–272. https://doi.org/10.2307/41713875

Beugré, C. D., & Baron, R. A. (2001). Perceptions of systemic justice: The effects of distributive, procedural, and interactional justice. *Journal of Applied Social Psychology, 31*(2), 324–339. https://doi.org/10.1111/j.1559-1816.2001.tb00199.x

Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., & Shadbol, N. (2018). It's reducing a human being to a percentage - Perceptions of justice in algorithmic decisions. *ACM Conference on Human Factors in Computing Systems*, Montréal, https://doi.org/10.1145/3173574.3173951

Binns, R. (2019). On the apparent conflict between individual and group fairness. *Conference on Fairness, Accountability, and Transparency*, Barcelona, https://doi.org/10.1145/3351095.3372864

Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1), 1–12. https://doi.org/10.1177/2053951715622512

Carr, C. L. (2007). The FAIRSERV model: Consumer reactions to services based on a multidimensional evaluation of service fairness. *Decision Sciences, 38*(1), 107–130. https://doi.org/10.1111/j.1540-5915.2007.00150.x

Chabria, S., & Someya, M. (2020). *Cracking the code for voice commerce – Barriers and solutions to accelerate V-Comm adoption*. Ipsos Views, pp. 1-12, Retrieved January 10, 2023, from https://www.ipsos.com/sites/default/files/ct/publication/documents/2020-11/cracking-the-code-for-voice-commerce-2020.pdf

Chen, Y., & Chou, T. (2012). Exploring the continuance intention of consumers for B2C online shopping – Perspectives of fairness and trust. *Online Information Review, 36*(1), 104–125. https://doi.org/10.1108/14684521211209572

Cheng, H., Wang, R., Zhang, Z., O'Connell, F., Gray, T., Harper, F.M., & Zhu, H. (2019). Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholder. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Glasgow, pp. 1–12, https://doi.org/10.1145/3290605.3300789

Chin, W. W. (2010). How to write up and report PLS analyses. In V. Esposito Vinzi, W. W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of Partial Least Squares: Concepts, Methods and Applications, 2* (pp. 655–690). Springer Handbooks of Computational Statistics Series, https://doi.org/10.1007/978-3-540-32827-8_29

Cole, A. D., Maxwell, S. E., Arvey, R., & Sala, E. (1993). Multivariate group comparisons of variable systems: MANOVA and structural equation modeling. *Psychological Bulletin, 114*(1), 174–184. https://doi.org/10.1037/0033-2909.114.1.174

Colquitt, J. A., & Rodell, J. B. (2015). Measuring justice and fairness. In R. S. Cropanzano & M. L. Ambrose (Eds.), *The Oxford Handbook of Justice in the Workplace* (pp. 187–202). Oxford University Press.

Dattalo, P. (2013). Analysis of multiple dependent variables. *Oxford University Press*. https://doi.org/10.1093/acprof:oso/9780199773596.001.0001

Diamantopoulos, A., & Siguaw, J. A. (2006). Formative vs. reflective indicators in measure development: Does the choice of indicators matter? *British Journal of Management, 13*, 263–282. https://doi.org/10.1111/j.1467-8551.2006.00500.x

Dodge, J., Liao, Q.V., & Bellamy, R.K.E. (2019). Explaining models: An empirical study of how explanations effect fairness judgement. *Proceedings of the 24ᵗʰ International Conference on Intelligent User Interfaces*, Marina del Rey, pp. 1–10, https://doi.org/10.1145/3301275.3302310

Dolata, M., Feuerriegel, S., & Schwabe, G. (2021). A sociotechnical view of algorithmic fairness. *Information Systems Journal, 32*(4), 754–818. https://doi.org/10.1111/isj.12370

Ebrahimi, S., & Hassanein, K. (2019). Empowering users to detect data analytics discriminatory recommendations. *Proceedings of the 40ᵗʰ International Conference on Information Systems*, Munich, pp. 1–9, Retrieved December 15, 2021, from https://aisel.aisnet.org/icis2019/cyber_security_privacy_ethics_IS/cyber_security_privacy/39

Ebrahimi, S., & Matt, C. (2023). Not seeing the (moral) forest for the trees? How task complexity and employees' expertise affect moral disengagement with discriminatory data analytics recommendations. *Journal of Information Technology, 0*(0), 1–26. https://doi.org/10.1177/02683962231181148

Fang, Y., Qureshi, I., Sun, H., McCole, P., Ramsey, E., & Lim, K. H. (2014). Trust, satisfaction, and online repurchase intention. *MIS Quarterly, 38*(2), 407–427. https://doi.org/10.25300/MISQ/2014/38.2.04

Favaretto, M., Clercq, E. D., & Elger, B. S. (2019). Big data and discrimination: Perils, promises and solutions. A systematic review. *Journal of Big Data, 6*(12), 1–27. https://doi.org/10.1186/s40537-019-0177-4

Feuerriegel, S., Dolata, M., & Schwabe, G. (2020). Fair AI: Challenges and opportunities. *Business & Information Systems Engineering, 62*(4), 379–384. https://doi.org/10.1007/s12599-020-00650-3

Fidell, L.S., & Tabachnikc, B.G. (2003). Preparatory Data Analysis. In Schinka, J.A., & Velicer, W.F. (Eds.) *Handbook of Psychology, 2, Research Methods in Psychology*, John Wiley & Sons: New Jersey, pp. 115–142, https://doi.org/10.1002/0471264385.wei0205

Finch, H. (2005). Comparison of the performance of nonparametric and parametric MANOVA test statistics when assumptions are violated. *Methodology, 1*(1), 27–38. https://doi.org/10.1027/1614-1881.1.1.27

Følstad, A., & Kvale, K. (2018). Customer journeys: A systematic literature review. *Journal of Service Theory and Practice Journal of Service Theory and Practice, 28*(2), 196–227. https://doi.org/10.1108/JSTP-11-2014-0261

Friedrich, G. & Zanker, M. (2011). A taxonomy for generating explanations in recommender systems. *AI Magazine 32*, 90–98. https://doi.org/10.1609/aimag.v32i3.2365

Fuchs, C., Matt, C., Hess, T. & Hoerndlein, C. (2016). Human vs. algorithmic recommendations in big data and the role of ambiguity, *Proceedings of the 22nd Americas Conference on Information Systems (AMCIS)*, San Diego, USA, https://doi.org/10.7892/boris.105395

Gedikli, F., Jannach, D., & Ge, M. (2014). How should I explain? A comparison of different explanation types for recommender systems. *International Journal Human-Computer Studies, 72*, 367–382. https://doi.org/10.1016/j.ijhcs.2013.12.007

Greenberg, J. (1993). The social side of fairness: Interpersonal and informational classes of organizational justice. In Cropanzano, R., (Ed.). *Series in Applied Psychology. Justice in the Workplace: Approaching Fairness in Human Resource Management*, Lawrence Erlbaum Associates, New Jersey, pp. 79–103, https://doi.org/10.12691/jbms-3-5-1

Gretzel, U., & Fesenmaier, D. R. (2006). Persuasion in recommender systems. *International Journal of Electronic Commerce, 11*(2), 81–100. https://doi.org/10.2753/JEC1086-4415110204

Hair, J.F., Hult, G.T.M., Ringle, C., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)*. 3, Sage Publishing: Los Angeles, https://doi.org/10.1007/978-3-030-80519-7

Jannach, D., Manzoor, A., Cai, W., & Chen, L. (2021). A survey on conversational recommender systems. *ACM Computing Surveys, 54*(5), 1–36. https://doi.org/10.1145/3453154

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. *Proceedings of the 28ᵗʰ European Conference on Information Systems*, An Online AIS Conference, pp. 1–16, https://aisel.aisnet.org/ecis2020_rp/168

Kiesel, J., Meer, L., Potthast, M., & Stein, B. (2021). Meta-information in conversational search. *ACM Transactions on Information Systems, 39*(4), 1–44. https://doi.org/10.1145/3468868

Kim, H. (2014). Statistical notes for clinical researchers: Two-way analysis of variance (ANOVA)-exploring possible interaction between factors - Open Lecture on Statistics. *Restorative Dentistry & Endodontics, 39*(2), 143–147. https://doi.org/10.5395/rde.2014.39.2.143

Kim, S. S., & Malhotra, N. R. (2005). A longitudinal model of continued IS use: An integrative view of four mechanisms underlying post-adoption phenomena. *Management Science, 51*(5), 741–755. https://doi.org/10.1287/mnsc.1040.0326

Kim, J., Song, H., Kim, W., & Nam, J. (2020). I can feel your empathic voice: Effects of nonverbal vocal cues in voice user interface. *CHI '20 Extended Abstracts*, Honolulu, pp. 1–8, https://doi.org/10.1145/3334480.3383075.

Knote, R., Janson, A., Söllner, M., & Leimeister, J. M. (2021). Value co-creation in smart services: A functional affordances perspective on smart personal assistants. *Journal of the Association for Information Systems, 22*(2), 418–458. https://doi.org/10.17705/1jais.00667

Kordzadeh, N., & Ghasemaghaei, M. (2021). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems, 31*(3), 388–409. https://doi.org/10.1080/0960085X.2021.1927212

Kroll, J.A., Huey, J., Barocas, S., Felten, E.W., Reidenberg, J.R., Robinson, D.G, & Yu, H.L. (2017). Accountable algorithms. *University of Pennsylvania Review, 165*(3), pp. 633–705, Retrieved March 3, 2023, from https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3

Kuempel, A. (2016). The invisible middlemen: A critique and call for reform of the data broker industry. Northwestern Journal of International Law & Business, 36(1), pp. 207–34, Retrieved October 8, 2022, from https://scholarlycommons.law.northwestern.edu/njilb/vol36/iss1/4

Lee, K., Joshi, K., & Kim, Y. K. (2011). Identification of the four-factor structure of users' perceived fairness. *Journal of Targeting, Measurement and Analysis for Marketing, 19*(2), 113–126. https://doi.org/10.1057/jt.2011.12

Lee, M. K., Anuraag, J., Hae, J. C., Shashank, O., & Kusbit, D. (2019). Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation. *Proceedings of the ACM on Human-Computer Interaction, 3*(CSCW), 1–26. https://doi.org/10.1145/3359284

Lim, B. Y., & Dey, A. K. (2009). Assessing demand for intelligibility in context-aware applications. *UbiComp '09: Proceedings of the 11th International Conference on Ubiquitous Computing*, pp. 195–204, https://doi.org/10.1145/1620545.1620576

Limayem, M., Hirt, S. G., & Cheung, C. M. K. (2007). How habits limits the predictive power of intention: The case of information systems continuance. *MIS Quarterly, 31*(4), 705–737. https://doi.org/10.2307/25148817

Lowry, P. B., Vance, A., Moody, G., Beckman, B., & Read, A. (2008). Explaining and predicting the effect of branding alliances and web site quality on initial consumer trust of e-commerce web sites. *Journal of Management Information Systems, 24*(4), 199–22. https://doi.org/10.2753/MIS0742-1222240408

Mavlanova, T., Benbunan-Fich, R., & Koufaris, M. (2012). Signaling theory and information asymmetry in online commerce. *Information & Management, 49*(5), 240–24. https://doi.org/10.1016/j.im.2012.05.004

Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. The MIT Press.

Mehta, N., Steinman, D., & Murphy, L. (2016). *User success – How innovative companies are reducing churn and growing recurring revenue*. John Wiley & Sons.

Natale, S., & Cooke, H. (2021). Browsing with Alexa: Interrogating the effect of voice assistants as web interfaces. *Media, Culture & Society, 43*(6), 1000–1016. https://doi.org/10.1177/0163443720983295

Nguyen, T. K., & Hsub, P. (2022). More personalized, more useful? Reinvestigating recommendation mechanisms in e-commerce. *International Journal of Electronic Commerce, 26*(1), 90–122. https://doi.org/10.1080/10864415.2021.2010006

Nunally, J. C., & Bernstein, I. H. (1994). The assessment of reliability. *Psychometric Theory, 3*, 248–292.

Nunes, I. & Jannach, D. (2017). A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction 27*, 393–444. https://doi.org/10.1007/s11257-017-9195-0

Ochmann, J., Zilker, S., & Laumer, S. (2021). The evaluation of the black box problem for AI-based recommendations: An interview-based study. *Proceedings of the International Conference of Wirtschaftsinformatik*, Duisburg Essen, pp. 232–246, https://doi.org/10.1007/978-3-030-86797-3_16

Ocón Palma, M. C., Seeger, A., & Heinzl, A. (2020). Mitigating information overload in e-commerce interactions with conversational agents. In Davis, F.D., Riedl, R., vom Brocke, J., Léger, P., Randolph, A. (Eds.). *Information Systems and Neuroscience*, Springer: Cham, pp. 221–228, https://doi.org/10.1007/978-3-030-28144-1_24

Park, S. K., Han, Y., & Sela, A. (2020). Alexa, I want it now: How conversational artificial intelligence agents shape consumer decisions. *Advances in Consumer Research North American Advances, 48*, 578–579. https://doi.org/10.1002/mar.21654

Pu, P., & Chen, L. (2007). Trust-inspiring explanation interfaces for recommender systems. *Knowledge-Based Systems, 20*(6), 542–556. https://doi.org/10.1016/j.knosys.2007.04.004

Qiu, L., & Benbasat, I. (2009). Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems, 25*(4), 145–181. https://doi.org/10.2753/MIS0742-1222250405

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science, 48*, 137–141. https://doi.org/10.1007/s11747-019-00710-5

Rhee, C. E., & Choi, J. (2020). Effects of personalization and social role in voice shopping: An experimental study on product recommendation by a conversational voice agent. *Computers in Human Behavior, 109*, 106359. https://doi.org/10.1016/j.chb.2020.106359

Robert, L. P., Bansal, G., Melville, N., & Stafford, T. (2020). Introduction to the special issue on AI fairness, trust and ethics. *AIS Transactions on Human-Computer Interaction*, *Special Issue on AI Fairness, Trust, and Ethics, 12*(4), 172–177. https://doi.org/10.17705/1thci.00134

Robert, L. P., Pierce, C., Marquis, L., Kim, S., & Alahmad, R. (2020). Designing fair AI for managing employees in organizations: A review, critique, and design agenda. *Human-Computer Interaction, 35*(5–6), 545–575. https://doi.org/10.1080/07370024.2020.1735391

Saeed, K. A., & Abdinnour-Helm, S. (2008). Examining the effects of information system characteristics and perceived usefulness on post adoption usage of information systems. *Information & Management, 45*(6), 376–386. https://doi.org/10.1016/j.im.2008.06.002

Sarstedt, M., Hair, J. F., Cheah, J. H., Becker, J. M., & Ringle, C. M. (2019). How to specify, estimate, and validate higher-order constructs in PLS-SEM. *Australasian Marketing Journal, 27*(3), 197–211. https://doi.org/10.1016/j.ausmj.2019.05.003

Shin, D. (2020). User perceptions of algorithmic decisions in the personalized AI system: Perceptual evaluation of fairness, accountability, transparency, and explainability. *Journal of Broadcasting*

& Electronic Media, 64(4), 541–565. https://doi.org/10.1080/08838151.2020.1843357

Sia, C.L., Lim, K.H., Leung, K., Lee, M.K.O., Huang, W.W., & Benbasat, I. (2009). Web strategies to promote internet shopping: Is cultural-customization needed? *MIS Quarterly, 33*(3), pp. 491–512, Retrieved October 12, 2023, from https://www.jstor.org/stable/20650306

Springer, A. & Whittaker, S. (2019). Making transparency clear: The dual importance of explainability and auditability. *Joint Proceedings of the International Conference on Intelligent User Interfaces, Workshops*, Los Angeles, Retrieved October 23, 2022, from https://ceur-ws.org/Vol-2327/IUI19WS-IUIATEC-5.pdf

Statista (2022). *Forecast of e-commerce transactions value via voice assistants worldwide in 2021 and 2023*. Retrieved February 5, 2023, from https://www.statista.com/statistics/1256695/ecommerce-voice-assistant-transactions/

Sweller, J., Kalyuga, S., & Ayres, P. (2011). *Cognitive load theory – Explorations in the learning sciences*. Instructional Systems and Performance Technologies, Springer.

Tiihonen, J., & Felfernig, A. (2017). An introduction to personalization and mass customization. *Journal of Intelligent Information System, 49*, 1–7. https://doi.org/10.1007/s10844-017-0465-4

Tintarev, N., & Masthoff, J. (2007a). Explaining recommendations: Design and evaluation. In Ricci, F., Rokach, L., Shapira, B. (Eds.) *Recommender Systems Handbook*, Springer: New York City, pp. 470–474, https://doi.org/10.1007/978-1-4899-7637-6

Tintarev, N., & Masthoff, J. (2007b). Effective explanations of recommendations: User-centered design. *Proceedings of the 2007b ACM Conference on Recommender System*s, Minneapolis, pp. 153–156, https://doi.org/10.1145/1297231.1297259

Valentino-DeVries, J., Singer-Vine, J., & Soltani, A. (2012). *Websites vary prices, deals based on users' information*. The Wall Street Journal. Retrieved May 10, 2022, from https://www.wsj.com/articles/SB10001424127887323777204578189391813881534

Van Berkel, N., Goncalves, J., Russo, D., Hosio, S., & Skov, M.B. (2021). Effect of information presentation on fairness perceptions of machine learning predictors. *Proceedings of the 2021 CHI Conference on Human Factors in Computer Systems*, Virtual Conference, pp. 1–13, https://doi.org/10.1145/3411764.3445365

Vimalkumar, M., Gupta, A., Sharma, D., & Dwivedi, Y. (2021). Understanding the effects that task complexity has on automation potential and opacity: Implications for algorithmic fairness. *AIS Transactions on Human-Computer Interaction, 13*(1), 104–129. https://doi.org/10.17705/1thci.00144

Von Zahn, M., Feuerriegel, S., & Kuehl, N. (2021). The cost of fairness in AI: Evidence from e-commerce. *Business & Information Systems Engineering, 64*(3), 335–348. https://doi.org/10.1007/s12599-021-00716-w

Wang, Y., Ma, W., Zhang, M., Liu, Y., & Ma, S. (2023). A survey on the fairness of recommender systems. *ACM Transactions on Information Systems, 41*(3), 1–43. https://doi.org/10.1145/3547333

Watkins, T. (1984). Consumer purchasing of low-involvement goods: Routine or impulse? *Marketing Intelligence & Planning Marketing, 2*(2), 51–66. https://doi.org/10.1108/eb045700

Weiler, S., Matt, C., & Hess, T. (2022). Immunizing with information – Inoculation messages against conversational agents' response failures. *Electronic Markets, 32*(1), 239–258. https://doi.org/10.1007/s12525-021-00509-9

Weith, H. & Matt, C. (2022). When do users perceive artificial intelligence as fair? An assessment of AI-based B2C e-commerce. *Proceedings of the 55th International Conference on System Sciences*, Hawaii, pp. 4336–4345, https://doi.org/10.24251/HICSS.2022.529

Wiens, S., & Nilsson, M. E. (2017). Performing contrast analysis in factorial designs: From NHST to confidence intervals and beyond. *Educational and Psychological Measurement, 77*(4), 690–715. https://doi.org/10.1177/0013164416668950

Wu, Z., Yang, Y., Zhao, J. & Wu, Y. (2022). The impact of algorithmic price discrimination on consumers' perceived betrayal. *Frontiers in Psychology*. Article 825420, 1-12, https://doi.org/10.3389/fpsyg.2022.825420

Xiao, B., & Benbasat, I. (2007). E-commerce product recommendation agents: Use, characteristics, and effect. *MIS Quarterly, 31*(1), 137–209. https://doi.org/10.2307/25148784

Xu, J., Benbasat, I., & Cenfetelli, R. T. (2014). The nature and consequences of trade-off transparency in the context of recommendation agents. *MIS Quarterly, 38*(2), 379–406. https://doi.org/10.25300/MISQ/2014/38.2.03

Yoo, K.H & Gretzel, U. (2011). Creating more credible and persuasive recommender systems: The influence of source characteristics on recommender system evaluations. In Ricci, F., Rokach, L., Shapira, B. & Kantor, P.B. (eds.). *Recommender Systems Handbook*, Springer: New York, pp. 455–477, https://doi.org/10.1007/978-0-387-85820-3_14

Yoshikawa, T., Wang, Y., Kawai, Y. (2019). A product recommendation system based on user complaint analysis using product reviews. IEEE 8th Global Conference on Consumer Electronics (GCCE), Osaka, Japan, pp. 710–714. https://doi.org/10.1109/GCCE46687.2019.9015542

Zanker, M., & Ninaus, D. (2010). Knowledgeable explanations for recommender systems. *International Conference on Web Intelligence and Intelligent Agent Technology, Toronto,*. https://doi.org/10.1109/WI-IAT.2010.131

Zednik, C. (2021). Solving the black box problem: A normative framework for explainable artificial intelligence. *Philosophy & Technology, 34*, 265–288. https://doi.org/10.1007/s13347-019-00382-7

Zhao, R., Benbasat, I., & Cavusoglu, H. (2019). Do users always want to know more? Investigating the relationship between system transparency and users' trust in advice-giving systems. *Proceedings of the 27th European Conference on Information Systems*, Stockholm & Uppsala, Retrieved July 9, 2022, from https://aisel.aisnet.org/ecis2019_rip/42

Zhou, J., Verma, S., Mittal, M., & Chen, F. (2021). Understanding relations between perception of fairness and trust in algorithmic decision making. *The International Conference on Behavioral and Social Computing*, https://doi.org/10.48550/arXiv.2109.14345