Improving User's Confidence to Act when Using Advice Algorithms Through Interactive Use of Counterfactuals

JEROEN VAN GRONDELLE,[1]
MARLIES VAN STEENBERGEN,[1] ALETTA SMITS,[2]
MARCEL STALENHOEF,[2] KOEN VAN TURNHOUT,[2]
TIJS TIMMERMAN,[3] HARALD POL[4]

[1] HU University of Applied Sciences Utrecht, Digital Ethics, Utrecht, Netherland
jeroen.vangrondelle@hu.nl, marlies.vansteenbergen@hu.nl
[2] HU University of Applied Sciences Utrecht, Human Experience & Media Design, Utrecht, Netherland
aletta.smits@hu.nl, marcel.stalenhoef@hu.nl, koen.vanturnhout@hu.nl
[3] HU University of Applied Sciences Utrecht, Marketing & Customer Experience, Utrecht, Netherland
tijs.timmerman@hu.nl.
[4] Hanze University of Applied Sciences Groningen, Marketing and Entrepreneurship, Groningen, Netherland
h.g.pol@pl.hanze.nl

In this paper, we explore the design of web-based advice robots to enhance users' confidence in acting upon the provided advice. Drawing from research on algorithm acceptance and explainable AI, we hypothesise four design principles that may encourage interactivity and exploration, thus fostering users' confidence to act. Through a value-oriented prototype experiment and value-oriented semi-structured interviews, we tested these principles, confirming three of them and identifying an additional principle. The four resulting principles: (1) put context questions and resulting advice on one page and allow live, iterative exploration, (2) use action or change oriented questions to adjust the input parameters, (3) actively offer alternative scenarios based on counterfactuals, and (4) show all options instead of only the recommended one(s), appear to contribute to the values of agency and trust. Our study integrates the Design Science Research approach with a Value Sensitive Design approach.

Keywords:
advice robot, confidence to act, interactive, design science research, value sensitive design

University of Maribor Press

DOI https://doi.org/10.18690/um.fov.4.2024.12
ISBN 978-961-286-871-0

1 Introduction

As part of their digital transformation, public organisations increasingly offer personalised information and advice on complex services digitally and through algorithms (Chandra et al., 2022). However, organisations note that, when offered such digital advice, many citizens hesitate to act upon the advice. Before choosing or applying for the service, citizens often look for confirmation through the human channels first. In an earlier study, we argue that this phenomenon of lacking 'confidence to act' should, specifically in a public service context, be seen as more than a user experience (UX) or online conversion problem (Van Grondelle et al., 2023). We describe confidence to act as the extent to which a recipient feels that the service or advice offered is a solid basis to act upon, trusting both their own part and that of the service provider in the interaction. Confidence to act, or lack thereof, ties in with larger values, such as trust, agency and accessibility. The literature on acceptance of algorithmic outcomes suggests that being in control and able to override the algorithm (Dietvorst et al., 2018) and being part of a contrastive, interactive dialog rather than one-way communication (Miller, 2019; Smits & Van Turnhout, 2023) help to accept the outcomes. The research question we address in this paper is: *How can a web-based advice robot be designed to instill confidence to act by supporting interactive exploration*? To answer this question, we use concepts and insights from the field of explainable AI (XAI) and human-AI-interaction, and specifically existing knowledge on the role of counterfactuals in understanding cause-effect underneath algorithmic decisions. Based on existing theory, a Design Science Research (DSR) approach is combined with Value Sensitive Design (VSD) techniques to create a design artefact, evaluate it in a concrete organisational context, yet also offer a generalisation path towards similar problem contexts. Combining the VSD techniques 'value-oriented prototype' and 'value-oriented semi-structured interview' (Friedman & Hendry, 2019), we aim to involve people that may be less comfortable or even unable to contribute in a more conceptual, cognitive interview approach, while addressing the breadth and conceptual nature of the confidence to act construct.

In the next section we discuss the theoretical background to our study, which is primarily rooted in explainable AI. Section 3 discusses how we employed a combination of Design Science Research and Value Sensitive Design as our research method. The findings of our study are presented in section 4, followed by discussion and conclusion in section 5.

2 Theoretical background

The degree to which people use available digital tools has been widely studied by the research field of technology adoption (Ajzen, 1991; Venkatesh et al., 2003; Rogers, 2003). In this study, however, we do not focus on the initial intention to use advice tools, but the intention, after using the tool, to follow up on the advice provided. This 'confidence to act' emerged in a previous study as a characteristic of what citizens regard personal public service (Van Grondelle et al., 2023), but is as yet underconceptualized in literature. It seems to combine elements of trust in the advice provided, and self-confidence of having provided the right information to the organisation and of having understood the advice correctly.

Users being hesitant to adopt the outcomes of algorithms is a documented issue. A number of factors contribute to this. Logoni et al. (2019) introduced the construct of uniqueness neglect to capture the phenomenon that users expect algorithms to be less able to address their unique, individual situations. Castelo et al. (2019) found that the perception of high subjectivity in the task diminishes the level of trust in algorithms. Also, users may expect algorithms to "just" maximize a one-dimensional utility function and therefore trust algorithms less in ethically complex domains (Dietvorst & Bartels, 2022). Similarly, users are less likely to trust algorithms in uncertain domains where the task includes an element of prediction (Dietvorst & Bharti, 2020). There appears to be a level of debate to the question whether algorithm aversion is non-compensatory or can be compensated by other factors, such as objectively delivering better results and accuracy (Pezzo & Beckstead, 2020). Also, allowing users to exert some control over the algorithm, allowing them to (slightly) override or course-correct its decisions helps overcome algorithm aversion to an extent (Dietvorst et al., 2018; Hekman et al, 2022; Kleemann & Ziegler, 2023; Tintarev & Masthoff, 2015).

In the context of XAI, much recent work focuses on explaining complex decisions taken by algorithms. Miller (2019) draws from social sciences when he indicates 4 criteria for successful explanation as it occurs in a human-to-human context: it is contrastive, in the sense that it takes into account and addresses the expectations of the explainee; it is social, in that it is part of an interactive dialog aiming to transfer knowledge between two parties; it is often selective, in the sense that it does not provide a complete list of factors, but instead presents a subset of relevant and sufficient causes for the decision that is explained; and it should not try to convey underlying probabilistic relations, as they are often too subtle to be comprehended by many people. In addressing these objectives, different techniques are used to identify the attribution of different factors to a decision, identify the most decisive factors and use those as the basis for explanation. Determining the relative importance of factors in physical and economical models has often been done using variance-based sensitivity analysis. Techniques like Sobol (2001) and its successors introduce random disturbances in the inputs and quantify their impact on the result. Other techniques are based on the work of Shapley (1953) in economics on quantifying the contribution of a team member to the performance of the team. It has been applied for instance in explaining medical diagnostics (Ibrahim et al., 2020) and in financial forecasting (Jabeur et al., 2021). In machine learning context, often the gradients of underlying models are used to quantify the rate of change in the outcome in relation to the different inputs. In addition to identifying decisive factors, many of these techniques are also used to convey a measure of confidence: to which extent for instance a weather prediction can be expected to hold under slightly different conditions (Parker, 2010). A group of explanation techniques that heavily focus on conveying cause-effect relations underneath an algorithm-derived prediction or decision, is counterfactuals. Counterfactuals are scenarios where one or more of the inputs are different, and that (given the underlying model that is to be explained) have a different outcome. Although they have their background in philosophical logic concerning causality, more recently they have been studied and applied in the context of XAI, conveying underlying causality as part of explanations. Various criteria have been found for the successful application of counterfactuals in explaining decisions. Successful counterfactuals have *proximity* to the actual situation of the explainee, presenting smaller rather than larger deviations while still reaching a different outcome (Verma et al., 2019). Related, successful counterfactuals are *sparse* in the number of factors they change to reach a different outcome (*ibid.*). And successful counterfactuals are found to be *actionable*, in the sense that they offer the explainee a realistic course of action if an undesirable decision is explained (Poyiadzi et al., 2020).

3 Research Method

We applied a DSR approach (Hevner et al., 2004; Peffers et al., 2007) combined with VSD techniques (Friedman et al., 2006) to develop an artefact in the form of UX design principles that address the lack of confidence to act. We conducted the following steps: (1) Formulate design principles based on literature, (2) Develop a prototype based on the design principles, distinguishing four scenarios, (3) Conduct think-aloud sessions followed by interviews, looking for underlying values, (4) Analyse the sessions using a combination of open coding and template coding, and (5) Refine the design principles.

From the literature as described in the theoretical background, we derived four basic principles. Next, we operationalised these principles in a web-based advice robot prototype that gives advice on public transport subscriptions for travelling in a large city in the Netherlands. To better distinguish the individual effects of the principles, we developed various versions of the prototype. Over two days we conducted 11 think-aloud sessions with current clients of the public transport organisation broadly ranging in situation and education. We opted for think-aloud sessions followed by interviews because that allowed us to explore the underlying values and beliefs motivating the participants' actions. The participants were recruited from the transport organisation's client panel. In this way we achieved a near-real life setting. Twelve persons were invited, but one person did not turn up. This number was primarily motivated by practical reasons. Each session lasted 60 minutes, starting with a brief introduction in which the setting was explained, including the independent role of the researcher, after which the participant was asked to find the best public transport subscription for three different cases. The cases were brief and written on cards. An example is: "You live on Mozartlaan in Rhoon and travel 4 days a week to the Melanchtonweg, for work. You are considering working one day less soon, but that is not yet certain. Find the most suitable product for you". Each case referred to a different version of the prototype. Each participant was presented with three out of the four versions. The versions were presented in varying combinations and orders. Each case was followed by a brief semi-structured interview, following an interview guide. In the think-aloud exercises and interviews we applied the VSD techniques value-oriented prototype and value-oriented semi-structured interview (Friedman & Hendry, 2019), focusing not on the UX factors themselves, but on the effect they have on the participants in terms of human values.

We did so mainly by asking open 'why'-questions. The sessions were conducted by two researchers, taking turns. The sessions were recorded on video. The recordings were transcribed and coded using Atlas.ti. First open coding was applied. The codes were then grouped into groups representing the design principles. In addition, a group was made for codes pertaining to confidence to act. Based on this an analysis was performed on the relation between each of the design principles and the participants' confidence to act on the advice provided.

4 Results

4.1. Design principles

From the literature we derived four design principles for interactive advice algorithms (table 1). Our first design principle addresses the lack of influence on the algorithm (Dietvorst et al, 2018) and lack of interactivity (Miller, 2019). Many current online advice modules apply a two-page approach: on the first page, all relevant context factors are entered through a form. On submittance, the algorithm computes an advice based on those factors and the second page presents the outcomes. Often, a back button is offered to return to the first page, to modify the factors if mistakes were made. Our first design principle proposes to remove the separation between input and output by combining them on one page. This allows changes to the inputs to be made more easily, encouraging users to explore the algorithm by trying out alternative scenarios.

Our second design principle is to formulate change-oriented questions to modify what essentially are "static" properties, instead of asking for the current situation. Often, questions in advice algorithms focus on the current situation, and users may be hesitant to change those values to fictitious ones, especially in the context of government websites. To address this hesitation, questions can be worded in terms of potential future changes. For instance, the context factor "number of days per week you work" can be modified using a change-oriented question "do you anticipate your number of workdays to change?". Design principle 3 uses sensitivity analysis to analyse the boundaries of the current advice. It distinguishes factors that require little change to impact the advice from factors that require a large change or that do not impact the advice at all. Visualising the boundaries of the input parameters in relation to the current advice guides users in effective scenario exploration. Design principle 4 uses counterfactual techniques to actively present alternative scenarios to the user, combining a potential change in input values and the new advice this change would generate.

Table 1: Design Principles for Confidence to Act

Design principle	Description	Example
1. Put context questions and resulting advice on one page and allow live, iterative exploration.	The questions about the current context are on the same page as the advice, allowing for interactive exploration of the effects of different inputs.	A change in the input parameters on the left-hand side of the page immediately leads to an update of the list of recommended subscriptions on the right-hand side of the page.
2. Use action or change oriented questions to change the context	Context can be asked initially, after which the questions are formulated "incrementally" or as a delta to the base context.	"Do you consider working more or fewer days?" 3 days less >> 3 days more (instead of having to correct the number of days you actually work)
3. Guide exploration of context factors based on advice boundaries Based on sensitivity analysis	Options that lead to a changed advice are highlighted. OR options that lead to the same advice are disabled.	If the current advice is valid regardless of the number of days you work, that change control is grayed out.
4. Guide exploration of context factors based on alternative scenarios Based on counterfactuals	A counterfactual algorithm runs in the background and offers advice: If your situation was X, our advice would be Y.	"If you go to your workplace one extra day each week, a 3-star subscription would be cheaper."

Our second design principle is to formulate change-oriented questions to modify what essentially are "static" properties, instead of asking for the current situation. Often, questions in advice algorithms focus on the current situation, and users may be hesitant to change those values to fictitious ones, especially in the context of government websites. To address this hesitation, questions can be worded in terms of potential future changes. For instance, the context factor “number of days per week you work” can be modified using a change-oriented question “do you anticipate your number of workdays to change?”. Design principle 3 uses sensitivity analysis to analyse the boundaries of the current advice. It distinguishes factors that require little change to impact the advice from factors that require a large change or that do not impact the advice at all. Visualising the boundaries of the input parameters in relation to the current advice guides users in effective scenario exploration. Design principle 4 uses counterfactual techniques to actively present alternative scenarios to the user, combining a potential change in input values and the new advice this change would generate.

4.2 Operationalisation

To test the design principles, we operationalised them in a prototype of a web-based advice robot for public transport subscriptions in a large city in the Netherlands. Although selecting a public transport subscription might seem a relatively simple task, in this particular case the complexity was considerable. The range of subscription models includes a geographical zoning system, pay-as-you-go components and discounts above a certain amount of travel. When combined, this leads to some unexpected, non-linear effects in total transportation cost, making the choice non-trivial for customers. Due to the zoning system, small changes of the origin or destination of the commute could have a high impact on the resulting total cost. The existing web application followed the typical scenario of a page with questions and a consecutive page with advice. The questions explored the need for public transport, on workdays and for leisure/other trips. The advice recommended a subscription, based on optimal total cost per month. Although great care was given to offer optimal advice, many online users still contacted the service centers on the stations for confirmation.

To implement principle 2 in the prototype, users are not asked to change source or destination address, but how many minutes they are willing to walk to and from the station. Similarly, they are asked whether they might decide to work more or less days in future. In the algorithm, this is used to shorten or lengthen the trip and/or change the frequency, and thus (depending on zone limits) increase or lower the subscription cost. Changing the answers to these questions leads to immediate updating of the advice portion of the screen (principle 1). To establish the advice boundaries, a naive, per-factor sampling of adjacent inputs is generated, and the outcome of the algorithm is compared to the outcome of the actual inputs. Based on this the input form shows, with a blue bar, which changes will not impact the outcome and are therefore less useful to explore (principle 3). Based on the same simulation a chatbot suggests input changes that are close to the current context yet lead to a different advice (principle 4). To be able to study the contribution of the four principles separately, four versions of the prototype were developed. In all versions input was provided by using sliders. The baseline version 0 reflected the current state, with two separate screens. All other versions offered questions and advice on a single page (principles 1 and 2). Version 2 additionally visualised with blue bars the advice boundaries in the input sliders (implementing principle 3). In version 3 a chatbot offered advice on how changes in input would lead to an alternative advice (principle 4). Figure 1 illustrates how all four principles were implemented in the versions of the prototype.

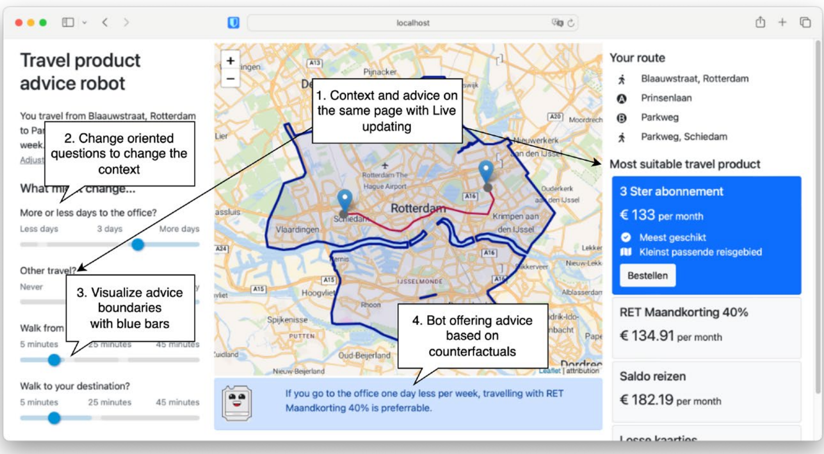

Figure 1: The prototype, with all 4 principles implemented

Source: Own printscreen

4.3 Evaluation

Below, we discuss how the participants (referred to as P#) experienced the advice robot in relation to the design principles and their confidence to act. All but one of the participants used version 0 and 1, while six participants used version 2 and five participants used version 3. One participant only used version 1 and 2.

Input and output on one page

Most participants appreciated the fact that their input and the robot's advice appeared on the same page. In version 0 participants do use the back button, but it is an obstacle. Participant P1 understands the purpose of the back button and remarks that normally they would make notes and then go back to change the parameters. When confronted with version 1 later, P1 appreciates the ease of use of the fact that changing the sliders back and forth immediately results in changes in advice on the same screen. Various participants in version 1 immediately focus on the initial list of subscriptions and reason about the possible effects of changing parameters instead of changing them in the prototype to see the effects. The back button is mainly used for confirmation of their own reasoning. P5, for instance, only starts to actively use the parameters in the one-page versions, subsequently expressing great enthusiasm for 'playing with the sliders'.

Playing with the sliders and seeing the immediate effects makes some participants reflect that there are more options than they had realized, which stimulates them to have a closer look at what they want and what subscription fits that need. Apart from just entering the parameters as they currently are, they can make conscious choices, such as choosing to walk farther, that may affect the price of their transportation costs. Some express great curiosity as to the effects of changing particular parameters. P2, for instance, freely plays with the sliders and uses the sliders to detect what parameters lead to a particular subscription, i.e. they reason from subscription to parameters instead of the other way round.

The fact that all options are shown instead of only the most suitable is appreciated by all participants. The fact that they can see the differences between options for themselves instills confidence. It disperses many doubts about whether the proposed subscription is really the best for them.

Change oriented questions

Though only one participant explicitly remarked on the fact that they liked the fact that they could indicate the number of minutes they were prepared to walk instead of having to select a station, the wording of the sliders seemed to encourage the participants to better think about their options. Especially the walking distance engendered deliberation. Not only about the time participants are prepared to walk, but it is also transferred to the option of taking the bicycle to the station. The precise wording of the sliders is extremely important though, as some participants had trouble interpreting some of them. Especially the formulation 'any other trips?' generated some doubts about its meaning.

Participants often used a mixed approach of reasoning about the list of subscriptions using their own previous experiences with public transport as well as experimenting with the input parameters. They use the list to compare the subscriptions mentioned, often using their own experiences and knowledge to reflect on the differences and suitability in different circumstances in addition to playing with the sliders. Some use the parameters to generate new outcomes, others use them primarily for confirmation of their own reasonings. The balance between reasoning and exploring differs between participants, ranging from accepting the top advice without further ado to extensively exploring all parameters and their effects on the list of options.

Passive presentation of advice boundaries

Even though a legend was given at the bottom of the page, the showing of advice boundaries in the form of a blue bar was noticed by none of the participants except P2, who did not understand what it was for and found it irritating. All participants had to be alerted to the blue bar and even after explanation it was not perceived as being helpful. On the contrary, if anything, it was experienced as confusing. P1 remarks that it is yet another aspect to consider, without adding much value. It makes a clear interface more complicated. Only P3 thought that marking the boundaries of the advice might be useful.

Active offering of alternative scenarios

The bot messages indicating that a particular change in parameters would lead to a cheaper subscription was better appreciated than the boundary markers in the sliders. P4 and P5 both spontaneously remarked on missing the messages when they moved from version 3 to version 1. P5 appreciates the bot messages especially because of their open phrasing. This is also connected to the format of the bot: it is not a pop-up that must be clicked away but a small panel integrated in the page and the wording is neutral, not trying to direct the user in a particular direction. Some participants, though, find the messages superfluous, confusing, or not applicable to their situation, and thus not actionable.

Confidence to act

Asked whether they would be confident to take a subscription based on the advice robot, most participants answered confirmatively. Various participants expressed that they feel confident to act upon the advice because they had been able to explore and check the various parameters by themselves. P2 appreciates having experienced for themselves that changing certain parameters doesn't make a difference as to the best subscription for them. Also, the fact that all options are shown seems to instill confidence. As P5 remarks, the fact that the options are shown in a simple list, instead of one option being shown 'full screen with a few other options tucked away in a corner' makes it more trustworthy for them. The fact that the option of not taking a subscription but buying individual tickets is included instills confidence, too. As do the tips provided by the bot in version 3, because the bot indicates what is needed to arrive at a cheaper subscription, instead of trying to promote a particular choice.

The most frequent reason for participants hesitating to take a subscription is the wish for further information on the exact subscription conditions, which were not included in the prototype. Various ways of acquiring this additional information are suggested, ranging from an information button to speaking to a person via chat or telephone.

5 Discussion and Conclusion

The experiments provide some support for the usefulness of three of the four design principles to instill confidence to act. Factors that seem to increase confidence to act are: being able to explore the effects of changing parameters (enabled by design principles 1 and 2), being presented with all options, not only the recommended one, in a neutral manner (enabled by principle 1), and being actively informed about how changes in input might affect the advice (principle 4). Despite being grounded in the literature on sensitivity analysis in XAI, no support is found for design principle 3. This may be due to the way it is implemented in the prototype. This requires further study. Showing all options in a neutral manner is enabled by principle 1, but its importance may warrant a separate design principle. The most important factor that inhibits confidence to act seems to be the lack of detailed information on what the recommended subscription entails. Most participants seem not to doubt that the recommended subscription is indeed the best option for their situation, but need more information about what exactly it allows or not. This gap seems easily mended by providing an information button for each of the subscriptions.

Combining the VSD techniques 'value-oriented prototype' and 'value-oriented semi-structured interview' allowed us to investigate the underlying values that influence confidence to act with a wide range of citizens. Using the prototype allowed for a diverse range of participants, selected from the actual clientele of the public transport system. By asking open questions about experiences, intentions, and motivations we were able to elicit impact on underlying values, thus moving beyond a merely functional UX study. The values that emerged most prominently were agency and trust. Agency was positively affected by design principle 1 and 2: being stimulated to play with input parameters while immediately seeing changes in recommendations gave a feeling of being in control of the selection process. Trust was positively affected by design principle 1 and 4: being presented with all options and being actively informed of how to arrive at cheaper options helped prevent mistrust about being manipulated to a particular unfavorable outcome. As may be expected, the wording of the parameters and information messages proves to be extremely important.

In this study we aimed to answer the research question '*How can a web-based advice robot be designed to instill confidence to act by supporting interactive exploration*?'. By deriving four design principles from literature followed by empirical validation, we arrived at three confirmed design principles: (1) put context questions and resulting advice on one page and allow live, iterative exploration, (2) use action or change oriented questions to adjust the input parameters, (3) actively offer alternative scenarios based on counterfactuals. In addition, from the experiments we derived a fourth principle: (4) show all options instead of only the recommended one(s).

This study has its limitations. The design principles were operationalized in a single use case, with a limited number of participants. Applying the design patterns to a different advice algorithm could solidify the current qualitative outcomes. Finally, this could give rise to a quantitative study where the impact on confidence to act is measured in terms of the number of people that no longer seek human confirmation after receiving digital advice.

We hope that our study contributes to the growing body of knowledge on the design and implementation of digital services that are of true service to citizens.

Acknowledgements

We would like to thank all citizens participating in our experiments. This research is co-financed by the Taskforce for Applied Research SIA, part of the Netherlands Organization for Scientific Research (NWO).

References

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-t

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. Journal of Marketing Research, 56(5), 809-825.

Chandra, S., Verma, S., Lim, W. M., Kumar, S., Donthu, N. (2022). Personalization in personalized marketing: Trends and ways forward. Psychology & Marketing, 39(8), 1529-1562.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. Management science, 64(3), 1155-1170.

Dietvorst, B. J., & Bharti, S. (2020). People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. Psychological science, 31(10), 1302-1314.

Dietvorst, B. J., & Bartels, D. M. (2022). Consumers object to algorithms making morally relevant tradeoffs because of algorithms' consequentialist decision strategies. Journal of Consumer Psychology, 32(3), 406-424.

Friedman, B., Kahn, P. H., & Borning, A. (2006). Value sensitive design and information systems. In P. Zhang & D. Galletta (Eds.), Human-computer interaction in management information systems:
Foundations (pp. 348–372). M.E. Sharpe.
Friedman, B., Hendry, D. G. (2019). Value Sensitive Design: Shaping Technology with Moral Imagination. The MIT Press.
Hekman, E., Nguyen, D., Stalenhoef, M. & Van Turnhout, K. (2022). Towards a Pattern Library for Algorithmic Affordances. Joint Proceedings of the IUI 2022 Workshops (pp. 24-33).
Hevner, A., March, S., Park, J., Ram, S. (2004). Design Science in Information Systems Research. MIS Quarterly 28, pp. 75—105
Ibrahim, L., Mesinovic, M., Yang, K. W., & Eid, M. A. (2020). Explainable prediction of acute myocardial infarction using machine learning and shapley values. Ieee Access, 8, 210410-210417.
Jabeur, S. B., Mefteh-Wali, S., & Viviani, J. L. (2021). Forecasting gold price with the XGBoost algorithm and SHAP interaction values. Annals of Operations Research, (pp. 1-21).
Kleemann, T. & Ziegler, J. (2023). Blending Conversational Product Advisors and Faceted Filtering in a Graph-Based Approach. In: Abdelnour Nocera, J., Kristín Lárusdóttir, M., Petrie, H., Piccinno, A., Winckler, M. (eds) Human-Computer Interaction – INTERACT 2023. INTERACT 2023. Lecture Notes in Computer Science, vol 14144. Springer, Cham. https://doi.org/10.1007/978-3-031-42286-7_8
Kumar, I. E., Venkatasubramanian, S., Scheidegger, C., & Friedler, S. (2020, November). Problems with Shapley-value-based explanations as feature importance measures. In International Conference on Machine Learning (pp. 5491-5500). PMLR.
Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. Journal of Consumer Research, 46(4), 629-650.
Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial intelligence, 267, 1-38.
Parker, W. S. (2010). Predicting weather and climate: Uncertainty, ensembles and probability. Studies in history and philosophy of science part B: Studies in History and Philosophy of Modern Physics, 41(3), 263-272.
Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of management information systems, 24(3), 45-77.
Pezzo, M. V., & Beckstead, J. W. (2020). Algorithm aversion is too often presented as though it were non-compensatory: A reply to Longoni et al. (2020). Judgment and Decision Making, 15(3), 449-451.
Poyiadzi, R., Sokol, K., Santos-Rodriguez, R., De Bie, T., & Flach, P. (2020, February). FACE: feasible and actionable counterfactual explanations. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (pp. 344-350).
Rogers, E.M. (2003). Diffusion of Innovations (5th ed.). New York: Free Press.
Smits, A. & Van Turnhout, K. (2023). T owards a Practice-Led Research Agenda for User Interface Design of Recommender Systems. In: Abdelnour Nocera, J., Kristín Lárusdóttir, M., Petrie, H., Piccinno, A., Winckler, M. (eds) Human-Computer Interaction – INTERACT 2023. INTERACT 2023. Lecture Notes in Computer Science, vol 14144. Springer, Cham. https://doi.org/10.1007/978-3-031-42286-7_10.
Tintarev, N. & Masthoff, J. (2015). Explaining Recommendations: Design and Evaluation. In: Ricci, F., Rokach, L., Shapira, B. (eds) Recommender Systems Handbook. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7637-6_10.
Shapley, L. S. (1953). A value for n-person games.
Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and computers in simulation, 55(1-3), 271-280.

van Grondelle, J., van Steenbergen, M., Smits, A., Timmerman, T., van Turnhout, K. & Pol, H. (2023). The Anatomy of a Personal Service: Eight dimensions of 'Personal'. Proceedings of the 36th Bled eConference, 483–495.
Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425-478. doi 10.2307/30036540
Verma, S., Boonsanong, V., Hoang, M., Hines, K. E., Dickerson, J. P., & Shah, C. (2020). Counterfactual explanations and algorithmic recourses for machine learning: A review. arXiv preprint arXiv:2010.10596.