Mind the (Persuasion) Gap: Contrasting Predictions of Intelligent DSS with User Beliefs to Improve Interpretability

Michael Chromik
LMU Munich
Munich, Germany
michael.chromik@ifi.lmu.de

Florian Fincke
LMU Munich
Munich, Germany
florian.fincke@campus.lmu.de

Andreas Butz
LMU Munich
Munich, Germany
butz@ifi.lmu.de

ABSTRACT

Decision support systems (DSS) help users to make more informed and more effective decisions. In recent years, many intelligent DSS (IDSS) in business contexts involve machine learning (ML) methods, which make them inherently hard to explain and comprehend logically. Incomprehensible predictions, however, might violate the users' expectations. While explanations can help with this, prior research also shows that providing explanations in all situations may negatively impact trust and adherence, especially for users experienced in the decision task at hand. We used a human-centered design approach with domain experts to design a DSS for funds management in the construction industry and identified a strong need for control, personal involvement, and adequate data. To create an adequate level of trust and reliance, we contrasted the system's predictions with the values derived from an analytic hierarchical process (AHP), which makes the relative importance of our users' decision-making criteria explicit. We developed a prototype and evaluated its acceptance with 7 construction industry experts. By identifying situations in which the ML prediction and the domain expert potentially disagree, the DSS can identify a persuasion gap and use explanations more selectively. Our evaluation showed promising results and we plan to generalize our approach to a wider range of explainable artificial intelligence (XAI) problems, e.g., to provide explanations with arguments tailored to the user.

CCS CONCEPTS

• **Human-centered computing → HCI design and evaluation methods**.

KEYWORDS

decision support systems; decision-making; interpretability; explainable artificial intelligence; analytical hierarchical process

ACM Reference Format:
Michael Chromik, Florian Fincke, and Andreas Butz. 2020. Mind the (Persuasion) Gap: Contrasting Predictions of Intelligent DSS with User Beliefs to Improve Interpretability. In *ACM SIGCHI Symposium on Engineering Interactive Computing Systems (EICS '20 Companion), June 23–26, 2020, Sophia Antipolis, France.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3393672.3398491

EICS '20 Companion, June 23–26, 2020, Sophia Antipolis, France

ACM ISBN 978-1-4503-7984-7/20/06…$15.00
https://doi.org/10.1145/3393672.3398491

1 INTRODUCTION

The rapidly growing volume of data in many parts of the enterprise makes it necessary to structure and manage it in information systems. Those systems which help in decision-making are referred to as *decision support systems (DSS)* [30]. With the recent improvements in *machine learning (ML)* methods, DSS are becoming more and more intelligent. So-called *intelligent decision support system (IDSS)* augment the collected data with predictions that guide and (semi-) automate parts of the decision-making process [30]. However, these intelligent DSS also introduced new challenges because their rationale is often not interpretable and hence perceived as non-deterministic by their users.

The effectiveness of an intelligent DSS depends not only on the accuracy of its underlying ML model or algorithm. Instead, it is only effective if it serves the information needs of decision-makers and is also accepted and trusted by them. Jarvis describes DSS as the general idea of "*combining the computer's computational power with the decision maker's intuition and judgment in an interactive manner, [so that] better decisions will result than by either the computer or human taken separately*" [11]. To achieve such a symbiosis, we need to design user interfaces (UI) that communicate the rationale behind algorithmic predictions in human-understandable terms. The UI should help to calibrate the user's understanding of the system's capabilities and limitations to prevent both over-reliance (when users blindly trust system recommendations) and under-reliance (when users simply ignore system recommendations) [5].

We conducted a design study in the construction industry and asked decision-makers about their requirements regarding interpretability of a novel intelligent DSS module on addenda approval. We use the term *interpretability* to refer to measures provided by a DSS with the aim of enabling users to understand its inner workings. Interpretability is a broad concept that may imply other distinct ideas such as transparency, trust, and fairness [17]. It is often used to indirectly evaluate whether important requirements, such as reliability, trust, or control are met in a particular context [8]. Biran and Cotton consider intelligent systems interpretable "*if their operations can be understood by a human*" [3]. We followed a human-centered design process to understand how project managers and executives make decisions regarding validation and approval of budget addenda. Budget management in the construction industry is an interesting context to study for two reasons: First, the construction industry itself is one of the least digitized industries but digitization efforts (e.g., building information modeling (BIM)) are

gaining adoption despite decision makers skepticism [1, 4]. Secondly, the addenda approval process is a complex decision situation that requires decision makers to retrieve and interpret data from distributed sources and also consider their dependencies. To date this is a highly manual and subjective process

This paper investigates interpretability needs of human decision-makers in the field regarding an intelligent DSS. In particular, we propose an approach to align the level of trust and reliance by contrasting ML predictions with user beliefs. User beliefs can be extracted through multi-attribute decision making approaches such as the *analytic hierarchical process (AHP)*. Making the user beliefs explicit allows the system to better identify *persuasion gaps* [6], i.e., situations in which the system and user base their decision on different criteria. We think that this approach might be a valuable starting point for providing selective and personalized explanations to the field of explainable artificial intelligence (XAI). With this work, we put our suggested approach and formative evaluation up for discussion with our fellow researchers.

2 RELATED WORK

2.1 Intelligent Decision-Support Systems

Decision-making refers to the cognitive process of selecting a logical choice from many available alternatives. Decision-making problems are often structured into three phases: problem identification, model development and use, and action plan development [21]. In our work, we focus on the second phase that deals with eliciting user preferences and comparing alternatives in a consistent way. If a decision is rational it is typically based on facts instead of arbitrary choices. *Multi-attribute decision making (MADM)* describes approaches that leverage (potentially conflicting) attributes to select, compare, and rank a limited number of discrete alternatives [31]. The rationality of decision-making, however, is bounded as individuals are often not able to make optimal decisions in an economically rational way due to cognitive limitations and resource constraints [28]. Simon suggests that instead of maximizing (search for the best possible option), decision makers in the field are rather satisficing (stick to an option that is considered good enough) [28].

In many business-related contexts, *decision support systems (DSS)* organize relevant facts to assist users in decision-making and improve effectiveness of the decision outcome [30]. DSS can range from simple spreadsheets to complex data warehouse systems [30]. They are typically distinguished by their underlying technology, theory foundations, target users, and decision tasks [2]. So called, *intelligent decision support systems (IDSS)* use artificial intelligence methods to support the decision-making and exhibit some notion of "intelligent behavior" [30]. Such intelligent behavior may either by applied to the system's underlying data base (e.g., identifying relevant attributes), knowledge base (e.g., suggesting decision alternatives), or model base (e.g., choosing applicable formal decision-making methods) [22].

In our work, we focus on IDSS that recommend decision alternatives to the user (*model development and use* at the *knowledge base*). The first generation of IDSS (also called *knowledge-driven DSS*) leveraged domain knowledge encoded in rule-based reasoning modules [30]. Examples include MYCIN [27] for bacterial infection diagnosis or DENDRAL [16] for chemical analysis. In contrast, modern IDSS leverage machine learning (ML) methods that implicitly infer rules from observations and thus learn from experience. This implicit inference of rules may result in the *black-box problem* for decision-makers. A black box refers to a situation in which it is possible to observe the input and outputs of a model, but the internals remain disclosed or uninterpretable to humans. In ML, the black box behavior may arise either from complex algorithms (as with deep neural networks) or from proprietary models that may otherwise be interpretable (such as with the COMPAS recidivism model) [24]. As decision-makers were always considered an integral part of the DSS [22], special attention must be paid to the design of the user interaction. With ML-enabled DSS this interaction must include explanation facilities that result in usable interpretability for decision makers.

2.2 Interpretability and Task Expertise

Prior research shows that a lack of interpretability can lead to users that mistrust, misuse, or reject a system [15, 19]. Often these result from a perceived mismatch between users' expectations and the actual behavior of a system [9]. Chander et al. describe two reasons for the mismatch to occur in a business-related decision-making context [6]: (i) the system's underlying data lacks decision criteria relevant for this situation (*awareness gap*). For instance, the user might have relevant contextual information from a phone call with a client that the system has no access to; (ii) the system's prediction is not in line with the user's beliefs and the system fails to persuade the user to adjust their beliefs (*persuasion gap*). In such a situation, the user and the system have access to the same information but weight decision criteria differently. The gaps are even more pronounced in a business-related context, where domain experts often can draw upon rich prior knowledge and beliefs about the decision situation when assessing the system (*extrinsic setting*) [20]. Explanations about the factors that contributed to the system's prediction, e.g., in natural language or in the form of visualizations, are considered one way of addressing those gaps. However, in prior research, rational explanations were shown to be only effective for participants that are not familiar with a given task [26]. The effects of explanation drop as users' confidence with the task increases over time. As user get confident with the task and the system's prediction, they become less situation aware. Most explanation approaches assume that explanations are displayed with every system prediction.

3 USE CASE AND METHODOLOGY

In our work, we outline and probe an approach that provides system explanations only when a mismatch with the user's beliefs occurs (persuasion gap). Such an approach may increase the situation awareness of decision-makers. We focused on the use case of addenda approval and risk assessment in the construction industry. We cooperated with *Alasco*[1] and their clients. Alasco provides a web-based cost controlling system for the construction industry that connects stakeholders and digitizes processes around budgeting, reporting, addenda management, and payment. We followed a human-centered design process that consisted of three

[1] http://www.alasco.de

phases: (i) we interviewed executives about their current addenda-related decision-making and derived intelligibility needs for an (semi-)automated addenda approval process; (ii) we designed and developed an interactive prototype that reflects those intelligibility needs; (iii) we evaluated our prototype in a formative user study to understand the acceptance of the prototype workflow.

Use Case: Addenda Approval. Our use case targets project management (PM) executives in the construction industry. The PM is responsible for the fulfillment of the construction project and acts as a coordinator between contractors on behalf of the building owner [13]. During the initial budget planning, the overall budget is split into a hierarchy of cost groups (e.g., property or financing). Each cost group consists of one to many contract units. Each contract unit represents the budget planned for commissioning a contractor for a task. As a construction project advances, contract units might require budget addenda due to unforeseen incidents or flaws in the initial budget planning. After ensuring the plausibility of the addendum, PM executives need to redistribute budgets from other contract units to accommodate the addendum. While doing this, decision-makers need to take the addendum risk and cost forecast of the other contract units into account.

Phase I: Need-Finding. The goal of the first phase, was to identify decision criteria and interpretability needs for an intelligent DSS for the addenda approval process. To understand domain experts' current decision-making processes around addenda approval, we conducted semi-structured interviews with 3 project managers and 2 project controllers who are proficient users of the Alasco software. Their average industry experience were 2.8 years (min=1, max=6). The interviews were held in the regular work environments and took between 30 and 45 minutes. The interviews were recorded and transcribed. To enrich our qualitative insights, participants were surveyed after each interview with the *decision-making questionnaire (DMQ).* The DMQ is a validated psychological questionnaire that aims to examine factors important to a decision-maker in the moment of decision-making in a specific context [7]. It consists of 14 questions which correspond to 3 factors (and 10 subfactors) that characterize a decision-making situation: (i) the nature of the decision or task, (ii) the cognitive and affective abilities of the decision maker, and (iii) the environment of the decision.

Phase II: Prototyping. We integrated our prototype as a separate module on top of the Alasco software. We reused the general structure and user interface of the software as participants were already familiar with it. The prototyping process was informed by the results of the need analysis as well as prior work on DSS and interpretability. Financial data has strict privacy regulations. Also, the production data of the participant's organizations varied greatly and was often incomplete. Thus, we centered our prototype around an addenda approval scenario based on a synthetically created data set so that all participants could be evaluated on the same scenario. The scenario consisted of an onboarding phase and an addenda approval phase. We developed a functional front-end prototype while the back-end was mostly static around the evaluation scenario.

Phase III: Formative Evaluation. After the design phase, we conducted a formative user study to evaluate the prototype's acceptance regarding participant's sense of control and sense of information. As the use case and required domain expertise limited the number of potential study participants, we adopted a qualitative evaluation approach. We recruited 5 project managers and 2 project controllers with an average industry experience of 3.7 years (min=0.5, max=15). The user study included 3 participants from phase I as well as 4 new participants. This reduced the risk of receiving biased feedback from participants who had already thought about (semi-)automating addenda approvals. The participants were presented with the scenario and asked to complete an addenda approval task including the onboarding task. While doing so, participants were encouraged to think aloud. Completing the task took approximately 10 minutes. After the tasks, participants were interviewed using open-ended questions about their experience. The user study was audio-recorded, transcribed. The results were qualitatively analyzed according to Kuckartz [14] by two coders (with a Kappa coefficient of 0.86). We used the driving factors resulting from the DMQ as categories and, following Kuckartz, their gradual levels as subcategories. Table 1 presents our final coding system after multiple iterations.

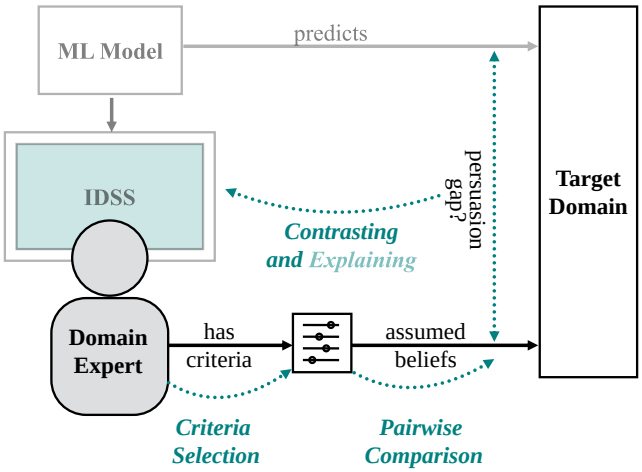

Figure 1: Explanations are triggered if there is a mismatch between the user's assumed beliefs (elicited through AHP) and the system's predictions. Blue parts relate to screens of our prototype. Muted parts relate to our proposed future work.

4 RESULTS

4.1 Interpretability Needs

The process for addendum validation was uniform for all participants. However, all participants agreed that there is no documented or formal way of deciding how to redistribute budgets. Instead, they base decisions on their personal experience and data derived from reporting features of the Alasco software. However, this approach has limitations. P2 asked for more structured workflow for addendum approval so that every stakeholder accomplishes the task in a predefined order to improve reporting. P1 would like have feedback on how well the initial budget distribution worked in

comparison to the final stage of a project. P1 and P5 asked for decision support that guides the user and recommends possible sources (e.g., based on the forecasted costs). P3 and P4 even suggested to (semi-)automate the allocation. The analysis of the DMQ indicated that control and personal involvement are important requirements for the participants. The most important subfactors were the need for *information and goals* (5 participants), *self-regulation* (4 participants), and *time/money pressure* (4 participants). It is important for the participants to have adequate and transparent data available that help them to plan, monitor, and evaluate results [7]. We leveraged these insights as guidelines for our prototype.

4.2 Prototype

We developed an IDSS interface with which participants could interact. The prototype consisted of two user flows. The first flow elicits the user's beliefs during the user onboarding through a widely accepted MADM approach. The second flow guides the user through the approval process once an addendum is requested and suggests options for budget transfer.

4.2.1 Belief Elicitation Flow. MADM approaches were used to make subjective user preferences explicit and, thus, make decision-making more transparent [21]. We leverage such an approach to elicit user beliefs about our target domain. A widely accepted and accessible MADM approach is the *analytic hierarchy process (AHP)* [10, 25]. AHP builds on a hierarchical representation of the decision problem. It leverages a user's judgments of the relative attribute importance to choose an alternative. The judgments and beliefs are elicited through pairwise comparisons of attributes. The decision criteria may be split into multiple hierarchy levels depending on the complexity. However, we limited our prototype to five decision criteria that are on the same level. We applied the wizard pattern to guide the decision maker through the steps of the AHP setup as part of a mandatory module onboarding [29]. First, users were introduced to the purpose of the flow and each step. Second, user had to select at least three criteria that they believe are important when withdrawing budget from a contract unit. Afterwards, they had to express the relative importance of each criteria through pairwise comparisons. We used the original AHP space consisting of a bidirectional Likert scale ranging from 9 (absolutely more important) to 1 (equally important) to 9. In a last step, we checked the judgments for inconsistencies and asked users to revise them if necessary. After the onboarding, users can revise their preferences anytime.

4.2.2 Intelligent Addenda Approval Flow. We enriched the manual approval flow with an intelligent overview that suggests contract units to withdraw budget from. First, the user is notified via email if a new addendum is to be reviewed. After confirming that the addendum is valid, the user sees an overview of possible contract units that may be used to accommodate the addendum. Each contract unit alternative is enriched with two types of information: (i) a score that reflects the user's beliefs. The score is calculated by AHP based on the user's relative importance of attributes as elicited during the onboarding; (ii) an intelligent suggestion that was said to take historical data into account. The suggestion may be derived through a machine learning model. Contrasting both information

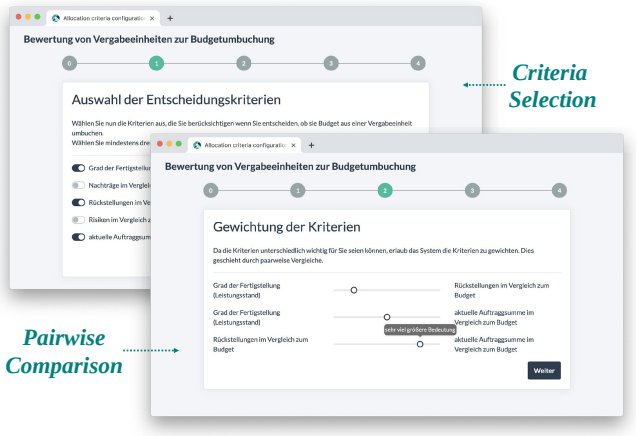

Figure 2: User's beliefs about the decision situation are elicited through AHP. In the first step, the user indicates which decision criteria are important for her. Afterwards, she compares those criteria pairwise express the relative importance.

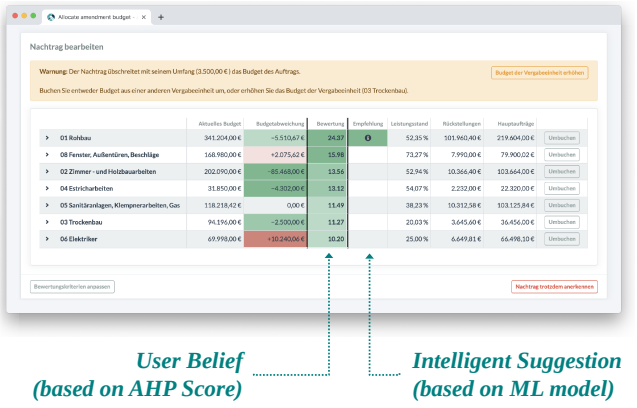

Figure 3: After an addendum is validated by the user, the IDSS gives an overview of contract units to transfer budget from. The alternatives are scored based on elicited user beliefs and contrasted with the system recommendation.

enables the user to grasp when their beliefs diverge from the system suggestion. Furthermore, it enables the system to identify and address a persuasion gap. Each column and the prediction have a tooltip that explains where the information is coming from. In our formative evaluation, the system suggestion and explanation were non-functional but based on static information. As participants were not provided with information about the underlying system logic, it resembles from a user perspective an IDSS.

4.3 Formative Evaluation

All participants were able to complete the given approval task. The results of the qualitative analysis show that all participants made

Table 1: (*Left*) Categories and subcategories derived from the results of the DMQ. (*Right*) Number of participants' statements during formative evaluation coded according to those (sub)categories.

Categories	*# of Statements*
Sense of Control	**36**
Full sense of control *(no doubts)*	12
Reinforced control *(feeling of guidance)*	8
Foreseeable behavior of the system *(no surprises)*	7
Expressed doubt/questioned system	8
Unclear statements regarding sense of control	1
Sense of Information	**23**
Improved experience due to information displayed	8
Satisfied with the amount of information	7
Desired additional information	7
Unclear statements regarding sense of information	1
Usability	**31**
Perceived increase in efficiency	7
User was hesitating/unclear	18
Expressed high mental effort	6

positive statements regarding their *sense of control* (relates to DMQ's self-regulation subfactor). 5 participants stated that their *sense of information* (relates to DMQ's information and goals subfactor) improved due to the information provided. However, 4 participants questioned the system at some point. 3 participants wished for additional information (e.g., emails or contract correspondences) or more detailed explanations (e.g., how their input affects the outcome). 4 participants perceived high mental efforts when choosing and comparing their relevant decision criteria during the onboarding. We attributed this to the fact, that they rarely had to articulate how they make addendum-related decisions before this study. However, these efforts paid off later on. 4 participants perceived increased efficiency during the addenda approval flow as they did not need to assess each alternative individually but instead could rely on the score and suggestion. Overall, we found that our prototype left the participants with an increased sense of control and information. However, the usability of the belief elicitation flow should be revised to reduce users' mental efforts. Table 1 presents a categorized summary of participants' statements.

5 LIMITATION AND FUTURE WORK

While our formative evaluation shows promising results, we acknowledge multiple limitations. Our work focuses on the limited use case of addenda approval in the construction industry. Our user studies were conducted under supervision in a controlled environment. Thus, actual user behavior and usage may be different in the field. Furthermore, our evaluation focused on the general acceptance of the approach by domain experts with a non-functional prototype. In future, we plan to conduct an experimental study that focuses on whether such an approach improves a user's understanding of an intelligent system. For this, we plan to transfer the approach to a human-grounded [8] evaluation scenario with lay users.

We believe that eliciting user beliefs and comparing them with intelligent predictions offers a promising basis for personalized explanations in XAI systems. ML algorithms take features and calculate their respective weights while optimizing a utility function. Post-hoc feature attribution methods, such as LIME [23] or SHAP [18], elicit the relative importance of a black box model's decision criteria. Similarly, decision-makers try to, explicitly or implicitly, optimize a utility function that is used to quantify their preferences regarding decision alternatives [12]. The difference is that decision-makers often do not know their utility function in advance and sometimes construct it ad-hoc during the decision-making situation. MADM methods, such as AHP, can make the user's beliefs explicit and accessible to explanation generating XAI systems. As part of our future work, we want to examine ways to relate the weights of post-hoc feature attribution methods to AHP's relative attribute importance. By this, XAI systems could adapt their explanation vocabulary (e.g., add or remove features to an explanation) or argumentation (e.g., argue with the user's expected outcome as the foil) based on the user's beliefs.

6 ACKNOWLEDGMENTS

We thank the participants who took their time to contribute their experiences and opinions from the field.

REFERENCES

[1] Rajat Agarwal, Shankar Chandrasekaran, and Mukund Sridhar. 2016. Imagining construction's digital future. *McKinsey & Company* (2016). https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/imagining-constructions-digital-future

[2] David Arnott and Graham Pervan. 2012. Design Science in Decision Support Systems Research: An Assessment using the Hevner, March, Park, and Ram Guidelines. *J. AIS* 13, 11 (2012), 1. http://aisel.aisnet.org/jais/vol13/iss11/1

[3] Or Biran and Courtenay Cotton. 2017. Explanation and justification in machine learning: A survey. In *IJCAI-17 workshop on explainable AI (XAI)*, Vol. 8. 1.

[4] JL Blanco, S Fuchs, M Parsons, and MJ Ribeirinho. 2018. Artificial intelligence: Construction technology's next frontier| McKinsey. https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/artificial-intelligence-construction-technologys-next-frontier

[5] A. Bussone, S. Stumpf, and D. O'Sullivan. 2015. The Role of Explanations on Trust and Reliance in Clinical Decision Support Systems. In *2015 International Conference on Healthcare Informatics*. 160–169. https://doi.org/10.1109/ICHI.2015.26

[6] Ajay Chander, Ramya Srinivasan, Suhas Chelian, Jun Wang, and Kanji Uchino. 2018. Working with Beliefs: AI Transparency in the Enterprise. In *Joint Proceedings of the ACM IUI 2018 Workshops co-located with the 23rd ACM Conference on Intelligent User Interfaces (ACM IUI 2018), Tokyo, Japan, March 11, 2018.* http://ceur-ws.org/Vol-2068/exss14.pdf

[7] María Luisa Sanz de Acedo Lizarraga, María Teresa Sanz de Acedo Baquedano, María Soria Oliver, and Antonio Closas. 2009. Development and validation of a decision-making questionnaire. *British Journal of Guidance & Counselling* 37, 3 (2009), 357–373. https://doi.org/10.1080/03069880902956959

[8] Finale Doshi-Velez and Been Kim. 2017. Towards A Rigorous Science of Interpretability. *CoRR* abs/1702.08608 (2017). arXiv:1702.08608 http://arxiv.org/abs/1702.08608

[9] Mary T. Dzindolet, Scott A. Peterson, Regina A. Pomranky, Linda G. Pierce, and Hall P. Beck. 2003. The role of trust in automation reliance. *Int. J. Hum. Comput. Stud.* 58 (2003), 697–718.

[10] Saul I. Gass. 2005. Model World: The Great Debate-MAUT Versus AHP. *Interfaces* 35, 4 (July 2005), 308–312. https://doi.org/10.1287/inte.1050.0152

[11] John J Jarvis. 1976. *Decision Support Systems: Theory.* Technical Report. Battelle Columbus Labs OH.

[12] Ralph Keeney, Howard Raiffa, and David Rajala. 1979. Decisions with Multiple Objectives: Preferences and Value Trade-Offs. *Systems, Man and Cybernetics, IEEE Transactions on* 9 (08 1979), 403 – 403. https://doi.org/10.1109/TSMC.1979.4310245

[13] Sascha Kilb and Markus Weigold. 2017. *Projektmanagement.* Springer Fachmedien Wiesbaden, Wiesbaden, 479–503. https://doi.org/10.1007/978-3-658-05368-0_20

[14] Udo Kuckartz. 2019. *Qualitative Text Analysis: A Systematic Approach.* Springer International Publishing, Cham, 181–197. https://doi.org/10.1007/978-3-030-

15636-7_8

[15] Brian Y. Lim, Anind K. Dey, and Daniel Avrahami. 2009. Why and Why Not Explanations Improve the Intelligibility of Context-Aware Intelligent Systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 2119–2128. https://doi.org/10.1145/1518701.1519023

[16] Robert K Lindsay, Bruce G Buchanan, Edward A Feigenbaum, and Joshua Lederberg. 1993. DENDRAL: a case study of the first expert system for scientific hypothesis formation. *Artificial intelligence* 61, 2 (1993), 209–261.

[17] Zachary C. Lipton. 2018. The Mythos of Model Interpretability. *Queue* 16, 3, Article 30 (June 2018), 27 pages. https://doi.org/10.1145/3236386.3241340

[18] Scott M. Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.). 4765–4774. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions

[19] Bonnie M Muir. 1994. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics* 37, 11 (1994), 1905–1922.

[20] Menaka Narayanan, Emily Chen, Jeffrey He, Been Kim, Sam Gershman, and Finale Doshi-Velez. 2018. How do Humans Understand Explanations from Machine Learning Systems? An Evaluation of the Human-Interpretability of Explanation. *CoRR* abs/1802.00682 (2018). arXiv:1802.00682 http://arxiv.org/abs/1802.00682

[21] M. Pavan and R. Todeschini. 2009. 1.19 - Multicriteria Decision-Making Methods. In *Comprehensive Chemometrics*, Steven D. Brown, Romá Tauler, and Beata Walczak (Eds.). Elsevier, Oxford, 591 – 629. https://doi.org/10.1016/B978-044452701-1.00038-7

[22] Gloria Phillips-Wren, Manuel Mora, Guisseppi A. Forgionne, Leonardo Garrido, and Jatinder N. D. Gupta. 2006. *A Multicriteria Model for the Evaluation of Intelligent Decision-making Support Systems (i-DMSS)*. Springer London, London, 3–24. https://doi.org/10.1007/1-84628-231-4_1

[23] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA) *(KDD '16)*. Association for Computing Machinery, New York, NY, USA, 1135–1144. https://doi.org/10.1145/2939672.2939778

[24] Cynthia Rudin. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1, 5 (2019), 206–215.

[25] Thomas L. Saaty. 2001. *Decision making for leaders : the analytic hierarchy process for decisions in a complex world* (new 3rd ed ed.). Pittsburgh, Pa., RWS Publications.

[26] James Schaffer, John O'Donovan, James Michaelis, Adrienne Raglin, and Tobias Höllerer. 2019. I Can Do Better than Your AI: Expertise and Explanations. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 240–251. https://doi.org/10.1145/3301275.3302308

[27] Edward Shortliffe. 1976. *Computer-based medical consultations: MYCIN*. Vol. 2. Elsevier.

[28] H.A. Simon. 1957. *Models of man: social and rational; mathematical essays on rational human behavior in society setting*. Wiley. https://books.google.de/books?id=lEZdwwEACAAJ

[29] Jenifer Tidwell. 2011. Designing Interfaces (Vol. 2).

[30] Efraim Turban, Ting-Peng Liang, and Jay E Aronson. 2005. *Decision Support Systems and Intelligent Systems:(International Edition)*. Pearson Prentice Hall.

[31] Stelios H. Zanakis, Anthony Solomon, Nicole Wishart, and Sandipa Dublish. 1998. Multi-attribute decision making: A simulation comparison of select methods. *European Journal of Operational Research* 107, 3 (1998), 507 – 529. https://doi.org/10.1016/S0377-2217(97)00147-1