



HAL
open science

Trust in AI-assisted Decision Making: Perspectives from Those Behind the System and Those for Whom the Decision is Made

Oleksandra Vereschak, Fatemeh Alizadeh, Gilles Bailly, Baptiste Caramiaux

► To cite this version:

Oleksandra Vereschak, Fatemeh Alizadeh, Gilles Bailly, Baptiste Caramiaux. Trust in AI-assisted Decision Making: Perspectives from Those Behind the System and Those for Whom the Decision is Made. CHI '24: CHI Conference on Human Factors in Computing Systems, May 2024, Honolulu HI USA, United States. pp.1-14, 10.1145/3613904.3642018 . hal-04768643

HAL Id: hal-04768643

<https://hal.sorbonne-universite.fr/hal-04768643v1>

Submitted on 6 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Trust in AI-assisted Decision Making: Perspectives from Those Behind the System and Those for Whom the Decision is Made

Oleksandra Vereschak

Sorbonne Université, CNRS, Institut des Systèmes
Intelligents et de Robotique, ISIR
Paris, France
vereschak@isir.upmc.fr

Gilles Bailly*

Sorbonne Université, CNRS, Institut des Systèmes
Intelligents et de Robotique, ISIR
Paris, France
gilles.bailly@sorbonne-universite.fr

Fatemeh Alizadeh

University of Siegen
Siegen, Germany
fatemeh.alizadeh@uni-siegen.de

Baptiste Caramiaux*

Sorbonne Université, CNRS, Institut des Systèmes
Intelligents et de Robotique, ISIR
Paris, France
baptiste.caramiaux@sorbonne-universite.fr

Abstract

Trust between humans and AI in the context of decision-making has acquired an important role in public policy, research and industry. In this context, Human-AI Trust has often been tackled from the lens of cognitive science and psychology, but lacks insights from the stakeholders involved. In this paper, we conducted semi-structured interviews with 7 AI practitioners and 7 decision subjects from various decision domains. We found that 1) interviewees identified the prerequisites for the existence of trust and distinguish trust from trustworthiness, reliance, and compliance; 2) trust in AI-integrated systems is strongly influenced by other human actors, more than the system's features; 3) the role of Human-AI trust factors is stakeholder-dependent. These results provide clues for the design of Human-AI interactions in which trust plays a major role, as well as outline new research directions in Human-AI Trust.

CCS Concepts: • Human-centered computing → HCI theory, concepts and models.

Keywords: trust, artificial intelligence, decision making, qualitative study, AI practitioners, decision subjects

1 Introduction

Decision making assisted by artificial intelligence (AI) has become more widespread in high-stakes domains, where decisions have real impacts on people's lives such as public safety [47], hiring [3] or loan approval [75]. Typically, the AI-based systems considered are based on automated processes (such as data-driven machine learning techniques) that provide assistance to human decision makers in a form of recommendations. Because Human-AI trust plays an important role in the adoption of these technologies [41] and

the improvement of decision making [9], it has become a priority for their design and development, as well deployment and regulation [73]. To understand how to achieve appropriate levels of human trust in these systems, more research at the intersection of Human-Computer Interaction and social study of AI is needed.

Trust is a complex and multifaceted concept [56, 61] and several studies have focused on a better understanding of the factors that can affect Human-AI trust (e.g., [64, 74, 106, 112, 116]). In these studies, trust is predominantly investigated through the lens of users, who are the persons interacting with the AI-assisted decision making system and its recommendations in order to deliver their decision [54]. Less is known about the perspectives from which other stakeholders involved in, and impacted by the design, deployment and use of these systems, view the notion of trust in AI, while this outlook on Human-AI trust can be shaped by their role. Jakesch et al. [46] demonstrate that the ethical values embedded in AI-assisted decision making systems can hold varied significance and interpretations for different groups. For example, people working on AI, on average, considered responsible AI values less important than general public and crowdworkers that contributed to the training of AI models. Such differences might also be reflected in the understanding of and opinions on Human-AI trust of the various stakeholders. For example, Lockey et al. [60] identify that different types of users do not encounter the same issues related to Human-AI trust: trust in AI of domain experts, e.g., doctors in medical decision-making, might be particularly affected by the factors that challenge their professional knowledge, skills, identity, and reputation. In contrast, fairness-related factors might impact general users' and society's trust in AI. Therefore, examining how stakeholders other than AI users view the definitions and factors of Human-AI is essential to advance the understanding of how trust is accounted for in

*Both authors contributed equally to this research.

the development and design of AI-embedded systems assisting decision-making and whether the existing approaches match the varying needs of different stakeholders.

In this article, we investigate how two groups of stakeholders - AI practitioners and decision subjects - understand Human-AI trust. Exploring the views of these groups on Human-AI trust definitions and factors allows to understand to which extent they prioritize and value the same aspects of Human-AI trust. **AI practitioners** are involved in system design and deployment (from AI developers to project managers). Given that they make decisions that influence the shape of human-AI interaction, impacting trust in AI-based technology and its acceptance for decision-making, understanding AI practitioners' views on trust can shed light on the factors they prioritize to build trust in AI among different stakeholders. Therefore, we explore the following first two research questions: RQ1a) *According to AI practitioners, what are the critical elements of human-AI trust in decision-making?*; RQ1b) *What do AI practitioners think influences the trust of various stakeholders in AI in the context of decision-making?*

The second group is **decision subjects**, i.e., people who do not interact directly with the systems incorporating AI but are affected by the decisions made by users based on the recommendations of these systems. For example, doctors are users, and patients are decision subjects in the medical context. Although decision subjects do not generally interact with AI-based systems the same way as users, they may nevertheless want to decide whether or not they wish to be impacted by the system [36]. For example, if a patient decides that the doctor's AI-based recommendation is not fair or trustworthy, they may want to change doctors or clinics. Therefore, we are investigating the following research questions: RQ2a) *According to decision subjects, what are the critical elements of trust between humans and AI in decision-making?*; RQ2b) *What factors influence decision subjects' trust in AI?*

We thus conducted semi-structured interviews with 7 AI practitioners related to AI-assisted decision making and 7 decision subjects from various risk-sensitive contexts (finance, law, management, medicine). The questions revolved around defining trust and trustworthiness when related to AI, and what they think can affect Human-AI trust. Using thematic analysis [14, 24], we established three themes: 1) definition of trust through three prerequisite elements and differentiation from other related concepts. The interviewees define Human-AI trust similarly to the literature with vulnerability and positive expectations, and additionally propose task complexity as a trust prerequisite. Moreover, AI practitioners distinguish trust from trustworthiness and trust-related behaviors such as reliance and compliance; 2) the effect of relationships between various stakeholders on Human-AI trust. We found that the extent to which decision subjects, AI practitioners, and users trust each other has an impact on

their trust towards AI and can moderate the effect of some factors on Human-AI trust; 3) stakeholder-dependency of the role and effects of some factors on Human-AI trust. We found that AI transparency, AI literacy, and interactivity of the system affect Human-AI trust differently for different stakeholders. Based on our findings, we provide a set of implications for academic researchers in HCI and AI practitioners. In particular, we recommend investigating the breaking and calibration points of trust between humans and AI beyond direct interaction with the system, and re-examining the techno-centric trust factors between humans and AI from a socio-technical point of view, as well as from the point of view of stakeholders other than users.

2 Related Work

In this section, we provide a brief overview of the methods to study Human-AI trust in assisted decision making and the different stakeholders at play with such systems.

2.1 Background on AI-embedded Systems Assisting Decision Making

While there is no universally accepted definition of AI [30], in this paper, we follow the definition provided by the European Commission: AI is a system capable of “*perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal*” [81]. Therefore, what we refer to as “AI-embedded systems assisting decision making” are the systems that analyse data to derive information used to facilitate human decision making [23, 88]. Usually, such systems provide assistance to human decision makers in a form of one or multiple recommendations, and when the system is not fully automated, it is the human who has the last word while making decisions. If the AI's recommendation differs from the decision maker's initial opinion, the decision maker finds themselves in conflict between her initial opinion and the new information received, which means that she has to choose between their opinion and the recommendation in order to make a better decision [110].

Making a better decision based on a recommendation means to be able to interpret the quality of the recommendation. However, it can sometimes be difficult to understand how a system arrived to a certain conclusion due to their “black box” nature [1, 53]. This, in turn, obfuscates understanding why a certain AI recommendation was produced, anticipating potential biases in decision making, and identifying the reasons for wrong predictions [87, 113]. When one is uncertain about how to correctly assess the quality of a recommendation [95], one can rely on their level of **trust** towards it to decide whether to stick to one's own opinion or to follow the system [100, 105]. As AI-embedded systems are becoming more widespread for assisting in making decisions

have real impacts on people’s lives, such as public safety [47], hiring [3] or loan approval [75], to name a few, the need for considering what contributes to human trust in the design of AI has arisen [13, 26, 34, 35, 51, 71, 92, 97, 104].

2.2 Human-AI Trust

Human-AI trust literature has two major themes of interest: defining what trust is and what factors affect it. The first line of research builds primarily on theoretical works, e.g. [45], taking a top-down approach to understanding Human-AI trust. A systematic literature review on Human-AI trust in the decision-making [103] defines Human-AI trust through three prerequisite elements, all encompassed in the trust **definition** by Lee and See [55]: *An attitude that an agent will achieve an individual’s goal in a situation characterized by uncertainty and vulnerability*. These three prerequisites are: vulnerability (or risk) of humans to the actions of the AI-based system, positive expectations of humans with respect to the AI-based system outcomes, and attitude as opposed to a behavior. Some scholars further define more granular facets of trust such as affective and cognitive trust [57, 70], weak and strong trust [10], warranted and unwarranted trust [45] or differentiate between trust in a particular AI tool, in people who built this tool and in AI in general [80].

The second major theme investigates Human-AI trust through a bottom-up approach, empirically studying what **factors** can affect users’ trust. Glikson and Woolley [37] in a literature review of studies empirically investigating Human-AI trust factors identify that the main ones for trust in AI are: tangibility, transparency, performance (reliability), task characteristics, anthropomorphism, and socially-oriented behaviors of the system. While they did not propose any classification of the factors, almost all of them belong to a category related to the system characteristics (a category present in trust frameworks from the fields other than Human-AI interaction [2, 11, 40, 41, 84, 85]). Type of task is the only trust factor that is related to the context of interaction, rather than the interaction with the system itself. Another framework on Human-AI trust factors in the medical context [17] calls for expanding the current literature’s focus on the contexts other than users’ interaction with the system. Browne et al. [17] argue that considering trust factors in the contexts beyond use reflects better the entire clinical AI deployment process in the real settings and, thus, opens up the floor to new trust calibration points. As most of the work on Human-AI trust targets a single type of stakeholder - direct users of the systems, we expand the analysis of Human-AI trust definition and factors to the stakeholders other than users that are related to the Human-AI decision making systems.

2.3 Human-AI Trust and Stakeholders Other Than Users

In this article, we focused on stakeholders that are the most linked to the development or the use of AI-assisted decision making systems: **AI practitioners**, people who develop the systems; **users**, people who use these systems to make decisions; and **decision subjects**, people who are affected by those decisions (see Table 1). Additional stakeholders, however, exist, such as regulators and policy makers, whose contributions, although interesting, are out of the scope of this paper (the reader can refer to different taxonomies [8, 29, 39, 46, 90, 115] for more information).

The stakeholders that have received the most attention in the literature on Human-AI trust are the **users** of the systems [54]. Researchers have repeatedly pointed to the need to explore and assess users’ trust in these systems to facilitate their adoption (see, for instance, [12, 89, 91]). It is not surprising that the literature focuses on system users, as understanding what affects their trust in the AI algorithms embedded in these systems can inform the development of interfaces and interactions that would facilitate the emergence of trust. However, different stakeholders may have different needs, expectations or roles when it comes to trust between humans and AI, and may also have an implicit impact on users’ trust in systems. The research on AI with human-centered values has investigated the perspectives and needs stakeholders other than users, notably AI practitioners (e.g. [7, 28, 49, 101, 102, 111]) and decision subjects (e.g. [36, 59, 62, 63, 68, 114]). Here we first present a set of previous works that have explicitly demonstrated differences between these stakeholders when it comes to concepts such as AI ethics, explanations, or fairness. While these results are not directly about Human-AI trust, they are related to our domain and motivate our approach.

Regarding responsible AI, which aims to deploy AI-based systems in line with ethical and legal frameworks, previous work shows the importance of including the AI practitioners’ perspectives to ensure that the system is designed to meet the actual needs of business and industry [42]. Typically, *AI practitioners* are pushed to quickly develop a service or a product that one can sell, which sometimes conflicts with ethical practices valued by the *users* [4, 65, 78, 107]. Regarding Explainable AI (XAI), which aims to propose means to explain AI-based predictions and help their interpretation, the usefulness of the explanation of a recommendation given by AI can vary depending on who sees it [32, 88]. A *user* might want to learn to which extent a recommendation can help them save money for instance [67], while *decision subjects* might want to know to which extent this recommendation is biased against a certain population in which they may belong [16, 108]. Regarding fairness, Smith et al. [94] take the case of microlending and show that depending on the different strategies to achieve fairness, stemming from




Icon	Acronym	Stakeholder	Definition
	U	Users	Individuals directly interacting with the system
	P	AI practitioners	Individuals who design, develop and deploy AI-based solutions
	DS	Decision subjects	Individuals affected by an AI-assisted decision-making system

Table 1. The different stakeholders related to the AI assisted decision making systems. This article focuses on AI practitioners and Decision subjects, two stakeholders who received less attention in the Human-AI trust literature.

Id	Role	Background	Organization	Type of AI	AI Application
P1	XAI R&D	CS and Maths	Large	CNNs	Transport, paleontology
P2	XAI R&D	Eng. and Maths	Small	OR	Task planning
P3	CEO	Maths	Small	Supervised ML	Evaluation of law cases
P4	Research mgr.	HCI	Large	OR, supervised and unsupervised ML	Project-based
P5	Research mgr.	Human Factors	Large	Not specified	Project-based
P6	CPO	Engineering	Small	ML (not specified)	Finance and business
P7	CEO	Bio. Eng. & Research	Small	Deep learning	Medical

Table 2. Characterization of AI practitioners, their companies, and AI they work with as reported by the interviewees themselves. “Small” refers to the companies with less than 20 employees, “Large” - with over 1000 employees. Explanation for abbreviations: *XAI* - explainable AI, *R&D* - research and development, *mgr.* - manager, *CEO* - chief executive officer, *CPO* - chief product officer, *CS* - computer science, *eng.* - engineering, *CNNs* - convolutional neural networks, *OR* - operations research, *ML* - machine learning.

its different definitions, Human-AI decisions favor *decision subjects*, *direct users* or *the organization behind the system*. Finally, regarding power relations in interaction, *users* and *AI practitioners* might see AI recommendations as tools, assistants or servants [50], while *decision subjects* might see the same AI recommendations as coming from someone in a more powerful position than they are. Such difference in perceived hierarchical roles between different stakeholders and AI can influence their attitude towards the system and interaction with it [21, 43, 83].

Previous work thus demonstrates the importance to study different stakeholders in the context of Human-AI interaction. In the context of Human-AI trust, Passi and Jackson [76] investigate how AI practitioners establish trust among themselves while working with data. Ammitzbøll Flügge et al. [5] and Okolo et al. [72] emphasize the importance of trust between users and decision subjects. Ferrario and Loi [32] analyze the importance of XAI for decision subjects’ trust in AI. Lastly, Ramesh et al. [79] show that decision subjects overtrust AI due to seeing it as a higher authority for financial decisions. These works tend to focus on a small set of factors influencing Human-AI trust. A more global perspective of how AI practitioners and decision subjects build and perceive Human-AI trust is yet to be explored.

3 Methodology

We adopted an interview-based qualitative methodology to answer our research questions about what trust is and what that trust depends on in the context of AI-assisted decision-making from the perspective of the real-world stakeholders. The project started in 2021.

3.1 Participants

We recruited participants through a convenience sampling technique combined with snowballing among colleagues and friends, and through announcements at events and on the project’s social media channels. We had two selection criteria to find interview participants: 1) they either work (as practitioners) on AI-embedded systems that support risk-sensitive decision making (e.g., in health, law, finance)¹ or they have been affected by their decisions (as decision subjects), 2) the system is used in the real world. We did not focus on any particular corporate position nor on any specific AI application in order to obtain a diversity of perspectives among interviewees. In total, we conducted 14 semi-structured interviews (7 with AI practitioners², 7 with AI decision subjects).

¹Risk in risk-sensitive applications is understood as defined by the European Union (EU) regulatory framework proposal on AI: <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

²We initially contacted 14 AI practitioners, 5 of them did not reply, and 2 did not have availability for an interview

Id	Background	Decision Context	The interview protocol consisted of four parts (Table 4) evolving around: the context with respect to their interaction with AI, Human-AI trust definitions, trust factors, and trust evaluation. In this article, we focused on the data regarding definitions and factors in the analysis. Where possible, we kept the formulation of questions identical (see Trust Definition in Table 4) for both groups of the participants. We adjusted the formulation of the questions related to the personal experiences to reflect the role of each group (example in Trust Factors, Table 4). We asked the questions around the Human-AI and trustworthiness definitions to understand what participants consider to be prerequisites of trust, that is in what contexts it is appropriate to consider Human-AI trust. We asked AI practitioners about their strategies to establish trust in their AI tool to understand what factors AI practitioners think influence trust of other stakeholders and which factors and stakeholders they prioritize. We did not explicitly refer to any group of stakeholders in our questions to let the AI practitioners spontaneously name the stakeholders relevant to the discussions around Human-AI trust. We asked decision subjects to share their experiences with receiving Human-AI decisions and, notably, what made them trust these decisions to identify the factors that influence their trust in AI. We also wanted to know whether decision subjects thought they trusted these decisions or AI in general too much or too little to gain more insights about what factors they prioritized to calibrate their trust towards more appropriate levels.
DS1	Software developer	Job application	
DS2	Medical student	Access to services	
DS3	Mechanical engineer	Job application	
DS4	Business economics researcher	Loan application	
DS5	Mechanical engineer	Job application	
DS6	Accounting and project management	Job application	
DS7	Computer engineer	Job application	

Table 3. Characterization of decision subjects, notably their background and in what context they received a Human-AI decision.

The participation in the study was on a voluntary basis. The AI practitioners are based in Europe and Oceania, and each worked for a different company. Table 2 provides an overview of the AI practitioners’ backgrounds, their roles in the company, and the application areas of AI. Three participants work on XAI (two are responsible for implementation and research, and another is the company’s chief executive officer - CEO). Three other participants are senior project and product managers. The AI decision subjects are all based in Europe and had been affected by AI decision making in three different risk-sensitive areas: job application, access to services, loan application. **The decision subjects we interviewed were not the people affected by the AI tools developed by the AI practitioners who participated in our study.** Table 3 provides an overview of the decision subjects’ backgrounds and in what context they received a Human-AI decision. Although 4 out of the 7 interviewed decision subjects have a background in computer science and engineering, we did not explicitly evaluate their prior experience with AI or their level of expertise in the field.

3.2 Interview Protocol

We conducted semi-structured interviews [66] of the recruited participants. The questions were compiled by the two first authors. They were independently reviewed by the other two authors and approved by the ethics committee of the research institution. In addition, we conducted a mock interview with an AI practitioner and a decision subject and adjusted the wording of the questions to improve their understanding. These data were not used for analysis. The questions were designed in English and translated to French and German for those participants preferring one of these languages. Interviews took place either by telephone or videoconference, whichever participants preferred. Participants could choose to allow us to record the interviews for note-taking purposes. All 14 participants agreed to do so. A total of 685 minutes were recorded, and each interview lasted an average of 50 minutes. Participants had access to our written notes before we used them in the article to ensure that their anonymity was maintained. All participants allowed us to quote them in the study.

There were 8 questions in total as approximate guidance for the interviewers (Appendices A and B). When needed, we deepened the topic with follow-up questions about all the stakeholders involved in an anecdote, clarifying theoretical terminology, possible solutions to a described challenge, and whether a proposed factor always has effect on Human-AI trust.

There were 8 questions in total as approximate guidance for the interviewers (Appendices A and B). When needed, we deepened the topic with follow-up questions about all the stakeholders involved in an anecdote, clarifying theoretical terminology, possible solutions to a described challenge, and whether a proposed factor always has effect on Human-AI trust.

3.3 Analysis of the Interviews

The first and second authors transcribed all interviews, removed all personal information (name of team, company, city, etc.) from the text, and assigned a code name to each interviewee, **P** for AI practitioners and **DS** for decision subjects. After transcription, the researchers deleted the audio files and allowed participants to review the interview text if they wished. The two researchers also translated the French and German texts to English and validated the translation with native speakers of the respective languages. Subsequently, the two researchers independently read all interviews at least twice, first without taking any notes and the second time highlighting the phrases or words related to people’s experiences and needs with AI, to get familiarized with the data.

The further data analysis was based on the inductive thematic analysis [14, 24], that is a bottom-up approach to coding and analysis driven by the data itself. The two authors independently assigned to each highlighted phrase a code

	AI practitioners	Decision Subjects
Context	How would you describe your role in the company? What is the main objective of your system?	Could you please tell me about your experience with Human-AI decision making?
Trust Definition	How would you define Human-AI trust in your own words? How would you define Trustworthy AI in your own words?	
Trust Factors	What is your strategy to establish trust of various stakeholders in your AI?	Have you ever trusted AI too much / too little?
Trust Evaluation	How would you know if someone trusts your AI?	Do you think AI developers consider human trust?

Table 4. Structure and examples of questions per each group of participants. Data analysis of this paper mostly relies on answers around the definitions and factors of Human-AI trust. A full list of questions is in Appendices A and B.

that encapsulates the best its main message, focusing on the semantic content of the data. They then compared the list of highlighted phrases and their codes, discussed whether to include or not the phrases highlighted only by one of the researchers, and fine-tuned the wording of the codes for the finalized list of the selected phrases. After three iterations, the first author organized the codes in a series of sub-themes. They were further reformulated or merged with the consensus of all four authors in the process of writing the paper, and organized, under three main themes: one on the definition of trust, one on the role of interpersonal relations, and one on the divergent opinions between AI practitioners and decision subjects on the factors affecting trust (further described in the next section).

4 Findings

The thematic analysis yielded three main themes discussed in this section. We first explore the definitions of Human-AI trust from the perspectives of AI practitioners and decision subjects. Secondly, we find that both groups of respondents attribute significant importance to trust in interpersonal relationships, rather than in system characteristics. We conclude the results section by emphasizing some differences in opinions between the groups regarding the impact of AI transparency, AI literacy, and interactivity on Human-AI trust.

4.1 On the Definition of Trust

When prompted to define Human-AI trust in decision making, the interviewees identified three prerequisites for trust: positive expectations that AI will be beneficial in achieving the goals, the perceived risk associated with a decision, and the complexity of the task at hand. Importantly, the interviewees differentiated between trust, trust-related behavior, and trustworthiness.

4.1.1 Positive expectations and perceived risk are prerequisites for the emergence of trust, but the nature of risk is debated. The interviewees state that for trust to emerge, people must have **positive expectations** that AI will help them achieve their goal and is aligned with their interest. They defined goal as “the best answer in the shortest

time” (DS5, DS7). P6 also highlights that AI recommendations must be aligned with the goal of people interacting with or affected by the system as opposed to the technology owner’s interest: “It is important that the owner [of an AI-embedded system] does not recommend something in the company’s interest” (P6).

Moreover, the interviewees refer to the **perceived risk associated with a decision** as another prerequisite for the emergence of trust: “When my physical integrity or money is at risk, trust becomes a consideration, especially when something important is at stake for me” (P4)³. Several participants associate risk with health (DS2, DS4, DS5, DS6) or financial stability (DS4, DS5, DS6). P4 refers to risks related to economic loss or threats to life and health as universal, stating, “... a foundation [for defining risk] would be the physical needs and individual and social integrity from the Maslow’s Hierarchy.” However, some, like P5 and P2, broaden the concept of risk to include “vulnerability” (P5) or “responsibility” (P2), showing that risk extends beyond just financial or health concerns. P4 notes that what is considered risky varies from person to person, as “not everyone has the same priorities”. For example, DS4 found even Tinder recommendations could induce vulnerability, recounting moments when “the algorithm says that I am ugly, something about myself that I do not want to accept” (DS4). Therefore, DS4’s experience of feeling vulnerable when their appearance was judged by AI indicates that the associated risk goes beyond monetary losses or health hazards and is closely related to one’s personal vulnerabilities and priorities. This points to the situatedness of the risks involved and suggests that the mere application domain of the AI-assisted decision is not enough to indicate the level of risk associated; rather, it is the perceived risk based on individual vulnerabilities and priorities that matters.

4.1.2 Task complexity as a new prerequisite for the emergence of Human-AI trust. Besides positive expectations and perceived risk as prerequisites for human trust in AI to emerge, some interviewees (P2, P4-P6, DS5) also mention **task complexity**. P2 and P6 describe “complex task” as a situation when a person cannot determine the quality of AI

³In this quote, AI practitioner P4 refers to their general reflection about what could trigger one’s trust to emerge, not taking a particular perspective as an AI practitioner nor a decision subject

661 recommendation and, as a result, has many doubts around
 662 the final decision. DS5 agrees with P2 and P6, citing data
 663 analysis as an example of a task that is complex because: “it
 664 is very difficult for a human to perform calculations and test
 665 the system.” A task is also perceived as more complex if the
 666 decision to make is a long-term one (P4). P5 suggests that
 667 when users face a complex task, trust emerges as a tool to
 668 mitigate the complexity: “Sometimes you can’t evaluate every-
 669 thing, you sort of use that quick «I just trust you, I just trust you
 670 to do the right thing».” Interestingly, while the interviewees
 671 reported task complexity as one prerequisite for trust, it is
 672 not present in the usual definitions of trust [45, 103], which
 673 typically considers two prerequisites: “positive expectations”
 674 and “vulnerability”.

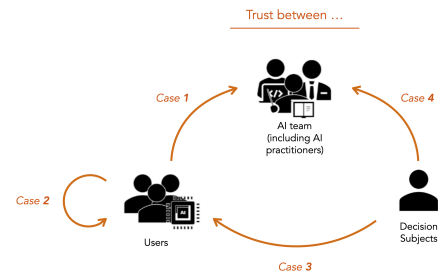
677 **4.1.3 Trust is differentiated from trust-related behav-**
 678 **iors and trustworthiness.** Some interviewees differentiate
 679 between trust (which is defined as an attitude [45]) and trust-
 680 related behaviors. For instance P4, P5, and P6 postulate that
 681 inferring users’ level of trust in AI from simply observing
 682 their behaviors could be misleading. Because users “can have
 683 a complex and elaborate way of thinking [about AI-embedded
 684 systems and recommendations]” (P4). P3 indicates that a user
 685 might follow AI recommendations not out of trust, but be-
 686 cause they “have no other solutions” (P3). The interviewees
 687 thus clarify that it is trust-related behaviors, not trust itself,
 688 that are in action. But trust-related behaviors are useful as
 689 they can serve as “indicators” of trust. As P2 notes, “as long
 690 as there aren’t too many complaints, no negative comments,
 691 [...] and the user uses the solutions, we can consider that trust
 692 is not broken” (P2).

693 Additionally, four interviewees (P2, P4, P5, P7) explicitly
 694 differentiate trust in AI from AI trustworthiness. Contrary to
 695 trust, which is seen as “human reaction” (P5), trustworthiness
 696 relates to features of the system (P2, P5), e.g., “whether the job
 697 has been well done” in designing and developing the system
 698 (P7). Such distinction further supports the stance that it is
 699 important to focus not only on what makes AI trustworthy,
 700 but also on what makes people trust AI [58]. Interestingly,
 701 two interviewees associated trustworthiness with AI govern-
 702 ance, i.e., the institution or organization behind the AI. P4
 703 states: “For me, it [trustworthiness] is not so much a question
 704 of AI, it’s more between the individual and the entity or the
 705 organization that makes the system.”

708 4.2 The Role of Inter-personal Relations on Trust

709 We discovered that trust between humans and AI is influ-
 710 enced by the trust among various stakeholders involved in
 711 the creation, use, and evaluation of the AI-based decision
 712 support system. Furthermore, AI certification, as a potential
 713 solution, is also contingent on trust within an interpersonal
 714 relational network.

716 **4.2.1 The team behind AI plays an important role in**
 717 **(Human-AI) trust.** The interviewees indicated that an indi-
 718 vidual’s trust in AI is closely linked to the level of trust they
 719 have in other stakeholders (Human-Human trust) within the
 720 socio-technical ecosystem. This ecosystem includes interac-
 721 tions between various stakeholders (such as AI practitioners,
 722 users, and decision subjects) and even the technical charac-
 723 teristics of the system. We found four cases illustrated in
 724 Figure 1. AI practitioners (P2, P4, P7) particularly emphasize
 725 Case 1: trust between the **users and the AI team**, where
 726 the AI team is the group of people behind the creation of the
 727 system. If users trust the AI team, their trust in AI “[...] is
 728 established before the system exists. [...] Trust is very strong
 729 in the co-design phase [between users and the AI team]” (P4).
 730 Interestingly, previous work on Human-AI trust does not
 731 generally consider trust between users and the AI team as a
 732 factor in Human-AI trust, even though it is likely to occur
 733 in real-world scenarios.



735
736
737
738
739
740
741
742
743
744
745
746 **Figure 1.** Schematic representation of the extent to which
 747 trust between different stakeholder groups discussed in
 748 relation to how it can affect Human-AI trust in the context
 749 of decision making.

751 AI practitioners (P3 and P6) also talk about the trust be-
 752 tween the **users and other users** of the same system (Case 2).
 753 They claim that previous experiences of other users influence
 754 users’ trust in AI: “We have 10,000 users, and 90% of them say
 755 «the feedback from the AI was very interesting», now [knowing
 756 this, current users] will tend to trust the AI” (P6). This trust in
 757 AI is further strengthened if “a domain expert confirms what
 758 the AI recommends” (P6).

759 Case 3 examines the trust that **decision subjects** place in
 760 the **users** of the system, concerning their usage and purpose.
 761 For instance, DS3 noted that trust in the system and its users
 762 are intertwined: “There is trust in the system and trust in
 763 those who use the system [...]. They [the users] should at least
 764 tell you they are using such a system [embedding AI] so you
 765 don’t lose your chance, just because you don’t know how it
 766 works [...]”. DS7 highlighted the complexity of this trust
 767 dynamic involving both humans (users) and machines from
 768 the perspective of the person being impacted by the decision:
 769 “I don’t trust mixing humans and machines. Either the decision
 770

771 *should be entirely made by a machine or a human. If you have*
 772 *only one machine, then you know what to expect. But if you*
 773 *have a machine and a human, then it would be very unfair*
 774 *because the users' roles are not defined, and the priority is not*
 775 *clear."*

776 Finally, one **decision subject** also talked about the role
 777 of **trust in AI team** for decision subjects (Case 4). DS7
 778 cites the example of Elon Musk and Tesla (at the time of the
 779 interviews), explaining that the trust of decision subjects
 780 in the company's high-level management influence their
 781 perceptions of and trust in the AI systems they develop:
 782 *"[he] is building trust with people through his own presence in*
 783 *the media [...]. People trust him and love his personality, so*
 784 *they trust his product even if it does not benefit them in the*
 785 *end."*

786 **4.2.2 The effect of AI certification on Human-AI trust**
 787 **depends on who is behind it.** The interviewees (P1, P4, P6,
 788 DS1-DS3, DS5, DS7) share the view that knowing that an AI
 789 system has been certified is a factor that influences trust in
 790 that system, because *"certification has always been a way to*
 791 *gain confidence in technological tools, whether they are AI [or*
 792 *not]"* (P6). This is especially true for critical systems: *"The ob-*
 793 *jective is clear - we [AI team] want certification"* (P1). P4 says
 794 that *"the certification alone should be enough [for Human-AI*
 795 *trust] if it is done well."* However, some interviewees high-
 796 light the importance of who is behind the certification, rather
 797 than the sole fact of AI having been certified (DS1-DS3, DS5,
 798 DS7): *"AI certificates are very important [for Human-AI trust]*
 799 *if there are organizations [that issue them] that people can*
 800 *trust"* (DS2).

801 Finally, P5 and DS7 are more suspicious about certification
 802 in general because they think there is not yet enough scien-
 803 tific evidence that *"certification will build trust [in AI], I am*
 804 *not quite convinced of that yet"* (P5) or because a certification
 805 does not warrant that everything will be alright *"if there is a*
 806 *hack or a problem"* (DS7).

808 4.3 Diverging Opinions on Three Factors impacting 809 Human-AI trust

810 Three factors playing a role for Human-AI trust were con-
 811 sidered differently by AI practitioners and decision subjects,
 812 and are important to be highlighted. These factors are: AI
 813 transparency, AI literacy, and the increase of interactivity on
 814 the system. We decided to focus on them, rather the factors
 815 that both groups of stakeholders agreed on (e.g. AI perfor-
 816 mance and errors, marketing of the system, expectations
 817 about the system), as we believe that the identified diverging
 818 opinions provide interesting insights and implications for
 819 the research community.

821 **4.3.1 AI practitioners and decision subjects do not**
 822 **share the same view on the role of AI transparency on**
 823 **trust.** AI transparency is one of the most discussed Human-
 824 AI trust factors in the interviews (P1-P7, DS3, DS4, DS7). The
 825

826 interviewees define two levels of transparency: a) explaining
 827 why a specific **AI recommendation** was shown and its
 828 quality, and b) explaining the **working processes** of AI
 829 development team.

830 The opinions about the effect of **explanations of AI**
 831 **recommendations** (a) on trust diverges not only between
 832 decision subjects and AI practitioners, but also among AI
 833 practitioners themselves. Some AI practitioners believe that
 834 explaining why a specific recommendation was shown can
 835 affect Human-AI trust, because it provides better understand-
 836 ing of how the recommendation was derived and, thus, lets
 837 estimate recommendation's quality (P2, P3, P6). At the same
 838 time, P4 strongly questions the necessity of understating
 839 for trust: *"One has to stop wondering how one can make tools*
 840 *that are more explainable, interpretable, or whatever, because*
 841 *sometimes there are tools that are not explainable in which*
 842 *we trust, a plane or a car, we don't know how it works inside,*
 843 *and yet we use them [...]"* (P4). Additionally, P1 and P7 raise
 844 concerns about the extent to which explanations can con-
 845 tribute to one's understanding of an AI recommendation:
 846 *"All the latest methods [of explainability] that have been de-*
 847 *veloped are often so complex that humans [laypeople] do not*
 848 *understand them, so the methods do not help them at all"* (P1).
 849 Decision subjects further disagree with the AI practitioners
 850 supporting usefulness of AI explanations (P2, P3, P6). They
 851 state that besides AI explanations being complex (DS3, DS4),
 852 they have limited contribution to understanding of AI rec-
 853 ommendations and, consequently, trust, because of the real
 854 world constraints: *"If people had the time to go through the*
 855 *explanations and review them in practice, they would have*
 856 *made the decision themselves in the first place"* (DS7).

857 Additionally, the AI practitioners (P3, P4, P7) seem to put
 858 considerable importance on transparency around the **work-**
 859 **ing processes** (b) of AI development team, while decision
 860 subjects did not mention this aspect of transparency at all
 861 in connection to trust. The AI practitioners believe that the
 862 working process is the most actionable means of AI trans-
 863 parency for their clients, i.e. users that request development
 864 of a specific AI algorithm either for their own business or
 865 for a third party. The examples of explaining the working
 866 processes could be explaining the data, e.g., *"you have to*
 867 *be very, very transparent about how you prepared the data,*
 868 *because any AI is biased just because of the quality of the data*
 869 *(and also the quantity)"* (P7) and explaining the specifics of
 870 the system and AI in general, e.g. *"when we [...] try to be as*
 871 *transparent as possible on how it [the AI-embedded system]*
 872 *works, we try to explain it to them [clients], because it can be*
 873 *sometimes quite technical, even mathematical, and then there*
 874 *are no more problems, no problem of trust..."* (P4).

875 **4.3.2 AI literacy: decision subjects perceived AI liter-**
 876 **acy as more specific and operational than AI practition-**
 877 **ers.** There have been diverging opinions about the role of
 878 AI literacy for Human-AI trust between AI practitioners and
 879

881 decision subjects: while AI practitioners emphasize the need
 882 for raising general public awareness around AI, decision sub-
 883 jects believe in system-related literacy, i.e. more specific and
 884 operational knowledge about AI. P5 believes public educa-
 885 tion on the general understanding of AI could be beneficial
 886 for calibrating human trust in AI, “because people will say
 887 «I do not trust AI», without really understanding what AI is”
 888 (P5). Similarly, P7 believes that users should understand the
 889 boundaries in AI performance - what AI can do and cannot
 890 do. However, for decision subjects, it is not enough to raise
 891 public awareness about how AI works in general, because it
 892 is not specific enough, e.g. “educational events [about AI] do
 893 not really make sense to me, because often nobody knows how
 894 the system really works” (DS4), or not actionable enough, e.g.
 895 “the educational sessions [about AI] do not make sense to me,
 896 how can they help?..” (DS6). Therefore, affecting Human-AI
 897 trust through AI literacy seems to be possible by accounting
 898 for needs of a specific stakeholder group. For example, for
 899 decision subjects to understand how a Human-AI decision is
 900 made to be able to act upon it, P7 provides training tailored
 901 for their decision subjects: “[we] create materials, [...] flyers,
 902 [...] content for patients so that they are informed, that they
 903 are not afraid of this new technology” (P7).

904
 905 **4.3.3 Interactivity: exploration tool for AI practition-**
 906 **ers, means to be included in the loop for decision sub-**
 907 **jects.** AI practitioners and decision subjects agree that in-
 908 teractivity is another factor impacting Human-AI trust in
 909 the context of decision making (P1-P4, DS2, DS3, DS4, DS6).
 910 They also agree that interactivity is often limited. For in-
 911 stance, “I give you [AI] input data - you [AI] send me back
 912 the solution, and I have no other contextual elements, elements
 913 of interaction with you” (P2), “I would like to have the op-
 914 portunity to negotiate and influence the [the AI’s] decision
 915 and say, «Hey, but look at this and that»” (DS4) or “these
 916 [AI] systems should be more tolerant to human error. Right
 917 now, it’s so strict” (DS6). However, their opinions differ when
 918 considering the consequences of this limited interactivity.
 919 For AI practitioners, it hampers one’s ability to explore the
 920 system, “asking for more explanations” (P3) and establish
 921 “a dialogue” (P4) or “cooperation” (P1) between users and
 922 AI. For decision subjects, the limited interactivity leads to
 923 more serious consequences. It provides a feeling of being
 924 excluded from the loop. Decision subjects feel they lose their
 925 sense of agency. They see themselves as “statistics” (DS2) or
 926 simply “filtered out” by AI (DS3) because the system is not
 927 “flexible” (DS3) or does not allow “to negotiate” (DS4).

928 5 Discussion

929
 930 In this paper, we investigated Human-AI trust from two per-
 931 spectives - what AI practitioners think is important for trust
 932 in AI of other stakeholders and what decision subjects think
 933 is important for their trust in AI. Combining these perspec-
 934 tives allows for understanding similarities and differences

935 in how these different stakeholders define Human-AI trust
 936 in the context of decision making and what factors affecting
 937 Human-AI trust they prioritize. In this section, we discuss
 938 what our results mean for 1) re-envisioning what factors
 939 affect Human-AI trust in the socio-technical ecosystem; 2)
 940 defining Human-AI trust and its key prerequisite elements
 941 for its existence; and (3) in terms of stakeholders’ agency over
 942 the system. Finally, we present some limitations of our study
 943 and propose future research directions that could address
 944 these limitations.
 945

946 5.1 On the Important Role of Inter-personal 947 Relations on Trust Within the Socio-technical 948 System 949

950 Our results revealed the important role of interpersonal re-
 951 lationships on trust. In other words, AI practitioners and
 952 decision subjects stressed the importance of trust links with
 953 other stakeholders involved in the system: its design, de-
 954 velopment, deployment or use in real applications. More-
 955 over, this importance seems to take precedence over the
 956 technical characteristics of the system. These results comple-
 957 ment recent findings on the under-explored concept of social
 958 transparency for AI-assisted decision-making [31]. Through
 959 highlighting the history of other users’ interactions with
 960 AI recommendations rather than the inner workings of AI,
 961 social transparency embraces the interviewees’ emphasis on
 962 trust factors related to social interactions, information ac-
 963 tionability, and expectations as a part of the system’s design.
 964 In this sense, our findings about the importance of interper-
 965 sonal relationships also support recent approaches arguing
 966 for trust calibration beyond direct interaction of people with
 967 AI [17].

968 From the different cases of trust links between stakehold-
 969 ers elicited in the findings, trust in the AI team (cases 1 and
 970 4 in Figure 1) is generally absent in the literature, while re-
 971 spondents believe that this plays an important role in the
 972 trust between humans and AI. So far, the literature suggests
 973 that the reputation of the organization that develops AI plays
 974 a role for doctors’ trust in AI recommendations [20, 93], and
 975 our study confirms this for the domains beyond medical deci-
 976 sion making. Another difference is that users’ trust in other
 977 users (case 2) is more emphasised in the academic literature
 978 than in the interviews [16, 31, 44, 72, 82]. Research shows
 979 that observing other users (especially colleagues) trusting
 980 the recommendations of the system can increase one’s own
 981 trust in AI [31, 44]. However, from the interviews, AI prac-
 982 titioners often serve as intermediaries between users and
 983 convey feedback as product reviews. Finally, the relation-
 984 ship of trust between decision subjects and other users (case
 985 4) is barely present in the interviews. The academic liter-
 986 ature shows that if decision subjects (e.g., a patient) trust
 987 the direct user (e.g., a clinician) and the direct user trusts
 988 the AI recommendations, then they would also trust the AI
 989 recommendations [72] and vice versa [16, 27].
 990

Our results suggest that these bonds of trust are either transversal (e.g. users to users) or upstream (e.g. users to AI team). We believe that trust, in this case, relates to the people who have either more expertise on the domain and technology or means of actions over the technology (such as the AI team). Therefore, these trust links might take an even more important role for decision subjects than other stakeholders. In fact, we saw this in the perception of the role of AI transparency on trust. Contrary to AI practitioners, decision subjects do not see how transparency can affect their trust in AI since explanations might be difficult to understand and the additional information about AI or a specific system is usually not actionable. Specifically, neither the interviewed AI practitioners, nor the literature provide ample reflections for the role of transparency for trust in AI of decision subjects. Transparency is, hence, viewed as a factor affecting primarily users' trust in AI, targeting their needs for quality evaluation of an AI recommendation and for refining their mental model about AI, which does not necessary encompass actionability and contestability - the needs of decision subjects [68, 114].

Research implications.

1. **Investigating the points of Human-AI trust breakdowns and calibrations beyond direct interaction with the system.** To this end, research needs to involve more fieldwork with the various stakeholders to understand how trust in AI is shaped and influenced within the complex web of relationships among AI practitioners, decision subjects, users, and other stakeholders, and identify key patterns and dynamics of trust flow among these stakeholders.
2. **Re-examining the techno-centric trust factors between humans and AI with a social lens.** Following the example of Ehsan et al. [31], who proposed the term of social transparency, moving away from providing more information about how AI works to more information about how other users make decisions with the system, we envision other Human-AI trust factors can be relooked in the same manner. For instance, in addition to reporting AI accuracy, one can inform users about how AI recommendations affected the performance of other users.

5.2 On the Prerequisites for the Existence of Trust

In order to understand what AI practitioners and decision subjects expect from a system they trust, we analysed how these stakeholders understand trust, i.e. what essential elements, or prerequisites, they associate with this notion. Both groups elicited the need for positive expectations and a situation of vulnerability. These two prerequisites are how theoretical work in the literature defines trust. This was unexpected, because trust is a complex and abstract theoretical concept that leads to frequent theoretical confusions [45, 58, 103]. It remains that we found a more nuanced

outlook on the key elements of trust in comparison with the academic literature. The interviewees' discussions highlight that vulnerability and positive expectations cannot be boiled down to monetary losses and high levels of accuracy as they are often presented in the empirical studies [103]. Vulnerability denotes a state in which someone feels the possibility of being emotionally attacked, and therefore finds themselves in a position of weakness. In our results, we had the example of a judgement based on physical appearance. So these prerequisites for the existence of trust depend on the individual or the community with which the individual identifies. Recent examples of the behaviour of algorithms that discriminate against a certain population, such as black women [19], place them in a vulnerable position more than other individuals. Additionally, decision subjects report to feel vulnerable, because they have no control over how the data they share about themselves for Human-AI decision making get interpreted by the users in charge of these decisions [27]. Sometimes, in order to appear cooperative, they provide more data about themselves than needed, which puts them at risk of "algorithmic stigmatization" [6, 83] - wrongfully assigned a certain label "at risk", e.g., risk of recidivism, child maltreatment, suicidal tendencies, based on the an algorithmic assemblage.

Our results also highlighted a new prerequisite for the existence of trust, namely the complexity of the task. Behind this prerequisite is the idea that if the task is simple, it can be easily solved by the person using the system or receiving a decision from it. Thus, if one knows the right answer, evaluating the quality of AI recommendation is straightforward, the conflict of between one's own opinion and the AI recommendation does not emerge, and consequently, neither does the state of trust. However, there is an ambiguity about the definition of complexity. Typically, we could envisage two scenarios. Firstly, complexity can arise from the impossibility for a human to process a large amount of information (for example, a large amount of data in a database) in order to produce a decision. Secondly, complexity can arise from a lack of expertise, either related to the decision domain, or related to the underlying AI techniques. If task complexity is a prerequisite for the emergence of trust, along with vulnerability and positive expectations, this implies that future research should study it and include it in the way experimental tasks are designed to focus on trust, rather than confidence [103].

Research implications.

1. **Understanding the role of vulnerability in AI-based decision-making systems.** As our findings indicate that feeling vulnerability to AI-based decision-making systems can go beyond monetary gains and losses, especially in the case of decision subjects, qualitative studies, such as interviews and case studies,

could be utilized to gain deeper insights into individual and community experiences of vulnerability towards AI in order to inform further research on trust.

2. **Incorporating complexity into experimental studies of trust.** Given that task complexity could be a key element in trust formation, it should be accounted for in designing experiments that study trust in AI to distinguish between trust and confidence in the system’s recommendations. Future research should investigate what aspects of task should be considered to vary the task complexity as well as to which extent it contributes to the formation of trust as a function of different levels of task complexity.

5.3 On the Notion of Agency over the System

AI practitioners and decision subjects both stress the importance of AI interactivity for trust, but their views on the purpose of interactivity differ. For AI practitioners, interactivity is a means to explore AI recommendations. From this point of view, they agree with what previous work has shown about the fact that interactivity contributes to explore to which extent nuances are accounted for AI recommendations [82]. Other works have shown, in addition, that interactivity also contributes to the refinement of the mental model about AI [22] and gives a sense of striving to improve decision making [72]. Decision subjects, on the other hand, see interactivity as a way of getting involved in the decision-making loop. In other words, they see interactivity first as a way of being represented in the decision-making process, before being able to formalise what this representation could bring in terms of understanding the system’s mechanisms and creating a mental model of its behaviour. We therefore see, in these different opinions between the stakeholders, a difference in power relationships. AI practitioners have the means to act on the system, and are therefore in a position to imagine what these means can bring them.

This interpretation suggests that interactivity is related to the notion of agency. In fact, decision subjects discuss the sense of agency and its relationship to Human-AI trust more than AI practitioners, which is expected considering the mentioned frustrations about their lack of actionability and power over the systems. This means that decision subjects value more the factors of trust linked to their inclusion in the decision-making loop in comparison with AI practitioners. It is an empowerment over Human-AI decisions so as not to feel solely “*part of the statistics*”, as put by DS2. These findings align with the prior work [46] showing that different groups of stakeholders prioritize ethical values differently. Our findings extend this line of research by demonstrating this for trust and underlines the importance of undertaking a multi-stakeholder approach [115] for Human-AI trust.

That being said, although the academic literature on Human-AI trust examining the interactivity of AI recommendations have led to certain results, as those mentioned above, this

research remains scarce [15, 22, 38, 72, 82]. Moreover, these studies are primarily about users rather than decision subjects. Similarly, while previous work has investigated the relationship between agency and trust in AI, it focuses exclusively on the agency of direct users (e.g., [18, 33, 48, 86, 96, 98, 109]). Additionally, in all these articles, participants are fully aware to which extent they have control over AI recommendations, and their level of agency remains unchanged throughout the experiment. Hence, the issue of varying levels of control over AI is not largely studied in the Human-AI trust literature in the context of decision making. Moreover, in the interviews, agency is mostly referred to as ability to contest a Human-AI decision, while in the literature, it is mainly represented as control over seeing an AI recommendation: full - AI recommendations are optional and appear on demand [18, 52, 52, 86, 96, 99], limited - mandatory AI recommendations that appear immediately [18, 33, 52, 77, 86, 96, 98, 99] or only after users’ initial decision [18, 33], and none - AI recommendations executed autonomously [52, 69, 77, 98, 99]. Therefore, it remains unclear to which extent the solution of “introducing four levels [of AI recommendations] instead of the binary [...]” proposed by P7 to increase the sense of agency for decision subjects would work.

Research implications.

1. **Investigating the role of different mechanisms of interactivity for trust in AI of various stakeholders.** Our findings indicate that interactivity plays a different role for decision subjects than for AI users, and thus might affect their trust in AI not through the same mechanisms. HCI researchers could conduct in-depth studies to examine how different interactive features (e.g., feedback loops, adjustable parameters) empower decision subjects or change trust links between them and practitioners or users.
2. **Investigating agency as human capability instead of a feature of the system.** Our results have shown the importance of human agency, particularly for decision subjects. While agency tends to be seen as a feature of the system (e.g., providing means to act on system behavior), it is also related to people’s perception of actions on the system and their representation by the system. In the same way as trust, this concept needs to be better understood from a human-centric point of view in the context of interactions with AI-based decision-making systems.

5.4 Future Work Directions

In this article, we interviewed representatives from a varied panel of decision domains (e.g. medicine, finance, recruitment). Although our main objective was to study the factors of trust between humans and artificial intelligence for risk-sensitive applications, each domain may nuance the effects on trust due to the diversity of decision-making flows, types

of stakeholders involved, etc., **which is one of the limitations in the interpretation of the study's results**. Considering that type of task and level of risk also have impact on Human-AI trust, it could be interesting to conduct a cross-domain comparison to see to which extent they put importance on the same Human-AI trust factors. Understanding the differences and similarities between various task domains can inform researchers and policy makers on higher level classification of domains [54]. Additionally, we did not account for individual differences such as gender, age, and explicitly assess prior experience with AI, and other demographic information in our analysis while these factors can further influence how certain ethical values are prioritized [46].

Secondly, we considered two types of stakeholders that are not users - AI practitioners and decision subjects. While there is no widely established categorization, some researchers propose a set of 11 stakeholders' groups [8] that are connected to the AI ecosystem, spanning from policy makers that work on high level strategies to hiring managers that recruit AI developers. An interesting research direction would be to extend the presented research to these stakeholders and inspect differences and commonalities in findings.

Lastly, we took an organization-focused approach to studying Human-AI trust when talking to AI practitioners. In other words, the AI-embedded systems that they are responsible for are developed, trained, designed, deployed, and monitored by the same company. However, nowadays AI technologies are often a product of "algorithmic supply chains" [25], that is multiple independent actors are responsible for commissioning different phases of production and deployment. As these actors have distributed responsibility over the outcomes of Human-AI decisions with imperfect control over how their work is used further down in an algorithmic supply chain, this can raise additional concerns over whether an AI recommendation produced by "many hands" can be trusted. Further investigating implications for Human-AI trust resulting from such a production set-up can shed light on new nuances about interpersonal dynamics between different stakeholders and identify new potential Human-AI trust breakdown points and factors that affect it.

Acknowledgments

This work was performed within the Labex SMART (ANR-11-LABX-65) supported by French state funds managed by the ANR within the Investissements d'Avenir programme under reference ANR-11-IDEX-0004-02. This research was supported by the ARCOL project (ANR-19-CE33-0001) - Interactive Reinforcement Co-Learning, from the French National Research Agency. We wish to acknowledge and thank everyone involved in this research, and express our sincere gratitude to the anonymous reviewers for their constructive comments.

References

- [1] Amina Adadi and Mohammed Berrada. 2018. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6 (2018), 52138–52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
- [2] B. D. Adams, L. Bruyn, S. Houde, and P. Angelopoulos. 2003. *Trust in Automated Systems literature review*. Technical Report. Defence Research and Development Canada, Toronto, Ontario, Canada. 124 pages. <https://cradpdf.drdc-rddc.gc.ca/PDFS/unc17/p520342.pdf>
- [3] Ifeoma Ajunwa, Sorelle Friedler, Carlos E. Scheidegger, and Suresh Venkatasubramanian. 2016. Hiring by algorithm: predicting and preventing disparate impact. *Available at SSRN* 2746078 (2016), 29.
- [4] Sanna J. Ali, Angèle Christin, Andrew Smart, and Riitta Katila. 2023. Walking the Walk of AI Ethics: Organizational Challenges and the Individualization of Risk among Ethics Entrepreneurs. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) (FAccT '23). Association for Computing Machinery, New York, NY, USA, 217–226. <https://doi.org/10.1145/3593013.3593990>
- [5] Asbjørn Ammitzbøll Flügge, Thomas Hildebrandt, and Naja Holten Møller. 2021. Street-Level Algorithms and AI in Bureaucratic Decision-Making: A Caseworker Perspective. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 40 (April 2021), 23 pages. <https://doi.org/10.1145/3449114>
- [6] Nazanin Andalibi, Cassidy Pyle, Kristen Barta, Lu Xian, Abigail Z. Jacobs, and Mark S. Ackerman. 2023. Conceptualizing Algorithmic Stigmatization. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 373, 18 pages. <https://doi.org/10.1145/3544548.3580970>
- [7] Narges Ashtari, Ryan Mullins, Crystal Qian, James Wexler, Ian Tenney, and Mahima Pushkarna. 2023. From Discovery to Adoption: Understanding the ML Practitioners' Interpretability Journey. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) (DIS '23). Association for Computing Machinery, New York, NY, USA, 2304–2325. <https://doi.org/10.1145/3563657.3596046>
- [8] Jacqui Ayling and Adriane Chapman. 2022. Putting AI ethics to work: are the tools fit for purpose? *AI and Ethics* 2 (08 2022). <https://doi.org/10.1007/s43681-021-00084-x>
- [9] Gagan Bansal, Besmira Nushi, Ece Kamar, Daniel S. Weld, Walter S. Lasecki, and Eric Horvitz. 2019. Updates in Human-AI Teams: Understanding and Addressing the Performance/Compatibility Trade-off. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence* (Honolulu, Hawaii, USA) (AAAI'19/IAAI'19/EAAI'19). AAAI Press, Washington D.C. USA, Article 300, 9 pages. <https://doi.org/10.1609/aaai.v33i01.33012429>
- [10] Riccardo Baratella, Glenda Amaral, Tiago Prince Sales, Renata Guizzardi, and Giancarlo Guizzardi. forthcoming. The Many Facets of Trust. In *Formal Ontology in Information Systems*. IOS Press, Amsterdam, Netherlands.
- [11] Jason M. Bindewald, Christina F. Rusnock, and Michael E. Miller. 2018. Measuring Human Trust Behavior in Human-Machine Teams. In *Advances in Human Factors in Simulation and Modeling*, Daniel N. Cassenti (Ed.). Springer International Publishing, Cham, 47–58.
- [12] Ann M Bisantz and Younho Seong. 2001. Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. *International Journal of Industrial Ergonomics* 28, 2 (2001), 85–97.
- [13] Defense Innovation Board. 2019. *AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense*. Technical Report. United

- States Department of Defense, Virginia, United States. 11 pages. [https://media.defense.gov/2019/Oct/31/2002204458/-1/1/0/DIBS_SAI_S\\$PRINCIPLES\\$PRIMARY_\\$DOCUMENT.PDF](https://media.defense.gov/2019/Oct/31/2002204458/-1/1/0/DIBS_SAI_S$PRINCIPLES$PRIMARY_$DOCUMENT.PDF)
- [14] Virginia Braun and Victoria Clarke. 2012. Thematic analysis. In *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological*. American Psychological Association, Washington, 57–71.
- [15] Tom Bridgwater, Manuel Giuliani, Anouk van Maris, Greg Baker, Alan Winfield, and Tony Pipe. 2020. Examining Profiles for Robotic Risk Assessment: Does a Robot’s Approach to Risk Affect User Trust?. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction (HRI ’20)*. Association for Computing Machinery, New York, NY, USA, 23–31. <https://doi.org/10.1145/3319502.3374804>
- [16] Anna Brown, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin, and Rhema Vaithianathan. 2019. Toward Algorithmic Accountability in Public Services: A Qualitative Study of Affected Community Perspectives on Algorithmic Decision-Making in Child Welfare Services. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI ’19)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300271>
- [17] Jacob Browne, Saskia Bakker, Bin Yu, Peter Lloyd, and S. Ben Al-louch. 2022. Trust in Clinical AI: Expanding the Unit of Analysis. In *HHAI2022: Augmenting Human Intellect*, Vol. 354. IOS Press, Amsterdam, Netherlands, 96–113.
- [18] Zana Bućinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 188 (April 2021), 21 pages. <https://doi.org/10.1145/3449287>
- [19] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (Proceedings of Machine Learning Research)*, Sorelle A. Friedler and Christo Wilson (Eds.), Vol. 81. PMLR, New York, NY, USA, 77–91. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- [20] Eleanor R. Burgess, Ivana Jankovic, Melissa Austin, Nancy Cai, Adela Kapuścińska, Suzanne Currie, J. Marc Overhage, Erika S Poole, and Jofish Kaye. 2023. Healthcare AI Treatment Decision Support: Design Principles to Enhance Clinician Adoption and Trust. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI ’23)*. Association for Computing Machinery, New York, NY, USA, Article 15, 19 pages. <https://doi.org/10.1145/3544548.3581251>
- [21] Federico Cabitza, Andrea Campagner, Riccardo Angius, Chiara Natali, and Carlo Reverberi. 2023. AI Shall Have No Dominion: On How to Measure Technology Dominance in AI-Supported Human Decision-Making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI ’23)*. Association for Computing Machinery, New York, NY, USA, Article 354, 20 pages. <https://doi.org/10.1145/3544548.3581095>
- [22] Carrie J. Cai, Emily Reif, Narayan Hegde, Jason Hipp, Been Kim, Daniel Smilkov, Martin Wattenberg, Fernanda Viegas, Greg S. Corrado, Martin C. Stumpe, and Michael Terry. 2019. Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI ’19)*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3290605.3300234>
- [23] Claude Castelluccia and Daniel Le Métaye. 2019. *Understanding algorithmic decision-making: Opportunities and challenges*. EU Publications, Strasbourg, France. 1–79 pages. <https://doi.org/10.2861/536131>
- [24] Victoria Clarke and Virginia Braun. 2013. Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *The Psychologist* 26 (02 2013), 120–123.
- [25] Jennifer Cobbe, Michael Veale, and Jatinder Singh. 2023. Understanding Accountability in Algorithmic Supply Chains. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (Chicago, IL, USA) (FAccT ’23)*. Association for Computing Machinery, New York, NY, USA, 1186–1197. <https://doi.org/10.1145/3593013.3594073>
- [26] European Commission. 2020. *On Artificial Intelligence - A European approach to excellence and trust*. Technical Report. European Commission, Brussels, Belgium. 27 pages. [https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_\\$en.pdf](https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_$en.pdf)
- [27] Vedant Das Swain, Lan Gao, William A Wood, Srikruthi C Matli, Gregory D. Abowd, and Munmun De Choudhury. 2023. Algorithmic Power or Punishment: Information Worker Perspectives on Passive Sensing Enabled AI Phenotyping of Performance and Wellbeing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI ’23)*. Association for Computing Machinery, New York, NY, USA, Article 246, 17 pages. <https://doi.org/10.1145/3544548.3581376>
- [28] Wesley Hanwen Deng, Nur Yildirim, Monica Chang, Motahhare Es-lami, Kenneth Holstein, and Michael Madaio. 2023. Investigating Practices and Opportunities for Cross-Functional Collaboration around AI Fairness in Industry Practice. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (Chicago, IL, USA) (FAccT ’23)*. Association for Computing Machinery, New York, NY, USA, 705–716. <https://doi.org/10.1145/3593013.3594037>
- [29] Advait Deshpande and Helen Sharp. 2022. Responsible AI Systems: Who Are the Stakeholders?. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (Oxford, United Kingdom) (AI/ES ’22)*. Association for Computing Machinery, New York, NY, USA, 227–236. <https://doi.org/10.1145/3514094.3534187>
- [30] Yangqing Duan, John S. Edwards, and Yogesh K Dwivedi. 2019. Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management* 48 (2019), 63–71. <https://doi.org/10.1016/j.ijinfomgt.2019.01.021>
- [31] Upol Ehsan, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz. 2021. Expanding Explainability: Towards Social Transparency in AI Systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (Yokohama, Japan) (CHI ’21)*. Association for Computing Machinery, New York, NY, USA, Article 82, 19 pages. <https://doi.org/10.1145/3411764.3445188>
- [32] Andrea Ferrario and Michele Loi. 2022. How Explainability Contributes to Trust in AI. In *2022 ACM Conference on Fairness, Accountability, and Transparency (Seoul, Republic of Korea) (FAccT ’22)*. Association for Computing Machinery, New York, NY, USA, 1457–1466. <https://doi.org/10.1145/3531146.3533202>
- [33] Riccardo Fogliato, Alexandra Chouldechova, and Zachary Lipton. 2021. The Impact of Algorithmic Risk Assessments on Human Predictions and Its Analysis via Crowdsourcing Studies. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 428 (oct 2021), 24 pages. <https://doi.org/10.1145/3479572>
- [34] AXA Research Fund. 2019. *Artificial Intelligence: Fostering Trust*. Technical Report. AXA. 45 pages. <https://www.axa-research.org/en/news/AI-research-guide>
- [35] G20. 2019. *G20 Ministerial Statement on Trade and Digital Economy*. Technical Report. G20, Brussels, Belgium. 14 pages. <http://trade.ec.europa.eu/doclib/press/index.cfm?id=2027>
- [36] Meric Altug Gemalmaz and Ming Yin. 2022. Understanding Decision Subjects’ Fairness Perceptions and Retention in Repeated Interactions with AI-Based Decision Systems. In *Proceedings of the 2022*

- 1431 AAAI/ACM Conference on AI, Ethics, and Society (Oxford, United Kingdom) (AIES '22). Association for Computing Machinery, New York, NY, USA, 295–306. <https://doi.org/10.1145/3514094.3534201>
- 1432
- 1433 [37] Ella Glikson and Anita Woolley. 2020. Human trust in artificial intelligence: Review of empirical research (in press). *The Academy of Management Annals* 14, 2 (August 2020), 62. <https://doi.org/10.5465/annals.2018.0057>
- 1434
- 1435
- 1436
- 1437 [38] Akshith Gupta, Debadeep Basu, Ramya Ghantasala, Sihang Qiu, and Ujwal Gadiraju. 2022. To Trust or Not To Trust: How a Conversational Interface Affects Trust in a Decision Support System. In *Proceedings of the ACM Web Conference 2022* (Virtual Event, Lyon, France) (WWW '22). Association for Computing Machinery, New York, NY, USA, 3531–3540. <https://doi.org/10.1145/3485447.3512248>
- 1438
- 1439
- 1440
- 1441
- 1442 [39] H. Güngör. 2020. Creating Value with Artificial Intelligence: A Multi-stakeholder Perspective. *Journal of Creating Value* 6, 1 (2020), 72–85. <https://doi.org/10.1177/2394964320921071> arXiv:<https://doi.org/10.1177/2394964320921071>
- 1443
- 1444
- 1445 [40] Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. de Visser, and Raja Parasuraman. 2011. A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction. *Human Factors* 53, 5 (2011), 517–527. <https://doi.org/10.1177/0018720811417254>
- 1446
- 1447
- 1448
- 1449
- 1450 [41] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors* 57, 3 (2015), 407–434. <https://doi.org/10.1177/0018720814547570>
- 1451
- 1452
- 1453 [42] Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé, Miro Dudik, and Hanna Wallach. 2019. *Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?* Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/3290605.3300830>
- 1454
- 1455
- 1456
- 1457 [43] Yoyo Tsung-Yu Hou, Wen-Ying Lee, and Malte Jung. 2023. “Should I Follow the Human, or Follow the Robot?” – Robots in Power Can Have More Influence Than Humans on Decision-Making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 114, 13 pages. <https://doi.org/10.1145/3544548.3581066>
- 1458
- 1459
- 1460
- 1461
- 1462
- 1463 [44] Maia Jacobs, Jeffrey He, Melanie F. Pradier, Barbara Lam, Andrew C. Ahn, Thomas H. McCoy, Roy H. Perlis, Finale Doshi-Velez, and Krzysztof Z. Gajos. 2021. Designing AI for Trust and Collaboration in Time-Constrained Medical Decisions: A Sociotechnical Lens. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 659, 14 pages. <https://doi.org/10.1145/3411764.3445385>
- 1464
- 1465
- 1466
- 1467
- 1468
- 1469
- 1470 [45] Alon Jacovi, Ana Marasović, Tim Miller, and Yoav Goldberg. 2021. Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) (FAccT '21). Association for Computing Machinery, New York, NY, USA, 624–635. <https://doi.org/10.1145/3442188.3445923>
- 1471
- 1472
- 1473
- 1474
- 1475 [46] Maurice Jakesch, Zana Bućinca, Saleema Amershi, and Alexandra Olteanu. 2022. How Different Groups Prioritize Ethical Values for Responsible AI. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) (FAccT '22). Association for Computing Machinery, New York, NY, USA, 310–323. <https://doi.org/10.1145/3531146.3533097>
- 1476
- 1477
- 1478
- 1479
- 1480 [47] Anil Kalhan. 2013. Immigration policing and federalism through the lens of technology, surveillance, and privacy. *Ohio St. LJ* 74 (2013), 1105.
- 1481
- 1482 [48] Shivani Kapania, Oliver Siy, Gabe Clapper, Azhagu Meena SP, and Nithya Sambasivan. 2022. “Because AI is 100% Right and Safe”: User Attitudes and Sources of AI Authority in India. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 158, 18 pages. <https://doi.org/10.1145/3491102.3517533>
- 1483
- 1484
- 1485
- 1486
- 1487
- 1488
- 1489
- 1490 [49] Jee Young Kim, William Boag, Freya Gulamali, Alifia Hasan, Henry David Jeffrey Hogg, Mark Lifson, Deirdre Mulligan, Manesh Patel, Inioluwa Deborah Raji, Ajai Sehgal, Keo Shaw, Danny Tobey, Alexandra Valladares, David Vidal, Suresh Balu, and Mark Sendak. 2023. Organizational Governance of Emerging Technologies: AI Adoption in Healthcare. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) (FAccT '23). Association for Computing Machinery, New York, NY, USA, 1396–1417. <https://doi.org/10.1145/3593013.3594089>
- 1491
- 1492
- 1493
- 1494
- 1495
- 1496
- 1497 [50] Taenyun Kim, Maria D. Molina, Minjin (MJ) Rheu, Emily S. Zhan, and Wei Peng. 2023. One AI Does Not Fit All: A Cluster Analysis of the Laypeople’s Perception of AI Roles. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 29, 20 pages. <https://doi.org/10.1145/3544548.3581340>
- 1498
- 1499
- 1500
- 1501
- 1502 [51] KPMG. 2019. *Controlling AI: The imperative for transparency and explainability*. Technical Report. KPMG. 28 pages. <https://advisory.kpmg.us/articles/2019/controlling-ai.html>
- 1503
- 1504
- 1505 [52] Matthias Kraus, Nicolas Wagner, and Wolfgang Minker. 2020. Effects of Proactive Dialogue Strategies on Human-Computer Trust. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (Genoa, Italy) (UMAP '20). Association for Computing Machinery, New York, NY, USA, 107–116. <https://doi.org/10.1145/3340631.3394840>
- 1506
- 1507
- 1508
- 1509
- 1510 [53] Francesca Lagioia and Giuseppe Contissa. 2020. The strange case of Dr Watson : liability implications of AI evidence-based decision support systems in health care. *Eur. J. Leg. Stud.* 12, 2 (2020), 241–289.
- 1511
- 1512
- 1513 [54] Vivian Lai, Chacha Chen, Alison Smith-Renner, Q. Vera Liao, and Chenhao Tan. 2023. Towards a Science of Human-AI Decision Making: An Overview of Design Space in Empirical Human-Subject Studies. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) (FAccT '23). Association for Computing Machinery, New York, NY, USA, 1369–1385. <https://doi.org/10.1145/3593013.3594087>
- 1514
- 1515
- 1516
- 1517 [55] John Lee and Katrina See. 2004. Trust in Automation: Designing for Appropriate Reliance. *Human factors* 46 (February 2004), 50–80. <https://doi.org/10.1518/hfes.46.1.50.30392>
- 1518
- 1519
- 1520 [56] Roy Lewicki and Chad Brinsfield. 2011. Measuring trust beliefs and behaviours. In *Handbook of Research Methods on Trust*, Fergus Lyon, Guido Möllering, and Mark Saunders (Eds.). Edward Elgar, Cheltenham, UK; Northampton, MA, USA, Chapter 3, 29–39. <https://doi.org/10.4337/9781781009246.00013>
- 1521
- 1522
- 1523
- 1524 [57] J. David Lewis and Andrew Weigert. 1985. Trust as a Social Reality. *Social Forces* 63, 4 (1985), 967–985. <http://www.jstor.org/stable/2578601>
- 1525
- 1526 [58] Q.Vera Liao and S. Shyam Sundar. 2022. Designing for Responsible Trust in AI Systems: A Communication Perspective. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) (FAccT '22). Association for Computing Machinery, New York, NY, USA, 1257–1268. <https://doi.org/10.1145/3531146.3533182>
- 1527
- 1528
- 1529
- 1530 [59] Gabriel Lima, Nina Grgić-Hlača, Jin Keun Jeong, and Meeyoung Cha. 2022. The Conflict Between Explainable and Accountable Decision-Making Algorithms. In *2022 ACM Conference on Fairness, Accountability, and Transparency* (Seoul, Republic of Korea) (FAccT '22). Association for Computing Machinery, New York, NY, USA, 2103–2113. <https://doi.org/10.1145/3531146.3534628>
- 1531
- 1532
- 1533
- 1534 [60] Steven Lockey, Nicole Gillespie, Daniel Holm, and Ida Asadi Someh. 2021. A Review of Trust in Artificial Intelligence: Challenges, Vulnerabilities and Future Directions. In *HICSS-54. Hawaii International Conference on System Sciences*, Hawaii, USA. <https://doi.org/10.1145/3531146.3534628>
- 1535
- 1536
- 1537
- 1538
- 1539
- 1540
- 1541
- 1542
- 1543
- 1544
- 1545
- 1546
- 1547
- 1548
- 1549
- 1550

- 1541 //doi.org/10.24251/HICSS.2021.664
- 1542 [61] Fergus Lyon, Guido Möllering, and Mark Saunders. 2015. *Handbook of*
1543 *Research Methods on Trust: Second Edition*. Edward Elgar Publishing,
1544 Cheltenham, United Kingdom. 1–343 pages. <https://doi.org/10.4337/9781782547419>
- 1545 [62] Henrietta Lyons, Tim Miller, and Eduardo Velloso. 2023. Algorithmic
1546 Decisions, Desire for Control, and the Preference for Human Review
1547 over Algorithmic Review. In *Proceedings of the 2023 ACM Confer-*
1548 *ence on Fairness, Accountability, and Transparency* (Chicago, IL, USA)
1549 (FAccT '23). Association for Computing Machinery, New York, NY,
1550 USA, 764–774. <https://doi.org/10.1145/3593013.3594041>
- 1551 [63] Henrietta Lyons, Senuri Wijenayake, Tim Miller, and Eduardo Vel-
1552 loso. 2022. What's the Appeal? Perceptions of Review Processes
1553 for Algorithmic Decisions. In *Proceedings of the 2022 CHI Conference*
1554 *on Human Factors in Computing Systems* (New Orleans, LA, USA)
1555 (CHI '22). Association for Computing Machinery, New York, NY, USA,
1556 Article 580, 15 pages. <https://doi.org/10.1145/3491102.3517606>
- 1557 [64] Shuai Ma, Ying Lei, Xinru Wang, Chengbo Zheng, Chuhan Shi,
1558 Ming Yin, and Xiaojuan Ma. 2023. Who Should I Trust: AI or
1559 Myself? Leveraging Human and AI Correctness Likelihood to Pro-
1560 mote Appropriate Trust in AI-Assisted Decision-Making. In *Proceed-*
1561 *ings of the 2023 CHI Conference on Human Factors in Computing*
1562 *Systems* (Hamburg, Germany) (CHI '23). Association for Comput-
1563 ing Machinery, New York, NY, USA, Article 759, 19 pages. <https://doi.org/10.1145/3544548.3581058>
- 1564 [65] Michael Madaio, Lisa Egede, Hariharan Subramonyam, Jennifer Wort-
1565 man Vaughan, and Hanna Wallach. 2022. Assessing the Fairness of
1566 AI Systems: AI Practitioners' Processes, Challenges, and Needs for
1567 Support. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 52
1568 (apr 2022), 26 pages. <https://doi.org/10.1145/3512899>
- 1569 [66] Danielle Magaldi and Matthew Berler. 2020. *Semi-structured Inter-*
1570 *views*. Springer International Publishing, Cham, 4825–4830. https://doi.org/10.1007/978-3-319-24612-3_857
- 1571 [67] Marco Marabelli and Sue Newell. 2019. Algorithmic Decision-making
1572 in the US Healthcare Industry: Good for Whom? *Academy of Man-*
1573 *agement Proceedings* 2019 (08 2019), 15581. <https://doi.org/10.5465/AMBPP.2019.15581>
- 1574 [68] Amelie Marian. 2023. Algorithmic Transparency and Accountabil-
1575 ity through Crowdsourcing: A Study of the NYC School Admission
1576 Lottery. In *Proceedings of the 2023 ACM Conference on Fairness, Ac-*
1577 *countability, and Transparency* (Chicago, IL, USA) (FAccT '23). As-
1578 sociation for Computing Machinery, New York, NY, USA, 434–443.
1579 <https://doi.org/10.1145/3593013.3594009>
- 1580 [69] Steffen Maurer, Rainer Erbach, Issam Kraiem, Susanne Kuhnt, Pe-
1581 tra Grimm, and Enrico Rukzio. 2018. Designing a Guardian Angel:
1582 Giving an Automated Vehicle the Possibility to Override Its Driver.
1583 In *Proceedings of the 10th International Conference on Automotive User*
1584 *Interfaces and Interactive Vehicular Applications (AutomotiveUI '18)*.
1585 Association for Computing Machinery, New York, NY, USA, 341–350.
1586 <https://doi.org/10.1145/3239060.3239078>
- 1587 [70] Daniel J. McAllister. 1995. Affect- and Cognition-Based Trust as
1588 Foundations for Interpersonal Cooperation in Organizations. *The*
1589 *Academy of Management Journal* 38, 1 (1995), 24–59. <http://www.jstor.org/stable/256727>
- 1590 [71] White House Office of Science and Technology Policy. 2020. *American*
1591 *AI Initiative: Year One Annual Report*. Technical Report. White House
1592 Office of Science and Technology Policy, Brussels, Belgium. 36 pages.
1593 <https://www.whitehouse.gov/ai/>
- 1594 [72] Chinasa T. Okolo, Srujana Kamath, Nicola Dell, and Aditya Vashistha.
1595 2021. "It Cannot Do All of My Work": Community Health Worker
1596 Perceptions of AI-Enabled Mobile Health Applications in Rural In-
1597 dia. In *Proceedings of the 2021 CHI Conference on Human Factors in*
1598 *Computing Systems* (Yokohama, Japan) (CHI '21). Association for
1599 Computing Machinery, New York, NY, USA, Article 701, 20 pages.
1600 <https://doi.org/10.1145/3411764.3445420>
- 1601 [73] High-Level Expert Group on AI (AI HLEG). 2019. *Building Trust in*
1602 *Human-Centric Artificial Intelligenc*. Technical Report. European
1603 Commission, Brussels, Belgium. 11 pages. <https://ec.europa.eu/jrc/communities/en/community/digitranscope/document/building-trust-human-centric-artificial-intelligence>
- 1604 [74] Jeroen Ooge, Shotallo Kato, and Katrien Verbert. 2022. Explaining
1605 Recommendations in E-Learning: Effects on Adolescents' Trust. In
1606 *27th International Conference on Intelligent User Interfaces* (Helsinki,
1607 Finland) (IUI '22). Association for Computing Machinery, New York,
1608 NY, USA, 93–105. <https://doi.org/10.1145/3490099.3511140>
- 1609 [75] Rachel O'Dwyer. 2018. Algorithms are making the same mistakes as
1610 humans assessing credit scores. Retrieved April 17 (2018), 2019.
- 1611 [76] Samir Passi and Steven J. Jackson. 2018. Trust in Data Science:
1612 Collaboration, Translation, and Accountability in Corporate Data
1613 Science Projects. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article
1614 136 (Nov. 2018), 28 pages. <https://doi.org/10.1145/3274405>
- 1615 [77] Bako Rajaonah, Franoise Anceaux, Nicolas Tricot, and Marie-Pierre
1616 Pacaux-Lemoine. 2006. Trust, Cognitive Control, and Control: The
1617 Case of Drivers Using an Auto-Adaptive Cruise Control. In *Pro-*
1618 *ceedings of the 13th European Conference on Cognitive Ergonomics:*
1619 *Trust and Control in Complex Socio-Technical Systems (ECCE '06)*.
1620 Association for Computing Machinery, New York, NY, USA, 17–24.
1621 <https://doi.org/10.1145/1274892.1274896>
- 1622 [78] Bogdana Rakova, Jingying Yang, Henriette Cramer, and Rumman
1623 Chowdhury. 2021. Where Responsible AI Meets Reality: Practitioner
1624 Perspectives on Enablers for Shifting Organizational Practices.
1625 *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 7 (April 2021),
1626 23 pages. <https://doi.org/10.1145/3449081>
- 1627 [79] Divya Ramesh, Vaishnav Kameswaran, Ding Wang, and Nithya Sam-
1628 basivan. 2022. How Platform-User Power Relations Shape Algorith-
1629 mic Accountability: A Case Study of Instant Loan Platforms and
1630 Financially Stressed Users in India. In *2022 ACM Conference on Fair-*
1631 *ness, Accountability, and Transparency* (Seoul, Republic of Korea)
1632 (FAccT '22). Association for Computing Machinery, New York, NY,
1633 USA, 1917–1928. <https://doi.org/10.1145/3531146.3533237>
- 1634 [80] Mark Ryan. 2020. In AI we trust: Ethics, artificial intelligence, and
1635 reliability. *Sci. Eng. Ethics* 26, 5 (Oct. 2020), 2749–2767.
- 1636 [81] S Samoilii, Cobo M Lopez, E Gomez Gutierrez, G De Prato, F Martinez-
1637 Plumed, and B Delipetrev. 2020. *AI WATCH. Defining Artificial In-*
1638 *telligence*. Technical Report KJ-NA-30117-EN-N (online). European
1639 Commission, Luxembourg (Luxembourg). [https://doi.org/10.2760/382730\(online\)](https://doi.org/10.2760/382730(online))
- 1640 [82] Devansh Saxena, Karla Badillo-Urquiola, Pamela J. Wisniewski,
1641 and Shion Guha. 2021. A Framework of High-Stakes Algorithmic
1642 Decision-Making for the Public Sector Developed through a Case
1643 Study of Child-Welfare. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2,
1644 Article 348 (oct 2021), 41 pages. <https://doi.org/10.1145/3476089>
- 1645 [83] Devansh Saxena, Erina Seh-Young Moon, Aryan Chaurasia, Yixin
1646 Guan, and Shion Guha. 2023. Rethinking "Risk" in Algorithmic Sys-
1647 tems Through A Computational Narrative Analysis of Casenotes in
1648 Child-Welfare. In *Proceedings of the 2023 CHI Conference on Human*
1649 *Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Asso-
1650 ciation for Computing Machinery, New York, NY, USA, Article 873,
1651 19 pages. <https://doi.org/10.1145/3544548.3581308>
- 1652 [84] Kristin E. Schaefer, Deborah R. Billings, James L. Szalma, Jeffrey K.
1653 Adams, Tracy Sanders, Jessie Y. C. Chen, and Peter A. Hancock. 2014.
1654 *A Meta-Analysis of Factors Influencing the Development of Trust in*
1655 *Automation: Implications for Human-Robot Interaction*. Technical
1656 Report. U.S. Army Research Laboratory.
- 1657 [85] Kristin E. Schaefer, Jessie Y. C. Chen, James L. Szalma, and P. A.
1658 Hancock. 2016. A Meta-Analysis of Factors Influencing the Devel-
1659 opment of Trust in Automation: Implications for Understanding
1660 Autonomy in Future Systems. *Human Factors* 58, 3 (2016), 377–400.

- 1651 <https://doi.org/10.1177/0018720816634228> PMID: 27005902.
- 1652 [86] James Schaffer, John O'Donovan, James Michaelis, Adrienne Raglin,
1653 and Tobias Höllerer. 2019. I Can Do Better than Your AI: Expertise
1654 and Explanations. In *Proceedings of the 24th International Conference*
1655 *on Intelligent User Interfaces (IUI '19)*. Association for Computing
1656 Machinery, New York, NY, USA, 240–251. <https://doi.org/10.1145/3301275.3302308>
- 1657 [87] Matthew U Scherer. 2015. Regulating artificial intelligence systems:
1658 Risks, challenges, competencies, and strategies. *SSRN Electron. J.* 29
1659 (2015), 48.
- 1660 [88] Timothée Schmude, Laura Koesten, Torsten Möller, and Sebastian
1661 Tschiatsek. 2023. On the Impact of Explanations on Understanding
1662 of Algorithmic Decision-Making. In *Proceedings of the 2023 ACM*
1663 *Conference on Fairness, Accountability, and Transparency* (Chicago,
1664 IL, USA) (FAccT '23). Association for Computing Machinery, New
1665 York, NY, USA, 959–970. <https://doi.org/10.1145/3593013.3594054>
- 1666 [89] Jakob Schöffner, Yvette Machowski, and Niklas Köhl. 2021. A Study
1667 on Fairness and Trust Perceptions in Automated Decision Making.
1668 In *Joint Proceedings of the ACM IUI 2021 Workshops, April 13–17, 2021,*
1669 *College Station, USA (CEUR Workshop Proceedings)*, Vol. 2903. RWTH
1670 Aachen, Aachen, Germany, 170005.
- 1671 [90] Ian A Scott, Stacy M Carter, and Enrico Coiera. 2021. Exploring stake-
1672 holder attitudes towards AI in clinical practice. *BMJ Health & Care In-*
1673 *formatics* 28, 1 (2021), 7. <https://doi.org/10.1136/bmjhci-2021-100450>
1674 arXiv:<https://informatics.bmj.com/content/28/1/e100450.full.pdf>
- 1675 [91] Younho Seong, Ann M Bisantz, and Ann M Bisantz. 2002. Judg-
1676 ment and Trust in Conjunction with Automated Decision Aids:
1677 A Theoretical Model and Empirical Investigation. *Proceedings*
1678 *of the Human Factors and Ergonomics Society Annual Meeting* 46,
1679 3 (2002), 423–427. <https://doi.org/10.1177/154193120204600344>
1680 arXiv:<https://doi.org/10.1177/154193120204600344>
- 1681 [92] Accenture Federal Services. 2019. *Responsible AI: A Framework for*
1682 *Building Trust in your AI Solutions*. Technical Report. Accenture.
1683 13 pages. [https://www.accenture.com/us-en/insights/us-federal-](https://www.accenture.com/us-en/insights/us-federal-government/ai-is-ready-are-we)
1684 [government/ai-is-ready-are-we](https://www.accenture.com/us-en/insights/us-federal-government/ai-is-ready-are-we)
- 1685 [93] Venkatesh Sivaraman, Leigh A Bukowski, Joel Levin, Jeremy M. Kahn,
1686 and Adam Perer. 2023. Ignore, Trust, or Negotiate: Understanding
1687 Clinician Acceptance of AI-Based Treatment Recommendations in
1688 Health Care. In *Proceedings of the 2023 CHI Conference on Human*
1689 *Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Asso-
1690 ciation for Computing Machinery, New York, NY, USA, Article 754,
1691 18 pages. <https://doi.org/10.1145/3544548.3581075>
- 1692 [94] Jessie J. Smith, Anas Buhayh, Anushka Kathait, Pradeep Ragothaman,
1693 Nicholas Mattei, Robin Burke, and Amy Voida. 2023. The Many Faces
1694 of Fairness: Exploring the Institutional Logics of Multistakeholder
1695 Microlending Recommendation. In *Proceedings of the 2023 ACM Con-*
1696 *ference on Fairness, Accountability, and Transparency* (Chicago, IL,
1697 USA) (FAccT '23). Association for Computing Machinery, New York,
1698 NY, USA, 1652–1663. <https://doi.org/10.1145/3593013.3594106>
- 1699 [95] Janet A. Sniezek and Lyn M. Van Swol. 2001. Trust, Confidence, and
1700 Expertise in a Judge-Advisor System. *Organizational Behavior and*
1701 *Human Decision Processes* 84, 2 (2001), 288–307. <https://doi.org/10.1006/obhd.2000.2926>
- 1702 [96] Steven C. Sutherland, Casper Harteveld, and Michael E. Young.
1703 2015. The Role of Environmental Predictability and Costs in Re-
1704 lying on Automation. In *Proceedings of the 33rd Annual ACM Con-*
1705 *ference on Human Factors in Computing Systems* (CHI '15). Associ-
1706 ation for Computing Machinery, New York, NY, USA, 2535–2544.
1707 <https://doi.org/10.1145/2702123.2702609>
- 1708 [97] AI Taskforce. 2019. *Report of Estonia's AI Taskforce*. Technical Re-
1709 port. Republic of Estonia Government Office and Republic of Estonia
1710 Ministry of Economic Affairs and Communications, Estonia. 47
1711 pages. [https://ec.europa.eu/knowledge4policy/ai-watch/estonia-ai-](https://ec.europa.eu/knowledge4policy/ai-watch/estonia-ai-strategy-report)
1712 [strategy-report](https://ec.europa.eu/knowledge4policy/ai-watch/estonia-ai-strategy-report)
- 1713 [98] Suzanne Tolmeijer, Markus Christen, Serhiy Kandul, Markus Kneer,
1714 and Abraham Bernstein. 2022. Capable but Amoral? Comparing
1715 AI and Human Expert Collaboration in Ethical Decision Making. In
1716 *Proceedings of the 2022 CHI Conference on Human Factors in Com-*
1717 *puting Systems* (New Orleans, LA, USA) (CHI '22). Association for
1718 Computing Machinery, New York, NY, USA, Article 160, 17 pages.
1719 <https://doi.org/10.1145/3491102.3517732>
- 1720 [99] Peter-Paul van Maanen, Francien Wisse, Jurriaan van Diggelen, and
1721 Robbert-Jan Beun. 2011. Effects of Reliance Support on Team Perfor-
1722 mance by Advising and Adaptive Autonomy. In *Proceedings of the*
1723 *2011 IEEE/WIC/ACM International Conferences on Web Intelligence and*
1724 *Intelligent Agent Technology - Volume 02 (WI-IAT '11)*. IEEE Computer
1725 Society, New York, NY, USA, 280–287. [https://doi.org/10.1109/WI-](https://doi.org/10.1109/WI-IAT.2011.117)
1726 [IAT.2011.117](https://doi.org/10.1109/WI-IAT.2011.117)
- 1727 [100] Lyn M Van Swol and Janet A Sniezek. 2005. Factors affecting the
1728 acceptance of expert advice. *Br J Soc Psychol* 44, Pt 3 (Sept. 2005),
1729 443–461.
- 1730 [101] Rama Adithya Varanasi and Nitesh Goyal. 2023. “It is Currently
1731 Hodgepodge”: Examining AI/ML Practitioners’ Challenges during
1732 Co-Production of Responsible AI Values. In *Proceedings of the 2023*
1733 *CHI Conference on Human Factors in Computing Systems* (Hamburg,
1734 Germany) (CHI '23). Association for Computing Machinery, New
1735 York, NY, USA, Article 251, 17 pages. [https://doi.org/10.1145/3544548.](https://doi.org/10.1145/3544548.3580903)
1736 [3580903](https://doi.org/10.1145/3544548.3580903)
- 1737 [102] Michael Veale, Max Van Kleek, and Reuben Binns. 2018. Fairness
1738 and Accountability Design Needs for Algorithmic Support in High-
1739 Stakes Public Sector Decision-Making. In *Proceedings of the 2018*
1740 *CHI Conference on Human Factors in Computing Systems* (CHI '18).
1741 Association for Computing Machinery, New York, NY, USA, 1–14.
1742 <https://doi.org/10.1145/3173574.3174014>
- 1743 [103] Oleksandra Vereschak, Gilles Bailly, and Baptiste Caramiaux. 2021.
1744 How to Evaluate Trust in AI-Assisted Decision Making? A Survey
1745 of Empirical Methodologies. *Proc. ACM Hum.-Comput. Interact.* 5,
1746 CSCW2, Article 327 (Oct. 2021), 39 pages. <https://doi.org/10.1145/3476068>
- 1747 [104] Cédric Villani, Yann Bonnet, Bertrand Rondepierre, et al. 2018. *For*
1748 *a meaningful artificial intelligence: Towards a French and European*
1749 *strategy*. Conseil national du numérique, France.
- 1750 [105] Xiuxin Wang and Xiufang Du. 2018. Why Does Advice Discounting
1751 Occur? The Combined Roles of Confidence and Trust. *Front Psychol*
1752 9 (Nov. 2018), 2381.
- 1753 [106] Xinru Wang and Ming Yin. 2023. Watch Out for Updates: Understand-
1754 ing the Effects of Model Explanation Updates in AI-Assisted Decision
1755 Making. In *Proceedings of the 2023 CHI Conference on Human Factors*
1756 *in Computing Systems* (Hamburg, Germany) (CHI '23). Association
1757 for Computing Machinery, New York, NY, USA, Article 758, 19 pages.
1758 <https://doi.org/10.1145/3544548.3581366>
- 1759 [107] David Gray Widder, Derrick Zhen, Laura Dabbish, and James Herb-
1760 sleb. 2023. It's about Power: What Ethical Concerns Do Software
1761 Engineers Have, and What Do They (Feel They Can) Do about
1762 Them?. In *Proceedings of the 2023 ACM Conference on Fairness, Ac-*
1763 *countability, and Transparency* (Chicago, IL, USA) (FAccT '23). Asso-
1764 ciation for Computing Machinery, New York, NY, USA, 467–479.
1765 <https://doi.org/10.1145/3593013.3594012>
- 1766 [108] Allison Woodruff, Sarah E. Fox, Steven Rousso-Schindler, and Jef-
1767 frey Warshaw. 2018. A Qualitative Exploration of Perceptions
1768 of Algorithmic Fairness. In *Proceedings of the 2018 CHI Confer-*
1769 *ence on Human Factors in Computing Systems* (CHI '18). Associ-
1770 ation for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3173574.3174230>
- 1771 [109] Jun Xiao, John Stasko, and Richard Catrambone. 2007. The Role
1772 of Choice and Customization on Users’ Interaction with Embodied
1773 Conversational Agents: Effects on Perception and Performance. In
1774 *Proceedings of the SIGCHI Conference on Human Factors in Computing*
1775 *Systems* (New York, NY, USA) (CHI '07). Association for Computing
1776 Machinery, New York, NY, USA, Article 100, 17 pages. <https://doi.org/10.1145/1280533.1280633>

1761				
1762				
1763	[110]	Ilan Yaniv and Eli Kleinberger. 2000. Advice Taking in Decision Making: Egocentric Discounting and Reputation Formation. <i>Organizational Behavior and Human Decision Processes</i> 83, 2 (2000), 260 – 281. https://doi.org/10.1006/obhd.2000.2909		
1764				
1765				
1766	[111]	Nur Yildirim, Mahima Pushkarna, Nitesh Goyal, Martin Wattenberg, and Fernanda Viégas. 2023. Investigating How Practitioners Use Human-AI Guidelines: A Case Study on the People + AI Guidebook. In <i>Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems</i> (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 356, 13 pages. https://doi.org/10.1145/3544548.3580900		
1767				
1768				
1769				
1770				
1771				
1772	[112]	Ming Yin, Jennifer Wortman Vaughan, and Hanna Wallach. 2019. Understanding the Effect of Accuracy on Trust in Machine Learning Models. In <i>Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems</i> (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300509		
1773				
1774				
1775				
1776	[113]	Kun-Hsing Yu and Isaac S Kohane. 2019. Framing the challenges of artificial intelligence in medicine. <i>BMJ Quality & Safety</i> 28, 3 (2019), 238–241. https://doi.org/10.1136/bmjqs-2018-008551		
1777				
1778				
1779				
1780				
1781				
1782				
1783				
1784				
1785				
1786				
1787				
1788				
1789				
1790				
1791				
1792				
1793				
1794				
1795				
1796				
1797				
1798				
1799				
1800				
1801				
1802				
1803				
1804				
1805				
1806				
1807				
1808				
1809				
1810				
1811				
1812				
1813				
1814				
1815				
			arXiv: https://qualitysafety.bmj.com/content/28/3/238.full.pdf	1816
	[114]	Mireia Yurrita, Tim Draws, Agathe Balayn, Dave Murray-Rust, Nava Tintarev, and Alessandro Bozzon. 2023. Disentangling Fairness Perceptions in Algorithmic Decision-Making: The Effects of Explanations, Human Oversight, and Contestability. In <i>Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems</i> (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 134, 21 pages. https://doi.org/10.1145/3544548.3581161		1817
				1818
				1819
				1820
				1821
				1822
				1823
	[115]	Mireia Yurrita, Dave Murray-Rust, Agathe Balayn, and Alessandro Bozzon. 2022. Towards a Multi-Stakeholder Value-Based Assessment Framework for Algorithmic Systems. In <i>2022 ACM Conference on Fairness, Accountability, and Transparency</i> (Seoul, Republic of Korea) (FAccT '22). Association for Computing Machinery, New York, NY, USA, 535–563. https://doi.org/10.1145/3531146.3533118		1824
				1825
				1826
				1827
				1828
	[116]	Yunfeng Zhang, Q. Vera Liao, and Rachel K. E. Bellamy. 2020. Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-Assisted Decision Making. In <i>Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency</i> (FAT* '20). Association for Computing Machinery, New York, NY, USA, 295–305. https://doi.org/10.1145/3351095.3372852		1829
				1830
				1831
				1832
				1833
				1834
				1835
				1836
				1837
				1838
				1839
				1840
				1841
				1842
				1843
				1844
				1845
				1846
				1847
				1848
				1849
				1850
				1851
				1852
				1853
				1854
				1855
				1856
				1857
				1858
				1859
				1860
				1861
				1862
				1863
				1864
				1865
				1866
				1867
				1868
				1869
				1870