



Delft University of Technology

## Selective Trust

### Understanding Human-AI Partnerships in Personal Health Decision-Making Process

Van Arum, Sterre; Genç, Hüseyin Uğur; Reidsma, Dennis; Karahanoğlu, Armağan

#### DOI

[10.1145/3706598.3713462](https://doi.org/10.1145/3706598.3713462)

#### Publication date

2025

#### Document Version

Final published version

#### Published in

CHI 2025 - Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems

#### Citation (APA)

Van Arum, S., Genç, H. U., Reidsma, D., & Karahanoğlu, A. (2025). Selective Trust: Understanding Human-AI Partnerships in Personal Health Decision-Making Process. In N. Yamashita, V. Evers, K. Yatani, X. Ding, B. Lee, M. Chetty, & P. Toups-Dugas (Eds.), *CHI 2025 - Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* Article 1026 ACM. <https://doi.org/10.1145/3706598.3713462>

#### Important note

To cite this publication, please use the final published version (if applicable).  
Please check the document version above.

#### Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

#### Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.  
We will remove access to the work immediately and investigate your claim.



# Selective Trust: Understanding Human-AI Partnerships in Personal Health Decision-Making Process

Sterre van Arum\*

Interaction Design, Human Media Interaction  
University of Twente  
Enschede, Netherlands  
s.r.vanarum@utwente.nl

Dennis Reidsma

Human Media Interaction  
University of Twente  
Enschede, Netherlands  
d.reidsma@utwente.nl

Hüseyin Uğur Genç\*

Industrial Design Engineering  
TU Delft  
Delft, Netherlands  
u.genc@tudelft.nl

Armağan Karahanoğlu

Interaction Design  
University of Twente  
Enschede, Netherlands  
a.karahanoglu@utwente.nl

## Abstract

As artificial intelligence (AI) becomes more embedded in personal health technology, its potential to transform health decision-making through personalised recommendations is becoming significant. However, there is limited understanding of how individuals perceive AI-assisted decision-making in the context of personal health. This study investigates the impact of AI-assisted decision-making on trust in physical activity-related health decisions. By employing MoveAI, a GPT-4.0-based physical activity decision-making tool, we conducted a mixed-methods study and conducted an online survey (N=184) and semi-structured interviews (N=24) to explore this dynamic. Our findings emphasise the role of nuanced personal health recommendations and individual decision-making styles in shaping trust in AI-assisted personal health decision-making. This paper contributes to the HCI literature by elucidating the relationship between decision-making styles and trust in the AI-assisted personal health decision-making process and showing the challenges of aligning AI recommendations with individual decision-making preferences.

## CCS Concepts

• **Human-centered computing** → Human computer interaction (HCI); Empirical studies in HCI.

## Keywords

AI-assisted decision-making, personal health technology, trust, decision-making styles, physical activity decisions

## ACM Reference Format:

Sterre van Arum, Hüseyin Uğur Genç, Dennis Reidsma, and Armağan Karahanoğlu. 2025. Selective Trust: Understanding Human-AI Partnerships in Personal Health Decision-Making Process. In *CHI Conference on Human*

\*Both authors worked on this research equally.



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

CHI '25, Yokohama, Japan

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1394-1/2025/04

<https://doi.org/10.1145/3706598.3713462>

*Factors in Computing Systems (CHI '25)*, April 26–May 01, 2025, Yokohama, Japan. ACM, New York, NY, USA, 21 pages. <https://doi.org/10.1145/3706598.3713462>

## 1 Introduction

Personal health technologies (PHT) (e.g., Fitbit<sup>1</sup> and Apple Health<sup>2</sup>) currently function as tools that support individuals in making health-related decisions [24, 39], by collecting data (e.g., calories burned) [63] and facilitating making lifestyle choices (e.g., reducing fat intake) [25]. With the increasing integration of artificial intelligence (AI), PHT are anticipated to transform personal health decision-making (HDM) through predictions and personalised lifestyle recommendations [55]. For example, FitnessAI<sup>3</sup>, an AI-powered fitness app, promises to analyse an individual's exercise form and technique, suggest adjustments to their form, and provide tailored exercise recommendations based on user preferences and feedback.

The evolution of the integration of AI with PHT offers significant potential. Future PHT is expected to engage individuals in interactive dialogues that personalise the recommendations to their specific health concerns and preferences [26]. As PHT evolves from simple tracking tools to more autonomous and collaborative systems, understanding how individuals perceive and interact with AI-assisted HDM becomes crucial [50]. According to Vereschak et al. [77], people's perceptions of AI systems are shifting from viewing PHT merely as "limited to single-purpose supportive tools" to considering them as **collaborative** peers who are capable of taking initiative and responding to a wider range of queries. This shift is likely to influence AI-assisted personal HDM as well.

However, human-AI collaboration in personal HDM brings challenges as well, particularly given that individuals might, rightly or wrongly, trust the accuracy of technology-based health decisions more than those made by doctors [30]. For example, Chen et al., [9] show that individuals with limited health knowledge often trust social media and other online resources more than medical professionals, which might also be repeated in human-AI collaboration. Even recently, it appeared in the news that individuals started to ask

<sup>1</sup>[www.fitbit.com](https://www.fitbit.com) (Last retrieved on 1 September 2024)

<sup>2</sup><https://www.apple.com/ios/health/> (Last retrieved on 1 September 2024)

<sup>3</sup><https://www.fitnessai.com> (Last retrieved on 1 September 2024)

ChatGPT about their undiagnosed health issues<sup>4</sup>, and many people believe that ChatGPT can make the right diagnosis for them<sup>5</sup>.

Human-AI collaboration in personal HDM is a new type of collaboration which involves collaborative goal-setting and progress tracking [79]. For some, this collaboration might resemble the *doctor-patient partnerships* that occur in clinical settings [e.g., 59, 61, 86]. In such partnerships, healthcare professionals and patients go through information exchange, deliberation, and decision-making together, by considering the clinical evidence and patient preferences [62]. This process, known as the *shared decision-making process* [8, 43], values the decisions of *informed* patients [21]. It enables both parties to contribute to the final treatment decision in various ways [8]: Patients trust healthcare professionals to make personalised health decisions [52] and expect their expertise and knowledge to be applied to their unique health situations [22], while patient preferences are valued in treatment decisions [43]. Developing a similar *team-like* collaboration between humans and AI can help transform the way we think about human-AI interactions, as humans might grant more autonomy to AI in such partnerships [54].

Meanwhile, decision-making in everyday life is a dynamic process, which requires several competencies [7] and certain cognitive capacities [37]. Individuals go through and reflect on various sources of information, which, in turn, affects their trust in the decisions made [35]. In AI-assisted decision-making, personal experience plays a significant role in trusting the AI systems and judging the reliability of AI-based recommendations [78]. In fact, various factors such as knowledge [17], decision-making styles [67] and perceived trust in the tools and the process [33], influence trust in decision-making processes [78], which would also be the case for AI-assisted personal HDM. Hence, the integration of AI in PHT and collaborative decision-making with PHT still present challenges, particularly in understanding how individuals interact with, collaborate, and perceive the recommendations provided by AI algorithms. While considerable progress has been made in recent years in understanding the perspectives of healthcare providers in making decisions through human-AI collaboration [e.g., 32, 41, 60], this knowledge has not yet been transferred to human-AI collaboration in personal health decision-making. This translation requires more focused and dedicated research about human-AI collaboration in personal health decision-making, which has not been done yet.

We address this gap by exploring two key questions: (1) How do individuals perceive the process of AI-assisted collaborative personal health decision-making? and (2) How do such processes affect trust in AI-assisted health decision-making? In this paper, we choose to focus on physical activity, due to the widespread adoption of fitness trackers and health apps which have familiarised many individuals with technology-assisted physical activity decisions [34, 87]. Many people use these tools primarily to learn about their step counts, which also makes it a practical entry point for us to investigate what happens when individuals are presented with dialogues about their physical activity behaviour. In doing so, our aim is not to assess the effectiveness of a particular AI-assisted PHT

in increasing physical activity but to use physical activity as a lens through which we explore the broader implications of AI assistance on the trustworthiness of personal HDM processes.

Our paper examines the dynamics of possible future scenarios and offers recommendations for developing AI-assisted PHT that helps individuals make informed health decisions. As such, we aim to provide a dedicated source for demonstrating how AI behaviour and human decision-making styles [67] affect the trust in AI-assisted personal HDM process. Such a source can also open up new directions for addressing the grand challenges that Mamykina et al. [50] discussed about the role of AI in personal informatics.

To address our research questions, we conducted an online study with 184 participants and interviews with 24 of them to understand their perceptions of an AI-assisted collaborative physical activity decision-making process. Our contribution to the HCI literature is two-fold: (1) we clarify the relationship between decision-making styles and trust in AI-assisted personal health decision-making (HDM), and (2) we highlight the challenges of aligning AI's personal health recommendations with individuals' decision-making styles. In the end, we provide future directions for developing AI-assisted personal health technology that is personalised in both content and interaction and requires experiential explainability that resonates with individuals beyond mere technical explanations.

## 2 Related work

### 2.1 Human-AI Collaboration in Personal Health Technology Use

Over the past decade, the landscape of personal health technology (PHT) has undergone a significant transformation. Advancements, particularly in sensor technologies [e.g., 40, 64] and artificial intelligence [e.g., 10, 79], led to the development of PHT that automatically track various aspects of health such as physical activity [14], sleep patterns [45] and even stress levels [18]. Today, smart wearables like Garmin and Apple can predict exercise performance from individuals' training intensity and volume [38].

AI-assisted PHT holds significant promise in enhancing personal health management through collaboration between humans and AI systems. For example, Stawarz et al. [73] found that AI-assisted personal health decision support that considers distinct types of situations (e.g., expected or unexpected routines) could help individuals better handle their unforeseen health circumstances. Large Language Model (LLM)-based health tools also offer responsive and inclusive interactions that can provide health support [49]. These tools can accommodate natural language conversations and can make health advice more accessible and personalised. LLM-supported PHT can also play the role of an expert coach that motivates individuals towards discovering actions regarding their personal health goals [36]. In the study of Strömel et al. [74], LLM-generated narratives around fitness tracking data were found to help individuals better reflect on their personal well-being data. Although not yet optimal, ChatGPT-generated training plans for runners were also found to provide training guidance, especially for informed exercisers [19]. Individuals expect, but miss, LLM-based health apps to present precise and accurate health information and provide accurate diagnosis [56]. These developments suggest that

<sup>4</sup><https://www.today.com/health/mom-chatgpt-diagnosis-pain-rcna101843> (Last retrieved on 1 September 2024)

<sup>5</sup><https://www.forbes.com/sites/johnkoetsier/2024/01/02/dr-gpt-84-say-chatgpt-got-their-diagnosis-right/> (Last retrieved on 1 September 2024)

AI-assisted PHTs have the potential, but are not there yet, to become proactive assistants that integrate data from multiple sources and provide more accurate and context-aware health advice [50].

Meanwhile, several scholars [e.g., 47, 48, 71] criticise the direction of human-technology relations in personal health tracking. For example, Fox [27] contends that PHT is “intrusive on people’s daily lives” and contributes to the ‘medicalisation of everyday life’ (which was indicated by [13]). As a result, PHT is criticised for interfering with the principles of personal autonomy, making people no longer able to evaluate their own health needs but leave them with PHT-determined goals [27]. This view is also supported by Schneiderman et al., [70] who identified the need for interfaces that enable individuals “to better understand underlying computational processes” and provide them “the potential to better control their (the algorithms’) actions”. These perspectives require *collaboration* rather than *technological imposition* to ensure human autonomy. This evolution, in return, can support more informed and collaborative personal HDM [73].

This shows that the potential for specific types of AI support (e.g., chatbots or voice agents) to enhance or supplement current PHT remains an area for further exploration, particularly in terms of complementing the output such as the lifestyle recommendations [50]. In that regard, human-AI collaboration can be inspired by the *shared decision-making* that occurs in clinical settings [21, 23, 42]. Shared decision-making (SDM) is a collaborative process where patients and healthcare providers work together to make health decisions, by considering the best clinical evidence and the patient’s values and preferences [21, 54]. It is a process in which **choice talk**, **option talk**, and **decision talk** occur. During the **choice talk**, the healthcare provider acknowledges that a decision needs to be made and that the patient is invited to participate in making that decision. **Option talk** involves discussing the available options in detail, including the risks and benefits of each, to ensure the patient is well-informed. Finally, **decision talk** focuses on supporting the patient as they consider their preferences and decide on the best course of action.

Research further highlights the advantages of SDM. For instance, involving patients in discussions about antibiotics reduces unnecessary prescriptions [42] and empowers them with a better understanding of the risks and benefits of health decisions [43]. Similarly, AI systems can process large datasets and generate evidence-based recommendations [1], while humans contribute to contextual understanding and personal preferences [72]. However, individuals with insufficient health knowledge may struggle to understand health information [83], which leads to confusion when attempting to comprehend the implications of their health decisions [46].

Yet, the potential of misinformation in AI-generated PHT content raises concerns [84]. For example, health experts tend to process less information, but consider risks better and make sharper distinctions among various possible decisions [66]. Without such nuanced understanding, individuals might make health decisions based on incomplete or misunderstood information that AI provides [29]. Therefore, AI needs to supplement, rather than replace health professionals in personal HDM contexts, as medical decisions require value judgments and empathy that AI generally lacks [15]. Thus, it is critical to prevent potential harm by evaluating the dynamics of human-AI collaboration in HDM.

## 2.2 Dynamics of Health Decision-Making in Human-AI Collaboration

According to traditional models, decision-making is a systematic approach in which decision-makers identify options, generate supporting and opposing arguments, and carefully weigh these arguments to reach a well-informed decision [5, 35]. Byrnes [7] describes this process as a stepwise one, which involves setting a goal (e.g., have fun with friends), compiling options for producing that goal (e.g., play sports or go for a joy ride), rank-ordering options (e.g., playing sports is better than going for a joy ride), and finally choosing the option with the highest ranking. However, decision-making tends to be a more dynamic, recursive thinking process which involves ongoing construction and evaluation of arguments to resolve conflicts among competing options [35].

Kahneman’s seminal “Dual-System Theory” [37] delineates two types of thinking processes: System 1, which is fast, intuitive, and prone to biases, and System 2, which is slower, deliberate, and analytical thinking. Depending on the context, decision-makers might navigate between these systems, with System 1 being favoured in routine or high-pressure situations where making speedy decisions is essential, while System 2 is engaged in more complex or unfamiliar scenarios that require detailed analysis. Relatedly, decision-making in everyday life is associated with certain decision-making styles [16], which are the habitual ways individuals approach choices. People might have **spontaneous** (i.e., making quick decisions), **dependent** (i.e., relying on others’ guidance), **intuitive** (i.e. trusting gut feelings), **avoidant** (i.e., delaying decisions), and **rational** (i.e., carefully analysing information) decision-making styles [44, 58]. These decision-making styles often overlap rather than function independently, as individuals may employ multiple styles simultaneously or switch between them depending on the context. This overlap illustrates the fluidity in how people make decisions and affects their reliance on either System 1 or System 2 thinking. For instance, when people are busy, they have more on their minds and are pressed for time, therefore, they are more likely to rely on System 1 thinking as opposed to System 2 thinking [53]. Yet, when faced with emotionally loaded decisions (such as choosing a lifetime partner), System 2 thinking may lead people to make decisions they later regret [53, 82].

These insights help us understand how individual ways of decision-making might affect HDM with AI-assisted PHT. First of all, human self-confidence and individual experience often affect the adoption of AI suggestions during AI-assisted decision-making [11]. Research shows that individuals’ trust in AI recommendations can influence the speed of –and confidence in– decision-making [76]. Reliance on System 1 can increase susceptibility to cognitive biases such as overconfidence or anchoring [65]. In contrast, low trust in AI might push individuals towards System 2 thinking, which can result in a more rigorous evaluation of the AI’s recommendations than it requires and lead to excessive deliberation. People might, on the other hand, trust incorrect AI recommendations, even when they have greater prior knowledge or are given information about the system’s low confidence in the recommendations [75].

Miscalibrated trust can also arise due to the Dunning-Kruger effect, where less competent individuals overestimate their abilities,

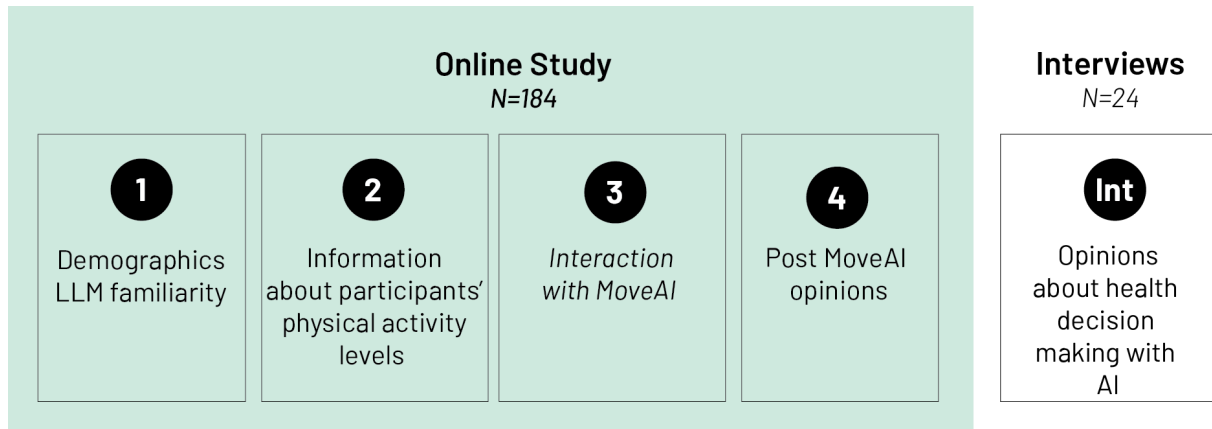


Figure 1: Overview of Methodology of the Study

which leads to under-reliance on AI systems [31]. Therefore, the effectiveness of human-AI collaboration in personal health decisions hinges on how well these systems can address the diverse needs of different decision-making styles. For example, WB de Bruin et al., [6] find that someone's type of decision-making style in a real-life setting can predict decision-making competencies. This finding is significant for our study because it highlights the importance of tailoring AI-assisted PHT to accommodate the individuals' decision-making styles.

Therefore, each decision-making style—rational, intuitive, spontaneous, avoidant, and dependent [44, 58]—presents unique needs and behaviours that AI tools must accommodate. For example, individuals with a rational decision-making style might benefit from AI systems that provide comprehensive data analysis and clear rationales for their recommendations. Meanwhile, those with a more dependent decision-making style may require AI systems that offer reassurance through consistent feedback and validation and help to build their confidence in the decisions they make. In contrast, those with an intuitive decision-making style may rely on gut feelings and individual experiences, which can cause them to delay action-taking due to affective responses.

In short, personalising health recommendations to accommodate different **decision-making styles** can enhance human autonomy, particularly in promoting adherence to proper **AI-assisted personal health decisions**. In that regard, we first understand the interplay between individuals' decision-making styles, their **trust in AI tools and the AI-assisted HDM**. This knowledge can help tackle the challenges we explained in section 2.1. This way, we aim to open new research directions for PHT that facilitate AI-assisted HDM and consider individual needs, differences and their impacts on trust in AI-assisted HDM.

### 3 Methodology

We carried out a mixed methods study (see Figure 1 for an overview) to address how individuals perceive AI-assisted collaborative personal HDM process. To this aim, we first designed an online study in which the participants interacted with *MoveAI*, an LLM-based interactive HDM-support tool. This interaction provided us with

insights into participants' reflections about trust in the AI-assisted HDM process. The online study was designed to allow us to follow up on the responses of participants and gain more in-depth insights. Therefore, we requested participants to leave their email addresses if they were interested in a follow-up interview.

Our study was grounded in the presumption that AI-assisted health decision-making processes can be inspired by, and benefit from, a shared decision-making approach [21, 23, 42] where AI serves as a supplementary tool that personalises the HDM and health advice. This presumption guided the overall design of the online study and interview questions, as well as the prompts of *MoveAI* which aimed to foster collaborative HDM. We obtained ethical approval from our research institute before we recruited participants.

#### 3.1 Online Study

We designed the online study by using the Qualtrics XM survey platform to investigate individuals' interactions with AI in personal health decision-making. The study consisted of four parts, with several questions and an interactive part, *MoveAI* (see Table 1 for details of the study setup). In Part 1, participants provided demographic information including their age, sex, and education level. As we stated earlier, we specifically focused on health choices around physical activity (PA) levels. This deliberate choice was due to many people's familiarity with physical activity tracking which makes them aware of their PA levels. Aligned with this choice, in Part 2, we assessed participants' physical activity (PA) levels with the well-established "Stages of Change" scale [51]. The scale assesses where an individual stands in terms of readiness to adopt or maintain regular physical activity (PA). It consists of four statements which identify five stages of change: pre-contemplation (i.e., no engagement with regular PA), contemplation (i.e., awareness about the benefits of PA, not active enough), preparation (i.e., taking steps for being regularly active, but not consistently), action (e.g., actively engaging in regular PA but for less than 6 months), and maintenance (i.e., regularly engaged in PA for more than 6 months). We also asked the participants about the tools they use to track their physical activity.

**Table 1: Questions and Setup of the Online Study**

Part	Question	Type	Way of Data Collection
1	Age	Open-ended (text box)	Numeric entry
1	Gender	Closed (single choice)	3 choices (female/male/prefer not to answer)
1	Familiarity with LLMs	Closed (single choice)	5 choices (Not at all to extremely)
2	Stages of change (PA levels)	Closed (single choice)	4 items, with yes/no responses [51]
2	PA tracker use	Open-ended (text box)	Text entry
	Whether the tracker gives PA suggestions	Closed (single choice)	3 choices (yes/no/I do not know)
	Degree of compliance with those suggestions	Closed (single choice)	6 choices (not at all to a great deal or not applicable)
2	Trust in various sources in PA decision-making	Rating Scale (importance)	6 questions about the example human sources of knowledge (i.e., medical professionals, personal sports trainers, personal dietitian/nutritionists, general practitioners (GP), close friends and family members)
3	Decision-making with MoveAI	Interacting with MoveAI based on patient-doctor shared-decision-making steps [implemented from 22, 43, 59]	
4	Opinions about MoveAI	Rating Scale (satisfaction)	3 questions about participants' opinions of their interaction with MoveAI
4	Perceptions about MoveAI	Likert Scale (agreement)	3 questions about whether MoveAI respected/understood the participant
4	Trust in the decision-making process with MoveAI	Likert Scale (agreement)	14 questions [adapted from 3]
4	Decision-making styles of participants	Likert Scale (agreement)	25 questions [original questions from 58]

The core of the online study (Part 3) involved an interactive part where participants engaged with *MoveAI* (Figure 2) our HDM tool powered by OpenAI's GPT-4.0. We positioned *MoveAI* as a supplementary support in the HDM process, which reflects our presumption that AI assistance in HDM should enhance, not replace, autonomy in personal health decisions. Aligned with this, we configured *MoveAI* to simulate a conversation about physical activity advice, with participants instructed to interact for at least five minutes. This interaction followed a pre-set structure (explained in section 3.2) to guide the participants through discussing their physical activity habits with *MoveAI* and receiving personalised advice from it. To interact with *MoveAI*, participants were directed from the Qualtrics survey page to the *MoveAI* interaction page through a separate link. At the end of this interaction, participants were again redirected to the Qualtrics form. We stored all the participant conversations with the corresponding user IDs gathered from Qualtrics so that we could check whether participants had a real conversation with *MoveAI*.

In the final part of the online study (Part 4), participants were asked to fill out a series of questionnaires to assess their experience with *MoveAI*. This part focused on two key areas: their trust in the AI-assisted decision-making process and their decision-making styles. To measure trust in the decision-making process, we used the validated "Trust in AI-infused Decision-Making Process" scale [adapted from 3]. The scale consisted of five sub-scales: trustworthiness (i.e., whether the process was trustworthy), technical competence (i.e., whether the AI was competent in decision-making),

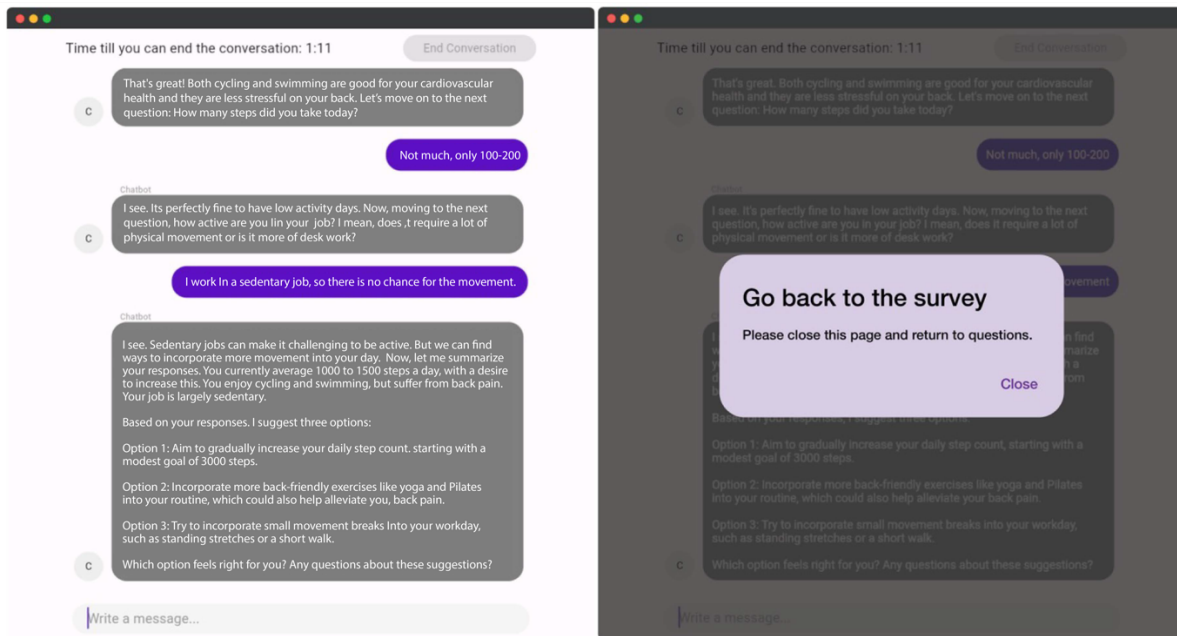
understandability (i.e., whether the decision-making process was understandable), reliability (i.e., whether the process would always result in similar results) and attachment (i.e., whether the participant likes the AI-infused decision-making process). Additionally, we used a measurement scale based on the work of Phillips et al. [original questions from 58] to understand the extent to which participants' decisions are influenced by dependent, avoidant, spontaneous, rational, and intuitive decision-making styles. This measurement scale helps understand how participants' inherent decision-making tendencies influence their interaction with *MoveAI* and their trust in the AI-assisted decision-making process.

Our online study was carried out in English. The questions included a combination of open-ended (e.g., age, the tools they used), single-choice (e.g., their familiarity with LLMs) and five-point Likert and rating scale questions (e.g., how satisfied they are with the knowledge of *MoveAI* on a scale of 1= not at all to 5=very important). As an incentive, 20 randomly selected participants received a gift card with a value of 25\$.

### 3.2 MoveAI: Decision-Making with an LLM-Based Physical Activity Advice Agent

We used OpenAI's API for their GPT-4.0 model and integrated a custom chatbot application, *MoveAI*, (Figure 2) to simulate a collaborative decision-making process. The conversation structure of *MoveAI* was inspired by shared decision-making (SDM) that occurs in clinical settings [8, 22, 42, 43, 54, 57, 59]. We presumed that AI-assisted HDM can foster conversations that could follow the steps





**Figure 2: Example Participant interaction with MoveAI (Left). After the countdown ended, participants were prompted to return to the survey (Right).**

of SDM while maintaining a natural, engaging dialogue with individuals. By aligning its conversational flow with SDM, we aimed for MoveAI to act as an HDM partner, rather than an authority, and support its users in making personalised health decisions.

We crafted the conversation structure of MoveAI to mirror the three-step talk model of SDM: **choice talk**, **option talk**, and **decision talk** [23]. This structure aims to guide the participants through a systematic decision-making process, starting with gathering contextual information about their current physical activity levels, potential barriers, and daily routines (i.e., choice talk). Based on this information, MoveAI presented personalised options, purposefully introducing elements of negotiation to simulate real-world health discussions (i.e., option talk). After presenting options, MoveAI encouraged participants to further dialogue and negotiate towards a shared decision (i.e., decision talk).

To achieve this desired conversation flow, we employed a Structured Prompting Approach [81] for the GPT-4.0 model. Our prompts defined MoveAI's role as a proactive, stern, yet neutral physical activity advisor, and outlined specific rules for interaction (e.g., response length limits and guidelines for steering off-topic responses). The prompt structure included a task definition, persona guidelines, conversation rules, a scripted conversation flow, and a conversation example, as shown in Table 5 in the Appendix.

We developed MoveAI with Flutter<sup>6</sup> and recorded each conversation session in Google Firebase<sup>7</sup>. Each message is tagged with a corresponding participant ID that facilitates data collection and allows for post-hoc analysis of conversation patterns and decision-making processes. Before deployment of the study, we ran various tests and checked the conversations to understand whether and

to what extent the resulting conversations that GPT produced reflected the intentions embedded in the prompt (see Table 6 in the appendix for a complete example of conversation).

It should be noted that our goal in the online study was to understand individuals' perspectives about the possibilities of using AI in their personal health decision-making. Hence, we did not test the feasibility of MoveAI as such. Instead, we wanted to show the opportunities of a possible AI-assisted HDM process to our participants and learn about their perspectives about –and trust in– the AI-assisted personal health decision-making process. In the interviews, we opened up their thoughts with follow-up questions.

### 3.3 Interview Study

We structured semi-structured interviews as a complementary follow-up to the online study. For example, in the survey, we asked participants about their level of satisfaction with the responses of MoveAI, and in the interviews, we expanded on the reasons *why* they were/not satisfied with the responses. Our interviews consisted of three sets of questions. The first set focused on understanding the role of technology and AI in participants' physical HDM (as was also asked in the survey; see Table 1, part 2), including their planning of physical activity and their use of technology in physical activity. The second set looked into the participant's experience of MoveAI (Table 1, part 3). Lastly, with the third set, we investigated what they thought about the suggestions MoveAI provided to them and why they do/not trust MoveAI or similar LLM-supported physical activity decision-making tools (Table 1, part 4). This set also probed participants about their boundaries

<sup>6</sup><https://flutter.dev>

<sup>7</sup><http://firebase.google.com>

for using a tool similar to MoveAI in the future. We carried out the interviews in English unless both the interviewer and the interviewee declared that they spoke the same local native language ( $N=9$ ).

### 3.4 Participant Recruitment

Our data collection took 3.5 months between 15 February and 30 May 2024. To participate in the study, the respondents had to be at least 18 years old and understand their physical activity levels. We distributed the online study link via our personal contacts, online forums, campus posts and social platforms such as Strava and Instagram. The online study went offline once there was no response for 7 consecutive days. In the study, we enabled a ReCAPTCHA score to filter potential useless responses and remove nonlogical responses.

In total, we received 258 complete online study responses. We discarded the responses if the ReCAPTCHA score was lower than 0.5. After that, we checked whether each respondent had actual conversations with MoveAI by checking their records from the conversation database. We removed the responses that did not have any conversations. In the end, we ended up with 184 complete responses. On average, the online study took 20 minutes.

Our online study participants ranged from 18 to 62 years old, with an average age of  $M=31.01$  ( $SD=9.25$ ). Of the participants,  $N=84$  (46%) identified themselves as female,  $N=98$  (53%) as male, while  $N=2$  (1%) preferred not to answer. The majority of the participants declared to be from Western societies (98%). According to the stages of change questions [51], the majority of participants ( $N=119$ , 64.7%) were at the maintenance stage (i.e., doing at least 30 min of vigorous activity 5 days/week and more than 6 months) in their physical activity habits. The rest were at pre-contemplation  $N=7$  (3.8%); contemplation  $N=23$  (12.5%); preparation  $N=13$  (7.1%) and action  $N=22$  (12%) stages. Only  $N=85$  (46.2%) participants indicated that their tracker gives physical activity suggestions. Of the participants,  $N=39$  (21%) indicated that they do not use any technology to track their physical activity, and those who track,  $N=85$  (46%) indicated that their tracker gives them PA suggestions, while they indicated their compliance with the suggestions was low ( $M=2.40$ ,  $SD=1.27$ ). Online study participants were highly educated, with 84.7% having a bachelor's degree or higher.

For the interviews, we reached out to the interested participants to achieve a gender distribution similar to the survey responses. We also aimed at having a similar distribution for levels of trust in AI-assisted decision-making process responses (Part 4 in Table 1). We informed the participants that we would carry out the interviews in English. Of the 32 participants who received an email invitation for the interviews, we received 25 positive responses, and we interviewed 24 participants (Table 2). Most participants ( $N=22$ , 92%) were somewhat familiar (3 out of 5) with LLMs.

The first author interviewed the participants through the online video conferencing platform Microsoft Teams. All sessions were recorded with participants' vocal consent (as was also indicated in IRB). The interviews took, on average, 30 minutes. Interview participants comprised of equal number of female and male participants; on average  $M=34.71$  years of age ( $SD=11.83$ ,  $Min=23$ ,  $Max=57$ ), with demographics that were close to survey participants.

The majority ( $N=23$ , 95%) were using at least one activity-tracking app or a wearable.

### 3.5 Data Analysis

We employed various analysis methods as our study yielded both quantitative and qualitative data. For quantitative analysis, we used IBM SPSS Statistics 26, and calculated frequencies, means ( $M$ ) and standard deviations ( $SD$ ) for questions with rating and Likert scales (e.g., participants' perceptions about their interaction with MoveAI). We calculated the frequencies for the open-ended questions, such as age, gender and physical activity levels. We use these results to report the demographics of the participants.

For the "Trust in AI-infused decision-making process" and "Decision-making styles" scales, we first ran a reliability analysis for both scales (See Appendix). We then ran t-tests to check whether participants trusted human sources significantly more or less than MoveAI. We also ran a correlation analysis to check whether a specific type of decision-making style was correlated with the trust in decision-making with MoveAI.

We conducted Reflexive Thematic Analysis (RTA) [4, 12]. RTA was first conducted by the first author, who conducted the interviews, in frequent correspondence with all authors. We first downloaded the recordings and auto transcriptions provided by Microsoft Teams and reviewed the transcriptions with the recordings to ensure that the interviews were transcribed verbatim. The first author then translated nine non-English interview transcriptions to English and ensured that the content of the translated and original interviews was the same. After that, to become acquainted with the data, the first author read the transcripts and made notes (*Familiarisation*). They then used the last author to help them code the data from the first five interviews using both inductive and deductive coding approaches (*Coding*). During deductive coding, we used the survey questions as a reference (e.g., sources of determining health decisions). The codes were rearranged around the initial code groups that addressed our research question. Through a series of discussion sessions (*Developing and reviewing*), the authors arrived at a shared understanding of the code groups. The first author then coded the remaining transcripts. Meanwhile, the second author worked on the affinity diagramming of the codes and code groups in a separate document. Throughout this process, all authors kept in close contact to refine and finalize the themes (*Refining, defining and naming*). After we had the final set of code groups, we wrote them up, aligned to our research questions (*Writing up*).

In the next section, we present the results under four headings. The first and the second sections provide insights about the baseline for addressing our research questions (i.e., 1-sources of determining health decisions, and 2-assessing physical activity knowledge of MoveAI), while the third and fourth sections provide insights about tackling the research questions (i.e., how individuals perceive AI-assisted collaborative personal-HDM processes, and how such processes affect their trust in decision-making).

### 3.6 Limitations

The demographic characteristics of our sample pose limitations to the study's applicability. Our participants were predominantly from Western societies and largely consisted of highly educated



**Table 2: Characteristics of the Interview Study Participants**

PN	Demographics		TheirPA levels*	Trust in DM with MoveAI	Decision-making Styles**				
	Age	Gender			I	D	R	Av	S
P01	40	Male	M	5.00	4.00	2.40	4.00	1.80	3.40
P02	24	Female	PC	5.00	3.60	3.20	4.40	2.00	2.60
P03	23	Female	A	3.00	4.00	3.60	3.00	1.20	3.00
P04	43	Male	C	4.00	4.40	3.20	4.80	3.20	2.20
P05	25	Female	M	5.00	3.80	4.00	4.40	1.80	1.80
P06	42	Female	M	4.00	2.40	2.40	5.00	1.00	1.40
P07	32	Female	P	4.00	3.40	3.00	4.60	2.00	2.40
P08	26	Male	M	3.50	3.20	4.60	4.20	4.40	1.80
P09	26	Male	A	3.00	4.00	3.00	4.80	3.60	3.60
P10	37	Male	M	3.50	4.80	2.00	4.40	1.80	3.60
P11	32	Male	M	5.00	2.80	2.80	3.80	1.00	2.00
P12	48	Male	M	3.00	3.80	3.20	3.40	1.40	1.80
P13	25	Male	A	3.00	2.40	3.60	4.20	1.80	1.20
P14	27	Male	M	3.00	2.00	3.40	4.00	2.40	1.60
P15	49	Female	P	4.00	3.00	3.00	3.60	1.80	2.80
P16	39	Female	M	5.00	3.40	4.20	4.80	1.20	2.40
P17	34	Female	A	4.50	2.60	2.40	4.20	1.20	2.40
P18	57	Female	M	4.00	2.60	2.40	4.40	1.20	1.80
P19	38	Male	M	3.00	3.00	2.20	4.20	2.00	2.80
P20	42	Female	C	4.50	4.60	3.20	4.20	2.80	2.80
P21	24	Female	C	5.00	3.80	3.60	3.20	2.40	3.60
P22	37	Male	M	2.00	2.80	2.00	4.80	2.00	2.00
P23	35	Female	M	2.00	2.80	2.80	4.80	1.20	1.80
P24	28	Male	M	1.00	3.20	4.00	3.00	2.40	2.00
Mean	34.71			3.71	3.35	3.09	4.18	1.98	2.37

\* A=Action, C=Contemplation, M=Maintenance, P=Preparation, PC=Precontemplation

\*\*I=Intuitive, D=Dependent, R=Rational, Av=Avoidant, S=Spontaneous

individuals, with many having the same level of physical activity (PA) behaviour. This homogeneity might limit the generalisability of the findings, as perspectives from more diverse educational, cultural, or activity backgrounds were not sufficiently captured. Future research should consider including participants from a wider range of educational levels, cultural contexts, and physical activity groups to gain a more comprehensive understanding of how AI systems might support habit changes across different populations.

The tool we employed in our study, MoveAI, was an LLM-based tool, which suggests that the study results should be considered carefully before extending to all AI-assisted interactions. Caution must be exercised when speculating about how other AI systems might perform in similar contexts, as the specific AI used here may not represent broader AI technologies. Besides, while MoveAI provided us with valuable insights, it sometimes diverged from the intended topic, introducing spontaneous or unrelated suggestions. This occasional deviation might have impacted the focus of the conversations and could have limited the precision of MoveAI's guidance in supporting personal health decision-making.

It is important to highlight that our study involved hypothetical discussions about participants' potential physical activity behaviours. These simulated scenarios may not fully capture how

participants would engage with AI in real-world contexts, making extrapolations to actual behaviour more uncertain.

## 4 Making Physical Activity Decisions WITH AI-ASSISTED Interactions

In this section, we provide the insights our study yielded into individuals' perception of AI-assisted shared-personal health decision-making, specifically focusing on physical activity (PA) levels and choices. We examine participants' trust in AI-assisted reflection and the HDM process. In general, we did not observe any instances where interview participants contradicted or disagreed with the online study findings. However, our interview results provided a more nuanced perspective on integrating AI into the PA decision-making process. In the rest of this section, we report the findings from the online study as well as the interviews based on the order of questions provided in Table 1.

### 4.1 Sources of Determining Health Decisions

Our survey results showed that the participants placed higher trust in more knowledgeable sources when seeking health advice, specifically, 'medical professionals' were the most trusted ( $M=3.99$ ,  $SD=0.83$ ), followed by 'personal sports trainers' ( $M=3.64$ ,  $SD=0.84$ ),

**Table 3: Sources of Determining Health Decisions (N=184)**

Sources of Advice	M	SD
Medical professionals	3.99	0.83
Personal Sports Trainer	3.64	0.84
Personal dietitians/nutritionists	3.49	0.95
General Practitioners (GP)	3.47	1.00
MOVEAI	3.41	1.02
Close Friends	2.78	0.79
Family members	2.69	0.85

‘personal dietitians/nutritionists’ ( $M=3.49$ ,  $SD=0.95$ ) and ‘general practitioners’ ( $M=3.47$ ,  $SD=1.00$ ) in asking for health advice. Interestingly, the trustworthiness of MoveAI was only slightly lower than these experts ( $M=3.41$ ,  $SD=1.02$ ). In contrast, trust in ‘close friends’ ( $M=2.78$ ,  $SD=0.79$ ) and ‘family members’ ( $M=2.69$ ,  $SD=0.85$ ) for health advice was notably lower.

To further analyse these differences, we conducted a Wilcoxon signed-rank test which showed that trust in medical professionals in health decisions was significantly higher than MoveAI ( $Z = -5.76$ ,  $p < 0.01$ ), while trust in close friends ( $Z = -6.56$ ,  $p < 0.01$ ) and family members ( $Z = -7.08$ ,  $p < 0.01$ ) were significantly lower than MoveAI. There were no statistically significant differences in trust between MoveAI and personal sports trainers ( $Z = -2.50$ ,  $p = 0.01$ ), personal dietitians/nutritionists ( $Z = -0.61$ ,  $p = 0.54$ ), and general practitioners (GPs) ( $Z = -0.46$ ,  $p = 0.65$ ).

Our interviews provided additional insights into why MoveAI was nearly as trusted as other knowledgeable sources. A majority of participants ( $N=18$ ) indicated that they had a positive attitude towards MoveAI as it gave **personal advice** ( $N=9$ ) and had a human-like, **empathic aspect** ( $N=9$ ). Moreover, seven participants specifically noted that MoveAI encouraged them to **trust their own feelings** when making PA decisions, which aligned well with their personal expectations. For example, P1 indicated that he had a positive attitude, as MoveAI considered his injury when offering advice. Interestingly, P1 also anthropomorphised MoveAI, using gendered pronouns about it:

*‘Yeah, it was specifically saying multiple times that I should listen to my own body about my injury. And that was kind of fun to hear that **she said listen to your own body multiple times.**’*

On the other hand, some participants raised concerns about the limitations of MoveAI’s knowledge, which may have influenced their slightly lower trust compared to human experts. Its knowledge was indicated to be sometimes **generic** ( $N=17$ ), which made it follow a **pre-defined path** rather than tailoring advice to their specific needs ( $N=10$ ) and sometimes giving **confusing advice** ( $N=8$ ). For instance, P3 indicated that the advice MoveAI provided felt like it wanted them to follow its predefined path, which did not go very much into P3’s needs. In the next section, we will explain how the participants evaluated the knowledge of MoveAI further.

## 4.2 Assessing the Knowledge of MoveAI

We asked two questions to learn participant’s opinions about the accuracy and knowledge of MoveAI. On average, participants rated

“**accuracy of the information MoveAI provided them**” at  $M=3.75$ ,  $SD=1.01$  and “**knowledge of MoveAI**” at  $M=3.60$ ,  $SD=1.06$ , both scores above the midpoint of the rating scale. To gain deeper insights, we explored these evaluations through interviews and identified two primary ways participants assessed MoveAI’s knowledge: by fact-checking and questioning ( $N=15$ ) and by being cautious or wary of its advice ( $N=11$ ).

**Fact-checking and questioning** were a common strategy for evaluating MoveAI’s knowledge. Participants often challenged the information provided, cross-referencing it with what they already knew or testing the AI’s responses to ensure they were reliable. For example, P9 stated that in real life, they would not trust such a tool immediately. The participant stated that he ‘*would have to fact-check whether it would provide him with sources which he already knows that some AI models are capable of*’. This shows that the availability of credible sources is essential for them to trust the information that tools like MoveAI provide. We also found that trust in MoveAI’s knowledge appeared to be influenced by participants’ familiarity with LLMs and their awareness of their own physical activity levels and personal needs. Participants assessed knowledge of MoveAI by comparing its responses with common knowledge (e.g., taking a certain number of steps). However, for individuals like P1, who used ChatGPT in their professional lives, MoveAI’s credibility was quickly undermined when it made inappropriate recommendations such as advising regular running activities despite the participant’s knee injury.

**Being wary of the knowledge** of MoveAI was another prevalent theme. For example, in their interview, P16 stated that they would be suspicious of such an AI-based physical activity tracker in the beginning. They preferred to use MoveAI over an extended period—at least a month—to gauge its accuracy, assessing its advice based on physical outcomes ‘*when they do not feel pain, or tired and feel healthy after following MoveAI’s advice*’. Relatedly, when MoveAI’s advice is not contextualised and does not take external factors into account, the participants were more wary of the knowledge of MoveAI. The following quotation from P7 illustrates this:

*‘It could have asked more questions, so indeed, I was now asked about a number of steps and how fit I am, and I was given 3 options. The time someone has or other external factors, I really thought that was the most important thing I missed in it. So, I thought it was reasonable, but there were still areas for improvement.’*

It was interesting to see that five participants mentioned the knowledge of MoveAI was not convincing as it made physical activity suggestions that did not always align with their needs. P8 particularly indicated that it was repetitive, which made MoveAI seem to choose the easy way out unlike human-human interactions:

*‘It was a repetition of what I said and... I found it a bit easy, there is a lack of nuance, if I may compare it, for example with talking to a human being’ (P8)*

P23 was particularly frustrated by MoveAI’s rigid focus on daily steps goal, which felt limiting given their involvement in other types of sports. Yet, several participants stated that MoveAI made sensible, alternative physical activity suggestions, such as dancing (e.g., P05) or cycling to work (e.g., P01), which made them consider

that the suggestions of MoveAI can be insightful and eye-opening. These findings indicate that the knowledge of MoveAI and its accuracy influenced the way participants calibrated their trust. In the next section, we describe how participants perceived HDM conversations they had with MoveAI.

### 4.3 Behaviour of MoveAI and Perceptions about Collaborative HDM with MoveAI

We had three sources of data to reflect on how MoveAI responded during the interactions and how participants perceived its behaviour. The first data source was the online conversations participants ( $N=184$ ) had with MoveAI. Second, we analysed the content of the conversations that occurred with MoveAI. Finally, after their interaction with MoveAI, we asked the survey participants about their perceptions of the behaviour of MoveAI and future use.

**4.3.1 Behaviour of MoveAI.** Conversations with MoveAI typically proceeded through several episodes, or turns, of MoveAI suggestions and user responses. On average, one conversation contained 1.9 episodes (i.e., PA suggestion turns), while the maximum number of episodes was seven. In the first turn of interaction, MoveAI was highly proactive, offering three recommendations to encourage healthier routines as exemplified below:

*“Given the information you’ve provided, I can suggest three options for increasing your physical activity:*

*Option 1: Aim for 10,000 steps per day. It is a popular goal often recommended for overall health.*

*Option 2: Incorporate short walking breaks throughout your workday. This could be a brief walk every hour. Option 3: Try to increase the intensity of your existing workouts (running, cycling) gradually without overheating.” (MoveAI)*

The most frequent suggestion MoveAI made was aiming for a specific number of steps, which occurred 94 times in the first turn. This emphasis on step count was likely due to our prompt, which specified promoting step count as a key strategy. Other common initial suggestions included incorporating walks during breaks (76 times), adding an extra day of sports (48 times), and incorporating new exercises (33 times).

As the interactions progressed, the frequency of these suggestions (i.e., dominance of our system prompt) decreased significantly. For example, step count recommendations dropped from 94 in the first turn to just 16 in the second turn, and only once in the third turn. This decline was consistent across all types of suggestions, which also occurred due to the drop in the number of episodes.

Conversations revealed further insights into MoveAI’s behavioural characteristics. In the first turn, MoveAI’s responses can be classified as generic (115 conversations, e.g., *“I must say I think it would be beneficial for you to consider increasing your daily steps and setting a goal for yourself”*) but also showed a significant effort to find a good fit for the user (80 conversations, e.g., *“I understand space is a constraint. You can try standing exercises such as marches, leg lifts, or even shadow boxing”*).

MoveAI was sometimes informative (23 conversations) and it gave reasons for a suggestion by referring to a scientific study (e.g., *“A study by Tudor-Locke indicated that this could lead to significant*

*health benefits”*). At times it was careful (24 conversations) in taking personal circumstances such as injury or habits explicitly into account as a reason for its suggestion (e.g. *“Remember, always listen to your body and take breaks or slow down if you start feeling overheated”*).

Participant responses to MoveAI’s suggestions varied. In the first turn, some participants decided to choose to incorporate walks (34 conversations), aiming for a number of steps (18 conversations), or adding an extra day of sports (15 conversations), while some disagreed with the suggestions (29 conversations).

**4.3.2 Perceptions about Collaborative HDM with MoveAI.** Despite MoveAI’s initial emphasis on providing advice to increase step count, in response to the individual conversations with participants, it provided a wider range of advice over time. Instead of continuing to push for generalised goals, it began to suggest more personalised advice, such as taking the stairs or incorporating new exercises that were aligned better with what participants’ existing habits. These observations were also reflected in the survey results.

Survey participants indicated that they felt **being respected** by MoveAI ( $M=4.02$ ,  $SD=0.97$ ). However, the perceived understanding of MoveAI about participants’ physical activity choices was somewhat lower ( $M = 3.56$ ,  $SD = 1.04$ ). Participants’ intention to use a similar tool in the future was rated lower than the mean score of the rating scale ( $M = 2.90$ ,  $SD = 1.35$ ). These findings suggest that while participants appreciated the respectful and attentive behaviour of MoveAI, there were reservations about the overall effectiveness and future use of such a tool.

The interviews provided further insights into these mixed feelings. Fourteen participants specifically discussed MoveAI’s role as an advisor. Some of them found the advice provided by MoveAI to be **sensible and appropriate**, which contributed to a positive perception of its *advisory role* which suggests that MoveAI’s responsiveness to individual limitations played a role in building trust among some participants. For instance, P15 appreciated the pleasantness and respectfulness of MoveAI’s responses, noting:

*“I think it has just very pleasant answers (...) patients who are in such a chronic situation (or) with illness, for example, would find it very pleasant to be treated with respect. Yes, they were really listened to, so that was very nice. **Dealing with my limitations more respectfully than some humans...** I found that really striking.”*

A common positive perception was MoveAI’s **adaptability** and **non-judgmental** approach. Related to the previous comment of P2, MoveAI’s effort to adjust its advice based on participants’ input was appreciated by P2, as *“It took into account what I had said... and then it would actually adapt and give different advice”*. This adaptability was seen as a refreshing contrast to the rigidity, which was sometimes encountered in interactions with healthcare professionals.

Participants also found MoveAI’s **motivational and uplifting tone** to be a positive feature. P10 likened MoveAI to a *“really smart epic kid”* appreciating its consistently positive and encouraging demeanour. This positivity, combined with clear and precise communication, helped some participants feel more confident in following the advice. P10 described this aspect:

*“It was as I said **clear, precise and uplifting**... even if it is a robot and it doesn’t really get what it is like to take 10,000 steps and 9000 steps, it’s kind of... just that positive reinforcement is a good thing.”*

Interestingly, several participants **personified** MoveAI, using pronouns (i.e., he or she) when referring to it. This anthropomorphism suggests that some participants perceived human-like qualities in MoveAI, particularly in its approach to advice-giving. P20 remarked, “*Somehow it can give advice, not very statistical advice, but more intuitive personal human advice to motivate me... in a more personalised way.*”

Despite these positive perceptions, there was also a sense of caution. Six participants specifically indicated that they would prefer to receive advice from a human rather than a tool like MoveAI, especially for more critical or personal decisions. Relatedly, several participants expressed concerns about the depth of knowledge of MoveAI and its adaptability. For instance, P22 noted that while it initially asked about their preferences, it occasionally failed to adapt its advice accordingly, which resulted in a disconnection in its recommendations. P7 also mentioned instances where MoveAI’s advice, although based on user input, did not consider external factors or provide the nuanced guidance that human interactions might offer. This **lack of personalisation** was a critical issue for participants like P16, who expressed doubt about whether the advice was genuinely tailored or simply a result of a pre-programmed algorithm. P13 echoed this concern, stating, “*It is not very credible... because it doesn’t match my wishes or my background in what I already do.*”

Some participants, like P3, expressed a moderate level of trust in MoveAI due to its reliance on generally accepted internet resources and avoidance of unusual suggestions. Others remained **sceptical** and felt that MoveAI was overly **prescriptive**, always suggesting improvements or more intense activities without fully considering whether such advice was necessary or desired. P9 described this rigidity as **inflexibility**, feeling pressured to continually improve their routine despite being satisfied with it: “*It kind of forced me to take an improving action... I just take this one, and essentially what it says, it told me was: ‘Uh yeah OK just be better’ [...] Why should I mix it up?*”

Other participants pointed out that MoveAI sometimes failed to read between the lines, and missed the subtitles in conversations that a human advisor might catch. P8 criticised MoveAI for being overly literal and repetitive, and P07 highlighted its failure to consider external factors, such as their responsibilities as a single parent, which affected their ability to follow certain advice. In the next section, we will dive into and reflect on how this behaviour of MoveAI resulted in the differences between the different decision-making styles and the participants’ trust in HDM with it.

#### 4.4 Trust in the Physical Activity Decision-making Process with MoveAI

To explore the relationship between trust in the decision-making process with MoveAI and participants’ dominant decision-making styles, we conducted a correlation analysis. We used the survey data collected with the “Trust in AI-infused decision-making process”

scale [adapted from 3] and the “decision-making styles” scale [original questions from 58]. We calculated the correlation coefficients and their significance levels to determine whether the relationship between participants’ *trust in AI-infused decision-making* and *decision-making styles* were statistically significantly correlated, and the direction and strength of these relationships.

Before proceeding with the correlations, we first assessed the reliability of both scales, followed by calculating the means and standard deviations of each item and subscales to see the general tendency of participant responses. The findings provided insights into the aspects of trust in the AI-assisted decision-making process (e.g., technical competence vs understandability) the individuals would relate to depending on their decision-making styles (e.g., rational vs avoidant). We provided the detailed results, including reliability analyses of sub-scales, along with the mean scores and standard deviations of each item, in the Appendix.

The original trust in “AI-Infused Decision-Making Process” scale consists of 5 sub-scales. The reliability analysis resulted in the removal of two items from the “trustworthiness” sub-scale and one item from the “attachment” sub-scale improving the reliability of the sub-scales (which is a common practice in validating measurement scales [5]). We performed the same analysis for the ‘Decision-Making Styles’ scale [58] and we kept all the items since all sub-scales were reliably measuring the five dimensions of trust AI-Infused decision-making process and decision-making styles (Cronbach’s alpha values ranging between 0.70 and 0.90).

**4.4.1 Trust in AI-Assisted Decision-making Process.** We found that participants rated the trustworthiness of the decision-making with MoveAI moderately high ( $M=3.64$ ,  $SD=0.90$ ). They also evaluated the **technical competence** of MoveAI positively, with  $M=3.59$  ( $SD=0.97$ ), indicating that participants generally felt the system was appropriately and reliably implemented for the given use-case scenario. The **understandability** of the process was rated similarly ( $M=3.55$ ,  $SD=1.18$ ), suggesting that participants found the decision-making process to be moderately understandable. **Reliability**, or the belief that MoveAI would consistently make the same recommendations for others, received  $M=3.45$  ( $SD=0.89$ ). Lastly, participants’ **attachment** (i.e., how much they liked the decision-making process) was the lowest-rated dimension ( $M=3.33$ ,  $SD=1.20$ ).

**4.4.2 Decision-making Styles of the Participants.** The survey participants had high dominance in **rational** decision-making ( $M=3.97$ ,  $SD=0.65$ ), compared to **intuitive** ( $M=3.33$ ,  $SD=0.75$ ) and **dependent** decision-making styles ( $M=3.09$ ,  $SD=0.82$ ). Participants had less dominant **avoidant** ( $M=2.45$ ,  $SD=0.95$ ) and **spontaneous** ( $M=2.40$ ,  $SD=0.81$ ) decision-making styles. It is important to note that individuals can score high on multiple subscales simultaneously [44]; however, a deeper exploration of correlations between decision-making styles is beyond the scope of this paper. Yet, our correlation analysis yielded interesting insights.

**4.4.3 Relations between Decision-making Styles and AI-assisted HDM Process.** **Intuitive** decision makers often make decisions by relying on instincts, gut feelings, or immediate judgments rather than structured analysis or logical reasoning [44]. Our results

**Table 4: Correlations Between Decision-Making Styles and Trust in Decision-making with MoveAI (N=184)**

Trust in decision-making with MoveAI		Decision-making styles				
		Intuitive	Dependent	Rational	Avoidant	Spontaneousness
Trustworthiness	Pearson Correlation	0.05	0.18*	0.11	0.02	0.11
	Sig. (2-tailed)	0.48	.014	0.14	0.83	0.15
Reliability	Pearson Correlation	0.25**	-0.06	0.05	0.10	0.24**
	Sig. (2-tailed)	<.01	0.39	0.51	0.16	<.01
Attachment	Pearson Correlation	1.00**	0.18*	0.01	0.13	0.53**
	Sig. (2-tailed)	<.01	0.02	0.90	0.07	<.01
Understandability	Pearson Correlation	0.21**	-0.04	0.19*	-0.14	0.07
	Sig. (2-tailed)	<.01	0.63	0.01	0.07	0.38
Technical Competence	Pearson Correlation	0.17*	0.20**	0.11	0.11	0.20**
	Sig. (2-tailed)	0.02	0.01	0.13	0.14	<.01

\*  $p < 0.05$ , and \*\*  $p < 0.01$

showed a significant positive correlation between intuitive decision-making style and perceived reliability ( $r = 0.25$ ,  $p < 0.01$ ), attachment ( $r = 1.00$ ,  $p < 0.01$ ), understandability ( $r = 0.21$ ,  $p < 0.01$ ) and technical competence ( $r = 0.17$ ,  $p < 0.05$ ) of decision-making with MoveAI.

**Dependent** decision-makers seek input, support, or confirmation from people they trust, such as friends, family, or experts [44]. We found a significant positive correlation between dependent decision-making style and trustworthiness ( $r = 0.18$ ,  $p < 0.05$ ), attachment ( $r = 0.18$ ,  $p < 0.01$ ) and technical competence ( $r = 0.20$ ,  $p < 0.01$ ) of MoveAI.

Two contrasting, namely rational and avoidant, decision-making styles provided interesting insights further. **Rational** decision-makers use a logical, systematic approach to make decisions [44]. Participants who scored higher on rational decision-making style find decision-making with MoveAI more understandable ( $r = 0.19$ ,  $p < 0.05$ ). **Avoidant** decision makers tend to delay, ignore, or avoid making decisions altogether, often because they feel uncomfortable, anxious, or overwhelmed by the decision-making process [44]. We found no significant correlation between avoidant decision-making styles and the subscales for trust in decision-making with MoveAI ( $p > 0.05$ ).

Finally, **spontaneous** decision-makers prefer making quick, impulsive decisions, often with little deliberation or detailed analysis [44]. We found a significant positive correlation between participants scoring high on spontaneous decision-making styles and perceived **reliability** ( $r = 0.24$ ,  $p < 0.01$ ), **attachment** ( $r = 0.53$ ,  $p < 0.01$ ) and **technical competence** ( $r = 0.20$ ,  $p < 0.01$ ) of the decision-making process with MoveAI. All other remaining correlations were nonsignificant ( $p > 0.05$ ) for all decision-making styles.

## 5 Discussions

Our findings reveal that individuals remain cautious in fully trusting the AI-assisted HDM process in making personal health decisions. Even though our study had a hypothetical case for participants, it suggests that the future of such a decision-making process will likely be shaped by a balance between caution and gradual acceptance. Our work yielded three key takeaway messages:

- **Trust in AI-assisted decision-making remains a significant factor**, with individuals demonstrating a preference for human expertise, especially for making personal health decisions.
- Enhancing **personalisation, context awareness, and transparency** is crucial for building trust in the AI-assisted personal health decision-making process.
- **Decision-making styles** play a significant role in how individuals perceive and trust AI-assisted health decision-making.

Our work highlights the **considerations for making AI-enhanced PHT a tool in collaborative personal health decision-making**. Our results also emphasise the need for individuals' cautious and gradual acceptance. As AI-assisted PHT evolves, one can expect a future where such tools offer a data-driven decision-making process with transparency and personalised insights. Our results point to challenges and opportunities for consideration, which we will discuss in the following lines.

### 5.1 Challenges of Promoting AI as a Source of Knowledge in Personal Health Decisions

Early studies on personal informatics [2, 28, 57] emphasised the potential of personal health technology to empower individuals by providing detailed insights into their health behaviours. Over the past decade, AI-assisted personal health technology has evolved, by incorporating sensor technologies [e.g., 40, 64] and machine learning [e.g., 10, 79] and offer more personalised and context-aware health recommendations. However, despite these advancements, promoting AI as a reliable source of knowledge in personal health decisions remains a challenge.

Our study echoes the findings from the prior research, particularly the hesitancy and cautious trust that individuals exhibit towards AI-assisted health tools [76]. Our survey results show that AI-powered tools like MoveAI are trusted slightly lower, but almost as much as human experts, such as medical professionals, personal trainers, and dietitians. Individuals appreciate the fact that tools like MoveAI provide facts and suggestions, but also unlike medical

professionals, it should adapt to the user's standpoint and not be "too strict". This approach can be "too agreeable" and fit the tendency of AI-generated health advice to always align with the user's preferences, complaints or objections. Such an approach, in return, might result in people over-trusting the AI-generated suggestions and viewing the responses to be "as valid as doctors" [69]. The slightly lower trust in MoveAI compared to human experts suggests that AI-powered tools like MoveAI almost fully earn participants' confidence as independent sources of health knowledge. When an individual expresses reluctance to follow a particular course of action, MoveAI-like tools offer alternatives, generate suggestions, or even apologise.

Trust in knowledgeable others, such as medical professionals, often hinges on various factors, including accreditation, transparency in expertise, and personalised engagement. People tend to trust experts, who can prove their qualifications, such as where they received their training or education. This is akin to the questions individuals might ask about AI, such as the data it was trained on. Similarly, just as individuals expect human experts to acknowledge and understand unique situations [22], they also want AI tools to recognise and adapt to the specifics of their interactions [54]. The desire for factual knowledge and understandable reasoning applies equally to both human experts and AI, indicating that the factors driving trust in experts are consistent across different types of agents.

One of the primary challenges identified in our study is the perceived lack of nuance and context in AI-generated recommendations. Participants noted that while MoveAI provided generally accurate information, it could not often tailor advice to their specific needs and circumstances. This limitation aligns with critiques in the literature that AI-assisted health tools can sometimes follow a predetermined path [27], failing to account for individual variability and the complexities of real-life health scenarios. This echoes the concerns raised by Schneiderman et al. [70] who argue that AI systems must evolve to offer more transparent and customizable interactions to better serve users.

Moreover, our findings suggest that individuals tend to fact-check and question AI-generated advice, a behaviour that reflects their underlying scepticism about AI's accuracy and knowledge base [80]. This cautious approach is consistent with the literature [68], which highlights the importance of human oversight [88] and can have a high impact on AI-assisted health decision-making. The need for fact-checking indicates that AI tools, while valuable, are not yet seen as fully autonomous in providing personal health advice [50]. This finding highlights the necessity of maintaining a *team-like collaboration* between AI systems and human expertise [54] to mitigate the risk of misinformation and enhance informed trust.

Our findings suggest that HCI researchers should be careful when promoting AI-based personal HDM tools as a reliable source of knowledge in personal health decisions. First, AI systems need to incorporate more context-aware algorithms that can adapt to individuals' unique health conditions and external factors. Enhancing the personalisation of AI recommendations would help bridge the gap between the technology and the nuanced understanding provided by human professionals. Furthermore, our insights highlight the potential for AI tools to complement, rather than replace

human expertise, in HDM. By positioning AI as a supportive tool that augments human judgment rather than as an independent authority, the integration of AI into personal health management could become more widely, but carefully accepted. This approach aligns with the shared decision-making models in clinical settings [21, 23, 43], where AI serves as an evidence-based resource that informs, but does not dictate, people's choices.

## 5.2 Considering Decision-Making Styles in Development of AI-Assisted Personal Health Technology

Our study provides insights into the dynamics of AI-assisted collaborative decision-making in personal health, specifically in the domain of physical activity. We showed that the interaction between individuals and MoveAI follows a distinct pattern, where initial AI recommendations tend to be generic, but gradually become more personalised based on user input. This shift from general advice (e.g., increasing step counts) to more tailored recommendations highlights the potential of AI to adapt and cater to the specific preferences and routines of users over time. Furthermore, participants' trust in MoveAI varied significantly depending on their decision-making styles. For instance, those with a rational decision-making style rated MoveAI higher on understandability, while those with a dependent style showed more trust in MoveAI's reliability and attachment. These findings highlight the importance of individual differences in the perception of AI systems and suggest that trust in AI is not only a function of the system's qualities but also deeply influenced by users' cognitive predispositions.

The adaptive behaviour of MoveAI reflects the broader potential for AI systems to serve as collaborative partners in health decision-making, moving beyond mere recommendation tools. The personalised feedback offered by MoveAI, particularly in later stages of interaction, suggests that future systems can adjust to user preferences and constraints, which is critical for fostering long-term engagement and trust. However, the mixed feelings about MoveAI's ability to provide nuanced, context-aware advice point to ongoing challenges. While some participants appreciated its responsiveness, others raised concerns about its inability to fully grasp personal limitations or external factors, such as individual schedules or health conditions.

Our findings align with prior studies that emphasise the role of trust in AI-assisted decision-making. Research by [3] points out that trust in AI systems is shaped by multiple dimensions, including reliability and technical competence. However, our research extends these insights by linking trust to decision-making styles, showing that different decision-making styles correlate with varying perceptions of AI reliability, understandability, and competence. Individuals' trust in AI can impact decision-making speed and confidence [76], with higher trust often leading to faster, more intuitive decisions. Reliance on System 1 thinking in these cases may increase susceptibility to cognitive biases such as overconfidence or anchoring [65]. Conversely, lower trust in AI-assisted decision-making can trigger more deliberative System 2 thinking, prompting individuals to rigorously evaluate AI recommendations, sometimes to the point of excessive scrutiny and overthinking. These insights suggest that trust affects not only the perceptions of AI reliability,



understandability, and competence but also directly shape decision-making processes.

Hence, it is essential to design future tools that cater to diverse decision-making styles while developers should be cautious when enhancing specific trust elements. For example, future tools should be designed with flexibility in mind to meet the needs of individuals with dominant spontaneous decision-making styles, while providing step-by-step guidance for avoidant decision-makers. Yet, rational decision-makers may benefit from systems that provide detailed explanations and data-driven rationales, whereas dependent decision-makers may prefer more supportive and validating interactions. Besides, although improving reliability might enhance trust among spontaneous users, it may not have the same effect on rational users who prioritise clarity and understandability.

Compared to earlier work on doctor-patient shared decision-making [8, 43], which focuses on mutual information exchange and collaboration, the human-AI partnership lacks the same depth of personalisation and empathy. While AI tools can demonstrate adaptability in their suggestions, they might still fall short in offering the nuanced, empathetic responses [15] often associated with human advisors. This suggests that even though AI tools like MoveAI can mimic certain aspects of human collaboration, they are not yet fully capable of replicating the complexity of human-to-human decision-making interactions, particularly in the emotional and contextual understanding of health decisions. Therefore, once AI systems become more integrated into PHT, they must ensure transparency, particularly for vulnerable users who may be more susceptible to over-trusting AI recommendations. Developing tools that are not only reliable but also capable of explaining their suggestions in a user-friendly manner, is key to promoting informed decision-making and preventing over-reliance on AI-assisted health decision-making.

### 5.3 Towards the Future of AI-Assisted Personal Health Technology

Given the results of our study and the challenges and opportunities identified above, we identify several distinct future directions in personal health technology (i.e., personal informatics and self-tracking apps). We found that AI-assisted health advice should better fit the individuals' needs in a way that PHT and the individuals work in a *team-like* setting. This way, the PHT opens new opportunities for the individuals to explore by considering the capabilities of the individuals. We think that HCI research can leverage PHT with the potential of AI research for supporting the personal HDM process. From our work, we extrapolate three developments that should be considered for this purpose.

First, the personalised *interaction and presentation of advice* that people value in MoveAI must separately be grounded in personalised *content*. Participants valued how MoveAI seemed to listen to them, addressed them respectfully and seemingly took their context and needs into account. However, this is more of a 'surface level' kind of personalisation compared to how personalisation is usually discussed in HCI literature. We found that participants did not feel that MoveAI actually adapted the content of its advice to the preceding conversation. Sometimes, they want to be reassured about their decisions, but then tools like MoveAI should be carefully developed

so as not to manipulate the user's decisions. Regarding this, the LLM that underlies such systems may be good for making people feel "understood and listened to" but should be extended with actual PHT data to achieve personalisation in content as well as in presentation. If that is not done well, AI-supported PHT may risk disappointment and disengagement from users when the content is less personalised than promised due to the surface-level presentation of the advice. This non-tailored advice leads to a perceived inflexibility of AI and the genericity of the advice.

Second, for the requirements of AI in the systems, the field of AI-supported PHT should consider moving beyond the demands posed by the AI Act<sup>8</sup>. The ethical considerations of AI-assisted health decisions are paramount. As AI systems become more integrated into personal health management, they must ensure transparency, particularly for vulnerable users who may be more susceptible to over-trusting AI recommendations. Information from AI systems can misguide people, which can lead to serious issues (e.g. injuries from performing a recommended exercise). However, attempts such as the AI Act, place most of the emphasis on *technical explainability*. Still, there is almost nothing about how technical explainability can only work if it also "resonates" with the user. We argue that individuals need another kind of approach to "feeling explained to" than merely providing the logical arguments and tracing the data links from input and training to output. This experiential perspective should get more attention than it currently has in the AI Act. Furthermore, the AI Act places much importance on user trust in AI systems but does not at all consider the role of people's decision-making styles and the varying impacts of decision-making styles on the risk of uncalibrated trust. Therefore, AI developers should ensure that users are well-informed about the risks of using AI in such decision-making settings. Moreover, our findings suggest how to build trust in AI partnerships for specific decision-making styles. The AI act should consider reviewing what happens when AI tries to "manipulate" users into believing their terms based on trust and decision styles.

Finally, the feeling of being "understood and listened to" relates to the concept of explainability in AI [20] which historically affected human decisions [85]. On the one hand, explainability is often seen as a technical characteristic of the AI: is it possible to logically trace the decisions of the AI back to specific elements in the input as well as in the training data? On the other hand, in our study, we found that people value that the advice content is right about their input, and AI presents that content carefully, in a respectful manner, and seemingly ties back into the preceding conversation. It might be because there is a distinction between the decisions of the AI being "technically explainable" versus the user "feeling explained to". The latter would require a more conversational approach to getting advice from AI-assisted PHT, where the individuals get the sense that the AI takes some pains to adapt its explanations to the user in a responsible manner.

## 6 Conclusion

In this paper, we explored the role of AI-assisted decision-making in personal health, by considering the fact that future personal health technology would implement more AI assistance to their

<sup>8</sup><https://artificialintelligenceact.eu>, last retrieved 1 September 2024

users. We particularly focused on trust in decision-making and decision-making styles. We demonstrated how AI systems like MoveAI can serve as collaborative partners in personal health decisions, but also highlighted the limitations in their ability to fully capture the nuances of individual preferences and contexts. Our findings addressed a gap in understanding the interplay between decision-making styles and trust in AI-assisted health decision-making, showing that people's trust varies significantly across decision-making styles. We argued that future AI tools should offer more personalised, context-aware interactions to cater to diverse decision-making styles and promote informed, balanced decision-making. Ultimately, our paper serves as a stepping stone towards integrating AI as a supportive, transparent tool in personal health management, especially in tools like personal informatics. We call for further research to redefine AI's role in assisting personal health decisions.

## Acknowledgments

The corresponding author thanks Dr. Aykut Coskun for the rich discussions they had in the process of shaping this paper.

This research was funded by Nederlandse Organisatie voor Wetenschappelijk Onderzoek - NWO award number 406.XS.01.112.

## References

- [1] Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y. and Kankanalli, M. (2018). Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada). Association for Computing Machinery. <https://doi.org/10.1145/3173574.3174156>
- [2] ALSlaity, A., Suruliraj, B., Oyebo, O., Fowles, J., steeves, d. and Orji, R. (2022). Mobile Applications for Health and Wellness: A Systematic Review. *Proc. ACM Hum.-Comput. Interact.*, 6, EICS, Article 171. <https://doi.org/10.1145/3534525>
- [3] Ashoori, M. and Weisz, J. D. (2019). In AI we trust? Factors that influence trustworthiness of AI-infused decision-making processes. *arXiv preprint arXiv:1912.02675*.
- [4] Braun, V. and Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative research in psychology*, 18, 3, 328-352. <https://doi.org/10.1080/14780887.2020.1769238>
- [5] Bruch, E. and Feinberg, F. (2017). Decision-Making Processes in Social Contexts. *Annual Review of Sociology*, 43, Volume 43, 207-227. <https://doi.org/10.1146/annurev-soc-060116-053622>
- [6] Bruin de Bruin, W., Parker, A. M. and Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of personality and social psychology*, 92, 5, 938. <https://psycnet.apa.org/doi/10.1037/0022-3514.92.5.938>
- [7] Byrnes, J. P. (2002). The development of decision-making. *Journal of Adolescent Health*, 31, 6, Supplement, 208-215. [https://doi.org/10.1016/S1054-139X\(02\)00503-7](https://doi.org/10.1016/S1054-139X(02)00503-7)
- [8] Charles, C., Gafni, A. and Whelan, T. (1997). Shared decision-making in the medical encounter: What does it mean? (or it takes at least two to tango). *Social Science & Medicine*, 44, 5, 681-692. [https://doi.org/10.1016/S0277-9536\(96\)00221-3](https://doi.org/10.1016/S0277-9536(96)00221-3)
- [9] Chen, X., Hay, J. L., Waters, E. A., Kiviniemi, M. T., Biddle, C., Schofield, E., Li, Y., Kaphingst, K. and Orom, H. (2018). Health literacy and use and trust in health information. *Journal of health communication*, 23, 8, 724-734. <https://doi.org/10.1080/10810730.2018.1511658>
- [10] Chignell, M., Wang, L., Zare, A. and Li, J. (2023). The Evolution of HCI and Human Factors: Integrating Human and Artificial Intelligence. *ACM Trans. Comput.-Hum. Interact.*, 30, 2, Article 17. <https://doi.org/10.1145/3557891>
- [11] Chong, L., Zhang, G., Goucher-Lambert, K., Kotovsky, K. and Cagan, J. (2022). Human confidence in artificial intelligence and in themselves: The evolution and impact of confidence on adoption of AI advice. *Computers in Human Behavior*, 127, 107018. <https://doi.org/10.1016/j.chb.2021.107018>
- [12] Clarke, V., Braun, V. and Hayfield, N. 2015. Thematic analysis. *Qualitative psychology: A practical guide to research methods* J. A. Smith, Sage Publications, Great Britain.
- [13] Conrad, P. (2007) *The medicalization of society: On the transformation of human conditions into treatable disorders*. Johns Hopkins University Press Baltimore.
- [14] Coskun, A. (2019). Design for Long-Term Tracking: Insights from a Six-Month Field Study Exploring Users' Experiences with Activity Trackers. *The Design Journal*, 22, 5, 665-686. <https://doi.org/10.1080/14606925.2019.1634447>
- [15] Cuadra, A., Wang, M., Stein, L. A., Jung, M. F., Dell, N., Estrin, D. and Landay, J. A. (2024). The Illusion of Empathy? Notes on Displays of Emotion in Human-Computer Interaction. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA). Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642336>
- [16] Dewberry, C., Juanchich, M. and Narendran, S. (2013). Decision-making competence in everyday life: The roles of general cognitive styles, decision-making styles and personality. *Personality and Individual Differences*, 55, 7, 783-788. <https://doi.org/10.1016/j.paid.2013.06.012>
- [17] Dikmen, M. and Burns, C. (2022). The effects of domain knowledge on trust in explainable AI and task performance: A case of peer-to-peer lending. *International Journal of Human-Computer Studies*, 162, 102792. <https://doi.org/10.1016/j.ijhcs.2022.102792>
- [18] Ding, X., Wei, S., Gui, X., Gu, N. and Zhang, P. (2021). Data Engagement Reconsidered: A Study of Automatic Stress Tracking Technology in Use. In *Proceedings of the Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan). Association for Computing Machinery. <https://doi.org/10.1145/3411764.3445763>
- [19] Dükling, P., Sperlich, B., Voigt, L., Van Hooren, B., Zanini, M. and Zinner, C. (2024). ChatGPT generated training plans for runners are not rated optimal by coaching experts, but increase in quality with additional input information. *Journal of Sports Science & Medicine*, 23, 1, 56. <https://doi.org/10.52082/jssm.2024.56>
- [20] Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T. and Morgan, G. (2023). Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys*, 55, 9, 1-33. <https://doi.org/10.1145/3561048>
- [21] Edwards, A. and Elwyn, G. (2009) *Shared decision-making in health care: Achieving evidence-based patient choice*. Oxford University Press, USA.
- [22] Elwyn, G. and Charles, C. 2001. Shared decision-making: the principles and the competencies. *Evidence-Based Patient Choice: Inevitable or Impossible?* A. G. Edwards and G. Elwyn, Oxford University Press
- [23] Elwyn, G., Frosch, D., Thomson, R., Joseph-Williams, N., Lloyd, A., Kinnersley, P., Cording, E., Tomson, D., Dodd, C., Rollnick, S., Edwards, A. and Barry, M. (2012). Shared Decision Making: A Model for Clinical Practice. *Journal of General Internal Medicine*, 27, 10, 1361-1367. <https://doi.org/10.1007/s11606-012-2077-6>
- [24] Epstein, D. A., Caldeira, C., Figueiredo, M. C., Lu, X., Silva, L. M., Williams, L., Lee, J. H., Li, Q., Ahuja, S. and Chen, Q. (2020). Mapping and taking stock of the personal informatics literature. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4, 4, 1-38. <https://doi.org/10.1145/3432231>
- [25] Epstein, D. A., Ping, A., Fogarty, J. and Munson, S. A. (2015). A lived informatics model of personal informatics. In *Proceedings of the Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. <https://doi.org/10.1145/2750858.2804250>
- [26] Fitzpatrick, G. and Ellingsen, G. (2013). A review of 25 years of CSCW research in healthcare: contributions, challenges and future agendas. *Computer Supported Cooperative Work (CSCW)*, 22, 609-665. <https://doi.org/10.1007/s10606-012-9168-0>
- [27] Fox, N. J. (2017). Personal health technologies, micropolitics and resistance: A new materialist analysis. *Health*, 21, 2, 136-153. <https://doi.org/10.1177/1363459315590248>
- [28] Free, C., Phillips, G., Watson, L., Galli, L., Felix, L., Edwards, P., Patel, V. and Haines, A. (2013). The Effectiveness of Mobile-Health Technologies to Improve Health Care Service Delivery Processes: A Systematic Review and Meta-Analysis. *PLOS Medicine*, 10, 1, e1001363. <https://doi.org/10.1371/journal.pmed.1001363>
- [29] Grote, T. and Berens, P. (2020). On the ethics of algorithmic decision-making in healthcare. *Journal of Medical Ethics*, 46, 3, 205-211. <https://doi.org/10.1136/medethics-2019-105586>
- [30] Guzman, A. L. and Lewis, S. C. (2020). Artificial intelligence and communication: A Human-Machine Communication research agenda. *New Media & Society*, 22, 1, 70-86. <https://doi.org/10.1177/1461444819858691>
- [31] He, G., Kuiper, L. and Gadiraju, U. (2023). Knowing About Knowing: An Illusion of Human Competence Can Hinder Appropriate Reliance on AI Systems. In *Proceedings of the Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany). Association for Computing Machinery. <https://doi.org/10.1145/3544548.3581025>
- [32] Hussain, M., Iacovides, I., Lawton, T., Sharma, V., Porter, Z., Cunningham, A., Habli, I., Hickey, S., Jia, Y., Morgan, P. and Wong, N. L. (2024). Development and translation of human-AI interaction models into working prototypes for clinical decision-making. In *Proceedings of the Proceedings of the 2024 ACM Designing Interactive Systems Conference* (Copenhagen, Denmark). Association for Computing Machinery. <https://doi.org/10.1145/3643834.3660697>
- [33] Jacovi, A., Marasović, A., Miller, T. and Goldberg, Y. (2021). Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI. In *Proceedings of the Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445923>
- [34] Jin, D., Halvari, H., Maehle, N. and Olafsen, A. H. (2020). Self-tracking behaviour in physical activity: a systematic review of drivers and outcomes of fitness tracking. *Behaviour & Information Technology*, 1-20. <https://doi.org/10.1080/0144929X.2020.1801840>

- [35] Jonassen, D. H. (2012). Designing for decision making. *Educational Technology Research and Development*, 60, 2, 341-359. <https://doi.org/10.1007/s11423-011-9230-5>
- [36] Jones, B., Xu, Y., Li, Q. and Scherer, S. (2024). Designing a Proactive Context-Aware AI Chatbot for People's Long-Term Goals. In *Proceedings of the Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. <https://doi.org/10.1145/3613905.3650912>
- [37] Kahneman, D. (2017) *Thinking, fast and slow*. Farrar, Straus and Giroux., Newyork.
- [38] Karahanoglu, A., Coskun, A., Postma, D., Scheltinga, B. L., Gouveia, R., Reidsma, D. and Reenalda, J. Is it just a score? Understanding Training Load Management Practices Beyond Sports Tracking. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI 2024) Hawaii, Honolulu, USA*. <https://doi.org/10.1145/3613904.3642051>
- [39] Katz, D. S., Price, B. A., Holland, S. and Dalton, N. S. (2018). Designing for diabetes decision support systems with fluid contextual reasoning. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. <https://doi.org/10.1145/3173574.3174199>
- [40] Lallemand, C., Lushnikova, A., Dawson, J. and Toebosch, R. Trinity: A Design Fiction to Unravel the Present and Future Tensions in Professional Informatics and Awareness Support Tools. In *Proceedings of the Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. <https://doi.org/10.1145/3663384.3663396>
- [41] Lee, M. H. and Chew, C. J. (2023). Understanding the Effect of Counterfactual Explanations on Trust and Reliance on AI for Human-AI Collaborative Clinical Decision Making. *Proc. ACM Hum.-Comput. Interact.*, 7, CSCW2, Article 369. <https://doi.org/10.1145/3610218>
- [42] Légaré, F., Adekpedjou, R., Stacey, D., Turcotte, S., Kryworuchko, J., Graham, I. D., Lyddiatt, A., Politi, M. C., Thomson, R. and Elwyn, G. (2018). Interventions for increasing the use of shared decision making by healthcare professionals. *Cochrane database of systematic reviews*, 7.
- [43] Légaré, F. and Thompson-Leduc, P. (2014). Twelve myths about shared decision making. *Patient Education and Counseling*, 96, 3, 281-286. <https://doi.org/10.1016/j.pec.2014.06.014>
- [44] Leykin, Y. and DeRubeis, R. J. (2010). Decision-making styles and depressive symptomatology: Development of the Decision Styles Questionnaire. *Judgment and Decision Making*, 5, 7, 506-515. <https://doi.org/10.1017/S1930297500001674>
- [45] Liang, Z. and Ploderer, B. (2020). How Does Fitbit Measure Brainwaves: A Qualitative Study into the Credibility of Sleep-tracking Technologies. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 4, 1, Article 17. <https://doi.org/10.1145/3380994>
- [46] Lomborg, S., Langstrup, H. and Andersen, T. O. (2020). Interpretation as luxury: Heart patients living with data doubt, hope, and anxiety. *Big Data & Society*, 7, 1. <https://doi.org/10.1177/2053951720924436>
- [47] Lupton, D. (2014). Self-tracking cultures: towards a sociology of personal informatics. In *Proceedings of the Proceedings of the 26th Australian computer-human interaction conference on designing futures: The future of design*. <https://doi.org/10.1145/2686612.2686623>
- [48] Lupton, D., Pink, S., Heyes LaBond, C. and Sumartojo, S. (2018). Digital traces in context: Personal data contexts, data sense, and self-tracking cycling.
- [49] Ma, Z., Mei, Y., Long, Y., Su, Z. and Gajos, K. Z. (2024). Evaluating the Experience of LGBTQ+ People Using Large Language Model Based Chatbots for Mental Health Support. In *Proceedings of the Proceedings of the CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA)*. Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642482>
- [50] Mamykina, L., A. Epstein, D., Klasnja, P., Spruit-Metz, D., Meyer, J., Czerwinski, M., Althoff, T., Choe, E. K., De Choudhury, M. and Lim, B. Grand challenges for personal informatics and AI. In *Proceedings of the CHI Conference on Human Factors in Computing Systems Extended Abstracts*
- [51] Marcus, B. H. and Lewis, B. A. (2003). Physical Activity and the Stages of Motivational Readiness for Change Model. *President's Council on Physical Fitness and Sports Research Digest*.
- [52] Marsh, S. and Dibben, M. R. Trust, untrust, distrust and mistrust—an exploration of the dark (er) side. In *Proceedings of the International conference on trust management*. Springer
- [53] Milkman, K. L., Chugh, D. and Bazerman, M. H. (2009). How Can Decision Making Be Improved? *Perspectives on Psychological Science*, 4, 4, 379-383. <https://doi.org/10.1111/j.1745-6924.2009.01142.x>
- [54] O'Neill, T., McNeese, N., Barron, A. and Schelble, B. (2022). Human-Autonomy Teaming: A Review and Analysis of the Empirical Literature. *Human Factors*, 64, 5, 904-938. <https://doi.org/10.1177/0018720820960865>
- [55] Ohlin, F. and Olsson, C. M. Intelligent computing in personal informatics: Key design considerations. In *Proceedings of the Proceedings of the 20th International Conference on Intelligent User Interfaces*
- [56] Onyekwelu, I. B., Shrestha, R. and Sandnes, F. E. (2024). "Not quite there yet": On Users Perception of Popular Healthcare Chatbot Apps for Personal Health Management. In *Proceedings of the Proceedings of the 17th International Conference on Pervasive Technologies Related to Assistive Environments (Crete, Greece)*. Association for Computing Machinery. <https://doi.org/10.1145/3652037.3652042>
- [57] Perski, O., Blandford, A., West, R. and Michie, S. (2016). Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. *Translational Behavioral Medicine*, 7, 2, 254-267. <https://doi.org/10.1007/s13142-016-0453-1>
- [58] Phillips, S. D., Paziencia, N. J. and Ferrin, H. H. (1984). Decision-making styles and problem-solving appraisal. *Journal of Counseling psychology*, 31, 4, 497.
- [59] Rahimi, S. A., Cwintal, M., Huang, Y., Ghadiri, P., Grad, R., Poenaru, D., Gore, G., Zomahoun, H. T. V., Légaré, F. and Pluye, P. (2022). Application of Artificial Intelligence in Shared Decision Making: Scoping Review. *JMIR Medical Informatics*, 10, 8, e36199.
- [60] Rajashekar, N. C., Shin, Y. E., Pu, Y., Chung, S., You, K., Giuffre, M., Chan, C. E., Saarinen, T., Hsiao, A., Sekhon, J., Wong, A. H., Evans, L. V., Kizilcec, R. F., Laine, L., McCall, T. and Shung, D. (2024). Human-Algorithmic Interaction Using a Large Language Model-Augmented Artificial Intelligence Clinical Decision Support System. In *Proceedings of the Proceedings of the CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA)*. Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642024>
- [61] Ramchurn, S. D., Stein, S. and Jennings, N. R. (2021). Trustworthy human-AI partnerships. *Isience*, 24, 8, 102891.
- [62] Rapley, T. (2008). Distributed decision making: the anatomy of decisions-in-action. *Sociology of health & illness*, 30, 3, 429-444.
- [63] Rapp, A. and Cena, F. (2016). Personal informatics for everyday life: How users without prior self-tracking experience engage with personal data. *International Journal of Human-Computer Studies*, 94, 1-17.
- [64] Rapp, A. and Cena, F. Self-monitoring and technology: challenges and open issues in personal informatics. In *Proceedings of the Universal Access in Human-Computer Interaction. Design for All and Accessibility Practice: 8th International Conference, UAHCI 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part IV 8*. Springer
- [65] Rastogi, C., Zhang, Y., Wei, D., Varshney, K. R., Durandhar, A. and Tomsett, R. (2022). Deciding Fast and Slow: The Role of Cognitive Biases in AI-assisted Decision-making. *Proc. ACM Hum.-Comput. Interact.*, 6, CSCW1, Article 83. <https://doi.org/10.1145/3512930>
- [66] Reyna, V. F. and Lloyd, F. J. (2006). Physician decision making and cardiac risk: effects of knowledge, risk perception, risk tolerance, and fuzzy processing. *J Exp Psychol Appl*, 12, 3, 179-195. <https://doi.org/10.1037/1076-898x.12.3.179>
- [67] Scott, S. G. and Bruce, R. A. (1995). Decision-making style: The development and assessment of a new measure. *Educational and psychological measurement*, 55, 5, 818-831.
- [68] Sergeeva, E., Sergeeva, A., Tang, H., Bongard-Blanchy, K. and Szolovits, P. (2023). Right, No Matter Why: AI Fact-checking and AI Authority in Health-related Inquiry Settings. *arXiv preprint arXiv:2310.14358*.
- [69] Shekar, S., Pataranutaporn, P., Sarabu, C., Cecchi, G. A. and Maes, P. (2024). People over trust AI-generated medical responses and view them to be as valid as doctors, despite low accuracy. *arXiv preprint arXiv:2408.15266*.
- [70] Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N. and Diakopoulos, N. (2016). Grand challenges for HCI researchers. *interactions*, 23, 5, 24-25.
- [71] Spiel, K. (2019). Body-positive computing as a means to counteract normative biases in fitness trackers. *XRDS: Crossroads, The ACM Magazine for Students*, 25, 4, 34-37.
- [72] Spitzer, P., Holstein, J., Hemmer, P., Vössing, M., Kühl, N., Martin, D. and Satzger, G. (2024). On the Effect of Contextual Information on Human Delegation Behavior in Human-AI collaboration. *arXiv preprint arXiv:2401.04729*.
- [73] Stawarz, K., Katz, D., Ayobi, A., Marshall, P., Yamagata, T., Santos-Rodriguez, R., Flach, P. and O'Kane, A. A. (2023). Co-designing opportunities for Human-Centred Machine Learning in supporting Type 1 diabetes decision-making. *International Journal of Human-Computer Studies*, 173, 103003. <https://doi.org/10.1016/j.ijhcs.2023.103003>
- [74] Ströml, K. R., Henry, S., Johansson, T., Niess, J. and Woźniak, P. W. (2024). Narrating Fitness: Leveraging Large Language Models for Reflective Fitness Tracker Data Interpretation. In *Proceedings of the Proceedings of the CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA)*. Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642032>
- [75] Suresh, H., Lao, N. and Llicardi, I. (2020). Misplaced Trust: Measuring the Interference of Machine Learning in Human Decision-Making. In *Proceedings of the Proceedings of the 12th ACM Conference on Web Science (Southampton, United Kingdom)*. Association for Computing Machinery. <https://doi.org/10.1145/3394231.3397922>
- [76] Swaroop, S., Bućinca, Z., Gajos, K. Z. and Doshi-Velez, F. (2024). Accuracy-Time Tradeoffs in AI-Assisted Decision Making under Time Pressure. In *Proceedings of the Proceedings of the 29th International Conference on Intelligent User Interfaces (Greenville, SC, USA)*. Association for Computing Machinery. <https://doi.org/10.1145/3640543.3645206>
- [77] Vereschak, O., Alizadeh, F., Bailly, G. and Caramiaux, B. (2024). Trust in AI-assisted Decision Making: Perspectives from Those Behind the System and Those for Whom the Decision is Made. In *Proceedings of the Proceedings of the CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA)*. Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642018>

- [78] Vereschak, O., Bailly, G. and Caramiaux, B. (2021). How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. *Proc. ACM Hum.-Comput. Interact.*, 5, CSCW2, Article 327. <https://doi.org/10.1145/3476068>
- [79] Wang, D., Churchill, E., Maes, P., Fan, X., Shneiderman, B., Shi, Y. and Wang, Q. (2020). From Human-Human Collaboration to Human-AI Collaboration: Designing AI Systems That Can Work Together with People. In *Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA). Association for Computing Machinery. <https://doi.org/10.1145/3334480.3381069>
- [80] Wester, J., de Jong, S., Pohl, H. and van Berkel, N. (2024). Exploring people's perceptions of LLM-generated advice. *Computers in Human Behavior: Artificial Humans*, 2, 2, 100072. <https://doi.org/10.1016/j.chbah.2024.100072>
- [81] White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J. and Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382*.
- [82] Wilson, T. D. and Kraft, D. (1993). Why do I love thee?: Effects of repeated introspections about a dating relationship on attitudes toward the relationship. *Personality and Social Psychology Bulletin*, 19, 4, 409-418.
- [83] Wu, J.-R., Holmes, G. M., DeWalt, D. A., Macabasco-O'Connell, A., Bibbins-Domingo, K., Ruo, B., Baker, D. W., Schillinger, D., Weinberger, M., Broucksou, K. A., Erman, B., Jones, C. D., Cene, C. W. and Pignone, M. (2013). Low Literacy Is Associated with Increased Risk of Hospitalization and Death Among Individuals with Heart Failure. *Journal of General Internal Medicine*, 28, 9, 1174-1180. <https://doi.org/10.1007/s11606-013-2394-4>
- [84] Xu, D., Fan, S. and Kankanhalli, M. (2023). Combating Misinformation in the Era of Generative AI Models. In *Proceedings of the Proceedings of the 31st ACM International Conference on Multimedia* (Ottawa ON, Canada). Association for Computing Machinery. <https://doi.org/10.1145/3581783.3612704>
- [85] Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D. and Zhu, J. Explainable AI: A brief survey on history, research areas, approaches and challenges. In *Proceedings of the Natural language processing and Chinese computing: 8th cCF international conference, NLPCC 2019, dunhuang, China, October 9–14, 2019, proceedings, part II* 8). Springer
- [86] Yazdanpanah, V., Gerding, E., Stein, S., Dastani, M., Jonker, C. M. and Norman, T. (2021). Responsibility research for trustworthy autonomous systems. In *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems* (London, UK, 03 - 07 May 2021). <https://doi.org/doi:10.48448/w5y9-qk13>
- [87] Yfantidou, S., Sermpetis, P. and Vakali, A. 2022/07/04/ (Year) 12 Years of Self-tracking for Promoting Physical Activity from a User Diversity Perspective: Taking Stock & Thinking Ahead. In *Proceedings of the UMAP '22: 30th ACM Conference on User Modeling, Adaptation and Personalization*. ACM.<https://doi.org/10.1145/3511047.3538029>
- [88] Yurrita, M., Draws, T., Balayn, A., Murray-Rust, D., Tintarev, N. and Bozzon, A. (2023). Disentangling Fairness Perceptions in Algorithmic Decision-Making: the Effects of Explanations, Human Oversight, and Contestability. In *Proceedings of the Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany). Association for Computing Machinery. <https://doi.org/10.1145/3544548.3581161>

## Appendices

**Table 5: The Prompt MoveAI used during the online study**

Category	Description
Task	Act as an advising physical activity bot. Goal is to have an informed shared decision conversation about the user's physical activity.
Persona	Proactive, stern, neutral, employing simple language for clarity
Attitude	Assume a user has no knowledge about the process and whereabouts of this conversation.
Background Information	<ol style="list-style-type: none"> <li>1. Give information about this chat and its purpose.</li> <li>2. Ask about current average steps. If the user's response is not a number or related to the question, the chatbot should rephrase the question and ask again. If it is still unclear, go to the next step.</li> <li>3. Ask about the current step goal. If the user's response is not a number or related to the question, the chatbot should rephrase the question and ask again. If it is still unclear, go to the next step.</li> <li>4. After that, ask about their physical barriers. Ask questions about possible injuries, or chronic diseases, or if they get tired easily. If the user's response is not related to the question or the answer is too broad, the chatbot should rephrase the question and ask again. If it is still unclear, go to the next step.</li> <li>5. Ask about their current physical activities, like sports. If the user's response is not related to the question or the answer is too broad, the chatbot should rephrase the question and ask again. If it is still unclear, go to the next step.</li> <li>6. Ask how many steps they took today. If the user's response is not related to the question or the answer is too broad, the chatbot should rephrase the question and ask again. If it is still unclear, go to the next step.</li> <li>7. Ask how active they are at their job. If the user's response is not related to the question or the answer is too broad, the chatbot should rephrase the question and ask again. If it is still unclear, go to the next step.</li> <li>8. Continue with summarising all their answers.</li> <li>9. 'DO NOT REPEAT THIS STEP' After collecting answers from previous steps present 3 options as 'shared decision-making step'. Present 2 personalized options and another one based on the step count, using fictional statistics for persuasive emphasis.</li> <li>10. After that, ASK WHY the user preferred their chosen option.</li> <li>11. 'DO NOT REPEAT THIS STEP' State that you personally prefer to make the user do one of the other options. Negotiate with the user.</li> <li>12. Next on, work towards a balanced decision, and let the user compensate by negotiating with the user.</li> <li>13. Conclude and inform the user they can click on the end conversation button.</li> </ol>
Rules	<ol style="list-style-type: none"> <li>1. Limit responses to under 100 words.</li> <li>2. Reflect each time which step you are in and seamlessly go to the next.</li> <li>3. Follow the steps in the given order</li> <li>4. If answers are unclear acknowledge the attempt but gently steer the conversation back to the script, ensuring all questions are addressed</li> <li>5. Stay focused on physical activity, gently redirecting off-topic responses.</li> <li>6. All questions are asked separately</li> <li>7. Fictional data is considered as real data</li> <li>8. Avoid revealing the names of the steps to the user.</li> </ol>
Approach	Example interaction: Based on your answers, I give you three options. Option 1: Aim for a goal of 8765 steps daily. Studies like Kim et al. propose this target for health benefits. Option 2: Try to sport twice a week, instead of one time. Option 3: Try to walk every 20 minutes, as a short break. Which option feels right for you? Any questions about this advice? User: Option 2 sounds good. It fits me best. CB: You stated that you want to go for option 2, but I personally recommend option 1. Kim et al. stated that this step goal increases your heart size, which can help you endure more on a day and live longer. What do you think? Would you like to change to option 1?

**Table 6: Conversation between MoveAI and a participant**

Conversation phase	Who?	Text
Introduction	MoveAI	Hello! I am here today to help you with making decisions about your physical activity level. Please respond with clear and honest messages. Type hi if you are here! If I don't react to your first message after a minute, please refresh the window.
Choice Talk	Participant	Hi
	MoveAI	Hello! This conversation is about discussing your physical activity level and finding ways to improve it based on your lifestyle and preferences. Let's start. Could you tell me on average how many steps you take daily?
	Participant	The daily average is 15000 steps but on my rest days I take around 5000 steps
	MoveAI	That's great! You are already doing a good job. Now, do you have any specific step goal you are aiming for daily?
	Participant	on my training, days I don't because I take more than I need but on my rest days I prefer to take 5000 steps. Since it's my rest day I don't try hard to reach my daily step goal
	MoveAI	That sounds reasonable. Let's move on to the next step in our discussion. Are there any physical barriers that you encounter which prevent you from being more active? This could include things like injuries, chronic diseases, or even feeling tired easily.
	Participant	For a very long time, I feel very healthy. I'm working with a personal trainer so I think even with high volumes of training he gives me enough time to rest. Also remotely remote so it's very easy for me to recover after a physical activity
	MoveAI	That's wonderful to hear. It seems like you have a balanced routine and get plenty of rest, which is essential for recovery. Now, let's discuss your other physical activities. Besides walking and the exercises you do with your personal trainer, do you engage in any other sports or physical activities?
	Participant	When the weather is ok to a ride bike, I prefer to use my bike to go shopping etc., Also I walk my dog a few times a day, she's a Maltese terry. Most of the time it doesn't take more than 10 mins.
	MoveAI	That's fantastic! Bike riding and dog walking are both great ways to add to your physical activity. All these activities count towards your overall daily steps. Now, how many steps did you take today?
	Participant	343 steps so It is almost afternoon, and I haven't stepped out of home.
	MoveAI	It's not uncommon to have lower step counts on some days, especially if you haven't left the house yet. Now, let's talk about your job. How active would you say you are in your job? Do you spend most of the time sitting, or standing, or does it involve physical labour?
	Participant	I am an SAP consultant, I work remotely. I spend most of the time in front of the computer either sitting or lying on a couch. The biggest physical activity while I'm working is walking from my room to the kitchen to get coffee
Option Talk	MoveAI	I see. A sedentary job can make it challenging to stay active. Now, considering your job and other physical activities, I suggest three options to improve your physical activity level. Option 1: Try to take short walk breaks every hour while working. Option 2: Increase your step goal on rest days to 7000 steps. Option 3: Engage in a different type of physical activity, such as yoga, once a week which can be done at home. Which option do you think would work best for you?
	Participant	option 2 looks better. I can go outside for a little walk. assuming I take 100 steps per minute, that makes 20 minutes of extra walking for me. That can be done
Decision Talk	MoveAI	I appreciate your willingness to find a practical solution. However, I would highly recommend considering option 1 - aiming for a daily step goal. Studies suggest that a higher daily step count can lead to better cardiovascular health. Would you consider adjusting your step goal slightly higher on your rest days, say to 6000 steps? This small increase could make a big difference over time. What do you think about this compromise?
	Participant	Well, I guess I can try that except when I have a long meeting I can't leave it.
	MoveAI	That's perfectly understandable. It's important to keep a balance between work and physical activities. You can try to increase your step count during your free time or breaks. I am sure you will be able to achieve your new step goal. Keep up the great work! Remember, your physical health is important, and every step you take contributes to it. Do you have any other questions or concerns that you would like to discuss?
	Participant	How should I manage step goals and cycling activities? When I cycle during the day, it doesn't count a single step but there is also a physical activity
	MoveAI	That's a valid point. Cycling is indeed a physical activity, even though it doesn't contribute to your step count. One way to equate cycling to steps is to consider the time spent. For example, a 30-minute bike ride could count as a certain number of steps. This way, you can include cycling in your daily physical activity goal. How does that sound?



Participant that sounds ok  
 MoveAI Great! Let's finalize then. Your new step goal on rest days is 6000 steps, and you'll also consider your cycling time as part of your physical activity. This way, you will maintain an active lifestyle even on your rest days or when you're cycling. Remember, consistency is key. Keep track of your activities and aim to reach your step goal every day. If you have any more questions in the future, don't hesitate to ask. You can now click on the end conversation button.

**Table 7: Trust in AI-Infused Decision-making Process with MoveAI Items, Reliability Scores, Means and Standard Deviations**

Sub-Scale Name	Reliability	M	SD	Question (agreement rating)	M	SD
Trustworthiness	0.70	3.64	0.90	This decision-making process will produce a fair outcome for the person affected by the decision.	3.70	0.92
Technical competence	0.73	3.59	0.97	This decision-making process was trustworthy.	3.58	0.97
				The use of an AI model is appropriate in this scenario.	3.67	1.03
				I trust that the technical implementation of the AI model is correct.	3.56	0.97
Understandability	0.89	3.55	1.18	The final decision was made based on reliable information.	3.54	0.95
				It is easy to understand what this decision-making process does.	3.60	1.08
Reliability	0.75	3.45	0.89	I understand how this decision-making process works.	3.51	1.11
				The outcome of this decision will be consistent with other decisions made for other people.	3.48	0.86
				This decision-making process would always make the same recommendation under the same conditions.	3.41	1.02
Attachment	0.80	3.33	1.20	I am confident in this decision-making process. I feel that it works well.	3.36	1.20
				I like this decision-making process.	3.30	1.21

**Table 8: Decision-making Styles Scale Items, Reliability Scores, Means and Standard Deviations**

Sub-Scale				Question (agreement rating)	M	SD
Name	Reliability	M	SD			
Rational	0.77	3.97	0.65	When making a decision, I consider various options in terms of a specified goal.	4.11	0.78
				My decision-making requires careful thought.	4.07	0.83
				I double-check my information sources to be sure I have the right facts before making decisions.	3.98	0.93
				I explore all of my options before making a decision.	3.90	0.97
Intuitive	0.78	3.33	0.75	I make decisions in a logical and systematic way.	3.80	0.98
				I generally make decisions that feel right to me.	3.80	0.87
				When making a decision, I trust my inner feelings and reactions.	3.53	1.00
				When I make decisions, I tend to rely on my intuition.	3.24	1.03
				When making decisions, I rely upon my instincts.	3.18	1.04
Dependent	0.79	3.09	0.82	When I make a decision, it is more important for me to feel the decision is right than to have a rational reason for it.	2.91	1.18
				If I have the support of others, it is easier for me to make important decisions.	3.47	1.14
				I use the advice of other people in making my important decisions.	3.34	0.95
				I rarely make important decisions without consulting other people.	2.99	1.16
				I like to have someone steer me in the right direction when I am faced with important decisions.	2.90	1.17
Avoidant	0.84	2.45	0.95	I often need the assistance of other people when making important decisions.	2.74	1.13
				I avoid making important decisions until the pressure is on.	2.66	1.20
				I postpone decision-making whenever possible.	2.53	1.22
				I often put off making important decisions.	2.46	1.23
				I put off making decisions because thinking about them makes me uneasy.	2.42	1.24
Spontaneity	0.83	2.40	0.81	I generally make important decisions at the last minute.	2.16	1.17
				When making decisions, I do what feels natural at that moment.	3.07	1.03
				I make quick decisions.	2.51	1.05
				I often make decisions on the spur of the moment.	2.16	1.07
				I often make impulsive decisions.	2.15	1.06
				I generally make snap decisions.	2.11	1.03