MUNI FSS

JOURNAL OF PSYCHOSOCIAL RESEARCH ON CYBERSPACE
CYBERPSYCHOLOGY

Martens, M., De Wolf, R., & De Marez, L. (2024). Trust in algorithmic decision-making systems in health: A comparison between ADA health and IBM Watson. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 18*(1), Article 5. https://doi.org/10.5817/CP2024-1-5

Trust in Algorithmic Decision-Making Systems in Health: A Comparison Between ADA Health and IBM Watson Oncology

Marijn Martens[1,2], *Ralf De Wolf*[1,2], *& Lieven De Marez*[1,2]

[1] Department of Communication Sciences, Ghent University, Ghent, Belgium
[2] imec-mict-UGent, Ghent University, Ghent, Belgium

Abstract

Algorithmic decision-making systems (ADMs) support an ever-growing number of decision-making processes. We conducted an online survey study in Flanders (n = 1,082) to understand how laypeople perceive and trust health ADMs. Inspired by the ability, benevolence, and integrity trustworthiness model (Mayer et al., 1995), this study investigated how trust is constructed in health ADMs. In addition, we investigated how trust construction differs between ADA Health (a self-diagnosis medical chatbot) and IBM Watson Oncology (a system that suggests treatments for cancer in hospitals). Our results show that accuracy and fairness are the biggest predictors of trust in both ADMs, whereas control plays a smaller yet significant role. Interestingly, control plays a bigger role in explaining trust in ADA Health than IBM Watson Oncology. Moreover, how appropriate people evaluate data-driven healthcare and how concerned they are with algorithmic systems prove to be good predictors for accuracy, fairness, and control in these specific health ADMs. The appropriateness of data-driven healthcare had a bigger effect with IBM Watson Oncology than with ADA Health. Overall, our results show the importance of considering the broader contextual, algorithmic, and case-specific characteristics when investigating trust construction in ADMs.

Keywords: health ADMs; algorithms; trust; ABI-model; survey; SEM

Editorial Record

First submission received:
October 4, 2022

Revisions received:
June 16, 2023
November 6, 2023
November 30, 2023

Accepted for publication:
December 20, 2023

Editor in charge:
David Smahel

Introduction

We live in an era in which our everyday decision making, such as choosing what to read, whom to befriend, or what to watch, is being substituted or augmented by algorithmic decision-making systems (ADMs; Araujo et al., 2018). ADMs are becoming ubiquitous (Shrestha & Yang, 2019). The quantity of data and the variety of contexts in which personal data is used, continue to increase (Jackson, 2018; Livingstone et al., 2020).

In both academia and the industry, we are witnessing a shift from designing algorithmic systems that are optimal, efficient, and effective toward an increased focus on fairness, accountability, and transparency (FAT), so to minimize the risks originating from the usage of ADMs (Barocas & Selbst, 2016; Jackson, 2018; Shrestha & Yang, 2019). These risks include, but are not limited to, using biased data, making unfair predictions, or bolstering unfavorable power concentrations. Regulatory agencies, such as the European Parliament, therefore, aim to

empower end-users by implementing regulations (i.e., art. 22 GPDR) that, among others, provide the right to opt out of ADMs and the right to be informed about their inner workings (Goodman & Flaxman, 2017).

With a renewed focus on the choice of the end user to adopt or reject ADMs, several studies have found that trust in ADMs functions as a strong predictor for adoption (Araujo et al., 2020; Ghazizadeh et al., 2012; Kim, 2018; Taddeo, 2010). Most of these studies focused on people's trust in specific ADMs, such as algorithmic management (Kim, 2018) or news selection (Logg et al., 2019). Other studies have compared people's trust in multiple ADMs, for example, Araujo et al. (2020) investigated and compared how people trust ADMs in justice, health, and media contexts. However, little effort is devoted to making comparisons between different ADMs in a given context or field. Moreover, most of these studies focused on comparing levels of trust, rather than comparing the predictors of people's trust in specific ADMs.

In contemporary healthcare several ADMs are used (Hass, 2019), including health monitoring systems (Rejab et al., 2014), treatment decision-making systems (Agarwal et al., 2010; Behera et al., 2019; Morley et al., 2019) and contact tracing applications (Martens et al., 2021). Specifically in this context, a differentiation can be made between ADMs tailored to medical staff (Aljaaf et al., 2015; Garg et al., 2005; H. Lee, 2014) and those tailored to laypeople (K. Lee et al., 2017; Lupton & Jutel, 2015). In both categories, patients are affected by the system, while only in the latter situation they have the agency to interact and intervene with the system (e.g., reject the system). Although both types of health ADMs have similarities, they can also differ on various levels, such as the level of involvement of companies and people, the complexity of the system, the associated risks, the data that is used, and the goals that are pursued.

In this study, we want to contribute to the growing area of trust research in ADMs by exploring laypeople's trust and the construction of their trust in two specific health ADMs, ADA Health and IBM Watson Oncology, using a survey study (n = 1,082). IBM Watson Oncology is a specialist-oriented system that offers suggestions to specialists with cancer treatments. ADA Health is a self-diagnosis medical chatbot. Our approach is inspired by the ability, benevolence, and integrity framework for interpersonal trustworthiness proposed by Mayer and colleagues (1995). We put forward the following research questions: How and to what extent do laypeople trust health ADMs? (RQ1) How does laypeople's trust construction differ between ADA health and IBM Watson Oncology? (RQ2)

In the remainder of the paper, we first give an overview of earlier research on attitudes and trust in ADMs, and how the framework of Mayer et al. (1995) served as an inspiration. We then substantiate and operationalize our approach before examining our empirical work.

Literature Review and Theoretical Considerations

Evaluating and Trusting Algorithmic Systems

In 1995, Mayer et al. argued that trust is an essential component in formalizing human relationships. They defined interpersonal trust as the "willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (p. 712). Almost a decade later, J. D. Lee and See (2004) used similar terms to describe trust in automation. Trust has played, and continues to play, a central role in the adoption of technology and more specific ADMs (Shin, 2021).

Multiple scholars have used trust and trustworthiness interchangeably (Greenwood & Van Buren, 2010; Kiyonari et al., 2006). Nonetheless, in line with Sekhon and colleagues (2014), we argue that trustworthiness distinctly differs from trust as it refers to the characteristics upon which someone forms a judgment of trust. Similarly, Colquitt and Rodell (2011) stressed the distinction between trustworthiness and trust, as this distinction can increase the understanding of how trust relates to other concepts.

Mayer et al. (1995) focused on perceived ability, benevolence, and integrity (ABI) as the three variables a trustee considers when deciding whether a person is trustworthy. "[These constructs] are not trust per se, [but] these variables help build the foundation for the development of trust," they claimed (p. 717).

These dimensions of trustworthiness have already been used to better understand how people trust automation (Cho et al., 2015; Svare et al., 2020), even though Mayer et al. (1995) originally developed this framework to study interpersonal trust. De Visser and colleagues (2018) and Glikson and Wooley (2020), for example, found parallels between interpersonal trust and automation trust. Langer and colleagues (2023) already used the ABI model to

measure trustworthiness as a predictor of trust in the context of algorithmic personnel selection. In what follows, we take a further look at how the traditional interpersonal ability, benevolence, and integrity framework by Mayer et al. (1995) could inspire perceived accuracy and fairness to be elements of the trustworthiness of an algorithmic system, in our case ADMs.

From Ability to Perceived Accuracy

Ability was originally defined as the skills and competence to perform a task effectively (Mayer et al., 1995). Competence or ability is often equated with accuracy and has been widely investigated in the context of algorithmic systems as an element of trustworthiness (Araujo et al., 2018; Hancock et al., 2011; K. Yu et al., 2016). Similarly, traditional algorithmic trust research has focused mainly on performance measures, such as accuracy (Al-Emran et al., 2018; Chatterjee & Bhattacharjee, 2020; Kennedy et al., 2022; Langer et al., 2023; Yin et al., 2019; K. Yu et al., 2016). In most of these studies, perceived accuracy is expected to positively relate to one's attitude or trust. It should be noted, however, that this relationship has not been investigated in many contexts in which algorithmic systems are used, let alone considering the variety of these systems. For example, in the context of education, Chatterjee and Bhattacharjee (2020) found that performance expectancy did not correlate with an individual's attitude toward artificial intelligence. Similarly, Kennedy and colleagues (2022) found that having information on the accuracy of an algorithmic system did not influence the decision to use a system that predicts recidivism. In the context of M-payments (i.e., payments carried out via a mobile device), however, performance expectancy was the biggest predictor for the intention to use M-payments (Al-Emran et al., 2018). Equally, in the context of interpretable machine learning models, Yin and colleagues (2019) found that people's trust was affected by both the stated and experienced accuracy of a system. Another approach was proposed by K. Yu and colleagues (2016), where the accuracy of an algorithmic system was seen as a threshold variable for reliance. In their experimental design, they found that respondents with an accuracy expectancy below 80% had a lower reliance score.

In a health context, Aljarboa and Miah (2020) found a positive relationship between a general practitioners' acceptance of a clinical decision support system and their performance expectancy. Moreover, Rahi and colleagues (2021) found a positive relation between the acceptance and performance expectancy of patients of telemedicine health services during COVID-19. Based on this literature, we hypothesize that perceived accuracy positively predicts trust in health ADMs.

H1: Perceived accuracy positively predicts trust in health ADMs.

From Benevolence and Integrity to Perceived Fairness

In the ABI-model, integrity is defined as the belief that "the trustee adheres to a set of principles the trustor finds acceptable" (p. 719), that they act unbiased (Mayer et al., 1995). Benevolence is described as "the extent to which the trustee is believed to want to do good to the trustor" (Mayer et al., 1995, p. 718). Traditionally, integrity and benevolence have been found to be important elements of trustworthiness in the context of algorithmic decision making (Höddinghaus et al., 2021), especially when people are impacted by the decision (Langer et al., 2023). Both dimensions of trustworthiness could be considered elements that determine how fair a system is. In the conceptualization of Mayer and colleagues (1995), fairness is seen as an element of integrity (Colquitt & Rodell, 2011).

In the context of algorithmic trustworthiness, however, fairness is often seen as a standalone concept in the FAT principles, encompassing both benevolence and integrity (Shin, 2020; Shin & Park, 2019; Shin et al., 2020). Specifically, fairness in an algorithmic context means that algorithmic decisions should not create discriminatory or unjust consequences (Shin, 2020; Shin & Park, 2019; Yang & Stoyanovich, 2017). The link between fairness and trust has been empirically proven by multiple scholars in the context of news algorithms (Shin, Zaid, et al., 2022) and ADMs (Shin, 2021).

Specifically in the context of ADMs, Zarsky (2016) summarized different issues of fairness that could impact trustworthiness. Similarly, Araujo and colleagues (2018) found that perceived fairness was an important construct when evaluating ADMs. Therefore, we believe that the perceived fairness of an ADM system is an essential part in the evaluation of its trustworthiness. Woodruff and colleagues (2018) argued that an individual's concern toward algorithmic fairness could substantially affect their trust, not only in the system, but also in the company involved.

Moreover, other scholars have found that the perception of fairness positively affects the evaluation of the algorithmic system (Shin, Lim, et al., 2022).

In the context of health ADMs, Ozawa and Sripad (2013) found in their systematic review that fairness was least investigated in the domain of health system trust. K.-H. Yu and Kohane (2019), however, point to bias presented in historical data as one of the prominent issues when developing medical algorithmic systems. Fairness is thus underexposed while possible issues remain that would cause unfair situations in health ADMs. Therefore, we hypothesize that perceived fairness will positively predict trust in health ADMs.

H2: Perceived fairness positively predicts trust in health ADM.

Perceived Control as Supplement to Algorithmic Accuracy and Fairness

Apart from accuracy and fairness, the willingness to be vulnerable to a system is indissolubly connected to the amount of control people perceive in using a system, as being vulnerable implies a "willingness to relinquish control" (Chiou & Lee, 2023, p. 11). In the context of interpersonal trust in healthcare, perceived control was found to impact patient-physician trust (Gabay, 2015). In automation, there is also a clear relationship between the amount of control someone has and the perceived trustworthiness of the system (Schaefer et al., 2016). Araujo and his colleagues (2020) found that people who feel more in control of ADMs are more likely to consider these as fair and useful. Moreover, human control is considered to be one of the main elements that potentially make the design of algorithmic systems more reliable, safe, and trustworthy (Shneiderman, 2020).

Indeed, retaining some level of control on the goals, the outcome, and how a system uses your data could enable the process of trusting (Amershi et al., 2014; Chiou & Lee, 2023). Therefore, we argue that the perceived sense of control will positively predicts the trust people have in health ADMs.

H3: Perceived control positively predicts trust in health ADMs.

The Appropriateness of Data Driven Healthcare

Multiple scholars have already argued to be mindful of the specific context in which algorithmic systems are implemented to understand how much they trust it (Hoffman et al., 2013; Johnson & Bradshaw, 2021; J. D. Lee & See, 2004; Schaefer et al., 2016). Bansal and colleagues (2016) found that contextual characteristics, such as the sensitivity of the context influence how trust is constructed.

In 1995, Mayer et al. (1995) argued that the question should not be "do you trust them" but rather "do you trust them to do what?" (p. 729). Indeed, the use of algorithmic systems in a health context is often criticized for not being appropriate or suitable for a specific task or purpose (Scott et al., 2021). Not every decision, so the argument goes, should be augmented with, or replaced by ADMs. How people evaluate a specific algorithmic system is expected to be partly explained by how appropriate they evaluate the context to be augmented with algorithmic systems.

Araujo and colleagues (2020) compared different contexts (i.e., justice, health, and media) with low- and high-impact cases. They found significant differences between scenarios in the different contexts, illustrating the contextual dependence of trust in ADMs. Hence, we argue that how appropriate someone evaluates data-driven healthcare in general will predict the individuals' attitudes toward specific health ADMs, including their perceived accuracy, fairness, and control.

H4: The appropriateness of data-driven healthcare positively predicts the perceived accuracy of a specific health ADMs.

H5: The appropriateness of data-driven healthcare positively predicts the perceived fairness of a specific health ADMs.

H6: The appropriateness of data-driven healthcare positively predicts the perceived control of a specific health ADMs.

The Role of Algorithmic Concern

Recent research has shown that people are often consciously or unconsciously reluctant to use ADMs (Mahmud et al., 2022). Yeomans and colleagues (2019) found that some individuals may have a natural tendency to distrust algorithmic systems, regardless of their actual performance or accuracy. This initial feeling of distrust can be linked to a broader concept known as "initial learned trust", which refers to an individual's general evaluation of technology before any actual interaction (Hoff & Bashir, 2015). Hence, individuals with a general pre-existing negative bias toward algorithmic systems in general may be more likely to distrust any specific algorithmic system.

As algorithmic decision-making systems are often plagued with issues in terms of bias (Köchling & Wehner, 2020), unfairness (Grgic-Hlaca et al., 2018), or unwanted and unintended consequences (Marjanovic et al., 2018), we argue that the extent of these concerns impacts attitudes toward algorithmic systems and how people evaluate ADMs. Therefore, we hypothesize that perceived concern negatively predicts perceived accuracy, fairness, and control in health ADMs.

H7: Perceived concern about algorithms in general negatively predicts the perceived accuracy of a specific health ADM.

H8: Perceived concern about algorithms in general negatively predicts the perceived fairness of a specific health ADM.

H9: Perceived concern about algorithms in general negatively predicts the perceived control of a specific health ADM.

Trust Construction in Different Types of Health ADMs

Health ADMs can be distinguished on the basis of their target audience; health ADMs that enhance professional decision making in healthcare (i.e., clinical decision support systems; Alexander, 2006; H. Lee, 2014), and health ADMs that target patients themselves through chatbots, automated diagnosis apps, or even just Google search (Fink et al., 2018; K. Lee et al., 2017). We find this differentiation especially interesting because in both cases it is the patient that is vulnerable to the outcome. Arguably, how individuals construct their trust varies in both cases.

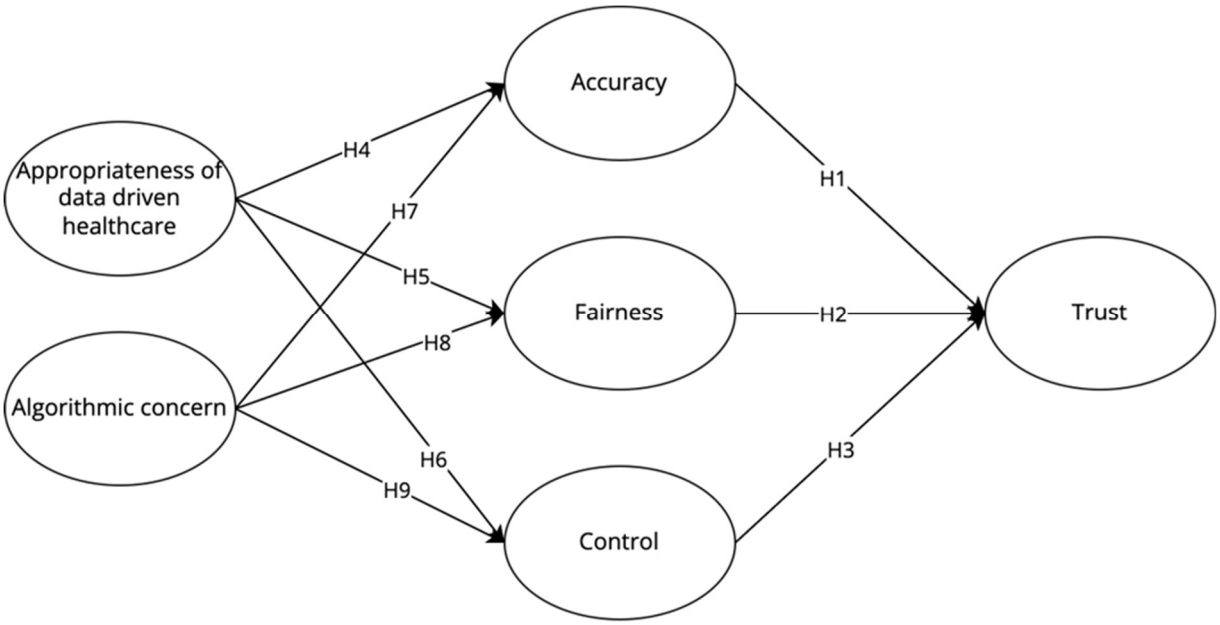
In our study we included one case of each type, IBM Watson Oncology (IBM Watson Health | AI Healthcare Solutions, 2022) and ADA Health (ADA Health, 2022). IBM Watson Oncology is an algorithmic system embedded in hospital software that can be used in hospitals by specialists to suggest treatments for cancer patients (from now on IBM Watson Oncology is sometimes referred to as IBM Watson to enhance readability). ADA Health, on the other hand, is a chatbot that works as an app on a smartphone. It suggests a diagnosis by asking the user questions about symptoms.

We expect differences in the construction of trust between both systems because they differ in terms of the type of system, system complexity, task difficulty, perceived risks, benefits, organizational setting, and framing. In other words, they differ on almost every level of external variability, as proposed by Hoff and Bashir (2015). Therefore, we propose an exploratory second research question that investigates the difference in trust construction in both health ADMs.

RQ2: How does laypeople's trust construction differ between ADA health and IBM Watson?

In summary, all the hypotheses for RQ1 (H1–H9) and our model for trust construction in health ADMs following these hypotheses are visualized in figure 1. We tested each hypothesis of RQ1 for IBM Watson (H_xa) and ADA Health (H_xb) and compared both use cases (RQ2).

Figure 1. *Research Model.*



Methods

Focus of Study: IBM Watson and ADA Health

ADMs can be categorized in multiple ways. For this study, we do not focus on how they technically function. Rather, and in line with Algorithm Watch (2019), we define an ADMs as "a socio-technical framework that encompasses a decision-making model" (p. 9) and are mindful of these socio-technical elements (such as the developers, the data that is used, and the goals that are pursued) when measuring trust in ADMs.

IBM Watson Oncology is a clinical decision support system that uses historic patient data, curated academic literature, and other medical guidelines to analyze patient data and recommend cancer treatment to specialists within hospital software. It supports first-line medical oncology treatment for a variety of cancers (e.g., breast, lung, colon, and rectal cancers; IBM Watson Health | AI Healthcare Solutions, 2022).

ADA Health is a health management smartphone app that takes the form of a chatbot. It asks for your symptoms and suggests possible causes and diagnoses based on that input. The system is based on academic research, WHO guidelines, and the insights of doctors and specialists. ADA Health considers various conditions or diagnoses including common, but also rare, and high-risk conditions (ADA Health, 2022). To start the questionnaire, each participant was given a short explanation of what IBM Watson Oncology and ADA Health are (see Table 1). All respondents had questions concerning both IBM Watson Oncology and ADA Health. The questions for IBM Watson Oncology always preceded the questions for ADA Health.

Data Collection

To answer our research questions, we collected data through an online questionnaire, which was distributed by a professional agency for three weeks in June 2020. The data was collected with a quota sampling method targeting respondents to gain representativeness for the Flemish population in terms of gender, age, and education. Of the 1,552 respondents who started the survey, 1,082 were retained after cleaning. Respondents who did not answer the control variable (*tick the box "disagree*") correctly and people who did not finish the survey were deleted from the sample. There were no significant relationships between sociodemographic variables and dropouts. The sample consisted of 51.1% women and 48.9% men, aged between 18 and 86 years old (M = 49.29, SD = 14.51). Of them, 36.8% had a bachelor's or master's degree, 48% graduated from secondary education, and 15.2% did not complete their secondary education.

Measures

People's perceptions of ADMs could be made up by their perceptions of each of the elements within the socio-technical ADMs. To be mindful of the specific elements of the ADMs, we chose to operationalize trust, accuracy, fairness, and control considering these elements. We, therefore, asked questions on the data being used (e.g., *to what extent do you think the data used is accurate?*), the people involved (e.g., *to what extent do you think the developers can be trusted?*) and the company (e.g., *to what extent do you think this company is fair?*). We modeled covariances between the questions focusing on the same element or stakeholder (e.g., all evaluations on the data used) to model for variances linked to only that element or stakeholder.

For algorithmic concerns, we considered how algorithmic systems could be risky, lead to bad results, be feared, or lead to general concerns. To investigate how people evaluate the appropriateness of data-driven healthcare, we differentiated between the appropriateness of using algorithms for three healthcare related goals, ranging from using algorithms to train doctors, develop medicines or improve healthcare. All questions used a 5-point Likert scale (*completely disagree—completely agree*), except for trust, which was measured using a more fine-grained 100-point scale.

To control for the influence of prior knowledge of the company IBM or ADA Health, two variables were included, asking whether they knew the companies (e.g., *I know the company IBM*). The analysis included people who had prior knowledge of IBM and those who did not. No meaningful differences in the results were found between these groups that would impact the findings of this study. The SEM-models for these two groups separately are provided in the appendix. For ADA, the group knowing ADA was too small (n = 30) to investigate further.

Table 1 contains a detailed description of the different variables included in this study, how they are operationalized, the means, standard deviations, Cronbach's alphas, and factor loadings of the confirmatory factor analysis for each construct. All factor loadings for the model on IBM Watson and ADA Health range between .613 and .920 thus exceeding the .6 threshold for newly developed items (Afthanorhan, 2013), except the item data accuracy for IBM Watson which only had a loading of .317. For ADA Health, data accuracy exceeds the .6 threshold with a factor loading of .603. To ensure that we measure the same constructs across both contexts, we excluded data accuracy from both contexts. Cronbach's alphas for all constructs range between .725 and .922.

Analysis

To answer and test the hypothesis of RQ1, we constructed two structural equation models (SEM) using the Lavaan package (Rosseel, 2012) in R, one for IBM Watson Oncology and one for ADA Health. For each latent construct, a confirmatory factor analysis was performed. We checked the TLI, CFI, and RMSEA indices for the measurement and structural model to ensure a good model fit. For RQ2, we first exploratively compared the measures across models using paired samples t-test. To compare the construction of trust between both contexts, we constructed a third SEM. In this SEM, the corresponding variables from IBM Watson Oncology and ADA Health were combined, and each context was rendered as a group of respondents. This way we could compare each context as if they were groups in our sample. We conducted a measurement invariance test to compare model fit when constraining factor loadings, intercepts, and regression weights across IBM Watson Oncology and ADA Health. Finally, we investigated possible differences by constraining each regression weight individually and comparing the model fit with the unconstrained model.

Table 1. *Overview of the Constructs.*

Construct	Items	*M*	*SD*	α	Factor loadings				
Algorithmic concern	I think that decisions made by AI could be risky.	3.53	0.636	.830	.616				
	I think that decisions made by AI could lead to bad results.				.657				
	I think that decisions made by AI could lead to concerns.				.858				
	I think that decisions made by AI could lead to fear.				.803				
Appropriateness of data- driven healthcare	An algorithm should be allowed to use my personal health data to reach more accurate results in the future.	3.68	0.805	.852	.813				
	An algorithm should be allowed to use my personal health data for research on new medicines.				.818				
	An algorithm should be allowed to use my personal health data to train doctors.				.753				
Explanation given	IBM Watson Oncology is an algorithmic system that could potentially be used in a hospital to treat cancer patients. The system will provide specialists with recommendations on the appropriate treatment for a specific person with cancer, based on data such as historical patient records, academic literature, and other medical guidelines. ADA Health is an algorithmic system in the form of a chatbot. ADA Health functions on a smartphone and provides a diagnosis by asking the user questions about their symptoms. The ADA diagnosis is based on academic research, WHO guidelines, and insights of doctors and specialists.								
					IBM Watson Oncology				ADA Health
Construct	Items	*M*	*SD*	α	Factor loadings	*M*	*SD*	α	Factor loadings
Accuracy	I feel like the medical data that X would use would be accurate. **(excluded)**	3.37	0.602	.725	.311	3.13	0.648	.766	.603
	I feel like the developers of X would develop an accurate system.				.640				.683
	I feel like the company X would accurately convert my data into insights.				.638				.764
Fairness	I feel like X would handle my personal health data fairly.	3.42	0.604	.789	.687	3.24	0.605	.811	.753
	I feel like the developers of X would handle fairly while developing the system.				.642				.677
	I feel like company X is a fair company.				.679				.739
Control	I feel like I would have enough control over my personal health data used in X.	2.76	0.791	.846	.692	2.56	0.832	.891	.777
	I feel like I would have enough control over how the developers of X would develop their system.				.809				.829
	I feel like I would have enough control over how company X handles my personal data.				.849				.917
Trust (on 100)	To what extent do you trust the data used in X?	62.1	16.2	.878	.778	54.7	17.3	.922	.862
	To what extent do you trust the developers who developed X?				.829				.897
	To what extent do you trust the company X?				.739				.869

Results

Model

Looking at the model fit indices for the measurement model presented in figure 1, the model has a reasonable fit for the context of ADA Health (CFI = .955, TLI = .943, RMSEA = .064) and IBM Watson Oncology (CFI = .933, TLI = .916, RMSEA = .070). Equally, for the general model the model fit is adequate for both ADA Health (CFI = .925, TLI = .896, RMSEA = .086) and IBM Watson Oncology (CFI = .918, TLI = .887, RMSEA = .086), approaching the cut-off points (see Table 2).

Table 2. *Summary of Goodness-of-Fit Indexes for Both IBM Watson Oncology and ADA Health.*

Fit indices (cut-off point)	IBM Watson Oncology		ADA Health	
	Measurements model	General model	Measurement model	General model
CFI (> .90)	.933	.918	.995	.925
TLI (> .90)	.916	.887	.943	.896
RMSEA (< .08)	.070	.086	.064	.086

The variables included in this study all have significant correlations with one another. The strongest correlation for both IBM Watson Oncology (r = .772, p < .001) and ADA Health (r = .689, p < .001) was found between fairness and accuracy. All correlations in the context of IBM Watson are included in Table 3, and for ADA Health they are included in Table 4.

Table 3. *Correlations Between Study Variables—IBM Watson Oncology.*

Variable	(1)	(2)	(3)	(4)	(5)
(1) Appropriateness of data-driven healthcare					
(2) Algorithmic concern	−.124**				
(3) Accuracy—IBM Watson	.427**	−.204**			
(4) Fairness—IBM Watson	.424**	−.208**	.772**		
(5) Control—IBM Watson	.230**	−.334**	.480**	.509**	
(6) Trust—IBM Watson	.472**	−.245**	.699**	.727**	.460**

Note. **p < .001.

Table 4. *Correlations Between Study Variables—ADA Health.*

Variable	(1)	(2)	(3)	(4)	(5)
(1) Appropriateness of data-driven healthcare					
(2) Algorithmic concern	−.124**				
(3) Accuracy—ADA Health	.277**	−.177**			
(4) Fairness—ADA Health	.386**	−.225**	.689**		
(5) Control—ADA Health	.180**	−.289**	.461**	.526**	
(6) Trust—ADA Health	.380**	−.246**	.652**	.684**	.519**

Note. **p < .001.

RQ1: The Construction of Trust in IBM Watson Oncology and Ada Health

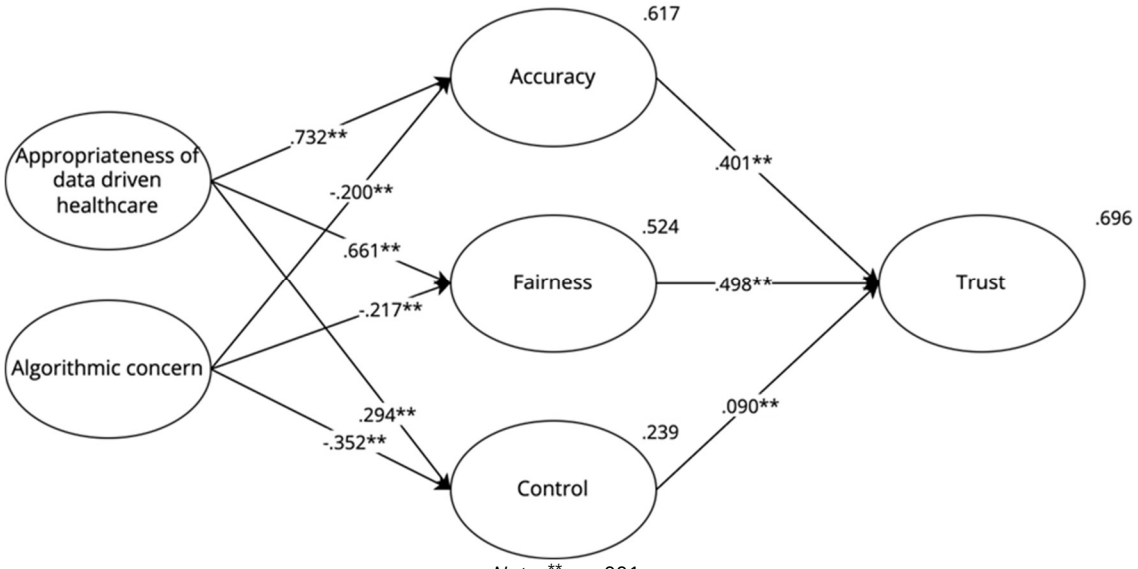
IBM Watson

In the context of IBM Watson Oncology, we see that the appropriateness of data-driven healthcare is a strong positive predictor for perceived accuracy (β = .732, p < .001), perceived fairness (β = .661, p < .001), and to a lesser extent perceived control (β = .294, p < .001) confirming H4a, H5a, and H6a. Algorithmic concern negatively predicts perceived control (β = −.352, p < .001), and to a lesser extent perceived fairness (β = −.217, p < .001) and perceived accuracy (β = −.200, p < .001), confirming H7a, H8a, and H9a. With both the appropriateness of data-driven

healthcare and algorithmic concern, 61.7% of the variance of perceived accuracy, 52.4% of the variance of perceived fairness, and 23.9% of the variance of perceived control can be explained.

69.6% of the variance of perceived trust of the IBM Watson Oncology system is explained by perceived accuracy (β = .401, p < .001), perceived fairness (β = .498, p < .001) and to a lesser extent perceived control (β = .090, p < .001), confirming H1a, H2a and H3a respectively. All these regression weights and the explained variances are included in figure 2. An overview of the hypotheses is given in Table 6.

Figure 2. *IBM Watson Oncology—Model for Trust.*



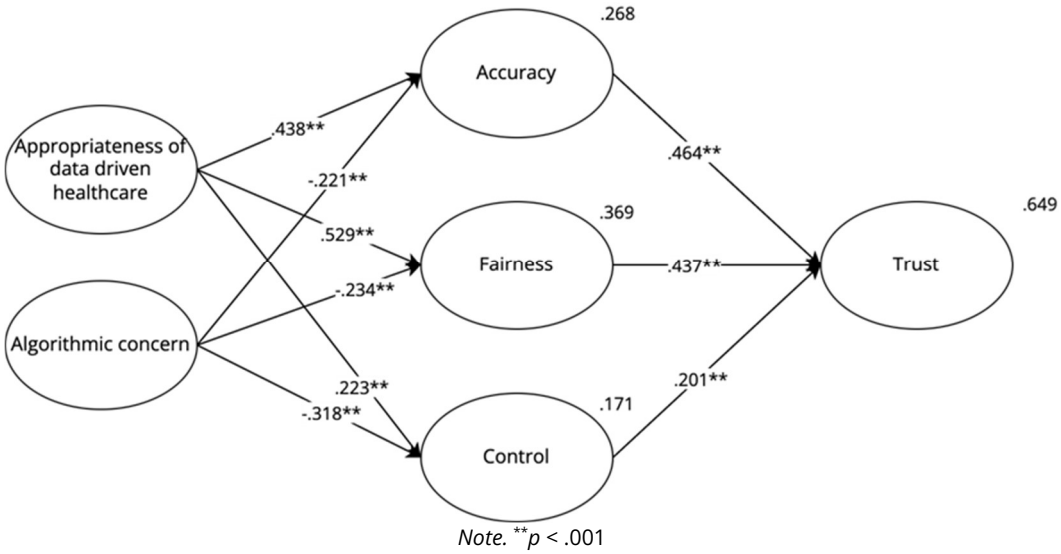
Note. $^{**}p$ < .001.

ADA Health

In the context of ADA Health, the variance in accuracy is explained for 26.8%. This is mostly explained by the appropriateness of data-driven healthcare (β = .438, p < .001), but also by algorithmic concern (β = –.221, p < .001) confirming H4b and H7b. 36.9% of the variance of perceived fairness is explained by the appropriateness of data-driven healthcare (β = .529, p < .001) and algorithmic concern (β = –.234, p < .001) confirming H5b and H8b. 17,1% of the variance of perceived control is explained by algorithmic concern (β = –.318, p < .001) and the appropriateness of data-driven healthcare (β = .223, p < .001) confirming H9b and H6b.

Perceived accuracy (β = .464, p < .001; H1b) and fairness (β = .437, p < .001; H2b) are also the strongest predictors for perceived trust in the context of ADA Health and explain together with perceived control (β = .201, p < .001) (H3b) 64,9% of the variance of trust in ADA Health. All the explained variances and regression weights are shown in figure 3. An overview of all hypotheses is given in Table 6.

Figure 3. *ADA Health—Model for Trust.*



Note. $^{**}p$ < .001

RQ2: Comparing the Construction of Trust Between IBM Watson Oncology and ADA Health

On average, the perceived trust in ADA Health is significantly lower than the perceived trust in IBM Watson Oncology, $M_{\text{IBM Watson}} = 62.08$, $SD_{\text{IBM Watson}} = 16.15$; $M_{\text{ADA Health}} = 54.70$, $SD_{\text{ADA Health}} = 17.30$; $t(1,081) = 17.969$, $p < .001$. The same is true for control, $M_{\text{IBM Watson}} = 2.76$, $SD_{\text{IBM Watson}} = 0.791$; $M_{\text{ADA Health}} = 2.56$, $SD_{\text{ADA Health}} = 0.832$; $t(1,081) = 10.496$, $p < .001$, fairness, $M_{\text{IBM Watson}} = 3.43$, $SD_{\text{IBM Watson}} = 0.604$; $M_{\text{ADA Health}} = 3.24$, $SD_{\text{ADA Health}} = 0.605$; $t(1,081) = 12.675$, $p < .001$, and accuracy, $M_{\text{IBM Watson}} = 3.37$, $SD_{\text{IBM Watson}} = 0.602$; $M_{\text{ADA Health}} = 3.13$, $SD_{\text{ADA Health}} = .648$; $t(1,081) = 12.544$, $p < .001$. Overall, people thus have a lower attitude toward ADA Health than toward IBM Watson Oncology on every construct we measured.

Following our multigroup measurement of invariance analysis (see Table 5) and a chi-square difference test, we found that the trust model for IBM Watson Oncology is metric invariant from the model for ADA Health. Thus, there is no significant difference in terms of their factor loadings. People interpreted the constructs we measured similarly for IBM Watson Oncology and for ADA Health, $\Delta\chi^2(12, n = 1,082) = 19.808$, $p = .071$. This test allows us to make a reliable comparison between ADA Health and IBM Watson Oncology in terms of the relationships in our model.

When testing for scalar invariances (see Table 5), we found a significant difference in model fit, $\Delta\chi^2(12, n = 1,082) = 95.258$, $p < .001$, between the model constraining loadings and intercepts and the model only constraining loadings. Indeed, some mean differences between IBM Watson Oncology and ADA Health in our constructs are not captured by the shared variance of the items and are caused by other elements defining the context.

When constraining the regression weights on top of the factor loadings and intercepts to be equal across use cases (see Table 5), there is a significant difference compared to the model constraining only the loadings and intercepts, $\Delta\chi^2(9, n = 1,082) = 22.867$, $p = .006$. The regression weights in the model to explain trust in IBM Watson Oncology thus significantly differ from ADA Health. To locate these differences, we constrained all factor loadings and intercepts, and then consecutively constrained each regression weight individually to determine if the model fit would be significantly different from the model that only constrains their factor loadings and intercepts.

Table 5. *Goodness-of-Fit Statistics for the SEM Models: IBM Watson Oncology and ADA Health.*

				Baseline test		Difference test		
	AIC	BIC	*n*	χ^2	*df*	$\Delta\chi^2$	Δdf	*p*
Configural invariance[1]	113492	114389	1,082	1990.2	220			
Metric invariance[2]	113487	114317	1,082	2010.0	232	19.808	12	.071
Scalar Invariance[3]	113559	114320	1,082	2105.2	244	95.258	12	< .001
Regression invariance[4]	113564	114274	1,082	2128.1	253	22.867	9	.006

Note. [1]No constrains, [2]constraining loadings, [3]constraining loadings and intercepts, [4]cosntraining loadings, intercepts, and regressions.

For the predictors of trust, constraining accuracy or fairness as a predictor did not impact the model fit. Constraining control as a predictor, however, did impact model fit, $\Delta\chi^2(1, n = 1,082) = 7.4376$, $p = .006$. Control thus plays a different role in both contexts. The explanatory power of perceived control is higher in the context of ADA Health than with IBM Watson Oncology ($\beta_{\text{ADA Health}_{\text{control}\rightarrow\text{trust}}} = .201$; $\beta_{\text{IBM Watson}_{\text{control}\rightarrow\text{trust}}} = .090$).

When investigating the appropriateness of data-driven healthcare as a predictor for accuracy, fairness, and control, our test showed that constraining the appropriateness of data-driven healthcare predicting accuracy impacted model fit, $\Delta\chi^2(1, n = 1,082) = 12.25$, $p < .001$, as well as the appropriateness of data-driven healthcare predicting fairness, $\Delta\chi^2(1, n = 1,082) = 3.8838$, $p < .05$. The appropriateness of data-driven healthcare is a stronger predictor of fairness and accuracy in the context of IBM Watson Oncology than in the context of ADA Health, ($\beta_{\text{ADA Health}_{\text{appropriateness}\rightarrow\text{fairness}}} = .529$; $\beta_{\text{IBM Watson}_{\text{appropriateness}\rightarrow\text{fairness}}} = .661$),

($\beta_{\text{ADA Health}_{\text{appropriateness}\rightarrow\text{accuracy}}} = .438$; $\beta_{\text{IBM Watson}_{\text{appropriateness}\rightarrow\text{accuracy}}} = .732$).

When investigating algorithmic concern as a predictor for accuracy, fairness, and control, our test showed that constraining algorithmic concern predicting accuracy, fairness or control did not significantly impact model fit.

These relationships should thus be considered equal between both use-cases. An overview of the hypotheses is given in Table 6.

***Table 6**. Overview of Hypotheses.*

Hypothesis	RQ1		RQ2
	IBM Watson Oncology	ADA Health	Comparing IBM Watson Oncology and ADA Health
H1: Accuracy -> Trust	✓	✓	ns
H2: Fairness -> Trust	✓	✓	ns
H3: Control -> Trust	✓	✓	ADA Health > IBM Watson
H4: Appropriateness of data-driven healthcare -> Accuracy	✓	✓	ADA Health < IBM Watson
H5: Appropriateness of data-driven healthcare -> Fairness	✓	✓	ADA Health < IBM Watson
H6: Appropriateness of data-driven healthcare -> Control	✓	✓	ns
H7: Algorithmic concern -> Accuracy	✓	✓	ns
H8: Algorithmic concern -> Fairness	✓	✓	ns
H9: Algorithmic concern -> Control	✓	✓	ns

Discussion

Automated decision-making systems are omnipresent and are employed in numerous contexts. A growing body of research has been conducted on how users adopt and accept such algorithmic systems (Araujo et al., 2020; Ghazizadeh et al., 2012; Kim, 2018; Martens et al., 2021; Taddeo, 2010). In this study, we focused on how and to what extent laypeople trust health ADMs, including ADMs in which laypeople are not the end-user. Our approach was inspired by the ABI framework of interpersonal trustworthiness put forward by Mayer and colleagues (1995). Moreover, we argue that trust in ADMs is highly context dependent. In our study we, therefore, included contextualizing elements from the broader health context (the appropriateness of data-driven healthcare), the algorithmic context (algorithmic concern), and compared the construction of trust between two specific health ADMs: ADA Health, and IBM Watson Oncology.

In line with the traditional ABI model (Mayer et al., 1995), our results indicated that accuracy, fairness, and control played a significant role in the trust evaluation of both health ADMs. Indeed, there could be a strong parallel between interpersonal trust, as modeled by Mayer and colleagues (1995), and ADMs' trust. This is consistent with the findings of multiple scholars in the field (Araujo et al., 2018; Cho et al., 2015; Hancock et al., 2011; Svare et al., 2020; K. Yu et al., 2016).

When looking at the broader contextual influence on accuracy, fairness, and control, we observed that the appropriateness of data-driven healthcare positively predicted accuracy, fairness, and control for both IBM Watson Oncology and ADA health. In line with Hoffman et al. (2013), Johnson & Bradshaw (2021), J. D. Lee & See (2004), and Schaefer and colleagues (2016), we thus found that the evaluation of a datafied context, positively predicted the evaluation of specific trustworthiness elements within a use-case of that context. A comprehensive evaluation of the broader context and how appropriate they evaluate the datafication within that context could help to identify other, not case-specific, factors that may impact the trustworthiness of a specific ADMs.

Furthermore, algorithmic concern negatively predicted accuracy, fairness, and control in each health ADMs. Indeed, levels of general concern toward technology influenced how specific applications of such technology were evaluated. This holds true even when the specific use-case has no direct relation to the concerns in algorithmic systems. Yeomans and colleagues (2019) equally found how people can distrust ADMs while not basing this sentiment on the system being evaluated. Indeed, general concerns toward algorithmic systems in general have consequences in trust construction in specific ADMs. Negative experiences with algorithmic systems, even in a non-related context, could thus negatively influence how people evaluate specific ADMs.

Zooming in on the different predictors of trust, perceived accuracy positively predicts people's trust in health ADMs. This is true for both ADA Health and IBM Watson Oncology. These findings confirm the findings of Aljarboa and Miah (2020) and Rahi and his colleagues (2021). Our results contrast with the findings of Chatterjee and Bhattacharjee (2020) in the educational context, where the expected accuracy or performance did not influence

the acceptability of the algorithmic system. Fairness equally proved to be an important predictor for trust in health ADMs, as was suggested by scholars implementing the FAT framework (Shin, 2020; Shin & Park, 2019; Shin, Lim, et al., 2022). This was also the case for both IBM Watson Oncology and ADA Health. Control was the weakest predictor of trust. However, it still played a significant role in both included health ADMs. This finding reinforces the often-validated relationship between control and trust. Indeed, just as in multiple other contexts (Araujo et al., 2020), and just as between patient-physician trust (Gabay, 2015), control influences the trust people have in health ADMs.

Taking these three relationships into account, we argue that when aiming for a trustworthy health ADMs, both expert-oriented (IBM Watson Oncology) and patient-oriented (ADA Health), we should inform the affected patients on the fairness of the system, the performance of the system, and give them some level of control. Moreover, by improving the performance, safety measures taken to ensure fairness, or control measures, we could improve how trusting people are toward health ADMs. The fact that these relationships were also found in the context of ADMs they were not the user of strengthens the idea that whenever laypeople are affected, they should also be considered and/or consulted.

In both contexts, the constructs are measured considering the algorithmic system and its actors. However, IBM Watson Oncology and ADA health significantly differ in their context of use. While IBM Watson Oncology is used by experts in a hospital setting, ADA health can be used by anyone with a smartphone anywhere. Consequently, we found differences between both contexts in how the constructs loaded and how trust was constructed.

While data accuracy did load sufficiently on accuracy in the context of ADA Health, it did not in the context of IBM Watson Oncology. This could indicate that data accuracy plays another role for IBM Watson Oncology than for ADA Health. Nonetheless, explicitly considering the data, developers, and the company involved when measuring accuracy, fairness, control, and trust envelops more elements of the algorithmic system, demystifying the opaqueness of the algorithm system.

Comparing both contexts, we found multiple significant differences between the trust construction in both use-cases. Overall, people evaluated the different factors for trustworthiness (= accuracy, fairness, and control) lower in the context of a patient-oriented system focused on automating symptom diagnosis (= ADA Health) than in the context of a specialist-oriented system that suggest cancer treatments (= IBM Watson Oncology). When investigating the difference between both contexts in their trust construction model, we found that perceived control played a stronger role in the construction of trust in the context of ADA Health. Moreover, the appropriateness of data-driven healthcare is a weaker predictor of accuracy and fairness in the context of ADA Health than IBM Watson Oncology. These findings show that while the constructs are equally constructed in both contexts, they play a different role in the construction of trust. Hence, a specific contextual approach is of paramount importance. We should be vigilant not to generalize within a context as there can be many differences between use-cases.

Limitations and Future Research

Our approach focused on the impacted of health ADMs, rather than only the users of these systems. Our findings show that even if an individual is not in a position to adopt a specific algorithmic system, they form a rationale for trusting this system. Their perceived accuracy, fairness, and control explained almost 70% of their trust in IBM Watson Oncology. We argue that in a society where evermore such algorithmic systems are used, and consequently evermore individuals are (unconsciously) impacted, research should also focus on the evaluation of the impacted, next to the users' evaluation. We want to stress, however, that our results show how the variation in algorithmic trust and trustworthiness can be explained, and it is not intended to investigate a causal relationship.

This research focused on the construction of trust in two different health ADMs. It should be noted, however, that while we gave the respondents an explanation of each health ADMs, we did not control for how thorough the respondents read these explanations. Moreover, the ADMs included in this study differ on multiple levels, including the complexity of the technology, target population, data used, and companies involved. Hence, it could be that other variables (not included in this study) caused differences in the trust construction in both systems. Moreover, we first asked all the questions on IBM Watson and then on ADA, and potential sequence effects were not accounted for. In future research, we recommend a more fine-grained comparison (e.g., a vignette study).

Equally, a qualitative research approach would allow us to discover why and how people perceive differences between both ADMs.

In our study, we provided a thorough explanation per ADMs on what data is used, who is involved, how it functions, and what its potential uses are. Future research could additionally make ADMs more tangible by using visualizations or giving the respondents a hands-on experience to increase the ecological validity.

Finally, our sample was representative for the population of Flanders in terms of gender, age, and education level. While this enables generalization, we did not focus on interpersonal differences. It could be interesting to investigate how socioeconomic status or digital literacy influences trust construction in health ADMs or ADMs in general.

Conclusion

Perceived accuracy, fairness, and control proof to be good predictors of algorithmic trust in health ADMs. Perceived accuracy and perceived fairness are the strongest predictors of trust in both IBM Watson Oncology and ADA Health. Perceived control plays a bigger role in explaining trust in ADA Health, exemplifying the difference in trust construction between both ADMs. Using the socio-technical elements (e.g., data, developers, company) to operationalize perceived accuracy, fairness, control, and trust helped to stress the complexity of the ADMs. Moreover, the appropriateness of data driven healthcare and how concerned people are with algorithms in general are good predictors for accuracy, fairness, and control. Overall, our results show the importance of considering broader contextual, algorithmic, and case-specific characteristics when investigating trust construction in ADMs.

Conflict of Interest

The authors have no conflicts of interest to declare.

Authors' Contribution

Marijn Martens: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, writing—original draft, writing—review & editing. **Ralf De Wolf:** conceptualization, funding acquisition, writing—review & editing, project administration, resources, supervision. **Lieven De Marez:** funding acquisition, project administration, resources, supervision.

Acknowledgement

This work was supported by the FWO (Grant No. 11F6819N). The funding body was not involved in the study design, the collection, analysis and interpretation of data, the writing of the report or the decision to submit the article for publication.

References

ADA Health. (2022). Health. *Powered by Ada.* https://ada.com/

Afthanorhan, W. (2013). A comparison of partial least square structural equation modeling (PLS-SEM) and covariance based structural equation modeling (CB-SEM) for confirmatory factor analysis. *International Journal of Engineering Science and Innovative Technology*, *2*(5), 198–205. https://www.ijesit.com/Volume%202/Issue%205/IJESIT201305_27.pdf

Agarwal, R., Gao, G., DesRoches, C., & Jha, A. K. (2010). Research commentary—The digital transformation of healthcare: Current status and the road ahead. *Information Systems Research*, *21*(4), 796–809. https://doi.org/10.1287/isre.1100.0327

Al-Emran, M., Mezhuyev, V., & Kamaludin, A. (2018). Technology acceptance model in M-learning context: A systematic review. *Computers & Education*, *125*, 389–412. https://doi.org/10.1016/j.compedu.2018.06.008

Alexander, G. L. (2006). Issues of trust and ethics in computerized clinical decision support systems. *Nursing Administration Quarterly*, *30*(1), 21–29. https://doi.org/10.1097/00006216-200601000-00005

Algorithm Watch. (2019). *Taking stock of automated decision-making in the EU*. Algorithm Watch and Bertelsmann Stiftung. https://algorithmwatch.org/wp-content/uploads/2019/02/Automating_Society_Report_2019.pdf

Aljaaf, A. J., Al-Jumeily, D., Hussain, A. J., Fergus, P., Al-Jumaily, M., & Abdel-Aziz, K. (2015). Toward an optimal use of artificial intelligence techniques within a clinical decision support system. In *Proceedings of the 2015 Science and Information Conference* (pp. 548–554). IEEE Xplore. https://doi.org/10.1109/SAI.2015.7237196

Aljarboa, S., & Miah, S. J. (2020). *Assessing the acceptance of clinical decision support tools using an integrated technology acceptance model.* ArXiv. http://arxiv.org/abs/2011.14315

Amershi, S., Cakmak, M., Knox, W. B., & Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. *AI Magazine*, *35*(4), 105–120. https://doi.org/10.1609/aimag.v35i4.2513

Araujo, T., de Vreese, C., Helberger, N., Kruikemeier, S., van Weert, J., Bol, N., Oberski, D., Pechenizkiy, M., Schaap, G., & Taylor, L. (2018). *Automated decision-making fairness in an ai-driven world: Public perceptions, hopes and concerns*. Digital Communication Methods Lab. http://www.digicomlab.eu/reports/2018_adm_by_ai/

Araujo, T., Helberger, N., Kruikemeier, S., & de Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY*, *35*(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

Bansal, G., Zahedi, F. M., & Gefen, D. (2016). Do context and personality matter? Trust and privacy concerns in disclosing private information online. *Information & Management*, *53*(1), 1–21. https://doi.org/10.1016/j.im.2015.08.001

Barocas, S., & Selbst, A. (2016). Big data's disparate impact. *California Law Review*, *104*(1), 671–729. http://doi.org/10.15779/Z38BG31

Behera, R. K., Bala, P. K., & Dhir, A. (2019). The emerging role of cognitive computing in healthcare: A systematic literature review. *International Journal of Medical Informatics*, *129*, 154–166. https://doi.org/10.1016/j.ijmedinf.2019.04.024

Chatterjee, S., & Bhattacharjee, K. K. (2020). Adoption of artificial intelligence in higher education: A quantitative analysis using structural equation modelling. *Education and Information Technologies*, *25*(5), 3443–3463. https://doi.org/10.1007/s10639-020-10159-7

Chiou, E. K., & Lee, J. D. (2023). Trusting automation: Designing for responsivity and resilience. *Human Factors*, *65*(1), 137–165. https://doi.org/10.1177/00187208211009995

Cho, J.-H., Chan, K., & Adali, S. (2015). A survey on trust modeling. *ACM Computing Surveys*, *48*(2), 1–40. https://doi.org/10.1145/2815595

Colquitt, J. A., & Rodell, J. B. (2011). Justice, trust, and trustworthiness: A longitudinal analysis integrating three theoretical perspectives. *Academy of Management Journal*, *54*(6), 1183–1206. https://doi.org/10.5465/amj.2007.0572

de Visser, E. J., Pak, R., & Shaw, T. H. (2018). From 'automation' to 'autonomy': The importance of trust repair in human–machine interaction. *Ergonomics*, *61*(10), 1409–1427. https://doi.org/10.1080/00140139.2018.1457725

Fink, C., Uhlmann, L., Hofmann, M., Forschner, A., Eigentler, T., Garbe, C., Enk, A., & Haenssle, H. A. (2018). Patient acceptance and trust in automated computer-assisted diagnosis of melanoma with dermatofluoroscopy. *JDDG: Journal Der Deutschen Dermatologischen Gesellschaft*, *16*(7), 854–859. https://doi.org/10.1111/ddg.13562

Gabay, G. (2015). Perceived control over health, communication and patient–physician trust. *Patient Education and Counseling*, *98*(12), 1550–1557. https://doi.org/10.1016/j.pec.2015.06.019

Garg, A. X., Adhikari, N. K. J., McDonald, H., Rosas-Arellano, M. P., Devereaux, P., Beyene, J., Sam, J., & Haynes, R. B. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes. *JAMA, 293*(10), 1223–1238. https://doi.org/10.1001/jama.293.10.1223

Ghazizadeh, M., Lee, J. D., & Boyle, L. N. (2012). Extending the technology acceptance model to assess automation. *Cognition, Technology & Work*, *14*(1), 39–49. https://doi.org/10.1007/s10111-011-0194-3

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine, 38*(3), 50–57. https://doi.org/10.1609/aimag.v38i3.2741

Greenwood, M., & Van Buren III, H. J. (2010). Trust and stakeholder theory: Trustworthiness in the organisation–stakeholder relationship. *Journal of Business Ethics, 95*(3), 425–438. https://doi.org/10.1007/s10551-010-0414-4

Grgic-Hlaca, N., Redmiles, E. M., Gummadi, K. P., & Weller, A. (2018). Human perceptions of fairness in algorithmic decision making: A case study of criminal risk prediction. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web* (pp. 903–912). ACM. https://doi.org/10.1145/3178876.3186138

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., Visser, E. J. D., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, *53*(5), 517–527. https://doi.org/10.1177/0018720811417254

Hass, N. C. (2019). *"Can I get a second opinion?" How user characteristics impact trust in automation in a medical screening task.* [Doctoral dissertation, University of Missouri]. https://mospace.umsystem.edu/xmlui/handle/10355/69666

Höddinghaus, M., Sondern, D., & Hertel, G. (2021). The automation of leadership functions: Would people trust decision algorithms? *Computers in Human Behavior*, *116*, Article 106635. https://doi.org/10.1016/j.chb.2020.106635

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434. https://doi.com/10.1177/0018720814547570

Hoffman, R. R., Johnson, M., Bradshaw, J. M., & Underbrink, A. (2013). Trust in automation. *IEEE Intelligent Systems*, *28*(1), 84–88. https://doi.org/10.1109/MIS.2013.24

IBM Watson Health | AI healthcare solutions. (2022). *IBM Watson Health.* https://www.ibm.com/watson-health

Jackson, J. R. (2018). Algorithmic bias. *Journal of Leadership, Accountability and Ethics*, *15*(4), 55–65. https://doi.org/10.33423/jlae.v15i4.170

Johnson, M., & Bradshaw, J. M. (2021). The role of interdependence in trust. In C. S. Nam & J. B. Lyons (Eds.), *Trust in human-robot interaction* (pp. 379–403). Elsevier. https://doi.org/10.1016/B978-0-12-819472-0.00016-2

Kennedy, R. P., Waggoner, P. D., & Ward, M. M. (2022). Trust in public policy algorithms. *The Journal of Politics*, *84*(2), 1132–1148. https://doi.org/10.1086/716283

Kim, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, *5*(1), 1–16. https://doi.org/10.1177/2053951718756684

Kiyonari, T., Yamagishi, T., Cook, K. S., & Cheshire, C. (2006). Does trust beget trustworthiness? Trust and trustworthiness in two games and two cultures: A research note. *Social Psychology Quarterly, 69*(3), 270–283. https://doi.org/10.1177/019027250606900304

Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, *13*(3), 795–848. https://doi.org/10.1007/s40685-020-00134-w

Langer, M., König, C. J., Back, C., & Hemsing, V. (2023). Trust in artificial intelligence: Comparing trust processes between human and automated trustees in light of unfair bias. *Journal of Business and Psychology, 38*(3), 493–508. https://doi.org/10.1007/s10869-022-09829-9

Lee, H. (2014). Paging Dr. Watson: IBM's Watson supercomputer now being used in healthcare. *Journal of AHIMA*, *85*(5), 44–47.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lee, K., Hoti, K., Hughes, J. D., & Emmerton, L. (2017). Dr Google is here to stay but health care professionals are still valued: An analysis of health care consumers' Internet navigation support preferences. *Journal of Medical Internet Research*, *19*(6), Article e210. https://doi.org/10.2196/jmir.7489

Livingstone, S., Stoilova, M., & Nandagiri, R. (2020). Data and privacy literacy: The role of the school in educating children in a datafied society. In D. Frau-Meigs, S. Kotilainen, M. Pathak-Shelat, M. Hoechsmann, & S. R. Poyntz (Eds.), *The handbook of media education research* (pp. 413–425). Wiley. https://doi.org/10.1002/9781119166900.ch38

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Lupton, D., & Jutel, A. (2015). 'It's like having a physician in your pocket!' A critical analysis of self-diagnosis smartphone apps. *Social Science & Medicine*, *133*, 128–135. https://doi.org/10.1016/j.socscimed.2015.04.004

Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, *175*, Article 121390. https://doi.org/10.1016/j.techfore.2021.121390

Marjanovic, O., Cecez-Kecmanovic, D., & Vidgen, R. (2018). Algorithmic pollution: Understanding and responding to negative consequences of algorithmic decision-making. In U. Schultze, M. Aanestad, M. Mähring, C. Østerlund, & K. Riemer (Eds.), *Living with monsters? Social implications of algorithmic phenomena, hybrid agency, and the performativity of technology* (pp. 31–47). Springer International Publishing. https://doi.org/10.1007/978-3-030-04091-8_4

Martens, M., De Wolf, R., Vadendriessche, K., Evens, T., & De Marez, L. (2021). Applying contextual integrity to digital contact tracing and automated triage for hospitals during COVID-19. *Technology in Society*, *67*, Article 101748. https://doi.org/10.1016/j.techsoc.2021.101748

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709–734. https://doi.org/10.5465/amr.1995.9508080335

Morley, J., Machado, C., Burr, C., Cowls, J., Taddeo, M., & Floridi, L. (2019). The debate on the ethics of AI in health care: A reconstruction and critical review. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3486518

Ozawa, S., & Sripad, P. (2013). How do you measure trust in the health system? A systematic review of the literature. *Social Science & Medicine*, *91*, 10–14. https://doi.org/10.1016/j.socscimed.2013.05.005

Rahi, S., Khan, M. M., & Alghizzawi, M. (2021). Factors influencing the adoption of telemedicine health services during COVID-19 pandemic crisis: An integrative research model. *Enterprise Information Systems*, *15*(6), 769–793. https://doi.org/10.1080/17517575.2020.1850872

Rejab, F. B., Nouira, K., & Trabelsi, A. (2014). Health monitoring systems using machine learning techniques. In *Intelligent systems for science and information* (pp. 423–440). Springer. https://doi.org/10.1007/978-3-319-04702-7_24

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Sekhon, H., Ennew, C., Kharouf, H., & Devlin, J. (2014). Trustworthiness and trust: Influences and implications. *Journal of Marketing Management, 30*(3–4), 409–430. https://doi.org/10.1080/0267257X.2013.842609

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors, 58*(3), 377–400. https://doi.org/10.1177/0018720816634228

Scott, I., Carter, S., & Coiera, E. (2021). Clinician checklist for assessing suitability of machine learning applications in healthcare. *BMJ Health & Care Informatics*, *28*(1), Article e100251. https://doi.org/10.1136/bmjhci-2020-100251

Shin, D. (2020). User perceptions of algorithmic decisions in the personalized AI system: Perceptual evaluation of fairness, accountability, transparency, and explainability. *Journal of Broadcasting & Electronic Media*, *64*(4), 541–565. https://doi.org/10.1080/08838151.2020.1843357

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, *146*, Article 102551. https://doi.org/10.1016/j.ijhcs.2020.102551

Shin, D., Lim, J. S., Ahmad, N., & Ibahrine, M. (2022). Understanding user sensemaking in fairness and transparency in algorithms: Algorithmic sensemaking in over-the-top platform. *AI & SOCIETY*. https://doi.org/10.1007/s00146-022-01525-9

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, *98*, 277–284. https://doi.org/10.1016/j.chb.2019.04.019

Shin, D., Zaid, B., Biocca, F., & Rasul, A. (2022). In platforms we trust? Unlocking the black-box of news algorithms through interpretable AI. *Journal of Broadcasting & Electronic Media*, *66*(2), 235–256. https://doi.org/10.1080/08838151.2022.2057984

Shin, D., Zhong, B., & Biocca, F. A. (2020). Beyond user experience: What constitutes algorithmic experiences? *International Journal of Information Management*, *52*, Article 102061. https://doi.org/10.1016/j.ijinfomgt.2019.102061

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Shrestha, Y. R., & Yang, Y. (2019). Fairness in algorithmic decision-making: Applications in multi-winner voting, machine learning, and recommender systems. *Algorithms*, *12*(9), Article 199. https://doi.org/10.3390/a12090199

Svare, H., Gausdal, A. H., & Möllering, G. (2020). The function of ability, benevolence, and integrity-based trust in innovation networks. *Industry and Innovation*, *27*(6), 585–604. https://doi.org/10.1080/13662716.2019.1632695

Taddeo, M. (2010). Modelling trust in artificial agents, a first step toward the analysis of e-trust. *Minds and Machines*, *20*(2), 243–257. https://doi.org/10.1007/s11023-010-9201-3

Woodruff, A., Fox, S. E., Rousso-Schindler, S., & Warshaw, J. (2018). A qualitative exploration of perceptions of algorithmic fairness. In *CHI '18: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). ACM. https://doi.org/10.1145/3173574.3174230

Yang, K., & Stoyanovich, J. (2017). *Measuring Fairness in Ranked Outputs.* Proceedings of the 29th International Conference on Scientific and Statistical Database Management, 1–6. https://doi.org/10.1145/3085504.3085526

Yeomans, M., Shah, A., Mullainathan, S., & Kleinberg, J. (2019). Making sense of recommendations. *Journal of Behavioral Decision Making, 32(*4), 403–414. https://doi.org/10.1002/bdm.2118

Yin, M., Wortman Vaughan, J., & Wallach, H. (2019). Understanding the effect of accuracy on trust in machine learning models. In *CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12). ACM. https://doi.org/10.1145/3290605.3300509

Yu, K., Berkovsky, S., Conway, D., Taib, R., Zhou, J., & Chen, F. (2016). Trust and reliance based on system accuracy. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization* (pp. 223–227). ACM. https://doi.org/10.1145/2930238.2930290

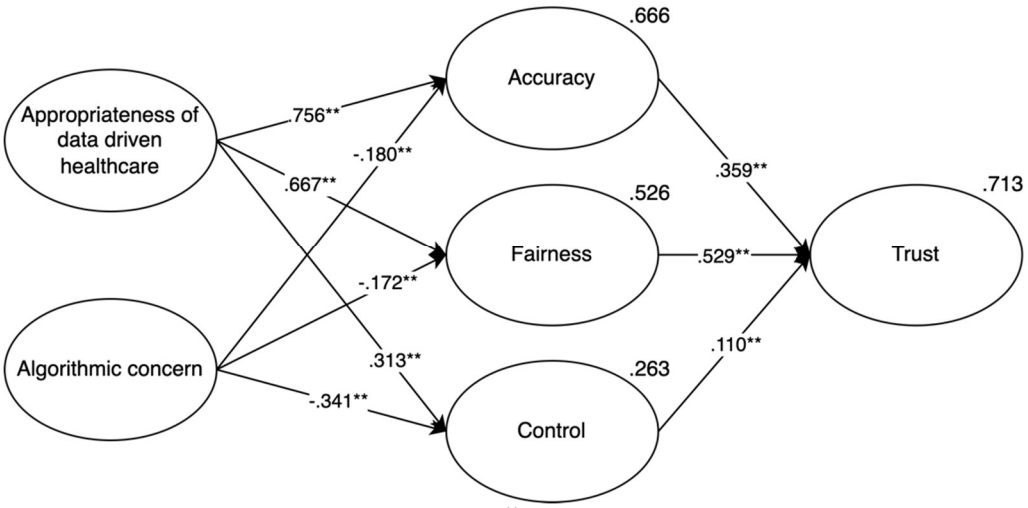
Yu, K.-H., & Kohane, I. S. (2019). Framing the challenges of artificial intelligence in medicine. *BMJ Quality & Safety*, *28*(3), 238–241. https://doi.org/10.1136/bmjqs-2018-008551

Zarsky, T. (2016). The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology, & Human Values*, *41*(1), 118–132. https://doi.org/10.1177/0162243915605575

Appendix

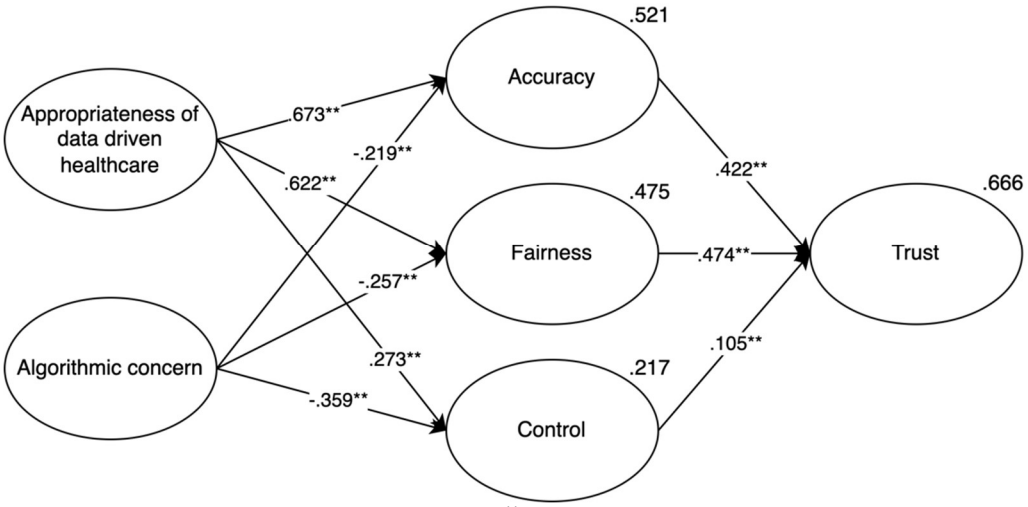
SEM Models Divided Between Respondents Knowing IBM and Not Knowing IBM

Figure A1. *IBM Watson Oncology Model for Trust (Only Respondents Who Know IBM).*



Note. $^{**}p < .001$.

Figure A2. *IBM Watson Oncology Model for Trust (Only Respondents Who Do Not Know IBM).*



Note. $^{**}p < .001$.

Table A1. *Fit Indexes for Measurement Model and General Model in the Context of IBM Watson Oncology, Divided Between Respondents Knowing IBM and Not Knowing IBM.*

Fit indices (cut-off point)	IBM Watson Oncology (only respondents who know IBM)		IBM Watson Oncology (only respondents who do not know IBM)	
	Measurements model	General model	Measurement model	General model
CFI (> .90)	.927	.918	.912	.938
TLI (> .90)	.908	.894	.888	.918
RMSEA (< .08)	.078	.086	.089	.072

About Authors

Marijn Martens is a postdoctoral researcher at the multidisciplinary imec research group for Media, Innovation and Communication Technologies (imec-mict-UGent).

https://orcid.org/0000-0001-6264-0343

Ralf De Wolf is an assistant professor in New Media Studies at the Department of Communication Sciences, Ghent University, Belgium.

https://orcid.org/0000-0002-2586-4150

Lieven De Marez is a professor 'Media, Technology & Innovation' & 'User-centric innovation research' at the Department of Communication Sciences, Ghent University, Belgium. At the department, he is heading the multidisciplinary research group for Media, Innovation & Communication Technologies (imec-mict-UGent).

https://orcid.org/0000-0001-7716-4079

Correspondence to
Marijn Martens, Dept. Communication sciences, Ghent University, Korte Meer 11, Ghent 9000, Belgium,
Marijn.Martens@UGent.be