BI Norwegian Business School - campus Oslo

GRA 19703

Master Thesis

Thesis Master of Science

What if AI is not that fair?

- Understanding the impact of fear of algorithmic bias and AI literacy on information disclosure -

Navn: Carlos Eduardo Caldas de Souza

Start: 15.01.2021 09.00

Finish: 01.07.2021 12.00

What if AI is not that fair?

- Understanding the impact of fear of algorithmic bias and AI literacy on information disclosure -

Hand-in date:
01.07.2021

Campus:
BI Oslo

Examination code and name:
GRA 19703 Master Thesis

Programme:
Master of Science in Strategic Marketing Management

Supervisor:
Matilda Dorotic

Acknowledgments

This journey would not have been possible without the support of my MSc. thesis supervisor, Matilda Dorotic who guided me throughout this process. Thank you, Matilda, for inspiring me with your curiosity about the world.

A big acknowledgement goes to Emanuela Stagno, always willing to listen to a researcher under construction and to find the best solutions to make the hard work enjoyable.

An honourable mention for my beloved Frederik Hanses for all his patience and love. I am happy and grateful to share my life with you. A big thanks to my siblings, João and Paula. Being away from Brazil has never been so difficult. Thank you for all your support.

Finally, deserved thanks go to my colleagues and professors at BI Norwegian Business School. It was a challenging year, and still, we managed to make it memorable.

Summary

Boosted by the COVID-19 pandemic, the use of AI technology to collect personal information regarding the population's health has been gaining traction globally. Public authorities worldwide pinned their hopes on developing disease contact-tracing apps to quickly identify and notify people who have come into contact with infected individuals. However, population engagement did not work as expected. Whereas some countries registered a moderate adoption in others, the adherence was shallow. Inspired by this discrepancy between nations, this thesis investigates the effect of fear that the algorithm may be inaccurate, reproduce bias and harm people (fear of algorithmic bias) on willingness to disclose information via an AI system. Based on risk perception and folk perception of algorithms' literature, this study hypothesises that the individuals' understanding and knowledge of AI technology and the nature of the organisation holding the data in interaction with socioeconomic conditions impact their disclosure intention. Data from an online experiment in Qualtrics with 800 adults living in high-income countries (400) and low-income countries (400) were analysed using between-subjects ANOVA, linear regressions, moderation and mediation with PROCESS, and GLM analysis. This study found strong evidence that willingness to disclose information decreases as the fear of algorithmic bias increases. This work also found statistical evidence of the mediating role of privacy concerns and the moderating role of trust in government. The results suggest that one way for policymakers to increase the acceptance of AI is to improve governance over data input in large databases to mitigate individuals' fear that the algorithms are not properly functioning.

Table of Contents

1 Introduction

Boosted by the coronavirus pandemic, the use of AI technology to collect personal information regarding the population's health has been gaining traction globally. Driven by health concerns, public authorities in many countries have pinned their hopes on developing disease contact-tracing apps to quickly identify and notify people who have come into contact with infected individuals. By using Bluetooth or GPS, it would be possible to track individuals and set stringent restrictions on the population as a whole while significantly reducing the risk of new waves of infections (Quilty-Harper, 2020). A similar objective is behind temperature measurement systems towards passengers in airports (Guardian Staff & Agencies, 2020) and public buildings.

Between March and August 2020, the first months after the COVID-19 outbreak, more than 30 countries registered efforts to turn smartphones into personal trackers of Corona (O'Neill, 2020; O'Neill, Ryan-Mosley & Johnson, 2020; Babones, 2020). South Korea, early on, created a public database of coronavirus cases that provided detailed information about every infected individual (Cellan-Jones, 2020). The database was updated continuously using location information from payment card transactions, mobile phone signal data, and closed-circuit TV footage (Kasulis, 2020). It means that not just the users' infection status was stored. In addition, the authority kept in their database information on the individuals' location and people they have recently been in touch with. These two are considered private and sensitive data because it is difficult to anonymise. Other countries soon followed in the same direction.

It did not take long before companies such as Apple and Google released their own mobile technology to notify users of coronavirus exposure (Paul, 2020). Proposing a decentralised model, where data does not remain under the control of one specific organisation, the tech giants from Silicon Valley pledged to give users a higher degree of privacy. This way, protecting them from hackers, other private institutions, or the state itself (Criddle & Kelion, 2020), thereby addressing a key concern of the population in countries where trust in government remains low.

However, many still have doubts about the app's efficiency, the safety of disclosing private data, and, mainly, about the good intentions of the organisations responsible for the data collection. Opinions differ on whether these apps are just a technocratic daydream, a potential weapon to control citizens, or a tool that, in the wrong hands, could cause irreversible damage (Gold, 2020).

In fact, population engagement has not worked as expected (Babones, 2020). People did not always disclose private information, even if it was to prevent the spread of a pandemic. Whereas the community’s engagement with the corona contact-tracing app was moderate in some nations, such as South Korea, Singapore and Norway, in others, the initiative has not found enough support. In countries like Brazil, the population’s adherence was shallow (Tagiarolly, 2021).

The Corona situation is solely the most recent example of how disclosing health-related information has been a sensitive topic over the past decades. Individuals, in general, feel frightened of facing negative consequences that could occur when giving away private health data (Staa et al., 2016). In addition, medical literature shows that the human distrust of new technologies plays an essential role in disclosure behaviour. People often fear technologies that they do not understand or consider complex to use (Brosnan, 1998). However, little is known on how the human perception of algorithmic decisions influences this disclosure behaviour. More specifically, how the fear that the algorithm is biased impacts the decision to disclose private information.

Furthermore, studies show that socioeconomic inequality negatively influences trust in institutions and positively influences different fears. For example, people living in poorer countries manifest higher fear of crime (Kujala, Kallio, & Niemelä, 2019) and fear of social decline (Gidron & Hall, 2017). Thus, given the differences in population engagement with the contact-tracing app, the countries’ socioeconomic conditions may have played a role in fear of algorithmic bias and, consequently, the willingness of information disclosure with the local public authorities.

Therefore, the current study aims to answer the following research question: *"How does fear of algorithmic bias affect people's willingness to disclose private information with an AI system?"*

More specifically, this research strives to understand how the fear of algorithmic bias affects people's willingness to disclose private information, considering whether (i) understanding and knowledge of AI technology and (ii) the nature of the organisation holding the data in (iii) interaction with socioeconomic conditions impact their disclosure intention.

2 Literature Review

2.1 Fear of AI bias and willingness to disclose private information

A broad scientific literature shows that information contained in a person’s medical record is the most sensitive kind of private data (Lazare, 1987). Since early history, the risk of being embarrassed, stigmatised, or discriminated against by the disclosure of medical records has led people to avoid sharing relevant information with physicians and relatives — and even prevent them from seeking medical help (Lyons & Dolezal, 2017). This is because they commonly perceive diseases as defects, inadequacies, or shortcomings (Swan & Andrews, 2010).

The advances in automation over the past decades increased society's ability to collect and store data, improving healthcare quality significantly. Nevertheless, research shows the technology has not increased people's level of trust nor their willingness to disclose private health information with other individuals or institutions (Kenny & Connoly, 2016). In most cases, individuals disclose their health data only when they feel they do not have a choice (Rinik, 2019). For example, a requirement from the government or a private company to access some specific service (e.g., appropriate medical care) (Rinik, 2019; Hall & Pesenti, 2017).

Smith et al. (2011) show that privacy is a concern not only in matters related to health. Year after year, privacy concerns continue to drive users away from online services and businesses (Kukar-Kinney Close, 2010). Among the most common reasons for that, the fear of surveillance and the fear of unauthorized secondary use of the data (Smith et al., 1996).

However, Jian et al. (1998) introduced another crucial factor for hesitation to disclose information, specifically when the relationship between individuals and organisations occurs through technological means: the fear of automation. It refers to how much people believe these systems will not perform effectively and how unreliable and inaccurate the technology is (Jian et al., 1998).

More recently, Lee (2018) showed that, in fact, perceptions of algorithms, regardless of the algorithms' actual performance, can significantly influence their adoption. People perceive algorithms as helpful tools, but they also see them as possible tracking mechanisms. Algorithmic decisions were perceived as less fair and trustworthy and evoked more adverse emotions than human decisions (Lee, 2018).

Such perceptions are not total paranoia. Over the past few years, a series of empirical work has revealed different sets of biases found in giant Artificial Intelligence databases. A paper from Kay, Matuszek & Munson (2015) found that searches for professions in high positions produce fewer women's images in search engines (e.g., Google). The work shows a Selection Bias of the algorithm that occurs when the dataset overrepresents one particular group and underrepresents another. In addition, Lum & Isaac (2016) found that software used in several states of the U.S. could unfairly lead police to target specific neighbourhoods. Further, ProPublica revealed that black defendants were far more likely than white defendants to be incorrectly judged or at a higher risk of recidivism (Angwin, Larson, Mattu & Kirchner, 2016). In this so-called Latent Bias, the algorithm incorrectly identifies something based on historical data or based on a stereotype that already exists in society (Angwin, Larson, Mattu & Kirchner, 2016).

The algorithmic bias is not only present in criminal datasets. It is also observed in health databases (Obermeyer et al, 2019). The study revealed that “Black patients are considerably sicker than White patients at a given risk score” (Obermeyer et al., 2019, p4). The result is in line with previous studies involving racial/ethnic bias in healthcare providers (Maina et al., 2017).

Although they do not directly mention concern about biased algorithms in large databases, Staa et al. (2016) show that people, indeed, are concerned about the proper use of their health data provided. The study examined the National Health

Service England's decision to close down its care.data programme, - in which English citizens could electronically share health data for research - after people did not give the programme permission to disclose their information. Their work pointed to three critical elements for the success of big health data projects: (i) public trust that records are held securely and anonymised appropriately, (ii) public awareness of how their data might be used, and (iii) data being used for high-quality science.

Nevertheless, there is still little scientific research on how the fear of algorithmic biases may lead people to not disclose private information. Therefore, the following hypothesis is suggested:

> **H1.** The willingness to disclose private information will decrease as the fear of algorithmic bias increases.

2.2 AI literacy as a moderator

People interact with AI systems on a daily basis, whether using facial recognition to unlock the mobile phone, using a navigation app to find the quickest route to the desired destination, or scrolling their feed on social media. However, the public knowledge and technical understanding of these technologies are often limited (Long & Magerko, 2020; Eslami et al., 2015). It is not rare that individuals do not even recognize that they are interacting with AI. (Eslami, 2019).

This lack of proper understanding of the technology limits people's ability to use and collaborate with AI (Long & Magerko, 2020). It often raises specific concerns (e.g. worries of loss of control of AI, ethical considerations) and leads to public frustration if expectations for development are not satisfied (Fast and Horvitz, 2017). Moreover, polls show a significant amount of public concern related to AI topics (Mozilla, 2019; Ipsos, 2017; British Science Association, 2016).

Hence, education seems to play a crucial role to advance AI understanding. Several initiatives, such as "AI for K12", have focused on teaching to improve people's AI literacy and reduce their misconceptions about technology. AI literacy is defined as a "set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace" (Long & Magerko, 2020, p2).

AI literacy is akin to other previously defined literacies, such as computational and digital literacy (Long & Magerko, 2020; Bawden et al., 2008). However, only digital literacy can be considered a prerequisite of AI literacy (Long & Magerko, 2020) since individuals need to understand how to use computers to make sense of AI systems.

Over the past decades, several studies have measured the digital literacy of young (Oblinger, 2005) and older adults (Oh et al., 2021), and previous literature shows that digital literacy moderates the relationship between privacy concern and disclosure behaviour (Park, 2013; Park, 2008; Hoffman, Novak & Peralta, 1999; Shehean & Hoy, 1998). However, to date, few studies have investigated AI literacy among groups and its association with fear and information disclosure.

Thus, considering the positive moderation of digital literacy on disclosure behaviour, one can argue that IA literacy might also behave as a moderator variable. Therefore, the following hypothesis is suggested.

> **H2.** The effect of the fear of algorithmic bias on the willingness to disclose private information decreases as AI literacy increases.

2.3 The nature of the organisation holding the data as a moderator

Whether a doctor, a family member or a medical authority, the trust in the counterpart holding the data appears to be a crucial determinant in disclosing or not personal health information (Platt, Jacobson & Kardia, 2017). Trust is defined as the belief that the other party "will behave responsibly and will not attempt to exploit the individual's vulnerabilities" (Pavlou, 2003, p102).

Despite the many benefits that large databases offer in health policy planning, the distrust towards the institution storing these data is a barrier to improving public service provision and scientific research (Marwick & Hargittai, 2017). This lack of trust appears even when the commitment to use the information only for the common good is explicitly stated to the public (Platt & Kardia, 2015). Examining data sharing in biobanks, a type of biorepository that stores biological samples for use in scientific research, Platt & Kardia (2015) found that information disclosure attitudes were positively associated with the trust in the organisation collecting the data.

Although trust is most commonly studied as a global, multidimensional construct, empirical studies show that people's overall trust level is structured in different dimensions, influenced by various factors (McKnight et al., 2002; LaVeist, 2009; Platt & Kardia, 2015).

The wide variety of descriptions developed in different periods and contexts has resulted in an assortment of labels and several overlapping typologies (McKnight, 2002). However, two trust dimensions seem preponderant across studies: trust in expertise and benevolence of the counterpart (Thorslund, 1976; Koller, 1988; Kasperson et al., 1992; McLain & Hackman, 1999; White, 2005). Trust in expertise relates to the perception that the counterpart is competent (1) to use the information provided to achieve an expected and positive outcome and (2) to keep the data safe, avoiding leakages or data stealing. Trust in benevolence relates to the perception that the counterpart is honest and integrous and will not use the information provided to harm the individual.

Literature shows that these two dimensions of trust vary according to whether the organisation making use of health information is public or private (Ostherr et al., 2017). In an investigation on people's willingness to disclose health information for scientific purposes, the general public considered public researchers - which were required to follow explicit protocols for data privacy - more suspicious than corporations that collected the same type of data without any supervision (Ostherr et al., 2017).

These asymmetric circumstances related to the organisation's nature - whether public or private, intensified during the Covid-19 outbreak. Edelman Trust Barometer report 2021 (Edelman, 2021) shows that people's trust in public institutions and businesses has plummeted worldwide. Still, in most countries, people seem to trust companies more than their own government and CEO's should step in when the political leaders do not fix societal problems (Edelman, 2021). Thus, one can argue that the nature of the organisation holding the data might moderate the relationship between fear of algorithmic bias and willingness to disclose information. Therefore, the following hypothesis is suggested:

H3. The effect of the fear of algorithmic bias on the willingness to disclose private information increases when the authority holding the data is the government and decreases when the authority holding the data is a company.

2.4 The interaction effect of the organisation holding the data and the country's income level on information disclosure

Although trust in governments has decreased worldwide in the first year of the pandemic, the fall was more pronounced in some places than in others. Low-income countries have registered significantly lower percentages of trust in public institutions than high-income countries (Edelman, 2021).

However, this does not seem to be an isolated incident caused by how public authorities dealt with the Covid-19 outbreak. Global comparisons of trust attitudes within countries over the last few decades suggest extensive time-persistent cross-country heterogeneity. Whereas some research focuses on the individual relationship between citizens (Welter & Nadezhda, 2011; Gambetta, 2000; Rotter, 1971), other studies examine the population's trust level in local authorities (Smallbone & Lyon, 2002; Ortiz-Ospina & Roser, 2016).

The Organisation for Economic Co-Operation and Development (hereafter called OECD), provides estimates of interpersonal trust and trust in public institutions across countries. Comparing the average ratings of trust in the (i) political system, (ii) police, and (iii) legal system, the report (OECD, 2015) shows that in countries such as Norway, Finland, Netherlands, and Switzerland more than 60% of respondents trust people and institutions. On the other hand, in low-income countries such as Colombia, Brazil, Ecuador and Peru, where the inequality gap between rich and poor people is high, less than 10% of respondents trust people and institutions (Ortiz-Ospina and Roser, 2016).

The extent to which trust is linked to economic development has been the subject of many academic papers on economic growth (Bjørnskov, 2018; Algan & Cahuc 2010; Guiso et al. 2006). Most studies find a substantial positive relationship between the country's internal trust and Gross Domestic Product level, meaning that in high-income countries, people show higher levels of trust towards public organisations than in low-income countries.

Furthermore, cross-country studies, within-country studies, and most experiments suggest that socioeconomic inequality negatively influences trust. It means that in countries with high inequality, people show lower levels of trust towards public organisations compared to countries with low inequality (Gould & Hijzen, 2016; Jordahl, 2007).

Therefore, it is hypothesised that the interaction of the organisation holding the data and the country's income level affects the relationship between fear of algorithmic bias and the willingness to disclose private information (moderated moderation). However, considering that countries with different socioeconomic conditions differ regarding trust in their government, the following two hypotheses are suggested:

> **H4a**. The effect of the fear of algorithmic bias on willingness to disclose private information in a low-income country is stronger when the government holds the data and weaker when a private company holds the data.
>
> **H4b.** The effect of the fear of algorithmic bias on willingness to disclose private information in a high-income country is stronger when a private company holds the data and weaker when the government holds the data.

The following research model has been developed based on the identified gaps in existing literature and list of identified hypotheses (Figure 1).

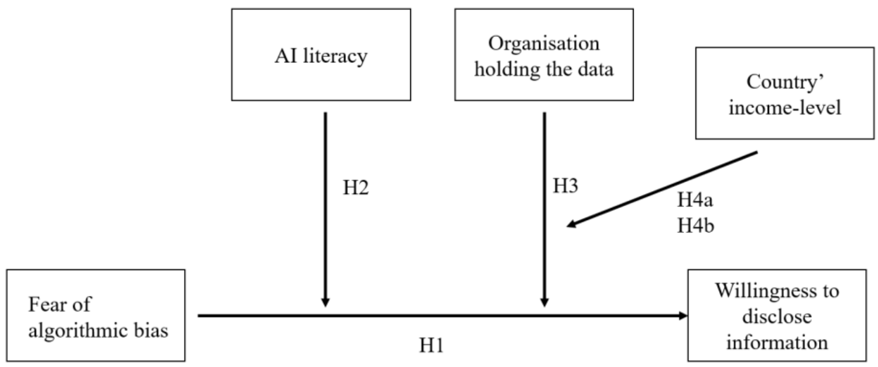

Figure 2.1. Conceptual Model of Proposed Moderation Role of AI Literacy and Authority holding the data.

In sum, prior research has not explained the role of the fear of algorithmic bias on people's willingness to disclose private information. Therefore, the current study aims to understand how fear of algorithmic bias affects people's willingness to disclose private information, considering whether (i) understanding and knowledge of AI technology and (ii) the nature of the organisation holding the data in (iii) interaction with socio-economic conditions impact the willingness to disclose private information.

3 Research Questions and Hypotheses

To summarize the discussion so far, the main research question of this thesis is the following:

"How does fear of algorithmic bias affect people's willingness to disclose private information with an AI system?"

Based on the existing literature and the gaps identified, the following hypotheses were developed:

> **H1.** The willingness to disclose private information will decrease as the fear of algorithmic bias increases.
>
> **H2.** The effect of the fear of algorithmic bias on the willingness to disclose private information decreases as AI literacy increases.
>
> **H3.** The effect of the fear of algorithmic bias on the willingness to disclose private information increases when the authority holding the data is the government and decreases when the authority holding the data is a company.
>
> **H4a**. The effect of the fear of algorithmic bias on the willingness to disclose private information in a low-income country increases when the government holds the data and decreases when a private company holds the data.
>
> **H4b.** The effect of the fear of algorithmic bias on the willingness to disclose private information in a high-income country increases when a private company holds the data and decreases when the government holds the data.

4 Research methodology and data collection

The following section describes the process used to gather and analyse the data needed to test the hypotheses. First, it gives details on the sampling method and data collection procedure. Then, it explains the operationalisation of variables, study design and procedure design.

4.1 Sample

A total of 812 subjects were engaged in this study: 405 participants living in High-income countries (Netherlands, Norway, Switzerland, and Finland) and 407 participants residing in a Low-income country (Brazil). The online surveys were conducted simultaneously during the second quarter of 2021.

The country selection observed the following criteria: they belong to two distinct groups both in terms of income level and social inequality indicators. When the Gross National Income per capita (GNI per capita) is considered, the gap between the two groups is meaningful. Similarly, the two groups differ in social inequality indexes. On the one hand, Netherlands, Norway, Switzerland, and Finland register some of the highest levels of human development when inequality is accounted for, appearing among the first places on the Inequality-adjusted human development index (IHDI). On the other hand, Brazil appears at the bottom in 88th place (United Nations, 2020).

Due to the limitations of the GNI as a measure of standard living, the combination of these two indexes seems appropriate and sufficient to define the country selection. The table 4.1 summarizes the countries positions in the two indexes (World Band, 2019; United Nations, 2020).

Table 4.1. Position of the countries concerning Gross National Income per capita (GNI) and the Inequality-adjusted human development index (IHDI).

Country	Gross National Income per capita (GNI) ranking position	Inequality-adjusted human development index (IHDI) ranking position
Switzerland	2	3
Norway	3	1
Netherlands	13	8
Finland	15	4
Brazil	72	88

Thus, the sample consisted of people in both groups of countries, aged 18 to 65+, with varied income and education backgrounds. There was no restriction regarding ethnic categories. The gender structure of the sample was 50,9% females and 48,2% males. Table 4.2 summarizes the characteristics of the sample:

Table 4.2. Demographics of Study Sample.

	Factor	Total sample	Low-income country	High-income country
	n	812	407	405
Gender				
	Female %	50.9%	61.4%	40.1%
	Male %	48.2%	38.1%	58.6%
	Other %	0.9%	0.5%	1.3%
Age				
	18-24	21.1%	9%	33.7%
	25-34	42.5%	43.8%	41.1%
	35-44	23.3%	26.9%	19.5%
	45-54	7.7%	10.9%	4.4%
	55-64	4.3%	7.7%	0.8%
	Over 65	1.1%	1.7%	0.5%
Education				
	Some School, no degree	2%	1%	3.1%
	High School graduate	10.2%	3.2%	20.6%
	Some College, no degree	12.9%	11.2%	14.7%
	Bachelor's degree	47.3%	61.7%	32.4%
	Master Degree	22.8%	18.4%	27.2%
	Doctorate's degree	4.8%	4.5%	5.1%

The participants were recruited online using both social media networks (50%) and Prolific (50%). Several studies show similarities between participants selected from convenience samples using online recruitment and large population samples (Coppock, 2017; Coppock & McClellan, 2019).

4.2 Operationalisation of Variables

Items from existing scales were used wherever possible to increase reliability and ease of comparison with previous work in this field. Besides, some items were modified, most of which were adaptations to increase their applicability to the local context.

Willingness to disclose information was measured in 10 items. First, a list of 9 items examines how likely/unlikely participants were to disclose different information with an AI system. The set with private items presented initially in Lo (2010) and developed by Mourey & Waldman (2020) was modified to include greater variety and more relevant examples. In addition to that a single-item regarding overall willingness to disclose was included.

Further, a set of five questions examined participants' fear of algorithmic bias. These items were adapted from the Cyber-Paranoia and Fear Scale (Mason, Stevenson, & Freedman, 2014).

As until today there is no AI literacy scale established, AI literacy was measured using items newly created. To do so, the format of the items followed the Privacy Literacy scale (Westin, 2003). The various competencies that define individuals' AI literacy were based on the work of Long & Magerko (2020), and the five "big ideas" of AI to guide the standards development (Computer Science Teachers Association, 2017). Table 4.3 shows the constructs, sources and scales for construct items established in the proposed model.

Table 4.3. Constructs' definitions, sources and scales for construction items.

Construct	Construct definition	Scales	Sources
Willingness to disclose information	Intention to disclose information to the AI system through an app.	Likert 1-5	Malhotra et al., 2004. Mourey ad Waldman (2020) Lo (2010)
Fear of algorithmic bias	Fear that algorithms may reproduce bias and cause harm against the individual or others.	Likert 1-7	Mason et al., 2014.
AI Literacy	Set of competencies that enables individuals to critically evaluate AI technologies.	Likert 1-7	Item format adapted from Privacy Literacy scale (Westin, 2003). Competencies of AI literacy based on Long
Country economic status	Country income-level' group per capita group according to the GNI index.	High-income / Low-income	World Bank

To exclude variance explained by potential confounding factors, and perform exploratory analysis in the future, information about variables that might have an impact on information disclosure was gathered as well. In addition, two variables were included as dependent variables.

Six control variables were measured: (1) Privacy Concern (Dinev et al., 2008), (2) Fear of Covid-19 (Ahorsu et al., 2020), (3) Trust in the Government (Poznyak et al., 2013), (4) Perceived Control over data (Lo, 2010), (5) Perceived Need for Governance Surveillance (Dinev et.al., 2008) and (6) Trust in AI (Lee, 2018).

In addition, a second dependent variable was measured: intention to Adopt the Application. The items are adapted from Bhattacherjee (2001) and Hamari & Koivisto (2015). Table 4.4 shows the constructs, sources, and scales for construct items.

Table 4.4. Constructs' definitions, sources and scales for construction items.

Construct	Construct definition	Scales	Sources
Privacy Concern	Individual's concern about possible loss of privacy due to voluntary or surreptitious information disclosure	Likert 1-7	Dinev et al., 2008.
Fear of COVID-19	Fear and worries relating to the coronavirus.	Likert 1-7	Ahorsu et al., 2020.
Trust in the government	Belief that the government is benevolent, competent, or honest.	Likert 1-7	Poznyak et al., 2013.
Perceived control over data	The extent to which someone perceives himself in charge of actively making privacy management choices.	Likert 1-7	Lo, 2010.
Perceived need for Governance Surveillance	Perceived need for the government to have greater access to personal information and to monitor personal activities.	Likert 1-7	Dinev et.al., 2008.
Trust in AI	Belief that the AI system is competent.	Likert 1-7	Lee, 2018.
Adoption Intention	Individual's willingness to adopt the app.	Likert 1-7	Bhattacherjee, 2001.

4.3 Study design

A quantitative online experiment based on hypothetical situations was conducted to test the hypotheses. As the purpose of the study was to examine the influence of fear of algorithmic bias on information disclosure and understand whether it varies as a function of the type of organisation holding the data, the experiment consisted of a 2 x 2 between-subjects design.

Common news stories inspired the stimuli and dependent variables' designs to give the current work greater ecological validity. The news stories' explicitness manipulated the design's key constructs: (1) the organisation responsible for holding the data (government vs private company) and (2) strength of threat of algorithmic bias (High Threat/Low threat). Everything else regarding the format, readability, font and text length was kept the same in all four conditions.

Furthermore, the survey was conducted simultaneously in five countries to assess a potential interaction between the type of organisation holding the data and country income-level. Initial survey development was conducted in English, as the source items had been previously published in English. Next, the survey was professionally translated. Finally, the survey was pre-tested (n = 10) to collect feedback regarding the wording and clarity of the questions and reduce potential misinterpretations and measurement inaccuracies (Ruel, Wagner & Gillespie, 2016).

4.4 Procedure

Participants were informed that the study aimed to assess their app usage habits and perceptions and began the survey by selecting what types of mobile applications (apps) they have used at least once in the last 12 months. Table 4.5 shows the entire list inspired by Mouray & Waldman (2020).

Table 4.5. List of mobile apps presented to the participants at the beginning of the study.

Select what types of mobile applications (app) you used at least once in the last 12 months:	Social media apps Weather forecast app Disease track app Bank app Photo editing app Games app Sports app

Once participants have selected the apps they used at least once, in the last 12 months, participants were informed that from their answer they would be randomly selected to read a news article and answer questions regarding the implementation of new technologies. However, the apps selection in the first section did not affect any measure. Regardless of which apps the participants selected, they answered questions related to launching a new AI system that tracks infectious disease spreading via a contact-tracing app.

The participants were then piped into the manipulations and dependent measures to make the survey relevant and meaningful. They were randomly assigned to read one of four news stories (Table 4.6). The four versions of the persuasive message used in the experiment can be found in Appendix 1.

Table 4.6. 2x2 Factorial design.

Treatments	Government holding the data	A company holding the data
Low Threat	The Health Ministry holding the data Low threat of algorithmic bias	A private big tech company holding the data Low threat of algorithmic bias
High Threat	The Health Ministry holding the data High threat of algorithmic bias	A private big tech company holding the data High threat of algorithmic bias

Participants could not advance in the survey until after a 1 min delay to ensure they read the news article. After that, participants proceeded to a new screen that read: *"In light of the news article you just read, please review each of the following items and indicate what you would do in this situation"*.

The first question consisted of a randomised list of 9 items. The participants had to indicate whether they would share different information items with the AI system presented in the scenario. The list below was inspired by the items presented initially in Lo (2010) and developed by Mouray & Waldman (2020). The set was adjusted to include greater variety and more relevant examples (Table 4.7).

Table 4.7. List of shareable items presented to the participants right after the stimuli.

	All items were measured on Likert 1-5
In light of the news article you just read, please review each of the following items and indicate whether you would share this information with the AI system presented before.	First and Last name E-mail address Name of family members Real-time Location Political views or preferences Whether you have a chronic disease Whether you have a contagious disease Pictures drunk or Drinking Sexual Orientation

Following these ratings, participants completed additional measures to assess their willingness to disclose personal health information, and the other relevant variables. This set consisted of a randomised list of 31 questions (Appendix 2).

Furthermore, three items were included in the set to (1) assess the effectiveness of a manipulation treatment, and (2) ensure participants follow instructions when completing the surveys. The questions were inspired by the Instructional Manipulation Check (Oppenheimer, 2009; Ejelöv & Luke, 2019) and adjusted for the current study (Appendix 3).

Finally, the online survey also captured appropriate demographic variables including gender, age, education, and income, inspired by Smith et al. (1996). The item Smartphone usage experience was introduced initially in Malhotra et al. (2004) and adapted to the context of the current study (Appendix 4).

5 Results

5.1 Data Preparation & Reliability

Respondents who (1) rejected consent to accept the questionnaire or (2) failed the attention check item were excluded from the study. This reduced the sample size from N=812 to N=788. After that, measure items that were required were reverse-coded, and new variables were computed based on the online survey measures conducted among the participants.

The willingness to disclose information was computed as the average score of ten disclosure-related items and the fear of algorithmic bias was computed as the average score of five algorithmic fear-related items for each participant. Similarly, AI literacy was computed by taking the reported average of item scales that measured AI literacy.

The overall scores for Privacy Concern, Fear of Covid, Trust in Government, Perceived Need for Governance Surveillance, and Adoption Intention were also computed. All scales had acceptable reliability (Coefficient alpha > .60) (Cortina, 1993). Table 5.1 summarizes the reliability statistics for computed variables.

A series of manipulation checks were performed to assess how effective the experimental manipulations were at increasing the (1) perception of algorithmic bias and the (2) perception of the organisation holding the data. Independent sample t-tests were statistically significant, so it can be concluded that the independent variables were effectively manipulated.

Table 5.1. Reliability statistics for computed variables.

Variable	Coefficient alpha	N of items
Willingness to disclose	.855	10
Fear of algorithmic bias	.816	5
AI literacy	.626	6
Privacy concern	.899	3
Fear of COVID-19	.859	3
Trust in Government	.919	2
Trust in AI	.762	2
Perceived Need for Governance Surveillance	.753	4
Adoption Intention	.871	4

5.2 Hypothesis Results

Hypothesis 1

H1. *The willingness to disclose private information will decrease as the fear of algorithmic bias increases.*

The willingness to disclose information scores averaged 4.84 (SD=1.53) in the low-threat algorithmic bias scenario (n=389), and lower at 4.59 (SD=1.54) in the high-threat algorithmic bias scenario (n=399). Table 5.2 shows the descriptive statistics.

Table 5.2. Dataset descriptive statistics.

	N	Mean	Std. Error	Std Deviation
Low Risk	389	4.84	1.53	0.78
High Risk	399	4.59	1.54	0.77

Linear regression analysis was used to test if fear of algorithmic bias significantly predicted respondents' willingness to disclose information. The results of the regression indicated the predictor explained 9% of the variance (R2 =.090, $F(1,804)=79.20$, $p<.01$). It was found that fear of algorithmic bias significantly predicted willingness to disclose (β =-.403, $p<.001$). Therefore, Hypothesis 1 is supported by statistical evidence.

The negative Unstandardized β indicates a negative effect of the predicator on the dependent variable. For every one-unit increase in the reported fear of algorithmic bias, the willingness to disclose information decreases by .403 due to the fear effect. Table 5.3 shows the model summary and table 5.4 summarizes the coefficient estimation.

Table 5.3. Model Summary for linear regression with the willingness to disclose information as the dependent variable and fear of algorithmic bias as the independent variable.

R	R squared	Adjusted R squared	Std Error for the Estimate
299	0.90	0.89	1.474

Table 5.4. Coefficient estimation for linear regression with the willingness to disclose information as the dependent variable and fear of algorithmic bias as the independent variable.

	Unstandardized B	**Coefficients Std Error**	**t**	**Sig.**
Constant	6.645	.223	29.815	<.001
Fear of algorithmic bias	-.403	.450	-.8.829	<.001

Figure 5.1 illustrates the relationship between the fear of algorithmic bias and willingness to disclose information.

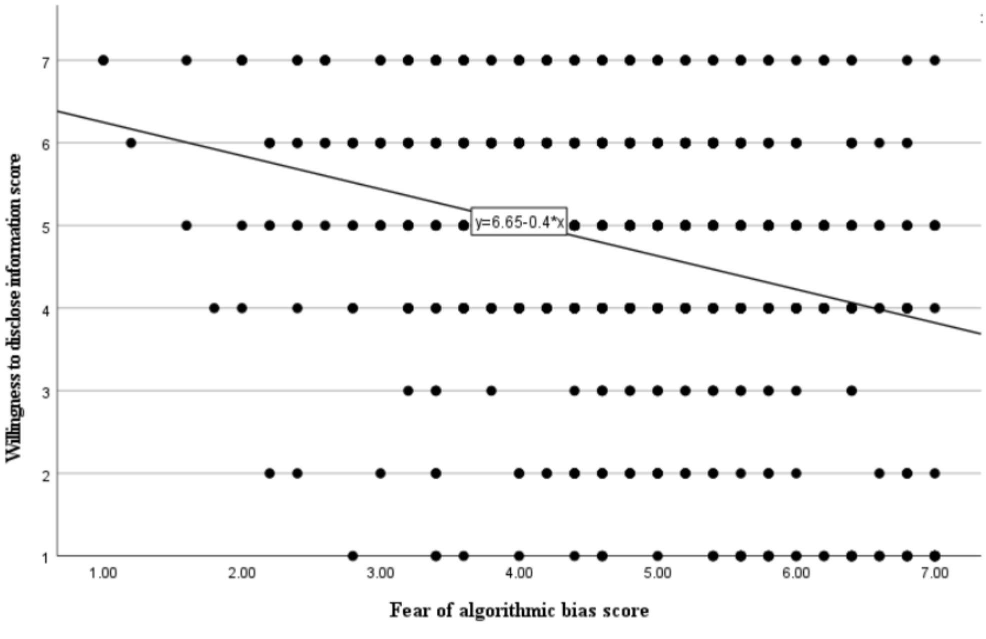

Figure 5.1. Regression plot of Willingness to disclose information and fear of Algorithmic bias scores.

Hypothesis 2

***H2:** The effect of the fear of algorithmic bias on the willingness to disclose private information decreases as AI literacy increases.*

PROCESS macro (model 1 in Hayes, 2017) was used to test for a moderating effect of AI literacy on the relationship between the fear of algorithmic bias and the willingness to disclose information. PROCESS is a freely available macro for SPSS that carries observed-variable moderation and conditional process analysis using a multiple regression approach (Hayes, 2017).

Results from a moderation analysis indicated that the fear of algorithmic bias, AI literacy, and the interaction term are not significant. Therefore, Hypothesis 2 is not supported by statistical evidence. Table 5.5 shows the Estimates of Fixed Effects.

Table 5.5. Estimates of Fixed Effects.

Parameter	Estimate	Std Error	t	Sig.
Fear of algorithmic bias	-1.715	2.939	-5.835	5.597
AI literacy	5.107	2.609	1.9574	.507
Interaction effect (Fear*AI literacy)	-.408	.521	-7.838	4.334

Hypothesis 3

***H3.** The effect of the fear of algorithmic bias on the willingness to disclose private information increases when the authority holding the data is a company and decreases when the authority holding the data is the government.*

A new test of moderation (model 1 in Hayes 2017) was conducted, this time having the fear of algorithmic bias as the predictor variable, the nature of the organisation holding the data as the mediator, and the composite willingness to disclose information as the dependent variable.

Results indicated that the fear of algorithmic bias is significant (b=-.4161, t (802) =-6.67, p=.000). However, the organisation holding the data (b=.0342, t (802) =.076, p=.9390) and the interaction between these two variables (b=.0230, t (802) =.254, p=.7996) are not significant. Therefore, Hypothesis 3 is not supported by statistical evidence. Table 5.6 summarizes the estimates of fixed effects.

Table 5.6. Estimates of Fixed Effects.

Parameter	Estimate	Std Error	t	Sig.
Fear of algorithmic bias	-4.161	.623	-6.67	.000
Organisation holding the data	.342	4.469	.766	9.390
Interaction effect (Fear*Organisation holding the data)	.230	2.540	2.540	7.996

Hypothesis 4a and Hypothesis 4b

***H4a**. The effect of the fear of algorithmic bias on the willingness to disclose private information in a low-income country increases when the government holds the data and decreases when a private company holds the data.*

***H4b.** The effect of the fear of algorithmic bias on the willingness to disclose private information in a high-income country increases when a private company holds the data and decreases when the government holds the data.*

Before testing hypothesis H4a and H4b to examine potential differences in the direction of the interaction effect, a test of moderated moderation (model 3 in Hayes 2017) was performed. The aim was to test whether the interaction of the organisation holding the data and the country's income level on the relationship between fear of algorithmic bias and the willingness to disclose information was significant or not.

The test that had the fear of algorithmic bias as the predictor variable (X), the nature of the organisation holding the data as the moderator (M), the country income-level as the moderated moderator (W), and the composite willingness to disclose information as the dependent variable (Y) was not significant (Table 5.7). Therefore, Hypothesis 4a and 4b are not supported by statistical evidence. Figure 5.2 shows the P-value of each variable.

Table 5.7. Estimates of Fixed Effects.

Parameter	Estimate	Std Error	t	Sig.
Fear of algorithmic	-5.329	.750	-71.050	.000
Organisation holding the data	-2.375	5.342	-4.446	6.568
Country income level	-.373	5.314	-.701	9.441
Fear * Organisation	.359	1.081	3.323	7.398
Fear* Country	-.151	1.086	-1.392	8.894
Organisation * Country	-4.775	7.818	-6.107	5.416
Fear *Authority *	1.618	1.587	1.019	3.081

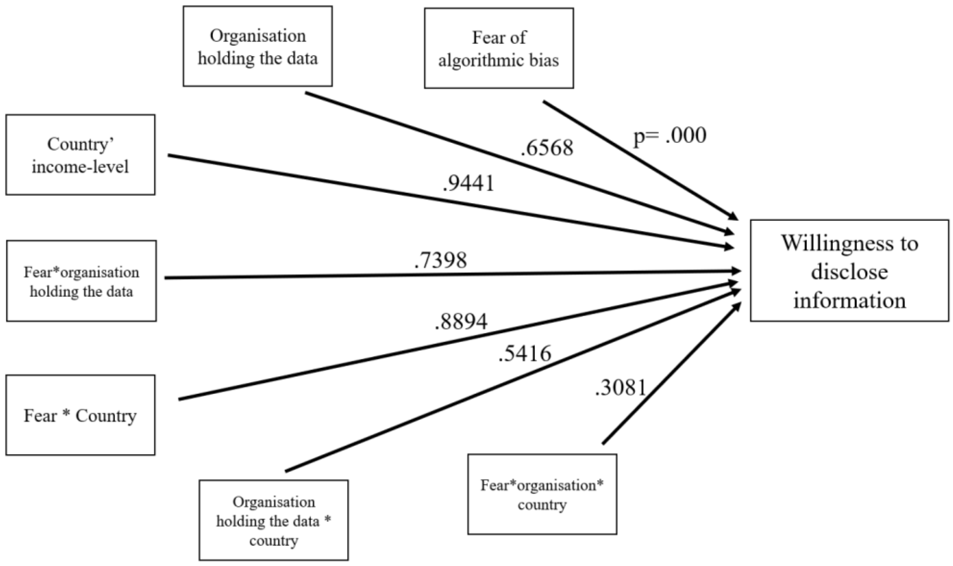

Figure 5.2. Variables' statistical significance.

5.3 Other results

Statistical differences among groups

Next, the analysis was expanded beyond the hypotheses to examine the significance and effect of variables outside the proposed model. These variables included Trust in Government, Privacy Concern, Fear of Covid-19 and Intention to adopt the App.

Firstly, a one-way between subjects ANOVA was conducted to examine statistical differences among the means of the two groups of countries considering the several variables collected in the study. A significant effect was found on Trust in Government [(F (1,786)= 902.48, p=<.001] and Fear of COVID [(F(1,786)= 402.76, p=<.001].

The low-income country presented a considerably lower level of Trust in Government (M=1.8, SD=1.1) than High-income countries (M=4.6, SD=1.4) as figure 5.2 shows.

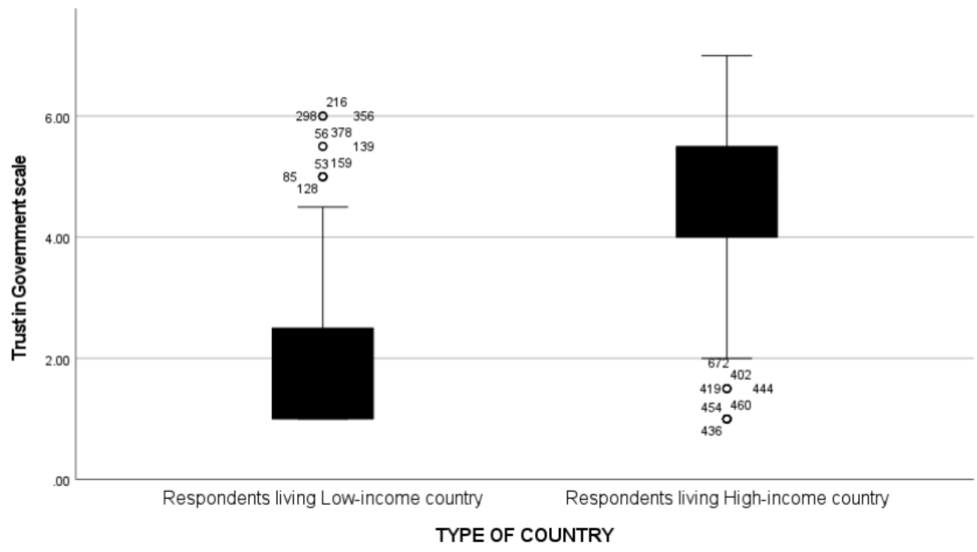

Figure 5.3. Boxplot of mean difference of Trust in Government in the Low-income country and High-income countries.

On the other hand, the low-income country (M=5.2, SD=1.3) presented a considerably higher Fear of Covid-19 than the high-income countries (M=3.3, SD=1.3) as Figure 5.2 shows.

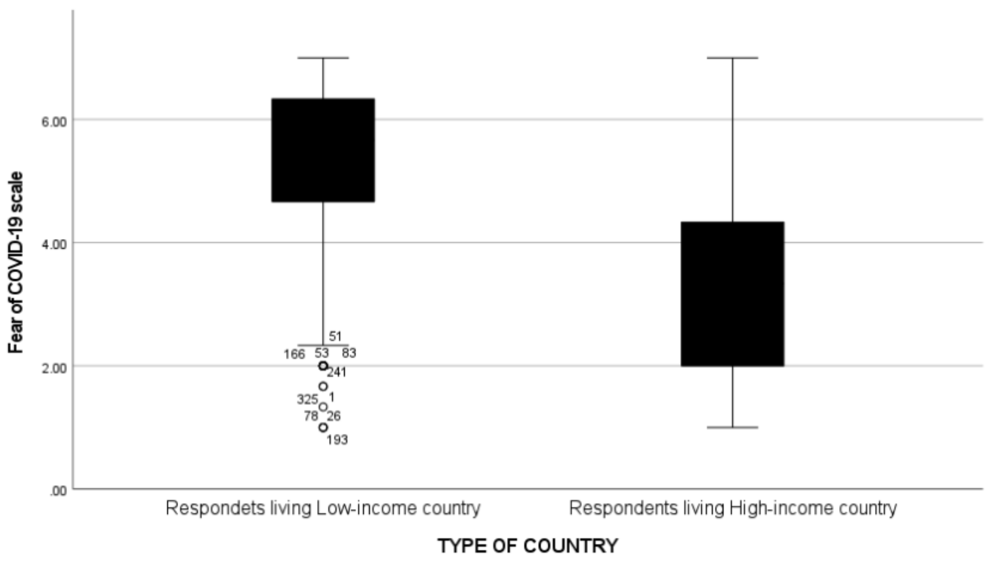

Figure 5.4. Boxplot of mean difference of Fear of COVID-19 in the Low-income country and High-income countries.

Next, a one-way ANOVA was conducted to examine statistical differences among the two threat conditions (Low threat/High threat of algorithmic bias) considering the several variables collected in the study.

Results indicated a significant effect on fear of algorithmic bias [(F(1,786)= 10.38, p=.001], privacy concern [(F(1,786)= 4.87, p=.028], willingness to disclose information [(F(1,786)= 5.06, p=.025], and adoption intention [(F(1,786)= 6.86, p=.009].

Similarly, a one-way ANOVA was conducted to examine statistical differences among the two organisations holding the data conditions (Government/Private company) considering the several variables collected in the study. However, results indicated no significant effect.

Trust in Government as a moderator

Further, results from a simple moderation analysis with PROCESS (model 1) indicated that Trust in Government moderates the relationship between the fear of algorithmic bias and intention to adopt the app. Both the independent variables and the interaction term are significant. Table 5.8 summarizes the parameter estimates.

Table 5.8. Parameter estimates.

Parameter	Coeff	Std Error	t	Sig.
Constant	7.8787	3.856	204.301	.000
Fear of algorithmic bias	-6.786	.754	-89.973	.000
Trust in Government	-2.469	.981	-25.180	.012
Fear*Trust	.704	202	3.486	.005

Results from the conditional effects of the focal predictor at values of the moderator shows that there is a relationship between trust in government and intention to adopt the app in all trust levels. For below average level of trust, average and above average, the effects are significant as table 5.9 shows below.

Table 5.9. Conditional effects of the focal predictor at values of the moderator.

Trust in Government	Effect	Std Error	t	Sig.
Below average	-6.082	.593	-102.597	.000
Average	-4.673	.402	-116.348	.000
Above Average	-2.912	.619	-47.055	.000

Privacy concern as a mediator

Further on, results from a simple mediation analysis with PROCESS (model 4) indicated that the fear of algorithmic bias is indirectly related to willingness to disclose information through its relationship with privacy concerns. Participants' privacy concerns score higher when fear of algorithmic bias is presented (b=.6850, p < .000). The privacy concerns, in turn, were subsequently related to a lower willingness to disclose information (b=-.1227, p < .000).

After considering the fear of algorithmic bias indirect effect through influencing privacy concern, the fear of algorithmic bias direct effect on willingness to disclose was also significant ((b=-.1939, p < .000). Besides, the total effect of the fear of algorithmic bias on willingness to disclose is also significant (b=-.2780, p < .000). Figure 5.5 illustrates the mediating effect. The full SPSS output can be found in Appendix 5.

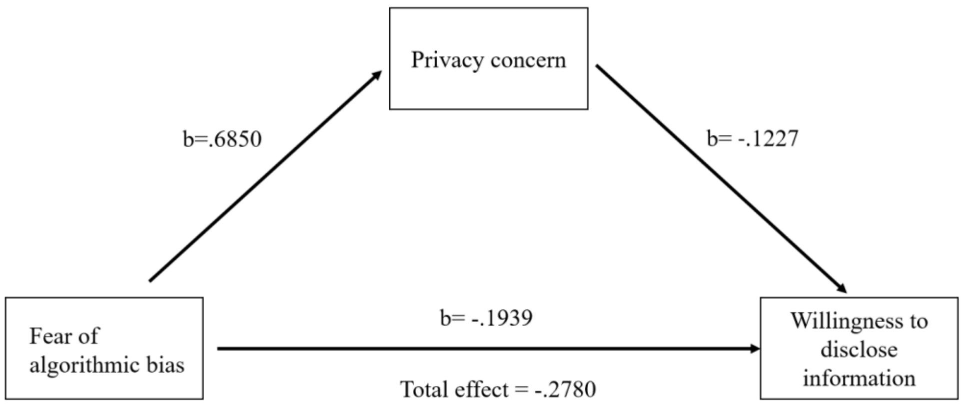

Figure 5.5: The mediating effect of privacy concern in the relationship between fear of algorithmic bias and willingness to disclose information.

Variables related to perceived risk as predictors of willingness to disclose information and intention to adopt the app

Finally, a general linear regression showed a significant relationship of the three variables related to perceived risk present in the study on willingness to disclose information ($R^2 = .250$, $F(3,784)=87.2$, $p<.01$). Fear of algorithmic bias ($\beta=-.207$, $p < 0.001$), privacy concern ($\beta=-.140$, $p < 0.001$), and fear of Covid-19 ($\beta=.151$, $p < 0.001$) are significant predictors. (Table 5.10).

The negative Unstandardized β of fear of algorithmic bias and privacy concern indicates a negative effect of both variables on the willingness to disclose. In contrast, the positive Unstandardized β of fear of Covid-19 shows a positive impact. The overall model-fit ($R^2 = .250$) shows that 25% of the variation in willingness to disclose information can be explained by the model containing the three predictors.

Table 5.10 Coefficient estimation for willingness to disclose as a dependent variable.

Variable	Unstandardized B	Coefficients Std. Error	t	Sig.
Constant	3.737	.135	27.660	<.001
Fear of algorithmic bias	-.207	.028	-7.402	<.001
Privacy concern	-.140	.022	-6.269	<.001
Fear of Covid-19	.151	.016	9.523	<.001

A follow-up model replacing the dependent variable by intention to adopt the app was also significant ($R^2 = .244$, $F(3,.784)=84,3$, $p<.01$). The table 5.11 summarizes the coefficient estimates.

Table 5.11. Coefficient estimation for adoption intention as a dependent variable.

Variable	Unstandardized B	Coefficients Std. Error	t	Sig.
Constant	7.180	.217	33.148	<.001
Fear of algorithmic bias	-.346	.045	-7.714	<.001
Privacy concern	-.249	.036	-6.946	<.001
Fear of Covid-19	.165	.025	6.492	<.001

6 Discussion

The outbreak of the COVID-19 pandemic led to severe challenges both for business and public authorities. On the one hand, companies have suffered from movement restrictions and lockdowns. On the other hand, governments around the globe were forced to find innovative solutions to better track and contain the spread of the virus.

Among the several alternatives presented, the contact tracing app was perhaps the most promising and controversial. Designed to be a tool to help policymakers to understand the behaviour of the new infection and thus enable them to allocate resources to the most exposed areas, the app was received with scepticism by the population. Despite the promise to use the technology for the common good, the public authorities have failed to convince the mass population to adopt the app and disclose information. It did not take long for the big tech companies from Silicon Valley (i.e., Google and Apple) to offer similar solutions.

Although it has not been massively adopted in any country, the contact-tracing app has been more successful in some areas than others. In fact, the level of engagement differed significantly in the democratic nations - ranging from moderate adoption rates to total disbelief. (European Observatory on Health Systems and Policies, 2020; Lee & Lee, 2020)

Inspired by the difference in engagement across countries, this study aimed to investigate how the fear of AI, especially the apprehension of algorithmic bias, affected the decision to disclose information. In particular, this study aimed to understand whether there is an interaction of this fear with the organisation holding the data and the country' income level.

The findings show that indeed the perception that the algorithms might lead to inaccurate decisions affects an individual's willingness to disclose information. As expected, the willingness to disclose falls as fear of algorithmic bias grows. This finding supports previous literature on people’s mental models on algorithms, automation and intelligent technologies (Lee and Baykal, 2017; Lee, 2018) by providing a deeper understanding of how people's perceptions of algorithmic bias affect their decisions. Although the goodness-of-fit measure is small (R^2=0.90), it does not represent a problem per se. First, because it is expected that studies that try to explain human behaviour have model-fit values less than 50%. Second, this field of study seems to have a great amount of unexplained variance. Thus, it is still possible to draw conclusions about the relationship between the model and the variables even though its strength is not robust.

On the other hand, the study did not find statistical support for the hypothesized negative relationship between AI literacy and fear of algorithm bias on willingness to disclose information. Contrary to predictions, the effect is not significant. There could be many possible explanations to that. Firstly, there is, at the point of this study being carried out, still no well-established scale to measure AI literacy (Long & Magerko, 2020). This study made an attempt to do so by adapting items from the Privacy Literacy scale (Westin, 2003) and using various competencies that define individuals' AI literacy (Long & Magerko, 2020; Computer Science Teachers Association, 2017). Although the scale has presented reliability considered acceptable (Cortina, 1993), this may have affected the result of the measurement.

Aside from that, a possible explanation is that both AI literate and non-AI literate people may fear algorithmic bias for different reasons. Being AI literate, the former group understands that the algorithm can be biased by the people coding them and thus fear algorithmic bias. The latter group, the non-literate one, because they do not know how the technology works and, therefore, fear AI because they fear new technologies (Brosnan, 1998).

The study also found that the organisation holding the data, whether the government or a private company, is not significant. Therefore, the hypothesis that the nature of the organisation works as a moderator between fear of algorithmic bias and willingness to disclose information did not find statistical support. Participants in both groups of countries do not seem to differentiate between governments and private companies holding the data.

However, further analysis showed that the two groups of countries indeed differ concerning Trust in Government. The respondents living in the low-income country conferred significantly lower trust in the public authority than people living in high-income countries. It supports previous findings in the literature that suggests causal mechanisms through which economic conditions may influence the expectations of cooperative behaviour and trust (Jordahl, 2007; Zak & Knack, 2001).

Further, a moderation analysis with PROCESS found that Trust in Government also moderates the relationship between fear of algorithmic bias and Intention to adopt the App. People have higher chances to adopt the app when they trust the government because this trust reduces the effect of fear of algorithmic bias on adoption intention. This falls in line with previous literature that shows that trust in the counterpart is crucial on behaviour change (Cook, Levi & Hardin, 2009; Farrel, 2009; Cook, Levi Hardin, 2005; Hardin, 2002).

Thus, more than a private company, the government seems to be crucial to define the level of fear of algorithmic bias. In two of the four scenarios of the 2x2 between-subjects design, the organisation holding the data was the government and in the other two scenarios, it was a private company. However, in all four scenarios the government was the one responsible for implementing the new AI system. Therefore, a possible explanation for the organisation holding the data being not significant is that people tend to care more about the organisation leading the implementation of the AI system than the organisation holding the data. Hence, when trust in the government is low, people do not disclose information even when a private company is responsible for holding the data. This might relate to the fact that in undeveloped countries, corruption cases involving public and private companies are more common (Olken & Pande, 2012). In this scenario, private companies would not represent a safeguard against corrupt governments.

Another study finding is that the interaction of the organisation holding the data and the country’s income level on the relationship between fear of algorithmic bias and the willingness to disclose information was not significant. The hypothesis of moderated moderation relationship was not supported by statistical evidence. The finding is reasonable given that the previous hypothesis was also not supported. This also might relate to the fact that participants appear to not differentiate between the organisation holding the data.

Another interesting finding is related to Privacy Concern that mediates the relationship between fear of algorithmic bias and Willingness to disclose information. A mediation analysis with PROCESS showed that respondents’ privacy concerns score higher when fear of algorithmic bias is presented, and this higher privacy concern relates to lower information disclosure. Similarly, a follow-up model using Adoption intention as dependent variable revealed that privacy concern as a mediator is also significant. This finding goes in line with Privacy Concern literature. For example, Bansal, Zahedi, and Gefen (2007) show that individuals' intention to disclose health information depends on their trust, privacy concern, health status, and risk beliefs. Similarly, White (2004) found that privacy concerns mediate disclosure behaviour and that the perceived benefit also influences self-disclosure.

The study also found that the three variables related to perceived risk (i.e. fear of algorithmic bias, privacy concern and fear of Covid-19) are significant predictors of willingness to disclose information and intention to adopt the app. However, whereas the fear of algorithmic bias and privacy concern has a negative effect on both information disclosure and adoption intention, fear of Covid-19 has a positive effect. This might be explained by the fact that people tend to disclose information when they feel they have no choice (Rinik, 2019). Thus, even though people might fear algorithmic bias, the Covid-19 perceived risk might represent a strong external factor that gives participants the need to disclose their data and avoid the infection spread.

In fact, the observed fear of Covid-19 in the low-income country was significantly higher than in the high-income countries. This might reflect the different scenarios in Brazil and Europe regarding Covid-19 at the point in time this study was conducted. While Brazil registered an average of 2,5 thousand deaths per day, in Europe the pandemic seemed to be receding (Johns Hopkins Coronavirus Resource Center, 2021). When breaking down the study results into the two groups of counties, the analysis shows that fear of Covid-19 has a strong positive effect on information disclosure in the low-income country (β =.092, $p < 0.006$) whereas in high-income countries this variable is not significant (β =.042, $p < .125$).

7 Limitations and Future Research

The current work presents initial evidence for the premise that the human fear of algorithmic bias has significant effects on disclosure behaviour. However, the study has limitations that future work should address.

One limitation of this research is that the human fear of algorithmic bias might be influenced by any number of contextual factors not included in the current investigation. From past traumatic experiences to demographics, many circumstances could affect the perception that the Artificial Intelligence systems are biased and, thus, influence the decision to disclose private information. Future studies can build on the current work by exploring these antecedents to understand when or why they elicit the human fear of algorithmic bias.

In addition, the study used a survey experiment based on hypothetical situations. Even though this scenario-based approach is commonly used in social psychology to study perceptions of decisions (Petrinovich et al., 1993), it requires the findings to be complemented with other studies that involve peoples' actual experiences. This is a crucial step in order to develop the study field on how people perceive algorithmic bias.

Considering the study sample demographics, the majority of participants are under 35 years of age, which means they may be more open to technological change and algorithmic decision-makers (Berinsky et al., 2012; Paolacci and Chandler, 2014). Therefore, an avenue for future research could be to conduct the study with other populations, especially older adults, to understand the influence of age on the perception of algorithmic bias.

Another suggestion for future research is to investigate how the actual outcomes of algorithmic decisions affect people's perceptions. For example, the current study investigated the individual's fear of algorithmic bias when participants are presented with two different threat levels; respondents did not directly evaluate whether the algorithm is biased or not biased from the assessment provided by the technology. Thus, whether the individual's perception of algorithmic bias changes when the outcome is beneficial to them is yet unknown.

Future work should investigate whether the fear of algorithmic bias is affected when the algorithmic assessment is delivered through human-like interactions (e.g., interactive chat) or not. Several studies (Shank, 2012) suggest that people perceive behaviours of computer agents differently than humans agents depending on the context.

Lastly, this study has made an attempt to measure AI literacy with a newly established scale by adapting items from the Privacy Literacy scale (Westin, 2003) and using various competencies that define individuals' AI literacy (Long & Magerko, 2020; Computer Science Teachers Association, 2017). Although this scale has presented reliability considered acceptable (Cortina, 1993), it should be further developed and refined in future studies.

8 Contributions

This is one of the first studies to explore the fear of algorithmic bias on information disclosure and to examine how different descriptions of algorithms - a neutral one and a biased one - may influence people's perception of AI technology.

The work contributes to emerging studies investigating the folk perception of algorithms (Ytre-Arne & Moe, 2020; Bucher, 2016) and social studies on Artificial Intelligence and Machine Learning. In particular, it joins the ongoing work that examines how the potential to affect oneself and others can influence people's reliance on algorithmic devices (Logg, 2017; Lee, 2018). Overall, the findings of this study suggest that the human fear that algorithms make biased decisions influence people's judgments. It affects not only their willingness to disclose information but also their intention to adopt new AI tools. In addition, this investigation highlights the significant relationship between fear of algorithmic bias and trust in the government and privacy concern.

Additionally, the current study explores a construct not much investigated in previous work on social understandings of algorithmic systems: AI literacy. To comprehend the relationship between the individual's literacy in AI and their perceptions of algorithms is a field still to be explored.

Finally, the research also offers implications for practice. The findings suggest that the general public does not fully understand, and trust algorithms and it influences disclosure behaviour. This perceived risk has many undesirable consequences for policymakers (Coravos et al., 2019; Corbett-Davies et al., 2017) and companies.

The most evident is perhaps the rising mistrust in AI technology. This issue becomes even more relevant in the General Data Protection Regulation (GDPR) context when customers have more power over their data. Users now have much more autonomy to decide what information they want to share – and with whom (Stefanouli & Economou, 2019). However, in order to function, Smart cities require a large amount of diverse and accurate data (Stefanouli, M.& Economou, 2019; Brauneis & Goodman, 2017) and this latent fear of algorithmic bias potentially decreases the diversity of data available below the required minimum. Thus, by shedding light on this problem, this thesis helps policymakers address this issue by implementing measures to ensure the algorithm's proper functioning. This investigation also highlights the need for improving governance over data input in large databases to assure that it is broad, diverse, accurate, and does not treat certain groups of people unfairly.

This thesis also has contributions for companies which work with AI and take advantage of it as point-of-differentiation from the competition. Understanding that the fear of algorithm bias affects information disclosure shows that businesses need to invest in campaigns to educate customers and reduce their distrust in technology.

9 Conclusion

Written by Philip Dick and first released in 1956, Minority Report is a classic in thriller literature. Over 112 pages, the author describes a promising future where police utilise cutting-edge technology to arrest and convict murderers before they commit their crime. Based on the machine's predictions - apparently objective and infallible - the government provides citizens with a safe city in exchange for mass surveillance. The unexpected twist happens when the machine – influenced by malicious individuals – makes a wrong prediction putting an innocent person under arrest.

Almost 70 years later, this is no longer science fiction. Today, not only the Justice system uses Artificial Intelligence to predict user behaviour, but also the Health system, the Social Security system, the Banking system, among others (Klein, 2020). Just to name a few applications, algorithms have been used to sort resumes for job applications (Tilmes, 2020), allocate social services, assess and allocate insurance and benefits, and decide who sees advertisements for open positions, housing, and products (Caplan et al., 2018). Both the public and private sectors strongly rely on AI's information to make decisions (Rovatsos, Mittelstadt & Koene, 2019). And just like in Philip Dick's novel, it has also generated inaccurate results and led to incorrect conclusions.

Even though there are many recent records of this so-called algorithmic bias in large Artificial Intelligence databases (Heaven, 2020; Lufkin, 2019; Buolamwini & Gebru, 2018) to date, people's fear of algorithmic bias remains poorly understood. The objective of this thesis was, therefore, to contribute to understanding the effect of this fear on information disclosure.

Inspired by the context of the contact-tracing app, this thesis hypothesized that the fear of algorithmic bias negatively affects information disclosure. The study also argued that variables such as AI literacy, organisation holding the data and country income-level would interact and affect willingness to disclose private data. These are the main significant findings of this study:

- Fear of algorithmic bias decreases individuals' willingness to disclose private information.

- Fear of algorithmic bias decreases willingness to disclose information indirectly by increasing individuals' privacy concern. There is also a direct negative effect of algorithmic bias on information disclosure.

- Similarly, the fear of algorithmic bias decreases App's adoption intention indirectly by increasing individuals' privacy concern. There is also a direct negative effect of algorithmic bias on intention to adopt the app.

- Trust in Government moderates the relationship between fear of algorithmic bias and intention to adopt the app. Thus, the higher is the individual's trust in the government, the less effective is the fear of algorithmic bias on information disclosure.

The folk perception of algorithms is a field of study that has been gaining relevance as the algorithms are increasingly being used to manage interaction among humans. Despite its limitations, this study offers preliminary support to claim that the fear of algorithmic bias affects information disclosure. It represents a small but still relevant step towards understanding how people's perception of AI impacts their behaviour and shapes human interaction with algorithms.

10 References

Ahorsu, D. K., Lin, C. Y., Imani, V., Saffari, M., Griffiths, M. D., & Pakpour, A. H. (2020). The Fear of COVID-19 Scale: Development and Initial Validation. International Journal of Mental Health and Addiction, 1–9.

Algan, Y. & Cahuc, P. (2010). Inherited trust and growth. *American Economic Review,100*(5), 2060-2092

Angwin, J., Larson, J., Mattu, S. & Kirchner L. (2016, May 23); Machine Bias. Pro Publica. Retrieved from: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Babones, S. (2020, May 12). Countries Rolling Out Coronavirus Tracking Apps Show WhyThey Can't Work. *Foreign Policy*. Retrieved from https://foreignpolicy.com/2020/05/12/coronavirus-tracking-tracing-apps-cant-work-south-korea-singapore-australia/

Bansal, G.; Zahedi, F., & Gefen, D. (2007). The Impact of Personal Dispositions on Privacy and Trust in Disclosing Health Information Online" (2007). *AMCIS 2007.Proceedings*. 57.

Bawden, David. (2008). Origins and concepts of digital literacy. Digital literacies: Concepts, policies and practices 30: 17–32.

Berinsky, AJ, Huber, GA, Lenz, GS (2012) Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. Political Analysis 20(3): 351–368.

Bhattacherjee, A. (2001). Understanding Information Systems Continuance: An Expectation-Confirmation Model. *MIS Quarterly, 25*(3), 351-370. doi:10.2307/3250921

Bjørnskov, C. (2018). Social Trust and Economic Growth Forthcoming. In *The Oxford Handbook of Social and Political Trust*. Eric M. Uslaner, ed.

Brauneis, R.& Goodman, E. (2017). Algorithmic Transparency for the Smart City. SSRN Electronic Journal. 10.2139/ssrn.3012499.

British Science Association and others. (2016). One in three believe that the rise of artificial intelligence is a threat to humanity. Retrieved from: https://www.britishscienceassociation.org/news/rise-of-artificial-intelligence-is-a-threat-to-humanity

Brosnan, M. J. (1998). Technophobia : The psychological impact of information technology. ProQuest Ebook Central https://ebookcentral-proquest-com.ezproxy.library.bi.no

Bucher, T. (2017) The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms, Information, Communication & Society, 20:1, 30-44, DOI: 10.1080/1369118X.2016.1154086

Buolamwini, J. & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pages 77-91, 2018.

Caplan R, Donovan J, Hanson L, Matthews J. Algorithmic accountability: A primer. Data and Society. April 4, 2018. Available at: https://datasociety.net/output/algorithmicaccountability-a-primer/. Accessed March 7, 2020

Cellan-Jones, R. (2020, May 15). Tech Tent: Can we learn about coronavirus-tracing from South Korea? *BBC*. Retrieved from https://www.bbc.com/news/technology-52681464

Computer Science Teachers Association (CSTA). 2017. CSTA K12 Computer Science Standards, Revised 2017. Retrieved from http://www.csteachers.org/standards

Cook, K. S., Hardin, R. and Levi, M. (2005). Cooperation without Trust? New York: Rus-sell Sage Foundation

Cook, K. S., Levi, M. and Hardin, R. (2009). Whom Can We Trust? How Groups, Networks, and Institutions Make Trust Possible. New York: Russell Sage Foundation.

Coppock, A (2017) Generalizing from survey experiments conducted on Mechanical Turk: A replication approach. Political Science Research and Methods. Epub ahead of print 27 March 2018. Available at: https://doi.org/10.1017/psrm.2018.10.

Coppock, A., & McClellan, O. A. (2019). Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. Research & Politics.

Coravos A, Chen I, Gordhandas A, Dora Stern A. (2019, February, 14) We should treat algorithms like prescription drugs. Quartz. Retrieved from:

https://qz.com/1540594/treatingalgorithms-like-prescription-drugs-could-reduce-ai-bias/. Accessed March 12, 2021.

Corbett-Davies, S. & Pierson, E. & Feller, A. & Goel, S. & Huq, A. (2017). Algorithmic decision making and the cost of fairness.

Cortina, Jose M. "What Is Coefficient Alpha?" *Journal of Applied Psychology* 78, no. 1 (1993): 98-104.

Criddle, C. & Kelion, L. (2020, May 7). Coronavirus contact-tracing: World split between two types of app. *BBC*. Retrieved from https://www.bbc.com/news/technology-52355028

Dinev, T., Hart, P., & Mullen, M. R. (2008). Internet privacy concerns and beliefs about government surveillance – An empirical investigation. *The Journal of Strategic Information Systems*, *17*(3), 214-233.

Edelman. (2021). Edelman Trust Barometer 2021. Retrieved from: https://www.edelman.com/trust/2021-trust-barometer

Ejelöv, E., & Luke, T. (2019). Rarely safe to assume: Evaluating the use and interpretation of manipulation checks in experimental social psychology. Journal of Experimental Social Psychology. 87.

Eslami, M., Rickman, A., Vaccaro, K., & Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., & Sandvig, C. (2015). " I always assumed that I wasn't really that close to [her] “: Reasoning about Invisible Algorithms in News Feeds. 10.1145/2702123.2702556.

Eslami, M., Vaccaro, K., & Lee, M. K., On, A., Gilbert, E., & Karahalios, K. (2019). User Attitudes towards Algorithmic Opacity and Transparency in Online Reviewing Platforms. 1-14. 10.1145/3290605.3300724.

European Observatory on Health Systems and Policies. (2020). Effective contact tracing and the role of apps: lessons from Europe. Eurohealth, 26 (2), 40 - 44. World Health Organisation. Regional Office for Europe. https://apps.who.int/iris/handle/10665/336294

Farrell, H. (2009). “Institutions and Midlevel Explanations of Trust.” In Whom Can WeTrust? How Groups, Networks, and Institutions Make Trust Possible, edited by K. SCook, M. Levi, and R. Hardin, pp. 127–48. Russell Sage Foundation Series onTrust, NY: Russell Sage Foundation.

Fast, E., & Horvitz, E. (2017). Long-Term Trends in the Public Perception of Artificial Intelligence. *AAAI*.

Gambetta, D. (2000). Trust: Making and Breaking Cooperative Relations. *British Journal of Sociology*, *13*(1).

Gidron, N. & Hall, P. A. (2017). The politics of social status: economic and cultural roots of the populist right. In: British Journal of Sociology, Vol. 68, Issue S1, pp. 57–84

Gold, H. (2020, June 5). Tracking apps were supposed to help beat the pandemic. What happened to them? *CNN Business*. Retrieved from https://edition.cnn.com/2020/06/05/tech/coronavirus-tracking-apps/index.html

Gould, E. D. & Hijzen, A. (2016). Growing Apart, Losing Trust? The Impact of Inequality on Social Capital. IMF Working Paper WP/16/176, IMF, Washington, DC. Retrieved from: https://www.imf.org/external/pubs/ft/wp/2016/wp16176.pdf

Guardian Staff and Agencies (2020, January 18). Coronavirus: what airport measures are in place to detect sick passengers? *The Guardian*. Retrieved from https://www.theguardian.com/world/2020/jan/18/coronavirus-what-airport-measures-are-in-place-to-detect-for-sick-passengers

Guiso, L., Sapienza, P., & Zingales, L. (2006). Does culture affect economic outcomes? Journal of economic perspectives, 20(2), 23-48.

Hall, W., & Pesenti, J. (2017). Growing the artificial intelligence industry in the UK. *Gov.uk.* Retrieved from https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk

Hamari, Juho, and Koivisto, Jonna. "Why Do People Use Gamification Services?" *International Journal of Information Management* 35, no. 4 (2015): 419-31.

Hardin, R. 2002. Trust and Trustworthiness. New York: Russell Sage Foundation.

Hayes, A.F. (2017). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. 2 nd edition.

Hoffman, D., Novak, T., Peralta, M. (1999). Information privacy in the marketplace: Implications for the commercial uses of anonymity on the web. Information Society, 15(2), 129-139.

United Nations.(2020). Human Development Reports. Retrieved from: http://hdr.undp.org/en/content/inequality-adjusted-human-development-index-ihdi

Ipsos. (2017). Revolution@Work: Fears and Expectations. Ipsos. Retrieved from https://www.ipsos.com/en/revolutionwork-fears-and-expectations

Jian, J.; Bisantz, A., & Drury, C. (2000). Foundations for an Empirically Determined Scale of Trust in Automated Systems. *International Journal of Cognitive Ergonomics*. 4. 53-71.

Johns Hopkins Coronavirus Resource Center. (2021, June, 25). Johns Hopkins University & Medicine. Retrieved from: https://coronavirus.jhu.edu/map.html

Jordahl, H. (2007). Inequality and Trust. *Institute of Industrial Economics*. Retrieved from: https://www.ifn.se/Wfiles/wp/wp715.pdf

Kasulis, K. (2020, April 9). S Korea's smartphone apps tracking coronavirus won't stop buzzing. Al Jazeera English. Retrieved from https://www.aljazeera.com/news/2020/04/korea-smartphone-apps-tracking-coronavirus-won-stop-buzzing-200408074008185.html

Kasperson, R. E., Golding, D., & Tuler, S. (1992). Social distrust as a factor in siting hazardous facilities and communicating risks. *Journal of Social Issues, 48*(4), 161–187.

Kay, M.; Matuszek, C. & Munson, S. (2015). Unequal Representation and Gender Stereotypes in *Image Search Results for Occupations*. 10.1145/2702123.2702520.

Kenny, G., & Connolly, R. (2016). Drivers of Health Information Privacy Concern: A Comparison Study. Retrieved from https://aisel.aisnet.org/amcis2016/Health/Presentations/3/

Klein, A. (2020, July 10). Reducing bias in AI-based financial services. Brookings Report. Retrieved from: ttps://www.brookings.edu/research/reducing-bias-in-ai-based-financial-services/

Koller, M. (1988). Risk as a Determinant of Trust, Basic and Applied Social Psychology, 9(4) 265–276.

Kujala, P., Kallio, J., & Niemelä, M. (2019). Income Inequality, Poverty, and Fear of Crime in Europe. Cross-Cultural Research, 53(2), 163–185.

Kukar-Kinney, M. and Close, A.G. (2010). The determinants of consumers online shopping cart abandonment. Academy of Marketing Science Journal 38 (2), 240–250.

LaVeist T.A., Isaac L.A., Williams K.P. (2009). Mistrust of health care organisations is associated with underutilization of health services. *Health Services Research*, 44:2093–2105.

Lazare, A. (1987). Shame and Humiliation in the Medical Encounter. *Arch Intern Med. 147*(9), 1653-1658.

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

Lee, M.K. and Baykal, S. (2017). Algorithmic mediation in group decisions: Fairness perceptions of algorithmically mediated vs. discussion-based social division. In: *Proceedings of the ACM conference on computer supported cooperative work and social computing*, Portland, USA, 25 February–1 March 2017, pp. 1035–1048.

Lee, T., & Lee, H. (2020). Tracing surveillance and auto-regulation in Singapore: ‘smart’ responses to COVID-19. Media International Australia, 177(1), 47–60. https://doi.org/10.1177/1329878X20949545

Lo, J. (2010). Privacy Concern, Locus of Control, and Salience in a Trust-Risk Model of Information Disclosure on Social Networking Sites. *American Conference of Information Systems Proceedings*, 1-12.

Logg, J. (2017) Theory of Machine: When Do People Rely on Algorithms? Harvard Business School NOM Unit Working Paper No. 17-086. Retrieved from: https://ssrn.com/abstract=2941774

Long, D., & Magerko, B. (2020). What is AI Literacy? Competencies and Design Considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*.

Lum, K. & Isaac, W. (2016). To predict and serve? *Royal Statistical Society*. Retrieved from: https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x

Lyons, B. & Dolezal, L. (2017). Shame, stigma and medicine. *Medical Humanities, 43*(4), 208-210.

Maddux, J., & Rogers, R. (1983). Protection Motivation and Self-Efficacy: A Revised Theory of Fear Appeals and Attitude Change. Journal of Experimental Social Psychology. 19. 469-479.

Maina, I., Belton, T., Ginzberg, S., Singh, A. & Johnson, T. (2017). A decade of studying implicit racial/ethnic bias in healthcare providers using the implicit association test. *Social Science & Medicine*. 199. 10.1016/j.socscimed.2017.05.009.

Malhotra, N. K., Kim, S. S. and Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research*,15(4):336–355.

Marwick, A., & Hargittai, E. (2017). Nothing to hide, nothing to lose? Incentives and disincentives to sharing information with institutions online. *Information, Communication & Society, 22*(12), 1697-1713.

Mason, O. J., Stevenson, C., & Freedman, F. (2014). Ever-present threats from information

Mcknight, D., Choudhury, V., & Kacmar, C. (2002). Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information Systems Research*.13. 334-359

McLain, D. L., Hackman, Z. K. (1999). Trust, risk, and decision-making in organisational change. *Public Administration Quarterly*. 23(2) 152–176

Mourey, J. A. & Waldman, A.E. (2020). Past the Privacy Paradox: The Importance of Privacy Changes as a Function of Control and Complexity, *Journal of the Association for Consumer Research*, 5, issue 2, p. 162 – 180.

Mozilla. (2019). We Asked People Around the World How They Feel About Artificial Intelligence. Here's What We Learned. Mozilla Foundation. Retrieved from: https://foundation.mozilla.org/en/blog/we-asked-people-around-the-world-how-they-feel-about-artificial-intelligence-heres-what-we-learned/

Obermeyer, Z.; Powers, B.; Vogeli, C. & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science. 366. 447-453.

Oblinger, Diana. (2005). Educating the Net Generation.

OECD (2015). *How's Life? 2015: Measuring Well-being*. OECD Publishing: Paris. Retrieved from: https://www.oecd-ilibrary.org/economics/how-s-life-2015_how_life-2015-en

Oh, S. Kim, K., Kim, M., Oh, J. & Chu, S. & Choi, J.. (2021). Measurement of Digital Literacy Among Older Adults: Systematic Review. Journal of Medical Internet Research. 23. 10.2196/26145.

Olken, B. A., & Pande, R. (2012). Corruption in Developing Countries. Annual Review of Economics, 4(1), 479–509. https://doi.org/10.1146/annurev-economics-080511-110917

Oppenheimer, D. M., Meyvis, T., and Davidenko, N. (2009). Instructional manipulation checks: detecting satisficing to increase statistical power. J. Exp. Soc. Psychol. 45, 867–872.

O'Neill, P. H. (2020, May 20). Apple and Google's covid-tracing tech has been released to 23 countries. *MIT Technology Review*. Retrieved from https://www.technologyreview.com/2020/05/20/1002001/apple-and-googles-covid-tracing-tech-has-been-released-to-22-countries/

O'Neill, P. H., Ryan-Mosley, T., & Johnson, B. (2020, May 7). A flood of coronavirus apps are tracking us. Now it's time to keep track of them. *MIT Technology Review*. Retrieved from https://www.technologyreview.com/2020/05/07/1000961/launching-mittr-covid-tracing-tracker/

Ortiz-Ospina, E. & Roser, M. (2016). Trust. *Our World in Data*. Retrieved from: https://ourworldindata.org/trust

Ostherr, K., Borodina, S., Bracken, R. C., Lotterman, C., Storer, E., & Williams, B. (2017). Trust and privacy in the context of user-generated health data. Big Data & Society. https://doi.org/10.1177/2053951717704673

Paolacci, G, Chandler, J (2014) Inside the Turk: Understanding mechanical Turk as a participant pool. Current Directions in Psychological Science 23(3): 184–188.

Park, Y. J. (2008). Privacy regime, culture, and user practices in the cyber marketplaces. Info, 10(2), 52-74.

Park, Y. J. (2013). Digital Literacy and Privacy Behaviour Online. Communication Research, 40(2), 215–236. https://doi.org/10.1177/0093650211418338

Paul, K. (2020, May 20). Apple and Google release phone technology to notify users of coronavirus exposure. *The Guardian.* Retrieved from https://www.theguardian.com/technology/2020/may/20/apple-google-phone-app- trace-coronavirus

Pavlou, P. (2003). Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. *International Journal of Electronic Commerce*, *7*(3), 101-134.

Petrinovich, L, O'Neill, P, Jorgensen, M (1993) An empirical study of moral intuitions: Toward an evolutionary ethics. Journal of Personality and Social Psychology 64(3): 467–478.

Platt, J., Jacobson, P., & Kardia, S. (2017). Public Trust in Health Information Sharing: A Measure of System Trust. *Health services research*. 53. 10.1111/1475-6773.12654.

Platt, J., & Kardia, S. (2015). Public trust in health information sharing: implications for biobanking and electronic health record systems. *Journal of personalized medicine*, 5(1), 3–21.

Poznyak, D., Meuleman, B., Abts, K. *et al.* Trust in American Government: Longitudinal Measurement Equivalence in the ANES, 1964–2008. *Soc Indic Res* 118, 741–758 (2014). https://doi-org.ezproxy.library.bi.no/10.1007/s11205-013-0441-5

Quilty-Harper, C. & Liverpool, L. (2020, June 11). Covid-19 death rates twice as high in England's most deprived areas. *New Scientist*. Retrieved from https://www.newscientist.com/article/2237475-covid-19-death-rates-twice-as-high-in- englands-most-deprived-areas/

Rayner, K., Schotter, E. R., Masson, M. E. J., Potter, M. C., & Treiman, R. (2016). So much to Read, So Little Time: How Do We Read, and Can Speed Reading Help? Psychological Science in the Public Interest, 17(1), 4–34.

Rinik, C. (2019). Data trusts: more data than trust? The perspective of the data subject in the face of a growing problem. *International Review of Law, Computers, & Technology*. 34. 1-22.

Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist, 26*(5), 443–452.

Rovatsos, M., Mittelstadt, B., & Koene, A. (2019). Landscape Summary: Bias in Algorithmic Decision-Making: What is bias in algorithmic decision-making, how can we identify it, and how can we mitigate it?. UK Government. Retrieved from: http://Inhttps://www.gov.uk/government/publications/landscape-summaries-commissioned-by-the-centre-for-data-ethics-and-innovation

Ruel, E., Wagner III, W. & Gillespie, B. (2016). Pretesting and pilot testing. In The practice of survey research (pp. 101-119). SAGE Publications, Inc, https://www.doi.org/10.4135/9781483391700

Shank, D.B. (2012). Perceived justice and reactions to coercive computers. *Sociological Forum* 27(2): 372–391.

Sheehan, K. B., Hoy, M. G. (1998). Flaming, complaining, abstaining: How online users respond to privacy concerns. Journal of Advertising, 28(3), 37-51.

Smallbone, D. & Lyon, F. (2002). A note on trust, networks, social capital and entrepreneurial behaviour. In F. Welter & H. Hans-Hermann (Eds.), *Entrepreneurial strategies and trust: structure and evolution of entrepreneurial behavioural patterns in East and West European environments*. Bremen: Forschungsstelle Osteuropa.

Smith H.J.; Milberg S.J., & Burke S.J. (1996). Information privacy: measuring individuals' concerns about organisational practices. MIS Quarterly 20 (2), 167– 196.

Smith H.J., Dinev, T. and Xu, H. (2011). Information privacy research: an interdisciplinary review. MIS Quarterly 35 (4), 992–1015.

Staa, T.V., Goldacre, B., Buchan, I., & Smeeth, L. (2016). Big health data: the need to earn public trust. *British Medical Journal, 354*.

Stefanouli, M.& Economou, C. (2019). Data Protection in Smart Cities: Application of the EU GDPR: Proceedings of 4th Conference on Sustainable Urban Mobility (CSUM2018), 24 - 25 May, Skiathos Island, Greece. 10.1007/978-3-030-02305-8_90.

Swan, S. & Andrews, B. (2010). The relationship between shame, eating disorders and disclosure in treatment. *British Journal of Clinical Psychology, 42*(4).

Tagiarolli, G. (2021, April 14) App do SUS que monitora avanço da covid fracassa por falta de uso. Uol Noticias. Retrieved from: https://www.uol.com.br/tilt/noticias/redacao/2021/04/14/falta-de-politica-nacional-faz-app-do-sus-flopar-no-rastreamento-de-contato.htm

Tilmes, Nicholas. (2020). The Limits of Fair Machine Learning: An Analysis of Structural Ableism in Algorithmic Hiring. 10.13140/RG.2.2.11718.50247.

Thorslund, C. (1976). Interpersonal trust: A review and examination of the concept. Goteborg Psych. Rep. 6(6) 1–21.

Welter, F. & Nadezhda, A. (2011). Researching trust in different cultures. Handbook of Research Methods on Trust. Edward Elgar, p. 50-60.

Westin, Alan. (2003). Social and Political Dimensions of Privacy. Journal of Social Issues. 59. 431 - 453. 10.1111/1540-4560.00072.

White, T. (2004). Consumer Disclosure and Disclosure Avoidance: A Motivational Framework. Journal of Consumer Psychology, 14(1-2), 41-51.

White, T. (2005). Consumer Trust and Advice Acceptance: The Moderating Roles of Benevolence, Expertise, and Negative Emotions. *Journal of Consumer Psychology*, 15. 141-148.

World Bank. (2019). GNI per capita, Atlas method. Retrieved from: https://data.worldbank.org/indicator/NY.GNP.PCAP.CD

Ytre-Arne, B., & Moe, H. (2020). Folk theories of algorithms: Understanding digital irritation. *Media, Culture & Society*. https://doi.org/10.1177/0163443720972314

Zak, P. J. & Knack, S. (2001), ‘Trust and growth’, Economic Journal, 111 (470), 295-321.

11 APPENDIX

APPENDIX 1. NEWS ARTICLES USED AS MANIPULATIONS IN THE STUDY

All messages followed the standard format for a fear appeal in that they consisted of a particular threat component (Maddux & Rogers, 1983). The four manipulations are similar in terms of format, typeface, size, and colour of the letter, only varying concerning (1) the one responsible for holding the data (Government vs Private company) and (2) the strength of threat of Algorithmic bias (High Threat/Low threat).

The message lengths are also similar (varying from 251 to 264 words). The text length allowed participants assigned in any of the conditions to read the text in about one minute (Rayner et al., 2016). Table 1 presents the exact wording for the Government holding the data conditions: (1) Low threat of AI bias (2) High threat of Algorithmic bias.

Table 1.1 Government holding the data conditions

Low Threat	High Threat
New AI system to track coronavirus cases With health authorities warning of a possible third wave of Covid-19, your country is now considering launching a new AI system to track infectious disease spreading more effectively in the future and avoid new lockdowns. Together with the government's technology department, the Health Ministry will develop the new system, which citizens will be able to use through a new app. The AI system will collect data on the users' infection, location, and movement. The government will be responsible for storing and analyzing the data. The data will be automatically deleted every 30 days. The adoption of the new mobile contact-tracing app would be voluntary. The app will work in the following way. Once an infected person is identified, the app will track other users who have been in the proximity of the infected patient and alert them. Moreover, the app will automatically notify all close contacts of the infected person. Through the AI system, the government will be able to better identify areas with a high spread of the disease and implement targeted health policies. However, the government indicates that the collected data will not be used to enforce policy adherence to quarantine for those with confirmed infections. According to Andrea Larsen, an expert from the Artificial Intelligence Institute, AI models are neutral tools that allow the authorities to collect and analyze data more efficiently. Therefore, AI models can lead policymakers to make appropriate decisions based on data analysis.	**New AI system to track coronavirus cases** With health authorities warning of a possible third wave of Covid-19, your country is now considering launching a new AI system to track infectious disease spreading more effectively in the future and avoid new lockdowns. Together with the government's technology department, the Health Ministry will develop the new system, which citizens will be able to use through a new app. The AI system will collect data on the users' infection, location, and movement. The government will be responsible for storing and analyzing the data. The data will be automatically deleted every 30 days. The adoption of the new mobile contact-tracing app would be voluntary. The app will work in the following way. Once an infected person is identified, the app will track other users who have been in the proximity of the infected patient and alert them. Moreover, the app will automatically notify all close contacts of the infected person. Through the AI system, the government will be able to better identify areas with a high spread of the disease and implement targeted health policies. However, the government indicates that the collected data will not be used to enforce policy adherence to quarantine for those with confirmed infections. According to Andrea Larsen, an expert from the Artificial Intelligence Institute, AI models can produce systematic errors that create unfair outcomes against some groups of individuals over others. Therefore, AI models can lead policymakers to make inaccurate decisions based on biased analysis.

Table 2 presents the exact wording for the Private company holding the data conditions: (1) Low threat of Algorithmic bias (2) High threat of Algorithmic bias.

Table 1.2. Government holding the data conditions.

Low Threat	High Threat
New AI system to track coronavirus cases With health authorities warning of a possible third wave of Covid-19, your country is now considering launching a new AI system to track infectious disease spreading more effectively in the future and avoid new lockdowns. The new AI system has been developed by a big high-tech company that offered to collaborate with the government to create the solution, which citizens will be able to use through a new app. The AI system will collect data on the users' infection, location, and movement, and the company will be responsible for storing and analyzing the data. The data will be automatically deleted every 30 days. The adoption of the new mobile contact-tracing app would be voluntary. The app will work in the following way. Once an infected person is identified, the app will track other users who have been in the proximity of the infected patient and alert them. Moreover, the app will automatically notify all close contacts of the infected person. Through the AI system, the company would help the government to better identify areas with a high spread of the disease and implement targeted health policies. However, the government indicates that the collected data will not be used to enforce policy adherence to quarantine for those with confirmed infections. According to Andrea Larsen, an expert from the Artificial Intelligence Institute, AI models are neutral tools that allow the authorities to collect and analyze data more efficiently. Therefore, AI models can lead policymakers to make appropriate decisions based on data analysis.	**New AI system to track coronavirus cases** With health authorities warning of a possible third wave of Covid-19, your country is now considering launching a new AI system to track infectious disease spreading more effectively in the future and avoid new lockdowns. The new AI system has been developed by a big high-tech company that offered to collaborate with the government to create the solution, which citizens will be able to use through a new app. The AI system will collect data on the users' infection, location, and movement, and the company will be responsible for storing and analyzing the data. The data will be automatically deleted every 30 days. The adoption of the new mobile contact-tracing app would be voluntary. The app will work in the following way. Once an infected person is identified, the app will track other users who have been in the proximity of the infected patient and alert them. Moreover, the app will automatically notify all close contacts of the infected person. Through the AI system, the company would help the government to better identify areas with a high spread of the disease and implement targeted health policies. However, the government indicates that the collected data will not be used to enforce policy adherence to quarantine for those with confirmed infections. According to Andrea Larsen, an expert from the Artificial Intelligence Institute, AI models can produce systematic errors that create unfair outcomes against some groups of individuals over others. Therefore, AI models can lead policymakers to make inaccurate decisions based on biased analysis.

Table 1.3 summarises the arguments used as low-threat appeals and high-threat appeals in each condition. The strength of threat manipulation appeared at the bottom of each message.

Table 1.3. Summary of the threat portions of the messages.

Low Threat	According to Andrea Larsen, an expert from the Artificial Intelligence Institute, AI models are neutral tools that allow the authorities to collect and analyze data more efficiently. Therefore, AI models can lead policymakers to make appropriate decisions based on data analysis.
High Threat	According to Andrea Larsen, an expert from the Artificial Intelligence Institute, AI models can produce systematic errors that create unfair outcomes against some groups of individuals over others. Therefore, AI models can lead policymakers to make inaccurate decisions based on biased analysis.

Table 1.4 summarises the arguments used as the authority holding the data appeals in each condition.

Table 1.4. Summary of the authority holding the data portions of the messages.

Government	Together with the government's technology department, the Health Ministry will develop the new system, which citizens will be able to use through a new app. The AI system will collect data on the users' infection, location, and movement. The government will be responsible for storing and analyzing the data.
Private company	The new AI system has been developed by a big high-tech company that offered to collaborate with the government to create the solution, which citizens will be able to use through a new app. The AI system will collect data on the users' infection, location, and movement, and the company will be responsible for storing and analyzing the data.

APPENDIX 2. MEASUREMENT SCALES USED FOR LATENT VARIABLES

Table 1 Measurement scales used for latent variables.

Variable	Code	Measurement
		All items measured on a 7 Likert scale
Willingness to disclose information	WDI	*To what extent would you share your personal health information with the AI system presented in the scenario if a new wave of pandemic breaks out?*
Fear of Technology bias	FTB1	*I am afraid that the AI system may reproduce bias against me via the people building them or through the data used to train them.*
	FTB2	*There would be a high potential for discrimination associated with AI system taking decisions based on population data.*
	FTB3	*The AI system may lead the authorities to make wrong decisions.*
	FTB4	*AI bias may harm specific groups of people.*
	FTB5	*I am willing to take additional actions to avoid being tracked by AI systems.*
AI literacy	AL1	*Artificial Intelligence systems are good at recognizing patterns in large amounts of data.*
	AL2	*Computers perceive the world using sensors.*
	AL3	*AI algorithms maintain models/representations of the world and use them for reasoning.*
	AL4	*Computers can learn from data.*
	AL5	*Making AI interact with humans is a substantial challenge for AI developers.*
	AL6	*AI applications can impact society in both positive and negative ways.*
Adoption Intention	AI1	*I would support the implementation of this AI solution, if such a system is launched in the future.*
	AI2	*I would oppose the introduction of such an AI system if my government decided to implement it. (r)*
	AI3	*I believe my participation in the new app could be advantageous to me and other people.*
	AI4	*I believe that downloading the new app could be mutually helpful to myself and other people.*

Privacy concern	PC1	*I am concerned that the information collected in this context could be misused by public authorities.*
	PC2	*I am concerned about allowing the collection of information in this context because the data could be used in a way I did not foresee.*
	PC3	*I am concerned about submitting information in this context, because of what others might do with it (e.g. if the data is hacked by third-parties).*
Trust in Government	TG1	*I trust the Government will do what is right for me.*
	TG2	*I can trust the government of the country where I live.*
Trust in AI	TAI1	*How much do you trust that the AI system in the scenario can make good-quality decisions?*
	TAI2	*To what extent do you feel the AI implementations will ultimately increase your well-being?*
Perceived control over data	COD1	*It is very important to me that I control how my personal data will be used and stored.*
Perceived Need for Governance Surveillance	PNS1	*The government needs to have greater access to personal information.*
	PNS2	*The government needs to have greater access to individual bank accounts.*
	PNS3	*The government needs broader surveillance authority.*
	PNS4	*The government needs to have more authority to use high tech surveillance.*
Fear of Covid-19	FCO1	*I am most afraid of coronavirus-19.*
	FCO2	*It makes me uncomfortable to think about coronavirus-19.*
	FCO3	*When watching news and stories about coronavirus-19 on social media, I become nervous or anxious.*

APPENDIX 3. MANIPULATION AND ATTENTION CHECKS

Table 1. Manipulation and attention checks.

Construct	Code	Measurement
Manipulation check 1 Governement vs private company condition	MC1	In the scenario you read, who was responsible for collecting, storing, and analyzing the data? A big tech company The government No data was collected
Manipulation check 2 Low Threat vs High Threat	MC2	In the article you read, what was the opinion of the expert in AI technology? AI might produce accurate results AI might produced biased results There was no expert in AI technology
Attention check 1	AC1	Please click on the item Disagree. This is an attention check item.

APPENDIX 4. DEMOGRAPHICS MEASURES

Table 1. Demographic measures.

Variable	Code	Measurement	Scale
Gender	GE	Which gender do you identify yourself the most?	Male Female Other
Age	AGE	Which age group describes you?	18–24 25–34 35–44 45–54 55–64
Education	EDU	What is the highest level of school you have completed?	Some school, no degree High school graduate Some college, no degree Bachelor's degree Master's degree Professional degree Doctorate degree
Income	INC	Which of these describes your income last year?	0-500 501-1000 1001-1500 1501-2000 2001-2500 2501-3000 3001-4000 4001-5000 5001-6000 6001-8000 8001-10.000 10.001 or greater
Smartphone usage experience	SUE	Which of these describes your experience using smartphones?	Less than a year; 1–less than 2 years; 2–less than 3 years; 3–less than 4 years; 4–less than 5 years; 5–less than 6 years; More than 6 years.

APPENDIX 5. OUTPUT FROM THE PROCESSPROCEDURE IN SPSS

```
Run MATRIX procedure:

**************** PROCESS Procedure for SPSS Version 3.5.3
****************

          Written by Andrew F. Hayes, Ph.D.       www.afhayes.com
    Documentation available in Hayes (2018). www.guilford.com/p/hayes3

**************************************************************************
*
Model  : 4
    Y  : Willing
    X  : biasfear
    M  : PrivConc

Sample
Size:  788

**************************************************************************
*
OUTCOME VARIABLE:
 PrivConc

Model Summary
          R       R-sq        MSE          F        df1        df2
p
      .5457      .2978     1.4561   333.3725     1.0000
786.0000      .0000

Model
              coeff         se          t          p       LLCI
ULCI
constant     2.1259      .1848    11.5011      .0000     1.7630
2.4887
biasfear      .6850      .0375
18.2585      .0000      .6114      .7587

**************************************************************************
*
OUTCOME VARIABLE:
 Willing

Model Summary
          R       R-sq        MSE          F        df1        df2
p
      .4044      .1636      .6330    76.7565     2.0000
785.0000      .0000

Model
              coeff         se          t          p       LLCI
ULCI
constant     4.2304      .1317    32.1158      .0000     3.9718
4.4890
biasfear     -.1939      .0295    -6.5702      .0000     -.2519
-.1360
PrivConc     -.1227      .0235    -5.2169      .0000     -.1688
-.0765

************************** TOTAL EFFECT MODEL
****************************
OUTCOME VARIABLE:
 Willing

Model Summary
```

```
          R       R-sq        MSE          F        df1        df2
p
      .3668      .1346      .6541   122.2206     1.0000
786.0000      .0000

Model
              coeff         se          t          p       LLCI
ULCI
constant     3.9696      .1239    32.0430      .0000     3.7264
4.2128
biasfear     -.2780      .0251   -11.0553      .0000     -.3273
-.2286

************** TOTAL, DIRECT, AND INDIRECT EFFECTS OF X ON Y
**************

Total effect of X on Y
     Effect         se          t          p       LLCI       ULCI
c_ps       c_cs
     -.2780      .0251   -11.0553      .0000     -.3273     -.2286
-.3200     -.3668

Direct effect of X on Y
     Effect         se          t          p       LLCI       ULCI
c'_ps      c'_cs
     -.1939      .0295    -6.5702      .0000     -.2519     -.1360
-.2232     -.2559

Indirect effect(s) of X on Y:
             Effect     BootSE   BootLLCI   BootULCI
PrivConc     -.0840      .0166     -.1172     -.0516

Partially standardized indirect effect(s) of X on Y:
             Effect     BootSE   BootLLCI   BootULCI
PrivConc     -.0967      .0190     -.1345     -.0595

Completely standardized indirect effect(s) of X on Y:
             Effect     BootSE   BootLLCI   BootULCI
PrivConc     -.1109      .0221     -.1551     -.0677

*********************** ANALYSIS NOTES AND ERRORS
************************

Level of confidence for all confidence intervals in output:
  95.0000

Number of bootstrap samples for percentile bootstrap confidence
intervals:
  5000

------ END MATRIX -----
```