IUL School of Social Sciences

Department of Social and Organizational Psychology

Artificial Intelligence and Human Values

Daniela Gaspar Couto

Dissertation submitted as partial requirement for the conferral of Master in Social and Organizational Psychology.

Supervisor:

Doutor Nelson Campos Ramalho

Assistant Professor, ISCTE-IUL

September, 2018

Acknowledgements

The accomplishment of my thesis means the end of one big important stage of my life, my master's degree in Psychology. For these five years of my life, I need to thank my parents and my brother, because they are, without a doubt, my biggest pillar of support. They always believe in me and in my capabilities to become a great professional. They are my everything and they will continue to be in my next life chapters.

In all these years, I've always had my friends with me, in good and bad moments. I have also made new friends, with whom I shared the bitterness of university and all the incredible things it gave us. To all my friends, thank you very much.

Equally, I need to thank all of my thesis participants, for their availability and for not giving up, even when I annoy them to not forget.

Last but not least, I am very grateful to all the support of professor Nelson Ramalho. He is undoubtedly an inspiring person for all the contagious energy and has always had an ambitious thinking throughout my thesis. Thank you for all the patience and for sharing so much knowledge with me. It was an honour to have you as my supervisor, thank you.

Resumo

A inteligência artificial (IA) é cada vez mais utilizada na vida quotidiana, em que onde as decisões e escolhas eram deixadas à gestão humana, a tecnologia assume agora um papel mais incisivo nessa questão. Esta temática tem sido motivo para várias opiniões divergentes e alarmísticas (e.g. Elon Musk e Stephen Hawkings), devido às várias suscetibilidades éticas que o desenvolvimento da IA abarca.

O presente estudo procura percecionar se os valores humanos influenciam as atitudes que os indivíduos têm para com a evolução da IA. Esta tecnologia é exposta em vários contextos: no caso da IA ameaçar cada um dos valores; no caso oposto, de a mesma os beneficiar; na necessidade ou não da presença de agentes reguladores, e se os mesmos de alguma forma alteravam a sua decisão inicial.

Com uma amostra de 205 participantes, através de uma metodologia quantitativa (questionário) e qualitativa (entrevistas semiestruturadas), concluiu-se que a igualdade, a liberdade, a saúde e a segurança nacional constituem um poder preditivo relativamente às atitudes que os indivíduos têm face à evolução da IA. Mais especificamente, se a IA ameaçar a igualdade surgem atitudes desfavoráveis à sua evolução, igualmente, no caso da liberdade as pessoas também são contra à evolução da IA, mesmo que ela beneficie ou ameace os valores humanos. A saúde se for beneficiada pela IA, a pessoas tendem a ser a favor da sua evolução, mas sempre com a presença de agentes reguladores. Por fim, se a IA beneficiar a segurança nacional, surgem atitudes positivas face à sua evolução, bem como, se a IA ameaçar o mesmo valor as pessoas continuam a demonstrar atitudes positivas, mas exigem a presença de agentes reguladores.

Palavras-chave: Inteligência Artificial; Valores Humanos; Atitudes; Agentes reguladores

Abstract

Artificial Intelligence (AI) is increasingly used in daily life. Where once decisions and choices were left to human management, technology now plays a much more incisive role. This topic has spawned several diverging and alarming opinions (e.g. Elon Musk and Stephen Hawkings), due to the various ethical susceptibilities that AI development spans.

The present study attempts to perceive whether human values influence individuals' attitudes towards AI evolution. This technology is exposed in several different contexts: in the chance that AI threatens each one of the values; the opposite case, where AI is beneficial; in the need (or lack thereof) for the presence of regulatory agents, and whether that changed people's initial decision.

With a sample of 205 participants, and through quantitative methodology (questionnaire), as well as qualitative (semi-structured interviews), the conclusion is that equality, freedom, health and national security constitute a predictive power when it comes to the attitudes that individuals nurture towards AI evolution. More specifically, in the event that AI threatens equality, people develop unfavourable attitudes towards its evolution. The same happens for freedom, where people are also against AI evolution, whether it benefits or threatens human values. People tend to be in favour of AI evolution if it benefits health, but require the presence of regulatory agents. Lastly, the attitudes towards AI evolution are positive if it benefits national security. People still demonstrate generally positive attitudes in the event that this value is threatened by AI, but require the presence of regulatory agents.

Key-words: Artificial Intelligence; Human Values; Attitudes; Regulatory Agents

Table of Contents

List of Tables

Introduction

Artificial Intelligence (AI) is a very antique notion originating since the dawn of history (McCorduck, 2004). However, it has only now been a subject of much discussion. It is applied in several fields, in diverse ways and there is an increasing interaction between human and AI in our day-to-day lives. Opinions on the benefits and harms on this matter are controversial, having also been propelled by the exposure of the opinions of several known figures of technological or similar areas, such as Stephen Hawking and Elon Musk.

Much of the efforts are directed today towards building more complex and accurate algorithms. However, there are dimensions that will be taken in the future that do not depend on the technical quality of an algorithm but instead of the subjective reading that people make of that algorithm. So, technical quality is not a sufficient condition for any algorithm to be taken as effective. The psychosocial dimension must be equated as it is relevant to understand expectations about the impact of AI is in people's lives. It is in this context that our object of study emerges.

Given the controversy of opinions established today, it is necessary to perceive the attitudes ordinary individuals have about the existing AI and its evolutions. Based on the attitudes, values are presented as guiding factors for a final decision, so it is pertinent to go a little further. In other words, to understand to which extent decisions and attitudes towards AI have a value-base rationale. Perceptions of threat or benefit towards human values may trigger decisions that one must understand.

The structure of this thesis will seek firstly to contextualize the AI topic, namely to define it and to describe its applications today. Also, since this technology envisages to operate autonomously, we will discuss whether it is pertinent to give it free arbitration or not. As previously mentioned, in order to explore the perceptions of individuals it is necessary to understand the terms of human values and everything it encompasses. In a technological era, privacy, as a value, has been gaining momentum. Additionally, if human values are the basis of attitudes and if one is interested to study the attitudes of individuals, a contextualization must also be given about this subject, as well as about the association of these two concepts and their interaction with AI. Due to the scarce literature on the topic, this section ends with the propositions to be addressed throughout the study.

The following section will show the methodological options concerning data analysis strategy, sampling, measures and procedure. Findings are showed and discussed them against the set of propositions. The thesis concludes and recognizes limitations as well as offers suggestions for future studies.

Chapter 1 – Literature Review

1.1. Artificial Intelligence

John McCarthy coined the concept of artificial intelligence (AI) in a workshop in Dartmouth, in 1956 (Copeland, 2015), where there were many people who were actively interested in this field, including some figures who are considered as the fathers of AI, such as Herbert Simon, McCarthy, Marvin Minsky, and Allen Newell (Gips, 1979). Its existence, albeit not names as AI, is much older though (McCorduck, 2004).

AI is characterised as a ramification of computer science, which uses heuristic techniques and symbolic processing, that is, specific procedures in programming languages to develop programs for performing tasks that function intelligently (Atkinson, 2016; Brent, 1988; Fox, 2016; Gips, 1979; Hillman, 1985). This has been referred to as the field that makes technology do things that require the same intelligence as mankind (Brent, 1988; Ginsberg, 2012; Gips, 1979; Hillman, 1985).

AI is not considered as a substantially defined intellectual field (Hillman, 1985) but it is generally characterised as a field of research that develops intelligent agents using deep and machine learning techniques to facilitate the interactions between human and machine/technology (Atkinson, 2016; Lisetti, & Schiano, 2000). AI is currently considered to be poor because it has been designed to perform specific tasks; however, due to its rapid development, it aims at performing every cognitive task (Lu, Li, Chen, Kim, & Serikawa, 2018).

1.1.1. AI Application

AI manages complex, data-intensive tasks (Dignum, 2017) that can be represented in many ways (Nilsson, 2014). We will highlight some main subfields of AI.

One relates to the expert systems (Brent, 1988; Hillman, 1985; Nilsson, 2014), that is, programs that carry out tasks within a certain limited field of expertise, such as games, medical diagnoses, treatment prescriptions, computer system configurations, computer visions, discovery of evidence regarding mathematical theorems, and oil prospection (Gips, 1979; Hillman, 1985; Lu et. al, 2018).

We can emphasise a few recent examples in this subfield, such as the scandal associated with Cambridge Analytica and the social network Facebook, where the company misused data collected

between 2014 and 2015 by an app called 'thisisyourdigitallife' from thousands of users of the social network Facebook. The use of these data and the determination of individual characteristics will have enabled the definition of voter profiles that could have been affected by specific messages later on, thus influencing the electorate. This scandal was associated with Donald Trump's 2016 US election campaign and the decision to implement Brexit. Hence, it is possible to define behavioural targeting through social networks. Given their current choices, these systems propose contents that a person will be likely to enjoy consuming, allowing for consumers to effortlessly discover their contents of interest.

Another example is related to health, the Alerte Digital Health's electrocardiogram platform using AI combines 'a non-invasive wearable electrocardiogram acquisition device, continuous wireless electrocardiogram data streaming to a mobile device, and real-time interpretation of the electrocardiogram data' (Playford, Jais, Weerasooriya, Martyn, Bollam, Turewicz, & Mohamad, 2017: 379). Pattern recognition for automated cardiac arrhythmia monitoring is possible through this program.

Another AI subfield is natural-language understanding (Brent, 1988; Hillman, 1985; Nilsson, 2014), which relates to programs that understand, read, and analyse natural language input. These are examples of its application: machine translation, which automatically translates texts from one language to another; speech understanding; holding intelligible conversations, answering questions and summarising stories (Gips, 1979; Hillman, 1985).

Finally, we highlight the subfield of robotics (Hillman, 1985; Nilsson, 2014): 'Software for automata capable of recognising physical objects and adapting devices to deal with them, for example, intelligent programs for linking perception and robot actions' (Hillman, 1985: 22). A practical example of this area is the 'killer robots', which are defined as machines that aim at replacing soldiers in battlefields with the purpose of serving a higher military and/or political goal (Coeckelbergh, 2011). Likewise, robots have been built for other non-military purposes such as education (Roll, & Wylie, 2016), agriculture (Bac et al., 2014), healthcare (Robinson, MacDonald & Broadbent, 2014; Prabhu & Urban, 2017; Huang, & Rust, 2018), recreation, work (Ramamoorthy & Yampolskiy, 2018), rescue (De Cubber et al., 2017), underwater operations (Bogue, 2015), milking cows (Unal & Kuraloglu, 2015) or even accompanying people walking side-by-side (Ferrer, Zulueta & Cotarelo, 2017) among many other (Fox, 2016). Robots represent an added value to humans in the sense that it can perform dangerous, hard, or boring work. They

can also bring more comfort to our daily lives, entertain us, and help us to cope with disasters, and save lives (Dignum, 2017). Globally, AI can be embodied as robots, in the form of transportation, service, military, and medical robots (Dignum, 2017).

AI is becoming quickly present in almost all dimensions of everyday life and its capabilities are expanding extremely fast, which is characterised by their interactivity, autonomy, and adaptability (Dignum, 2017; Fox 2016; Hillman, 1985). Nevertheless, current expectations and perceptions concerning the capabilities of AI are diverse, so there is not a consensus on the societal impact of AI (Dignum, 2017) especially as regards ensuring control on potentially life threatening robots (Verdiesen, 2017).

1.1.2. AI Regulations

AI systems are currently under the responsibility and control of their users or owners. However, there are breath-taking advances being made in AI development through its autonomy and machine learning focusing on the creation of systems that can act and make decisions without human control, which is continuously developing (Dignum, 2017; Helbing, Frey, Gigerenzer, Hafen, Hagner, Hofstetter, Hoven, Zicari, & Zwitter, 2017). In this sense, AI could affect the lives of people in greater or smaller ways in various areas of application (Makridakis, 2017) where the results of their actions and decisions will not be permanently possible to predict or direct.

There are many controversial arguments regarding the consequences of AI (Boyd, & Holton, 2017), since there could be extremely powerful machines, which would benefit humans, or they could end up being dangerous (Armstrong, Bostrom, & Shulman, 2016; Fox, 2016). However, potentially rapid advances in AI have led to signs of alarm, including governmental requests to restrict AI operations and regulations regarding AI development with law makers searching for best options yet to be found (Scherer, 2015). Relevant icons of science and technology like Elon Musk and Stephen Hawking also expressed concerns regarding the long-term future of AI (Babcock, Kramár, & Yampolskiy, 2017).

'*I think we should be very careful about artificial intelligence. If I had to guess at what our biggest existential threat is, it's probably that . . . I'm increasingly inclined to think there should be some regulatory oversight, maybe at the national and international level, just to make sure that we don't do something very foolish.*'

(Elon Musk in an interview at MIT's 2014 AeroAstro Centennial Symposium)

According to Ayoub and Payne (2016), readers might believe the threat to be imaginary, or at least more distant, than the technology icons, perhaps (by our logic reflection) due to the lack of knowledge about it. The authors (Ayoub, & Payne, 2016), consider that if the evolution of AI will make it capable of independently interpreting and setting and attaining its goals, then there will be a high level of uncertainty leading to implications for humankind (Scherer, 2016).

For this reason, through the eventual need for careful control (Armstrong, Bostrom, & Shulman, 2016) a concept of regulation arises to ensure the beneficial, safe, and fair use of AI, as well as the perception that humans remain in control (Dignum, 2017; Reed, 2018). Fundamentally, the regulation must focus on looking up to solve the controlling undesirable risks and define responsibilities, and it also does not prevent worthwhile advances in technology (Dickow, & Jacob, 2018; Reed, 2018). The regulatory agencies need to foresee the required changes and recognise that some of these risks will not be known yet; hence they need to convene AI experts (Bundy, 2017; Reed, 2018)

Thus, the following questions arise: Why are there not technological markets to regulate the production of innovation? Is it necessary to have regulators managing the evolution of AI? Do individuals feel the need for agent regulators in the evolution of technology?

1.2. Human Values

The concept of human values has been an important factor in the exploration of various psychological, social, political, and economic phenomena (Hitlin, 2003; Cheng, & Fleischmann 2010). There can be various definitions for human values, but their nature is similar with a slight difference (Zheng, 2015). Most scientists describe values as deeply rooted abstract motivations with an enormous influence on people's lives at a social and individual level representing a central element that guides, explains, and justifies norms, actions, attitudes, and opinions (Granjo, & Peixoto, 2013; Tuulik, Ounapuu, Kuimet, & Titov, 2016; Zheng, 2015).

This dissertation is founded on Rokeach's theory that defines values as: 'an enduring belief that a specific mode of conduct or end-state of existence is personally or socially preferable to an opposite or converse mode of conduct or end-state of existence' (Rokeach, 1973, p. 5). Rokeach (1973) proposed that values consisted in cognitive representations of three forms of universal

human demands: social interactional demands for interpersonal coordination, biologically based needs of the organism, and social institutional requirements for group wellbeing and subsistence.

Human values are therefore crucial in the organisation of individual belief systems (Pereira, Camino, & Costa, 2005). They occupy a central position in the cognitive part that underlies the organisation of people in societies being among the most important evaluative beliefs (Feather, 1990; Gouveia, Andrade, Milfont, Queiroga, & Santos, 2003; Pereira, Camino, & Costa, 2005, Rokeach, 1973). People can consciously characterise demands as values through cognitive development and, through socialisation, they become aware of the culturally shared terms that allow for them to communicate these values (Rokeach, 1973; Schwartz, & Bilsky 1987).

Human values can be considered as an association between oneself and the individual relationship with society as individuals develop and evolve in a social context (Cheng, & Fleischmann 2010; Rokeach, 1973). They are described as judgmental conceptions providing outcomes in terms of desirable or undesirable and good or bad actions revealing beliefs and standards that characterise people or group values (Al-Kahtani & Allam, 2013; Granjo & Peixoto, 2013; Sagiv, Roccas, Cleciuch, & Schwartz, 2017; Schwartz & Bilsky 1987). For that reason, this psychological construction has a predictive and explanatory potential of behaviour, attitudes, and decision-making at the individual and societal stage of analysis (Al-Kahtani & Allam, 2013; Hitlin, 2003; Rokeach, 1973).

The interest of individuals in the form of a social entity governs values guiding them towards organised actions according to their axiological prioritisation (Al-Kahtani & Allam, 2013; Costa, 2012; Granjo & Peixoto, 2013; Schwartz, 1994). Every individual has different value priorities and an unique value system. Moreover, values can reflect major social changes in societies and across nations (Rokeach, 1973; Tuulik Ounapuu, Kuimet, & Titov, 2016). As they depend on culture and human conditions, they may be altered in reaction to socio-cultural changes and extreme events (Prati, Pietrantoni & Albanesi, 2005).

Finally, Rokeach (1973) states that when a value is internalised, it consciously or unconsciously transforms into standards or criteria that will guide actions in order to develop and maintain attitudes towards relevant objects and situations, to morally judge oneself and others and to compare themselves to others. This helps to describe and explain the similarities and differences between people, groups, nations, and cultures.

1.2.1. The Organisation of Human Values

Attitude and behaviour are not usually guided according to an absolute value but they are rather set by priority and antagonistic values. In this sense, the organisation of personal values refers to the structure and association of values in a system of values (Costa, 2012; Pereira, Camino, & Costa, 2005; Rokeach, 1973; Schwartz, 1994).

Based on these assumptions, Rokeach created in 1973 the Rokeach Value Survey, which contains a list of 36 values. Respondents are expected to rank them as guiding principles in their lives. They are branched out as two sets of values, 18 instrumental values and 18 terminal values.

Instrumental values (means) refer to desirable modes of conduct or the means used to achieve individual preferences and can be divided into moral values (with interpersonal focus) or competence/self-actualisation values (particularly centred on the individual). They are:

- Ambitious: Hard-working, aspiring;
- Broadminded: Open-mind;
- Capable: Competent, effective;
- Cheerful: Lighthearted, joyful;
- Clean: Neat, tidy;
- Corageous: Standing up for your beliefs;
- Forgiving: Willing to pardon others;
- Helpful: Working for the welfare of others;
- Honest: Sincere, truthful;
- Imaginative: Daring, creative;
- Independent: Self-reliant, self-sufficient;
- Intellectual: Intelligent, reflective;
- Logical: Consistent, rational;
- Loving: Affectionate, tender;
- Obedient: Dutiful, respectful;
- Polite: Courteous, well-mannered;
- Responsible: Dependable, reliable.

Terminal values (end values) represent the final states of existence and comprehend two subcategories - personal and social, the first referring to the final state of the individual as a whole and the second to the final state at an interpersonal level. They are:

- A comfortable life: A prosperous life;
- A sense of accomplishment: A lasting contribution;
- A peaceful world: Free of war and conflict;
- A world of beauty: Beauty of nature and the arts;
- An exciting life: A stimulating active life;
- Equality: Brotherhood, equal opportunities for all;
- Family security: Taking care of loved ones;
- Freedom: Independence, free choice;
- Happiness: Contentedness;
- Inner harmony: Freedom from inner conflict;
- Mature love: Sexual and spiritual intimacy;
- National security: Protection from attacks;
- Pleasure: An enjoyable life of leisure;
- Salvation: Saved, eternal life;
- Self-respect: Self-esteem;
- Social recognition: Respect, admiration;
- True friendship: Close companionship;
- Wisdom: A mature understanding of life.

However, their order may need to be changed due to changes in culture, society, and the personal experience of individuals (Prati, Pietrantoni, & Albanesi, 2005; Rokeach, 1973). Thus, it is part of the concept of value that there are individual differences concerning its organisation and stability depending on variables such as identification with gender roles, among others (Prati, Pietrantoni, & Albanesi, 2005; Schwartz, 1994). Therefore, Rokeach (1973) has developed a cross-cultural survey following an order that can be reasonably comprehensive and universally applicable allowing to 'compare any country's values with those of another'.

1.2.2. Privacy as an Emergent Value

General Data Protection Regulation (GDPR) was recently approved (April 2016) by the European Parliament, which '*was designed to harmonise data privacy laws across Europe, to protect and empower all EU citizens data privacy and to reshape the way organisations across the region approach data privacy*'.

This Act emerged due to the increasing importance of maintaining privacy nowadays, given the late scandals involving data breaches and misuse of data (i.e. Cambridge Analytica). It seems, therefore, important to consider this concept as a central human value, since we are permanently dealing with technological contexts.

Privacy is considered as access limitation to oneself comprising a means of control or autonomy over significant personal matters (Kokolakis, 2015; Parent, 1983). It is possible to differentiate three aspects of privacy (Holvast, 1993; Kokolakis, 2015): the privacy of a person – individual protecting against undue interference –, territorial privacy – physical area surrounding an individual – and informational privacy – controlling whether or how personal data can be processed, stored, gathered, and disseminated.

Searing (1979: 159) defined privacy as "avoidance of intrusion and publicity" which is easily breached by inconspicuous IT means. The arrival of AI is seen to be a likely factor in the loss of privacy in a way that large amounts of personal data will be used in exchange to services (Cath, Wachter, Mittelstadt, Taddeo, & Floridi, 2018; Jin, 2018). The magnitude of the risk associated with privacy and its potential damage to consumers can be directly or indirectly related to AI and other data technologies (Cath, Wachter, Mittelstadt, Taddeo, & Floridi, 2018; Jin, 2018). An example of this risk is the expected increase in data value by AI, which encourages firms to collect and accumulate data, regardless of its use, which will also become a prime target for scammers and hackers (Jin, 2018).

When asking people's perceptions regarding the use of personal data, a study (Pew Research Centre, 2014) has revealed an extreme concern about the collection and use of their personal information and subsequently their privacy. However, a few studies show that, in spite of these concerns, they might abdicate of some privacy in order to obtain benefits – called a privacy paradox (Brown, 2001; Kokolakis, 2015; Sayre & Horne, 2000). Even though people were afraid that too much personal information was being collected and that their privacy could be infringed,

individuals were still willing to give personal data as long as they obtained some gain in return (Kokolakis, 2015).

In this sense, privacy reveals itself as an important added value on Rokeach's list concerning technological impacts on the daily life of individuals.

1.2.3. Biological Gender Difference in Values

In 1975, Rokeach showed significant differences in terms of biological gender between the terminal and instrumental value rankings. He justified these results through sexual roles, which were socialised in the society of the time and were played by men and women in very different ways. However, Rokeach (1975) highlights the various similarities (albeit fewer than the differences) between both biological genders in supposedly stereotyped values for men and women. Recent papers also found some differences between both genders and the ranking of human values (i.e. Chakraboxty, & Bagchi, 2018; Maslova, 2018; Weber, 2017).

1.3. Attitudes

In the field of Social Psychology, the concept of attitude is defined as the mediator process of the way individuals act and think, which is inferred and not directly observable (Lima, 2010). According to Allport (1935: 801), an attitude is "a mental and neural state of readiness, organised through experience, exerting a directive or dynamic influence upon the individual's response to all objects and situations with which it is related'.

Therefore, the attitude of an individual is not necessarily related to a particular belief, but rather the organisation of several beliefs (Fishbein & Ajzen, 1975; Rokeach, 1973). In other words, the attitude of an individual towards an object, a problem, a type of behaviour, or an event is determined by their salient beliefs when relating the object to various attributes and their assessment of those attributes (Fishbein & Ajzen, 1975).

Conclusively, an attitude is a lasting cognitive organisation where one adopts a favourable or unfavourable position of affection with a specific psychological object (Eagly & Chaiken, 1993; Greenwald, Brock & Ostrom, 1968; Rodrigues, 1978).

1.4. Attitudes and Values

Once the concepts have been defined, we can note that values and attitudes differ in a number of aspects. Whereas a value is a unique belief represented by a standard, an attitude is no longer a standard, since it refers to a set of beliefs that are specific to a particular object or situation (Rokeach, 1973; Sagiv, Roccas, Cleciuch, & Schwartz, 2017). Each individual can have several attitudes whereas the value numbers of one person are more restrict because they are transitional, that is, they apply to different situations (Fonseca, Porto, & Barroso, 2012; Rokeach, 1973).

Within the makeup and cognitive system of one's personality, values assume a more central position than attitudes, since they determine attitudes (Rokeach, 1973). These are therefore interconnected within an individual's cognitive system (Rokeach, 1981), and in the context of attitude analysis, we must consider underlying values not only in an attempt to explain them, but also to predict them (Fishbein & Ajzen, 1975; Schwartz, 2006). In that way and according to Reigh and Adcock (1976, cited by Fonseca, Porto, & Barroso, 2012), we can affirm that there is an intrapersonal coherence between values and attitudes, and value orientations can be good indicators of specific attitudes.

There are different possible processes connecting value priorities with people's attitudes (Sagiv, Roccas, Cleciuch, & Schwartz, 2017). The attention individuals pay to the world around them is affected by values (Schwartz, 2006). In this way, the values that become more important to an individual are activated and awarded being expressed in an action (Sagiv, Roccas, Cleciuch, & Schwartz, 2017; Zheng, 2015). Then values will bring awareness to possible actions, which may release the need and perceptions for making useful attitudes and interpretations of methods and reactions based on values (Zheng, 2015). Defining and reacting to a situation based on the weighting of its pros and cons (Zheng, 2015).

1.5 Attitudes and Values towards AI

There are some elements that are important to understand in human beings. This is because in future decisions relating to any technological evolution or society are the values that people somehow mobilise in order to make a decision. In this sense, it is pertinent to know whether any AI application can threaten human values. However, due to scarce literature on this explicit subject, we have focused on attitudes towards AI, since the values are the basis for attitudes and without them there is no understanding of behaviour or foresight of it.

Nomura (2006) shows that people tend to have extreme attitudes, either positive or negative, towards robots. As was previously explained, robotics is an AI subfield, so we can expect people to have the same tendency towards AI. However, robots are physical; humans can see and interact with them, which is not the case with AI systems, since they are not such an easy perception.

In the same way, Crowed and Friess (2013) came to the same conclusion, that individuals have either a positive extreme attitude or a negative extreme attitude towards AI. Explaining these results with the help of AI could make humans function better (i.e. help in detecting national security threats) or AI could take over a few human functions (i.e. robots replacing human jobs).

Fox (2016) also concludes that technological innovations can be central or a part of people's lives, or wholly rejected. On the other hand, a study (Tussyadiah, Zach & Wang, 2017) of 325 residents in the United States has concluded that individuals show a low level of negative attitudes towards technology, particularly computers and robots.

Due to the few existing studies on AI attitudes, we can conclude that the results are disparate. There is no common ground among the studies perhaps due to the fact that it is only now that the consequences of AI are becoming the subject of widespread talk and controversy. Therefore, we can conclude that attitudes can change in relation to AI, and taking into account the human values underlying them, the following question arises: can values influence people's attitude towards AI?

In this sense, emerge two prepositions:

1) Whenever values are favourably connected to AI, decisions will tend to be more in favour rather than against.

In this context, with and without regulation, individuals will also tend to be in favour of AI innovation.

2) Whenever values are negatively connected to AI, decisions will tend to be more against it than in favour.

In this context, with and without regulation, individuals will tend to be against AI innovation.

Chapter II - Method

2.1. Data Analysis Strategy

As this research deploys a mixed method approach, data analysis will vary according to the nature of the variables under study. For quantitative variables, in the case of one of the variables (human values) we will conduct confirmatory factor analysis to judge on its psychometric quality. Technically we can take as valid any factorial solution that observes the following criteria (Hair et al., 2010): CMIN/DF <3. with a non-significant p-value (although this fit index might be discarded due to sample size bias), CFI>.92, TLI>.92, and RMSEA<.07. Additionally, we will report also PCFI to judge on the factorial solution parsimony (which is taken as better when values are closer to 1). In case fit indices do not achieve acceptability levels we will make use of Lagrange Multipliers to identify possible biases from single items that should be excluded (Hair et al., 2010). As regards the values-statement, because the scale is ipsative (ask to rank order) factorial analysis is not applicable.

For qualitative measures of human values we will proceed differently. They will be quantified into presence vs absence to build a new set of variables, matching the original quantitative measure of human values depicted above.

This mixed method is necessary due to the possible complementary role they play is providing rich and meaningful information.

2.2. Sample

The current study comprehends a sample of 205 individuals, of which 106 are male (51.7%) and 99 are female (48.3%). The ages of the participants range between 18 and 36 years-old, averaging 23.7 years-old. The larger number of individuals' are single, representing 92.2% of the total sample. The professional area is divided into11 categories, as embodied in table 2.1.

Table 2.1 - Professions

	Frequency	Percent	Valid Percent	Cumulative Percent
Human and social sciences	41	20,0	20,0	20,0
Technology	56	27,3	27,3	47,3
Health	19	9,3	9,3	56,6
Management/ accounting	30	14,6	14,6	71,2
Law	4	2,0	2,0	73,2
Arts	10	4,9	4,9	78,0
Marketing/ tourism	14	6,8	6,8	84,9
Operators	12	5,9	5,9	90,7
Unemployed	5	2,4	2,4	93,2
Others (student)	8	3,9	3,9	97,1
Catering	6	2,9	2,9	100,0
Total	205	100,0	100,0	

The profile participants human values priorities we asked them to rank them. The frequency table below shows that the less important value for the majority of the participants are "world at peace", on the another hand the most important value ranked was "health".

Table 2.2 – Human values ranking

Importance level	True Friendship	Self-respect	Equality	Freedom	Privacy	Health	National Security	World at Peace
1 – nothing important	12.2%	6.8%	3.4%	1%	20%	0.5%	27.8%	28.3%
2	9.3%	11.7%	8.3%	7.3%	13.7%	5.4%	30.2%	14.1%
3	10.2%	15.6%	18.5%	7.8%	23.9%	6.3%	11.7%	5.9%
4	15.1%	14.1%	18.5%	10.2%	16.6%	7.3%	10.2%	7.8%

5	11.7%	10.7%	24.4%	16.1%	11.2%	8.8%	6.8%	10.2%
6	18.5%	11.2%	13.2%	22%	8.8%	8.8%	6.8%	10.7%
7	12.7%	18%	10.7%	22.9%	4.4%	16.6%	6.3%	8.3%
8 – very important	10.2%	11.7%	2.9%	12.7%	1.5%	46.3%	0%	14.6%
Total								100%

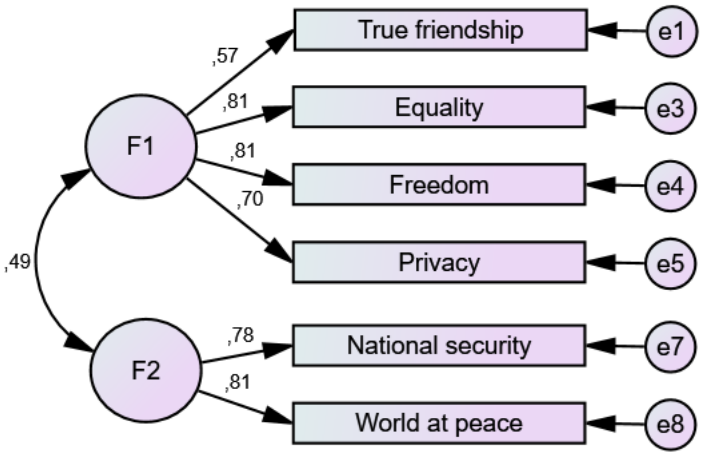
2.3. Measures

2.3.1. Quantitative Values

Personal values were measured with an adaptation from Rokeach's (1973) value survey, namely from the terminal values subscale. Due to its length, we conducted a previous analysis that highlighted seven terminal values to which we added one more (privacy, due to its salience in literature review on AI). The process was conducted by means of surveying 187 individuals, gathered as a convenience sample, and asking to select five values from the original Rokeach's terminal values subscale. This request was made twice: the first one pertained values that were personally relevant, and the second pertained values they thought were relevant for the AI subject. By counting frequencies and ranking values we kept the five top values from each list. We ended up with 10 human values but as two were shared, the final list comprehends eight human values: health, national security, world at peace, self-respect, true friendship, privacy, freedom, and equality. The option for retaining up to 10 items took into consideration the short-term memory retention standard capacity.

This scale was intended to be used twice in the research apparatus. The first measurement asked participants to rank the eight values based on personal importance. As it was treated as an ipsative scale no further psychometric analyses are required. The second measurement asked participants to state in a 6-point Likert scale (1-Not important at all, 6-Very important) their view on each value within the context of AI. An exploratory factor analysis showed a two-factor solution with one item showing insufficient commonality (health). After removal of this item the emerging solution is valid (KMO=.772, X^2 Bartlett = 569.179, 21 df, p=.000) and explains 67.3% after rotation (varimax). An ensuing confirmatory factor analysis show unacceptable fit indices (CMIN/DF=5.494, p<.001, CFI=.887, TLI=.831, PCFI=.591, RMSEA=.148). Judging from

Lagrange multipliers we removed one item (self-respect) which also made sense because it was not ranked at all in the AI-important values list. The resulting solution for the revised two-factor showed good fit indices (CMIN/DF=1.760, p=.07, CFI=.984, TLI=.973, PCFI=.590, RMSEA=.061) and thus comprehended: 1) Social factor (true friendship, freedom, equality, privacy) and 2) Peace factor (national security and world at peace). Both factors have good reliability (Cronbach alpha=.797 and r_{SB}=.777, respectively). The solution has also convergent validity as the AVEs for both factors exceed the cut-off value of .500, with .532 and .632, respectively.



Considering the process, we will use these two factors plus the item "health" as indicators of AI-related values.

2.3.2. Qualitative Values

We made a content analysis ruled by a priori categories of the quantitative values measure, these are: true friendship, equality, freedom, privacy, health, national security and world at peace. The way as the interviews were guided did not guarantee the same attention level for each value, for this reason we set all values as a dichotomy variable – presence or absence of the value because frequency counting could be misleading. For each category it is important to understand if mentions are positive or negative valence, i.e. if AI is taken as being aligned with each value (reinforcing it) or if it is against the nature of each value (threatening it). Exceptionally, the true friendship value was mentioned only in a positive way.

Examples of excerpts per category, per valence follow:

- Equality (in a positive way):

"I think AI can be fundamental, for example, to reach a situation of equality for genders, races, all of that, and something else."(P59)

- Freedom (in a positive way):

"We have the freedom to be able to navigate the world, to be somewhere else in the world without actually being there. It conveys a certain freedom."(P76)

- Privacy (in a positive way):

"Imagine you research for computers and it (AI) *automatically knows that you are interested in computers and will redirect you to contents about that context. (...) In some way that is not bad. The fact that someone knows what you know is helping you to bring contents that can help you with your decisions or in your research."*(P96)

- Health (in a positive way):

"I think it was only going to help a lot because they could find cures for diseases, which have not been discovered yet."(P110)

- National security (in a positive way):

"You can have more security because you will have better and more automated systems; you will not need as much human intervention to protect you and for you to feel safer in society..." (P21)

- World at peace (in a positive way):

"Detecting terrorist cell and making much more incisive attacks, which at the moment are a little arbitrary, make them much more incisive. This can translate into fewer human casualties."(P151)

- True friendship (in a negative way):

"In terms of friendship, maybe, as robots are beginning to appear, there may be the fact that they want… not to replace people but to start producing robots, as it has already appeared on Facebook, so that the female robot can replace women and accompany men."(P36)

- Equality (in a negative way):

"… the difference in society. AI will dramatically increase the gap between the poor and the rich. Drastically. (…) Even regulated, the rich are the ones who have the power, they are the ones who have the technology, and they are the ones who have access to this tool. And this tool gives an advantage like no other tool in society."(P66)

- Freedom (in a negative way):

"Because if you do not have privacy, then you may become more contained, then you may also feel less comfortable and free to do certain things. And if they start to control you, you will also begin to have less freedom."(P138)

- Privacy (in a negative way):

"… to invade a person's privacy. I think this is already happening and I think this is one of the values that we will see more and more invaded by a super intelligent AI that can know in five seconds where we are, as seen on the video. And I think they enter our cell phones, they know our texts, calls, and social networks. They know everything."(P23)

- Health (in a negative way):

"Namely, there are only personal interests and connected people in the health system, for example, pharmacists who misrepresent some health systems so that they can sell medicine A or B. This can be done through AI."(P18)

- National security (in a negative way):

"One problem regards the malfunction of machines or anything like that. I think this is enough to create lack of security" (P31)

- World at peace (in a negative way):

"A peaceful world, maybe. (...) Like I was saying, if we have a much more intelligent super intelligence than us, if it wants to take care of the whole world, we know it will surely endanger our lives and may generate great chaos here." (P23)

As mentioned, people ranked their values by importance level and we consider it is important to take account of its influence in the choice of human values being threatened or reinforced by AI. In this sense, we created new variables by weighting each qualitative originated value (both valences positive and negative) according to its rank in the ipsative scale. In this way, we obtained a weighting of the frequency of the value mentioned qualitatively depending on the priority level that participants raked their values.

2.4. Procedure

The research design comprehended two phases, the first quantitative and the second qualitative. In the first phase, the participants were requested to answer an online survey which started with acknowledging the nature of the study and asking the informed consent. Next, the participants were requested to order from more important to less important, an eight-item human values scale taking into consideration their personal beliefs. This was the first moment that intended to establish the baseline measures.

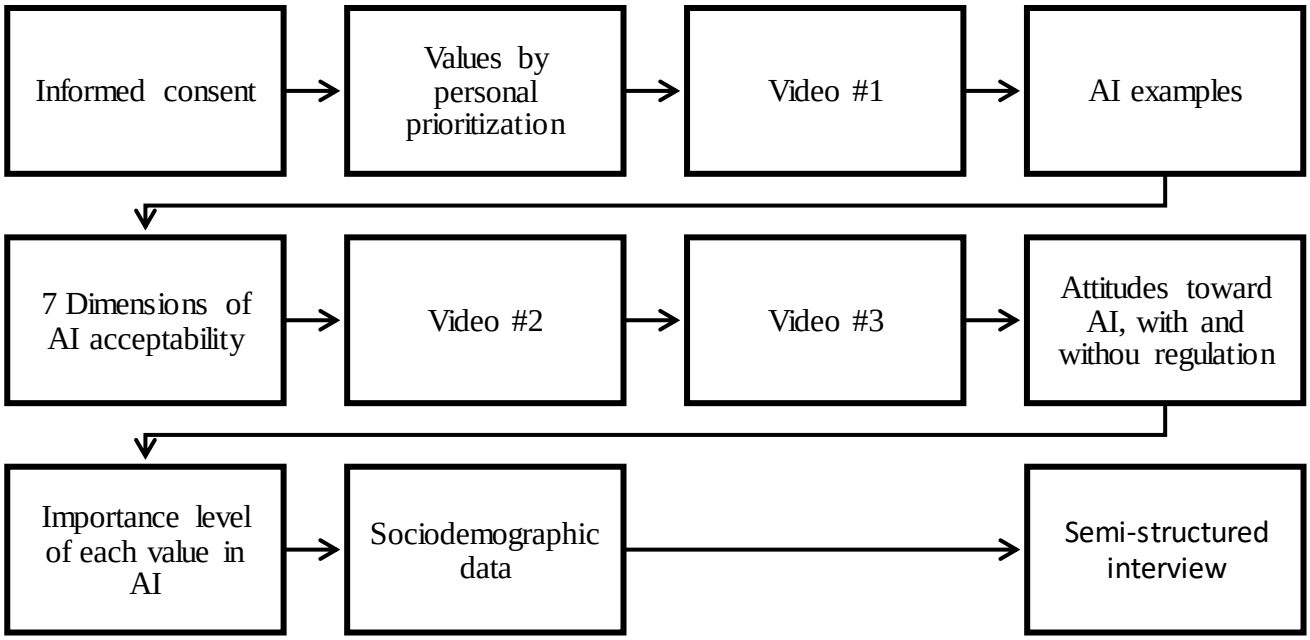
The second moment started with exhibiting video #1 that explains what AI is. After watching the video the participants were challenged to give examples of AI use in their daily life. This request was intended to involve the participant in the topic. After that, participants rated the ascribed degree of AI acceptance on seven dimensions.

The third moment started with video #2 offering examples of nowadays AI applications (e.g. in health, music, transportation etc). Next to this was exhibiting video #3 that showed three testimonies from three renowned individualities with divergent opinions concerning AI. These testimonies were intended to face the participant with a critical positioning towards the future of

AI. The first question after watching the video #3 concerned asking whether each participant has already seen any of the testimonies (this is a control variable). Lastly, we asked each participant to state their degree of agreement (attitude measurement) in a six-point Likert scale (1 Total disagree to 6 Total agree) regarding the evolution of AI, the evolution of AI with regulation, the evolution of AI without regulation.

After participants took a stand we asked them to fill in the human values scale showed in moment 1 (but in a six-point Likert scale instead of the ipsative scale) concerning the reasons that underlie their standpoint. Lastly, participants gave some sociodemographic data for sample characterization.

After completing the online survey the participants were subjected to a semi-structured interview. The questions addressed the following topics: perceive their final position on AI and the reasons behind it; whether any of the human values would be threatened or benefited by the evolution of AI; if there was any moment of the videos that stood out the most; final conclusion toward AI evolution.

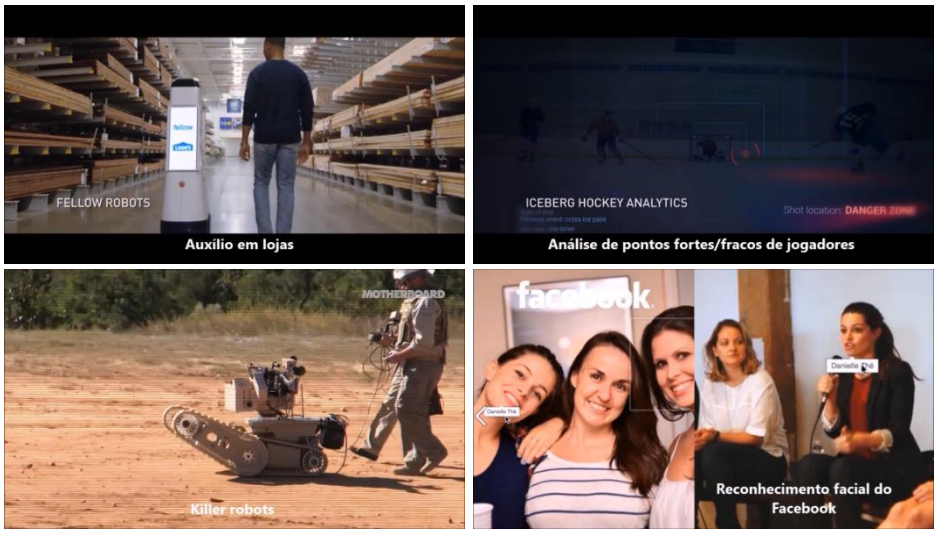


2.4.1. Stimulus

We made all the stimulus videos, with some previous research and taking into account the reliability of the sources.

- **Video #1 -** This video has 1'11'' of duration and aims to explain what is AI.

- **Video #2 –** This one has 1'30'' duration and gives some examples about AI uses until now.



- **Video #3 -** The last video shows three opinions about AI and their evolution with a duration of 6’57’’. Namely, a neutral one stating that it has advantages and disadvantages (Nick Bostrom), another with a caveat against AI without regulation (Elon Musk) and another one with a more hopeful vision of AI (Stephen Hawking).

Chapter III - Results

The results chapter initiates by presenting descriptive and bivariate analyses to provide an overview of the interrelations among variables and their corresponding magnitude in the sample. Then the analyses of relations between the variables under study will be shown.

3.1. Descriptive and Bivariate Statistics

Descriptive statistics report the sample number, minimum and maximum responses registered, average and standard deviation. The variables covered are the socio-demographics (biological gender and age), values weighting, and the three decisions concerning the evolution of AI. The values-variables show different maximum's because they are a weighting between two variables. As explained in the previous section, one variable was measured in a quantitative way versus its corresponding variable but measured by a qualitative analysis (and dummy coded to 0 absent and 1 present). For example, the privacy-positive variable ranges from one to six because none of the participants who mentioned it, rates privacy in more than six on a scale of 1-8. All descriptive and bivariate findings are reported in the table (3.1) below.

Table 3.1 – Descriptive statistics

	N	Min-Max ipsative scale	Min-Max weighted	Mean	Std. Deviation
Biological gender	205	1-2	1-2	-	-
Age	205	18-36	18-36	23.70	3.74
Against/Favor	205	1-6	1-6	4.01	1.14
Against/Favor with regulation	205	1-6	1-6	5.02	1.12
Against/Favor without regulation	205	1-6	1-6	1.79	1.19
Truefriendship_positive	204	1-8	0-0	0	0
Truefriendship_negative	204	1-8	1-8	0.66	1.84
Equality_positive	204	1-8	1-6	0.05	0.50
Equality_negative	204	1-8	1-8	0.82	1.98
Freedom_positive	204	1-8	1-8	0.13	0.86
Freedom_negative	204	1-8	1-8	1.88	2.93
Privacy_positive	204	1-8	1-6	0.10	0.61
Privacy_negative	203	1-8	1-8	2.43	2.21
Health_positive	204	1-8	1-8	3.81	3.55
Health_negative	204	1-8	1-8	0.61	2.04
NationalSecurity_positive	204	1-8	1-7	0.74	1.59
NationalSecurity_negative	204	1-8	1-7	1.58	1.87
WorldPeace_positive	204	1-8	1-8	0.26	1.30
WorldPeace_negative	204	1-8	1-8	1.76	2.62

The three decisions about the evolution of AI – against or favor, with and without regulation – have approximately the same variability on the sample. The decision about AI in against/favor with regulation has the highest average, meaning that the majority of the participants considered the regulation as a differentiator factor. In relation to the human values, the health-positive (being reinforced by AI) characterized the biggest variability between individuals. On the other hand, the true friendship-positive is not referred at all for none of the participants, and the equality-positive

(reinforced by AI), followed immediately by true privacy-positive and freedom-positive, represented the values with the lowest variability among participants.

Table 3.2 – Bivariate Correlation Statistics

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Gender	1																	
2. Age	,027	1																
3. Against/ Favor	-,290**	-,033	1															
4. With regulation	-,105	,007	,573**	1														
5. Without regulation	-,072	-,005	,214**	-,086	1													
6. True friendship positive	.b	.b	.b	.b	.b	.b												
7. True friendship negative	,016	,066	,008	-,009	,045	.b	1											
8. Equality positive	-,095	,060	-,018	-,019	,034	.b	-,035	1										
9. Equality negative	-,025	,048	-,171*	-,109	-,031	.b	,106	,137	1									
10. Freedom positive	,096	-,061	-,177*	,059	-,021	.b	-,053	,326**	,025	1								
11. Freedom negative	,046	,032	-,250**	-,037	-,083	.b	,061	,131	,313**	,031	1							
12. Privacy positive	-,060	,133	,014	,012	,111	.b	-,005	,368**	,043	,163*	,141*	1						
13. Privacy negative	-,042	,007	-,005	-,065	,014	.b	-,055	,087	,044	-,024	,102	-,024	1					
14. Health positive	,080	,049	,037	,146*	,038	.b	,056	-,055	-,046	,056	-,052	-,091	-,006	1				
15. Health negative	,086	,130	-,031	-,098	-,072	.b	,264**	-,029	,260**	-,044	,107	,141*	-,045	,158*	1			
16. National security positive	-,061	,048	,135	,047	,006	.b	-,011	,029	-,098	,079	-,049	,037	,080	,094	-,026	1		
17.National security negative	,067	,023	,017	,098	-,057	.b	,119	,022	,052	,092	,118	,023	,014	,000	,133	,265**	1	
18. world at peace positive	-,092	-,047	,052	,020	,110	.b	-,030	-,020	-,033	-,030	-,045	,041	-,022	,002	-,061	,174*	,056	1
19. World at peace negative	-,001	,007	-,082	,020	-,052	.b	,091	,054	-,061	,201**	,064	,085	-,022	,121	,029	,020	,200**	,217**

* p<0.05; ** p<0.01; b: Cannot be computed because at least one of the variables is constant

As showed in previous table (3.2) the bivariate statistics evidenced some relevant elements. Within sociodemographic variables biological gender and the decision against/ favor toward AI evolution shows a negative significant association. Instead, the age has not a significant relation with any of the variables.

The decision against or in favor of AI evolution presents significant associations with both decisions with regulation. Also, the decision about evolving AI is negatively related with equality-negative, and freedom.

When deciding whether in favor or against AI evolving with regulation, only one correlation was positive significant, the health-positive. Instead of this, when deciding whether in favor or against AI evolving without regulation, there was no significant correlation.

Lastly, there is 14 significant correlations, both positive and negative, between all the 14 human values variables. Equality-positive seems to be the highest correlated value as it shows a correlation of .33 with freedom-positive and .37 with privacy-positive.

3.2. Analysis of the Relations Between Variables Under Study

We used bootstrapping technique in all the analyses reported below. At first, we considered the factorial structure of the human values quantitative measure. So, in order to understand the extension of the association between the two factors (social factor: true friendship, freedom, equality and privacy; peace factor: national security and world at peace) plus the item "health" of the values' second quantitative measurement and attitudes towards AI evolution, with and without regulation, we have conducted a multiple linear regression analysis. However, due to some statistical assumptions not being fully observed, we did not advance with this analysis.

Through the extraction of categories and frequencies from content analysis it is relevant to understand the incidence of positive and negative occurrences, i.e. how frequently did respondents mentioned that AI reinforced some human values versus how frequently they mentioned it threatens those human values. Table 3.3 shows the frequencies per value.

Table 3.3 – Frequencies of value concerns

Human Values	AI reinforces value	AI threatens value	Not mentioned
True Friendship	0%	14.2%	85.8%
Equality	1%	16.2%	82.8%
Freedom	2.5%	31.4%	66.1%
Privacy	2.9%	69.0%	28.1%
Health	57.8%	8.8%	33.4%
National Security	25%	58.3%	16.7%
World at Peace	4.9%	44.1%	51%

An overview clearly shows that value threat is much more present in answers than value reinforcement. The most frequent value threat targets privacy, which it is followed by national security and world at peace. On the other hand, there is a remarkable frequency of health-positive. Also, it is patent that nobody refers that AI reinforces true friendship value, for that reason we will not consider this variable in the next analyses.

Regarding the factorial structure of the quantitative values measure, we categorized the qualitative' values in the same way and made multiple linear regressions too with the three attitudes toward AI. As frequencies of positive (value reinforcement) and negative (value threat) are so distinct, we reason that they should not be conceived in the same set of predictors. Therefore, we conducted the analyses separately.

Table 3.4 – Multiple linear regression with qualitative factors and the three AI decisions

	Qualitative values							
	Positive				**Negative**			
	Social	Peace	Health	R^2	Social	Peace	Health	R^2
Against/ Favor	-0.09	0.15*	-0.02	0.02	-0.24*	0.01	0.05	0.05
With regulation	0.01	0.00	0.15*	0.02	-0.10	0.10	-0.07	0.02
Without regulation	0.06	0.03	0.02	0.01	0.06	-0.17*	-0.08	0.03

Values shown are Beta; *p<0.05

For each decision concerning AI, there are some significant statistic differences between the two qualitative factors (social – true friendship, equality, freedom, and privacy; peace – national security and world at peace) and the health variable. Participants are more in favor of the evolution of AI when it reinforces peace, but if it threatens the social factor they are against AI. Conversely, if AI reinforces health respondents have a positive attitude towards its regulation. Finally, in case the evolution of AI threatens peace the participants take an attitude against AI without regulation.

In the next step, we group weighted qualitative values in the same factors as the previous quantitative values measure. In order to understand the extension of the association between impacted values and attitudes towards AI evolution (with and without regulation), we have conducted a multiple linear regression analysis taking values frequency as predictors and attitude towards AI as the criterion variable.

Table 3.5 – Multiple linear regression with weighted qualitative factors and the three AI decisions

	Weighted qualitative values							
	Positive				**Negative**			
	Social	Peace	Health	R^2	Social	Peace	Health	R^2
Against/ Favor	-0.11	0.13	0.03	0.03	-0.20*	-0.04	0.02	0.04
With regulation	0.04	0.03	0.14	0.02	-0.07	0.08	-0.09	0.02
Without regulation	0.04	0.07	0.04	0.01	-0.01	-0.06	-0.06	0.01

Values shown are Beta; *p<0.05

Only the first decision – against/ favor to AI – have a significant association with the factors, which is: if social factor is threatened by AI, participants are against its AI evolution.

However, because treating values as latent variables may lower explained variance as compared with treating them separately, we opted also to analyze per value the same

possible regressions. This is plausible because of the low explained variance of the models.

Therefore, we have conducted a multiple linear regression analysis taking values frequency as predictors and attitude towards AI as criteria variable.

Table 3.6 – Multiple linear regression with weighted qualitative and the three AI decisions

Human values		**Against/ Favor**	**With regulation**	**Without regulation**
True Friendship	Negative	0.03	0.01	0.08
Equality	Positive	0.04	-0.04	0.01
	Negative	-0.12*	-0.09	0.00
Freedom	Positive	-0.21*	0.06	-0.04
	Negative	-0.21*	0.01	-0.07
Privacy	Positive	0.03	0.03	0.12
	Negative	0.02	-0.07	0.02
Health	Positive	0.04	0.14*	0.05
	Negative	0.01	-0.10	-0.08
National Security	Positive	0.14*	0.03	-0.02
	Negative	0.06	0.11*	-0.04
World at peace	Positive	0.02	0.01	0.11
	Negative	-0.09	-0.01	-0.04
R^2	Positive	0.06	0.03	0.03
	Negative	0.08	0.03	0.02

Values shown are Beta; *p<0.05

The table shows various associations between human values and the decisions toward AI evolution. If AI threaten the equality value the individuals are against his evolution. Relatively to the freedom value, participants take a more against AI stance when AI reinforces and threatens their freedom. The health value only has one significant difference with the decisions, which are, if AI beneficiate the health the individual adopt a pro attitude to AI evolution, but it is needed the regulators. Finally, individual assume a pro attitude to AI when it beneficiate the national security, but if AI threaten the national security value they request the presence of regulators.

In this line of research, human values has been a known predictor for the final decisions to AI so, we would like to investigate if the addition of some demographic characteristics, age and biological gender, could be too an important predictor. Therefore, we made several hierarchical regressions. The first model (Model 1) includes demographic information such as age and biological gender. In the next step (Model 2), contains the previous variables plus the human values.

Table 3.7 – Hierarchical regression with positive weighted qualitative and demographic variables to the three AI decisions

	Against/ Favor		**With regulation**		**Without regulation**	
Predictor	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Biological gender	-0.29*	-0.27*	-0.11	-0.13	-0.07	-0.06
Age	-0.03	-0.05	0.01	0.01	-0.00	-0.02
Human Values						
Equality		0.01		-0.06		0.01
Freedom		-0.18*		0.08		-0.04
Privacy		0.03		0.02		0.12
Health		0.06		0.15*		0.06
National security		0.13*		0.02		-0.02
World at peace		-0.00		0.01		0.10
R^2	0.09	0.13	0.01	0.04	0.01	0.03
R^2 change	0.09	0.05	0.01	0.03	0.01	0.03

Values shown are Beta; *p<0.05

Table 3.8 – Multiple linear regression with negative weighted qualitative and demographic variables to the three AI decisions

	Against/ Favor		**With regulation**		**Without regulation**	
Predictor	**Model 1**	**Model 2**	**Model 1**	**Model 2**	**Model 1**	**Model 2**
Biological gender	-0.30*	-0.30*	-0.12	-0.12	-0.08	-0.06
Age	-0.03	-0.02	0.01	0.02	-0.00	0.00
Human Values						
True Friendship		0.03		0.01		0.08
Equality		-0.14		-0.10		-0.00
Freedom		-0.19*		0.02		-0.06
Privacy		0.01		-0.07		0.02
Health		0.04		-0.09		-0.07
National Security		0.07		0.12		-0.04
World at peace		-0.10		-0.01		-0.05
Adjusted R^2	0.09	0.17	0.01	0.05	0.01	0.02
Adjusted R^2 change	0.09	0.08	0.01	0.04	0.01	0.01

Values shown are Beta; *$p<0.05$

As we can see biological gender it is a good predictor of the AI decision against or in favor. Instead of the age, that there are not significant differences, so this variable does not represent any predictor in this model.

Per this conclusion, we made ANOVA to understand if that are differences between male or female in the decision-making toward the AI evolution.

Table 3.9 – ANOVA between biological gender and the three decisions toward AI evolution

		Sum of Squares	df	Mean Square	F	Sig.
Against/ Favor	Between Groups	22.01	1	22.01	18.60	0.00*
	Within Groups	239.00	202	1.18		
	Total	261.00	203			
With regulation	Between Groups	2,80	1	2.80	2.23	0.14
	Within Groups	253.12	202	1.25		
	Total	255.92	203			
Without regulation	Between Groups	1,47	1	1.47	1.05	0.31
	Within Groups	283.04	202	1.40		
	Total	284.51	203			

*$p<0.001$

From the previous chart, we can determinate that are significant differences between being a male (M=2.66, SD=1.08) or a female (M=3.33, SD=1.11) in relation to make the decision against or in favor of AI evolution.

In the same way, we want to understand the extension of the association between impacted values and biological gender differences, and hence, we conducted ANOVA.

Table 3.10 – ANOVA between biological gender and all weighted human values

		Sum of Squares	df	Mean Square	F	Sig.
True Friendship negative	Between Groups	.14	1	0.14	0.04	0.84
	Within Groups	683.41	201	3.40		
	Total	683.55	202			
Equality positive	Between Groups	0.47	1	0.47	1.85	0.18
	Within Groups	51.04	201	0.25		
	Total	51.51	202			
Equality negative	Between Groups	0.40	1	0.39	0.10	0.75
	Within Groups	794.49	201	3.95		
	Total	794.89	202			
Freedom positive	Between Groups	1.37	1	1.37	1.84	0.18
	Within Groups	149.31	201	0.74		
	Total	150.670	202			
Freedom negative	Between Groups	5.14	1	5.14	0.60	0.44
	Within Groups	1713.72	201	8.53		
	Total	1718.86	202			
Privacy positive	Between Groups	0.28	1	0.28	0.74	0.39
	Within Groups	75.75	201	0.38		
	Total	76.03	202			
Privacy negative	Between Groups	1.75	1	1.75	0.36	0.55
	Within Groups	985.96	201	4.91		
	Total	987.71	202			
Health positive	Between Groups	14.45	1	14.45	1.15	0.29
	Within Groups	2528.52	201	12.58		
	Total	2542.97	202			
Health negative	Between Groups	6.06	1	6.06	1.45	0.23
	Within Groups	840.20	201	4.18		
	Total	846.26	202			
National Security positive	Between Groups	2.03	1	2.03	0.80	0.37
	Within Groups	509.13	201	2.53		
	Total	511.16	202			
National Security negative	Between Groups	2.82	1	2.82	0.81	0.37
	Within Groups	702.42	201	3.50		
	Total	705.24	202			
World at Peace positive	Between Groups	3.00	1	3.00	1.76	0.19
	Within Groups	342.64	201	1.71		
	Total	345.64	202			
World at Peace negative	Between Groups	0.01	1	0.01	0.00	0.96
	Within Groups	1394.64	201	6.94		
	Total	1394.65	202			

An overview clearly indicated that there are not any significant differences between human values, even if they are threatened or reinforced by AI, between both biological genders.

We have conducted a multiple linear regression between human values, positive and negative, and the AI evolution attitudes, with and without regulation, with the sample

segmented by biological gender. We made this to perceive if the human values differ on being a male or a female, depending on their final decision to the AI.

Table 3.11 – Multiple linear regression between positive weighted values and the three decisions divided by biological gender

Biological gender	Predictor	Against/ Favor	With regulation	Without regulation
Male				
	Equality	0.14*	0.18*	-0.10*
	Freedom	-0.33*	-0.31*	0.08
	Privacy	0.17*	0.11*	0.23*
	Health	0.14*	0.21*	0.14
	National Security	0.06	0.02	-0.15*
	World at peace	-0.03	-0.03	0.20*
	R^2	0.06	0.06	0.11
Female				
	Equality	-	-	-
	Freedom	-0.21*	0.14*	-0.08
	Privacy	-0.10	0.02	-0.09*
	Health	0.01	0.12	-0.03
	National Security	0.22*	0.04	0.10
	World at peace	0.06	0.05	-0.05
	R^2	0.10	0.04	0.02

Values shown are Beta; *p<0.05

Table 3.12 – Multiple linear regression between negative weighted values and the three decisions divided by biological gender

Biological gender	Predictor	Against/ Favor	With regulation	Without regulation
Male				
	True Friendship	-0.02	-0.02	0.03
	Equality	-0.17*	-0.01	-0.08
	Freedom	-0.16*	-0.03	0.03
	Privacy	0.07	0.04	-0.06
	Health	0.21*	0.12*	0.02
	National Security	-0.01	0.07	-0.09
	World at peace	-0.08	0.02	-0.01
	R^2	0.10	0.02	0.02
Female				
	True Friendship	0.02	-0.03	0.11
	Equality	-0.14	-0.18*	0.07
	Freedom	-0.23*	0.06	-0.15*
	Privacy	-0.09	-0.23*	0.10
	Health	-0.08	-0.23*	-0.13*
	National Security	0.13	0.16*	-0.00
	World at peace	-0.10	-0.01	-0.08
	R^2	0.13	0.16	0.06

Values shown are Beta; *p<0.05

Once the equality value is reinforced by AI, only male gender represents a predictor to the three decisions. Men have an AI evolution pro-attitude and with regulation when AI reinforce the equality value, and an against attitude towards AI evolution without regulation. On other hand, when equality value is threatened by AI, men assume an against attitude and women adopt an against with regulation attitude.

Both biological genders have significant statistical differences on freedom value. If AI reinforces this value, men adopt an against attitude toward AI evolution and with the

regulation decision. Also, when AI threatens freedom, men take an against decision toward the evolution of AI. Women embrace an attitude against AI evolution also without regulation option, even if this type of technology threatens or reinforces the freedom value to AI evolution. But if the regulation is present, women take a pro-attitude toward AI evolution when it is beneficial.

Relatively to privacy value been reinforced by AI, men are in favor of his evolution in all three decisions. On other hand, women takes an against attitude toward AI evolution without regulation when it benefits the value and against with regulation when it threatens privacy.

Men always took a pro attitude to AI evolution and with regulation even if AI threatens or beneficiate the health value. Instead, female only assume a position when AI threaten this value, this is, the women are against AI with and without regulation when it impacts in a negative way on health value.

In the case of the national security value, if AI beneficiate it the men are against his evolution without regulation. Women adopt a pro attitude toward AI when it benefits the value, but when the opposite occurs – it threatens the value – they also assume a pro-favor attitude but only in presence of regulators.

Lastly, when AI reinforces the "world at peace" value, the male gender are in favor of its evolution without regulation. Instead, women are against AI evolution without regulation even if it benefits de "world at peace" value.

Chapter IV – Discussion and Conclusions

As previously mentioned, the literature regarding the association between human values and AI is quite scarce. Therefore, this chapter will first clarify the preposition test presented in the previous chapter through our logical thinking alongside the information obtained in the 205 interviews. Next, we will explore additional data from a control variable which proved to be essential. The chapter will finish with the acknowledgement of the limitations present during the completion of the study, as well as suggestions for future studies.

There are two moments of deliberation during the final decision in which the participant must make a decision regarding the evolution of AI. On a first instance the participants must conclude whether they are in favor or against the evolution of AI, taking into consideration whether said evolution benefits or threatens a specific human value. On a second level the participants must decide if the existence of regulatory agents is necessary, once again taking into consideration the impact they perceive AI has on each human value. From the results it becomes clear that there is a distinction between these two moments: on the one hand, the participants display strong stances on the first level; on the other hand, they are only able to adopt a less neutral stance on the second level.

There are four human values referred to as predictive, in order for the individuals do assume a certain attitude towards AI evolution. They are: equality, freedom, health and national security.

Regarding equality, the results have shown that in case the AI threatens this human value in any way, shape or form, the individuals adopt an unfavorable attitude towards its evolution. This can be explained through two different common attitudes mentioned by the participants: the possible divergence between social classes that can be aroused from AI development, and the fear that such technology might stand equal to humans.

Regarding the first justification, the participants tend to mention that with the evolution of AI the social and economic differences will be further emphasized. The upper class possesses a larger purchasing power which allows it to obtain several technological tools. It can also obtain the mastery of AI knowledge more easily and quickly, being able to adapt to this context, as opposed to a lower social class. The companies that sell their products cheaper in other countries but avoid purchasing there, for instance, make it so that the access to the free market and price transparency are no

longer guaranteed (Helbing, Frey, Gigerenzer, Hafen, Hagner, Hofstetter, Hoven, Zicari, & Zwitter, 2017).

« *[…] it is the difference in society. AI will increase the difference between the poor and the rich even further. Drastically. Because the rich will become more intelligent, and, becoming more intelligent, they will also become richer. And the poor will continue the same, and certainly will become even poorer. I think every other value is controllable. Not this one. Even if regulated, the rich are the ones who possess the power, they have the technology and the access to this tool. And this tool gives them an advantage in society like no other tool before. In the past, when kings lived, the kings and the nobility, who were 30 or 40, were superior to the rest of the population, who were millions in number. In the meantime we have decreased the gap because we have had access to better tools, but this is the one tool that drastically increases the difference in capacity between several tracks.* » P66

« *There are no doubts that if machines with power and freedom to act on their own start replacing us, there will be less people, less and less people. And, in the end, only the people up top will remain. And the ones down at the bottom. The ones up top will stay there because of a simple fact: money will always exist. And the ones at the bottom will surely find some, because someone will have to make the machines, right?! Until the machines start making themselves. Which is already happening, we are already that way. We are a chain functioning that way. What I think might happen is that the middle-class might disappear. Only the ones up top and the ones down at the bottom will remain. The gap separating the social classes will be much larger. It will be as large as the machine.* » P139

Considering the previous statements, we highlight the question of the decreasing number of jobs caused by AI development, as well as the mismatch between the rhythm of the interruption lead by the AI and the inertia of social readjustment, considering its impact on the economical world and on the workforce (Bundy, 2017; Cath, Wachter, Mittelstadt, Taddeo, & Floridi, 2018). Consequently, the possibility arises for the elimination or reduction of salaries, both for low qualification jobs and for jobs requiring higher professional qualifications (Bundy, 2017). Workers with low and medium wages represent, however, the higher chance of being negatively impacted by AI (Cath,

Wachter, Mittelstadt, Taddeo, & Floridi, 2018). Therefore, the implementation of measures that fight an increase in wealth discrepancy and influence, as well as widely decrease economic benefits, is increasingly necessary. (Bundy, 2017; Cath, Wachter, Mittelstadt, Taddeo, & Floridi, 2018). The dimension of equality is, for our sample, a highly relevant factor for the acceptance of AI evolution.

On the other hand, there are participants who understand that equality might be threatened by the evolution of AI, in the event that the AI itself achieves the same rights as a human being. One concrete example of such fear can be seen in the recent case of Sophia, the humanoid robot developed by Hanson Robotics, that in 2017 obtained citizenship in Saudi Arabia. Therefore, an AI system that possesses the same rights as a human citizen decreases the distinction between both.

« *Equality. I do not think they should be equal to humans.* » P123

« *Regarding equality, it is not the same. Because AI, as soon as it is developed to its fullest, has many more capabilities than the human brain. In that sense, it is unequal to human beings. The comparison between Humanity and AI is not a question of equality.* » P151

Freedom was another meaningful value that we came across. The conclusion is that, even if AI benefits or threatens this value, the individuals will always be against its evolution. In modern societies freedom is seen as a core value and a human right (Carlsson, & Weibull, 2018; Cohen, 2018), which means that any element that compromises it might generate some opposition. As we can see from this study, freedom can never be jeopardized or mentioned when discussing matters of AI. It also represents a strong attitude predictor.

« *If badly managed, freedom can be used to somewhat condition the society we live in. And that might cause a breach in our freedom, because you are subjugated to a force you do not control. And which you do not know how it will evolve.* » P151

« *We will stop having freedom the moment they begin. We stop having the freedom to think for ourselves. It is already happening often. Many times, without wanting to, or without thinking we are losing our freedom, we already are. Hypermarket x is already*

doing that. That one and others as well. But hypermarket x is perhaps the best example. What hypermarket x does with the promotions we get on our cards, based on our past purchases, is limiting our freedom to choose other products. They are directing us. I think there are many advantages in a machine that creates profiles of people, but the disadvantage is that they are channeling all the people to the same place, or to the place they want people to go. For me that is not the same way that goes to you, but we are now speaking of hypermarket x, we are talking about shopping. But when we start talking about the city where you live, or what you want to do with your life, they will channel all the people. They can channel all the people to a certain place and in a large scale. » P139

From this study one can conclude that, when talking about the development of AI systems, individual freedom can never be questioned. This means that the manipulative impact of AI when restricting freedom of choice becomes clear, which can be seen through the justifications provided by the participants (Helbing et al., 2017).

Individuals are exposed to persuasive communication in many different contexts: hypermarkets, governments, companies and political parties (Matz, Kosinski, Nave, & Stillwell, 2017). This kind of persuasive communication becomes more effective in influencing behavior when adapted to the unique characteristics and psychological motivations of people (Matz, Kosinski, Nave, & Stillwell, 2017; Shrap, Danenberg, & Bellman, 2018). It is possible to obtain more information about individuals through social networks and other similar mediums. Therefore, it becomes possible to influence the behavior of large groups through psychological targeting, through adaptation and persuasive appeals to each target audience's psychological needs (Matz, Kosinski, Nave, & Stillwell, 2017; Shrap, Danenberg, & Bellman, 2018). This technique allows for the possibility to manipulate individuals into behaving in certain ways that might go against their own best interest, or society's interest. Cambridge Analytica, mentioned in chapter one, is one of the recent cases that exemplifies this question of invasion of freedom. This company used several psychological profiles from millions of American citizens to influence the votes in the presidential elections.

Overall, it is in these questions of decision-making and choice manipulation that individuals express their stronger feelings about the violation of freedom. Such pattern of answers may be explained by these being the examples with which the participants are more familiarized with, as well as the fact that, at the time of data collection for the study, the Cambridge Analytica topic was being widely discussed by the media.

Similarly, the participants indicate the necessity for the existence of a strict control of the use of AI as one of their main concerns, as to avoid harm and, consequently, the conditioning of humanity itself. Therefore, before deliberating about the need for regulations in the development of AI, they assume that this technology cannot be, at any time, affected.

Another value which proved to possess a strong predictive value regarding the attitudes towards AI evolution was health. In the event that AI benefits healthcare, individuals have been in favor of its evolution. They perceive, however, the necessary existence of regulatory agents. In this situation, the individuals do not present a significant final decision in the first moment, but only on the second, in which regulation plays an

additional role in their decision-making process. They present a larger deliberative coherency when it comes to regulation.

« *Now, the truth is that the quality of life of people who have health limitations can be vastly influenced by AI. So, I am not the case, but for someone who has a large limitation and cannot communicate, AI might help that person being more proactive and accepted in society. Which means that in a life without said evolution, that person would either not be here, or would live severely hindered.* » P107

« *Health is an improvement, I think it is an improvement. Because it can vastly help with medical processes, detecting diagnostics, during treatment. I think that is a great advantage.* » P151

Following the opinions of the participants, Bynum (2007) claims that, due to the success of the machines, they represent a positive contribution to medicine. Stressing the positive effects of AI in health (Atkinson, 2016), it has been applied in several tools with the purpose of bettering human healthcare (e.g. Fernandez-Luque, & Imram, 2018; Hamet, & Tremblay, 2017; Mehta, & Devarakonda, 2018; Playford et. al, 2017).

However, the participants point out that AI should evolve within this context in a regulated fashion. It is argued that, since medicine has access to many health data obtained through AI systems, they should be carefully controlled as to avoid misuse (Helbing, 2017). Likewise, following the results of this study, AI should take into consideration the possible presence of power plays and private interests in the field of healthcare.

« *Namely, the fact that in the healthcare system there are large personal interests at play, and people connected with pharmaceutical companies, for instance, that can warp the system to push the sale of medication A or B. Such thing can be done through AI. And it is something that needs to be carefully dealt with, and some regulation is needed for that. […] Which means, the evil is not in AI itself, but in what can be done with AI and who can use AI. I believe that is where the problem resides. And that is it. It goes for healthcare, as well as for companies.* » P18

« *Health, with the usage of equipment. A surgery might be compromised if a certain machine is badly configured.* » P27

Lastly, the final value that provided a significant difference was national security. Here the participants assume a strong final decision in both moments, both when taking into consideration the regulation factor, and when not. This means that, when AI benefits and threatens national security, people are in favor of its evolution. But in the case AI threatens this value, people request the presence of regulation.

When it comes to the benefits of AI perceived by the participants within this value, they are mostly related to the protection of the country or the protection of human beings.

« *If we stop having soldiers dying in the war fronts, developing war traumas and so on, and start using machines for combat – that is, if war really is necessary, because in such case a lot of values are immediately excluded –, then it is positive: less death, less destruction.* » P75

« *Detecting terrorist cells, planning attacks that are much more direct, much more concise, which at this moment are a bit arbitrary. That can translate into less human casualties.* » P151

« *And I think that, through AI, the country can become safer.* » P163

When it comes to the second result concerning national security the conclusion is that, even if AI can be seen as a threat, the fact that it guarantees a higher amount of safety in

people's lives, as small as that might be, is enough for them to request its presence. This comes alongside the belief that regulating agents are able to control the hazards that AI may bring.

Perhaps some parallels can be drawn with the privacy paradox, which suggests that people may be inclined to renounce some privacy if in return they obtain some sort of benefit (Brown, 2001; Kokolakis, 2015; Sayre &, Horne, 2000). Ultimately, individuals are in favour of AI even if it may threaten national security, as long as regulating agents are also present, as they assume that AI may bring several benefits to their national security.

« *But regarding national security and that kind of security in many sectors of civil society, I think it is dangerous and should be managed with some precautions. But I still believe that the benefits far outweigh the possible hazards.* » P18

« *And I think that, through AI, the country may achieve a higher safety. But on the other hand not really, just imagine a scenario where another country's intelligence invades our own country's intelligence, and removes all the archives and exposes its citizens lives. I'm somewhat afraid of that happening. It can be good or it can be bad. I think it can turn bad if there's no one regulating it, that's more what I mean.* » P63

In sum, the obtained results do not agree with the scarce literature found about attitudes towards AI. Tussyadiah, Zach and Wang (2017) state that individuals tend to have a low level of negative attitudes towards technology, whereas, in our study, people are equal against and favor toward AI evolution. Likewise, Nomura (2006), Fox (2016) and Crowed and Friess (2013) demonstrated that people tend to nurture extremist attitudes, whether positive or negative, in reaction to technology. However, only four values proved to be attitude predictors, when it comes to AI evolution. This divergence

in results can be explained by the lack knowledge of the participants regarding the concept of AI. There is also the existence of diverse myths, many associated with the film industry, in which the AI element is involved in a negative way, thus helping to consolidate the shared assumption that it might be a threat.

5.1. Limitations and future research

As concluded, when it comes to AI evolution there are at least four human values that are affected during the final decisions, thus raising the relevance of those values as attitude predictors for this technology. However, our study is not conclusive due to the fact that R2 presents very low values. In this regard, we now present several limitations faced during the progress of the study which may be able to explain those values, as well as suggestions for future studies.

First of all, we emphasize the lack of literature about variable association, human values, attitudes and AI. As shown throughout the investigation, there are no studies about the impact of AI in human values, and only a few about the attitudes towards AI. But our study leans more towards the concerns for the future of this technology and does not address some specific AI system as much. The literature that was found focuses on the concept of AI, its applications, and ethical questions that might be implied. When it comes to values specifically, the studies revolve mostly around the placing of those values into the AI in such a matter as to avoid the existence of discrepancies.

There are a number of results obtained through this study which are not yet sufficiently well documented in literature. In that sense, it would be interesting to further explore this line in future studies – although we have found a significant difference in means, it would be important to verify if other samples maintain this result. And, consequently, to attempt to develop an explanatory model.

In methodological terms, there were some visible flaws. Initially we could quite feel the participants' lack of knowledge of AI, as well as their constant associations with productions enveloped in myths from the film industry. One of the most important was the fact that the participants grouped AI within their subfield of robotics, which lead to a very conditioned thinking on their part. Moreover, due to the complexity of the theme, the passing of knowledge becomes a difficult task. Therefore, it is necessary, in studies to come, to guarantee that the participant understands and can distinguish the concept of AI, in a way as to not condition clear and elongated answers on the theme.

When asked to fill out the online questionnaire about the possibility of AI somehow threatening some human values, some participants perceived the questions in a different spectrum. They understood that they had to evaluate the possibility of each value being successfully placed in an AI system. Thus, due to the abstract nature of the theme, it is necessary to have a larger control over the participants' comprehension, regarding what is being asked of them.

This takes us over to the next limitation, in which the values are not efficiently measured in a quantitative manner, but in a qualitative one. Since the individuals are not constantly contemplating their values in a conscious manner, it is necessary to evoke said values and allow some space for reflection, alongside a certain orientation relevant for the purpose of the study. The fact that we did not evoke each single value as a reminder prevented the attainment of more detailed information from the participants. We wanted to promote a free flow during the interviews and to avoid recognition, as not to influence an answer. In future studies we consider it necessary to reinforce a focusing on each single value, since it is absent when it comes to mnesic evocation. A large part of the decision is not accessible to consciousness unless purposely explored, which means that if each single value is separately selected the results might be much more useful, which will therefore translate into a better explicative value. The conclusion is that, in approaching this theme in future studies, a qualitative methodology should be chosen.

Due to the very subjective and abstract nature of the theme, the regulation of AI evolution raised many questions for the participants, especially when it came to the ways the regulatory agents would act. Therefore, it is necessary for a future study to consider more detailed information about this question, in order to help the participants form a clearer decision.

Last but not least, we highlight the result obtained regarding gender differences in attitude towards AI. Since there are no published studies touching this subject, we suggest that future research takes into consideration gender as something more than a control variable, as it can perform explicative functions. There might be other control variables still within socio-demographic data, namely the professional area, since the male gender represents the larger number of people connected to technology. That might turn into a variable with more relevance than others, and the upcoming studies should take that into account. We suggest that future studies compare people within the technology field with people outside of it, as to evaluate if knowledge on the subject might influence people's thoughts about AI evolution.

References

Al-Kahtani, N. S., & Allam, Z. (2013). Exploring value preferences among strudents_ An empirical study of salman bin abdulaziz university. *Journal of American Science, 9*(12), 44-53.

Allport, G. (1935). Attitudes. Carl Murchison (ed). *The Handbook of Social Psychology*, 798-844), Worcester: Clark University Press.

Armstrong, S., Bostrom, N., & Shulman, C. (2016). Racing to the precipice: A model of artificial intelligence development. *Artificial Intelligence & Society, 31*, 201-206.

Atkinson, R. (2016). 'It's going to kill us!' And other myths about the future of artificial intelligence. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3066182

Ayoub, K., & Payne, K. (2016). Strategy in the age of artificial intelligence. *Journal of Strategic Studies, 39*(5-6), 793-819.

Babcock, J., Kramár, J., & Yampolskiy, R. V. (2017). *Guidelines for Artificial Intelligence Containment*. arXiv preprint arXiv:1707.08476

Bac, C. W., van Henten, E. J., Hemming, J., & Edan, Y. (2014). Harvesting robots for high-value crops: State-of-the-art review and challenges ahead. *Journal of Field Robotics, 31*(6), 888-911.

Bardi, A. (2000). *Relations of values to behavior in everyday situations* (unpublished doctoral dissertation). The Hebrew University, Jerusalem.

Bogue, R. (2015). Underwater robots: a review of technologies and applications. *Industrial Robot: An International Journal, 42*(3), 186-191.

Boyd, R., & Holton, R. J. (2017). Technology, innovation, employment and power: Does robotics and artificial intelligence really mean social transformation?. *Journal of Sociology, 54*(3), 1-15.

Brent, E. (1988). Is there a role for artificial intelligence in sociological theorizing? The American Sociologist, 19(2), 158-166.

Brown, B. (2001). *Studying the internet experience*. HP Laboratories Technical Report HPL, 49.

Bundy, A. (2017). Preparing for the future of artificial intelligence. *AI & Society, 32*(2), 285-287.

Bynum, T. W. (2007). Ethicl challenges to citizens of 'the automatic age': Norbert Wiener on the information society. In J. Wecker (Ed.), *Computer ethics*. New York: Routledge.

Carlsson, U., & Weibull, L. (2018). *Freedom of expression in the digital media culture: A study of public opinion in Sweden*. University of Gothenburg, Sweden.

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the 'good society': The US, EU, ad UK approach. *Science and Engineering Ethics, 24*, 505-528.

Chakraboxty, N., & Bagchi, N. (2018). Green consumerism: Exploring the green buying behaviour of male and female consumers. *The Research Journal of Social Sciences, 9* (7), 26-39.

Cheng, A., & Fleischmann, K. R. (2010). Developing a meta-inventory of human values. *ASIS&T'10, 47*(3).

Coeckelbergh, M. (2011). From killer machines to doctrines and swarms, or why ethics of military robotics is not (necessarily) about robots. *Philosophy & Technology, 24*(3), 269-278.

Cohen, A. J. (2018). *Toleration and freedom from harm: Liberalism reconceived*. New York: Routledge.

Copeland, J. (2015). *Artificial intelligence: A philosophical introduction*. John Wiley & Sons.

Costa, N. A. (2012). *Priorização e (in)satisfação dos valores relativos ao trabalho conforme o género: Estudo transcultural entre Portugal e o Brasil* (Tese de mestrado). ISCTE-IUL, Lisboa.

Crowder, J. A., & Friess, S. (2013). Artificial psychology: The psychology of AI. *Systemics, Cybernetics and Informatics, 11*(8), 64-68.

De Cubber, G., Doroftei, D., Balta, H., Matos, A., Silva, E., Serrano, D., Govindaraj, S., Roda, R., Lobo, V., Marques, & Wagemans, R. (2017). *Operational Validation of Search and Rescue Robots. In Search and Rescue Robotics-From Theory to Practice*. InTech.

Dickow, M., & Jacob, D. (2018). The global debate on the future of artificial intelligence: The need for international regulation and opportunities for German foreign policy. *SWP Comments, 23*, 1-9.

Dignum, V. (2017). Responsible artificial intelligence: Designing AI for human values. *ITU Journal: ICT Discoveries, Special Issue, 1*, 1-8.

Eagly, A.H., & Chaiken, S. (1993). *The psychology of attitudes*. Texas: Harcourt.

EU General Data Protection Regulation (2016, april) Retrieved from https://www.eugdpr.org/

Feather, N. T. (1990). Bridging the cap between values and actions: Recent applications of the expectancy-value model. In E. T. Higgins & R. M. Sorrentino (Ed.), *The handbook of motivation and cognitions: Foundations of social behavior* (pp 151-192). New York: Guilford Press.

Fernandez-Luque, L., & Imram, M. (2018). Humanitarian health computing using artificial intelligence and social media: A narrative literature review. *International Journal of Medical Informatics, 114*, 136-142.

Ferrer, G., Zulueta, A. G., Cotarelo, F. H., & Sanfeliu, A. (2017). Robot social-aware navigation framework to accompany people walking side-by-side. *Autonomous robots, 41*(4), 775-793.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Massachusetts: Addison-Wesley.

Fonseca, A. M. O., Porto, J. B., & Barroso, A. C. (2012). O efeito dos valores pessoais nas atitudes perante estilos de liderança. *Revista de Administração Mackenzie, 13*(3), 122-149.

Fox, S. (2016). Domesticating artificial intelligence: Expanding human self-expression through applications of artificial intelligence in presumption. *Journal of Consumer Culture, 0*(0), 1-15.

Ginsberg, M. (2012). *Essentials of artificial intelligence*. California: Morgan Kaufmann Publishers.

Gips, J. (1979). Artificial intelligence. *Environment and Planning B, 6*, 353-364.

Gouveia, V. V., Andrade, J. M., Milfont, T. L., Queiroga, F., & Santos, W. S. (2003). Dimensões normativas do individualismo e coletivismo: É suficiente a dicotomia pessoal vs. social? *Psicologia: Reflexão e Crítica, 16*(2), 223-234.

Granjo, M., & Peixoto, F. (2013). Contributo para o estudo da escala de valores humanos de Schwartz em professores. *Laboratório de Psicologia, 11*(1), 3-17.

Greenwald, A. G., Brock, T. C., & Ostrom, T. M. (1968). *Psychological foundations of attitudes*. New York: Academic Press.

Hamet, P., Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism Clinical and Experimental, 69*, 36-40.

Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., Hoven, J. V. D., Zicari, R.V., & Zwitter, A. (2017). Will democracy survive big data and artificial intelligence? In *Towards Digital Enlightenment*(pp. 73-98). Springer, Cham.

Hillman, D. J. (1985). Artificial intelligence. *Human Factors, 27*(1), 21-31.

Hitlin, S. (2003). Values as the core of personal identity: Drawing links between two theories of Self. *Social Psychology Quarterly, 66*(2), 118-137.

Holvast, J. (1993). Vulnerability and privacy: Are we on the way to a risk-free society. *Facing the challenge of risk and vulnerability in an information society*, *33*, 267-279.

Huang, M., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research, 21*(2), 155-172.

Jin, G. Z. (2018). Artificial intelligence and consumer privacy. *NBER Working Paper No. 24253*, DOI: 10.3386/w24253.

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security, 64*, 122–134.

Lima, L. (2010). Atitudes: estrutura e mudança. In J. Vala, & M. Monteiro (8th ed.), *Psicologia Social* (pp. 187-225). Lisboa: Fundação Calouste Gulbenkian.

Lisetti, C. L., & Schiano, D. J. (2000). Automatic facial expression interpretation: Where human- computer interaction, artificial intelligence and cognitive science intersect. *Pragmatics & Cognition, 8*(1), 185-235.

Lu, H., Li, Y., Chen, M., Kim, H., & Serikawa, S. (2018). Brain intelligence: Go beyond artificial intelligence. *Mobile Networks and Applications, 23*(2), 368-375.

Makridakis, S. (2017). The forthcoming artificial intelligence (AI) revolution: Its impact on society and firms. *Futures, 90*, 46-60.

Maslova, O. V. (2018). Value shifts in Vietnamese students studying in Russia. *Psychology in Russia: State of the Art, 11*(2), 17-27.

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciiences of the United States of America, 114*(48). 12714-12719.

McCorduck, P. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. Natick: A K Peters.

Mehta, N., & Devarakonda, M. V. (2018). Machine learning, natural language programming, and electronic health records: The next step in the artificial intelligence journey? *Journal of Allergy and Clinical Immunology, 141* (6).

Nilsson, N. J. (2014). *Principles of artificial intelligence*. Morgan Kaufmann Publishers.

Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). Measurements of negative attitudes toward robots. *Interaction Studies, 7*(3), 437-454.

Pannu, A. (2015). Artificial intelligence and its application in different areas. *Artificial Intelligence, 4*(10), 79-84.

Parent, W. A. (1983). Privacy, morality, and the law. *Philosophy & Public Affairs, 12*(4), 269-288.

Pereira, C., Camino, L., & Costa, J. B. (2005). Um estudo sobre a integração dos níveis de análise dos sistemas de valores. *Psicologia: Reflexão e Crítica, 18*(1), 16-25.

Pew Research Center (2014), Public Perceptions of Privacy and Security in the Post-Snowden Era. Retrieved from http://www.pewinternet.org/2014/11/12/public-privacyperceptions/

Playford, D., Jais, P., Weerasooriya, R., Martyn, S., Bollam, L., Turewicz, M., & Mohamad, R. (2017). P1721: A validation study of automated atrial fibrillation detection using alerte gigital health artificial intelligence system. *European Heart Journal, 38*.

Prabhu, G. R. D., & Urban, P. L. (2017). The dawn of unmanned analytical laboratories. *TrAC Trends in Analytical Chemistry, 88*, 41-52.

Prati, G., Pietrantoni, L., & Albanesi, C. (2005). Human values and beliefs and concern about climate change: A Bayesian longitudinal analysis. *Quality & Quality, 52*, 1613-1625.

Ramamoorthy, A., & Yampolskiy, R. (2018). Beyond mad?: The race for general artificial intelligence. *ITU Journal: ICT Discoveries, Special Issue, 1*, 1-8.

Reed, C. (2018). How should we regulate artificial intelligence? *Philosophical Transactions Royal Society,* 1-12. DOI: 10.1098/rsta.2017.0360

Robinson, H., MacDonald, B., & Broadbent, E. (2014). The role of healthcare robots for older people at home: A review. *International Journal of Social Robotics, 6*(4), 575-591.

Rodrigues, A. (1978). *Psicologia Social* (7th ed.). Rio de Janeiro: Editora Vozes.

Rokeach, M. (1973). *The nature of human values*. New York: Free Press.

Rokeach, M. (1981). *Crenças, atitudes e valores*. Interciência.

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 26*, 582-599.

Sagiv, L., Roccas, S., Cleciuch, J., & Schwartz, S. (2017). Personal values in human life. *Nature Human Behaviour, 1*, 630-639.

Sayre, S., & Horne, D. (2000). Trading secrets for savings: How concerned are consumers about club cards as a privacy threat? *Advances in Consumer Research, 27*(1), 151-155.

Scherer, M. U. (2015). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. *Harvard Journal of Law & Technology, 29*, 353.

Scherer, M. U. (2016). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. *Harvard Journal of Law & Techonology, 29*(2), 353-400.

Schwartz, S. H. (1994). Are there universal aspects in the structure and contents of human values? *Journal of Social Issues, 50*(4), 19-45.

Schwartz, S. H. (2006). Les valeurs de base de la personne: théorie, mesures et applications. *Revue française de sociologie*, *47*(4), 929-968.

Schwartz, S. H., & Bilsky, W. (1987). Toward a universal psychological structure of human values. *Journal of Personality and Social Psychology, 53*(3), 550-562.

Searing, D. (1979). A study of values in the British House of Commons in Milton Rokeach (ed). *Understanding Human Values: Individual and societal*. (154-178). New York: Free Press.

Shrap, B., Danenberg, N., & Bellman, S. (2018). Psychological targeting. *Proceedings of the National Academy of Sciiences of the United States of America*, *115*(34).

Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. *Cutter Business Technology Journal, 31*(2), 47-53.

Tussyadiah, I. P., Zach, F. J., & Wang, J. (2017). Attitudes toward autonomous on demand mobility system: the case of self-driving taxi. In: R. Schegg, & B. Stangl (Ed.), *Information and communication technologies in tourism 2017*. Springer: Cham

Tuulik, K., Ounapuu, T., Kuimet, K., & Titov, E. (2016). Rokeach's instrumental and terminal values as descriptors of modern organization values. *International Journal of Organizational Leadership, 5*, 151-161.

Unal, H., & Kuraloglu, H. (2015). Determination of operating parameters in milking robots with free cow traffic. *Engineering for Rural Development*, *14*, 234-240.

Verdiesen, I. (2017). How do we ensure that we remain in control of our autonomous weapons?. *AI Matters, 3*(3), 47-55.

Weber, J. (2017). Discovering the millennials' personal values orientation: A comparison to two managerial population. *Journal of Business Ethics*, *143*(3), 517-529.

Zheng, S. (2015). *Exploring motivational value preferences among different occupations in Portugal* (Tese de mestrado). ISCTE-IUL, Lisbon.