ISCTE IUL
University Institute of Lisbon

IUL - School of Social Sciences

Department of Social and Organizational Psychology

Emotions and Acceptance towards Artificial Intelligence and its Evolution

Gabriela Morais dos Santos

Dissertation submitted as partial requirement for the conferral of

Master in Social and Organizational Psychology

Supervisor:

Doutor Nelson Campos Ramalho

Assistant Professor, ISCTE-IUL – University Institute of Lisbon

October, 2018

Acknowledgements

There have been five years since this journey through university has begun and this dissertation means the end of a cycle.

For that reason, I want to thank first to those without whom this dissertation would not have been possible: the participants. Thank you all for your time and all the patience, and willing to help me through the entire process.

Coming from this, I want to give a special thanks to two of my friends. José Nuno, there are no words to describe all the patience you had with me and all the help you gave me when bringing all your friends to participate in my experiment. I'm sure life will pay you back for such generosity. Magda Rocha, you've been my "thesis-fellow" for months and were there through all my *paranoia* moments. You surely deserve a place in heaven, thank you for everything.

Speaking of friends, there goes a very special thank you, from the bottom of my heart, to my *crew* from Aveiro: Joana Rolo, Sílvia Xavier and Adriana Costa – you girls rock!

To my soon-to-be Electrical and Computer Engineer, I want to thank all the advice, all the help, all the curiosities you shared about technology, all the patience and all the love. You are one of the fiercest people I know and your willing to fight for your dreams goes beyond the limits of human nature and that sure makes you a precious being in this wild World. So, thank you, Tiago Guedes, you truly inspired me to be writing these lines today.

Before my final thanks, I cannot forget the person who made all this possible. The person who was the source of all brainstorming right from the very beginning, the one with unimaginable thoughts, unmeasured ambition and contagious energy. Thank you so much, Professor Nelson Ramalho, for sharing all your wisdom and for all the unconditional support throughout the entire process.

And finally, I want to not only thank but to dedicate this Master's Dissertation to the two most important people of my life. One for providing me an exquisite education and all the means for me to achieve my dreams. You gave me everything I needed before you even knew me and you have always put myself first, no matter what – thank you, Mother. The other, for being my dearest role model since I was born. For helping me through a variety of situations and for always showing me the way, even before I knew I would get there – thank you, Brother. For different reasons, you are both the strongest people I know and inspire me every day to become a better version of myself. Sometimes I wonder if you are both superhumans, but then I realise you are just my family and that I am the luckiest person in the World.

Resumo

A Inteligência Artificial (IA) é um dos mais recentes e entusiasmantes enigmas da sociedade atual, e as suas aplicações aumentam de dia para dia. Sem nos apercebermos, lidamos com a IA nas formas mais subtis. O seu rápido desenvolvimento tem espoletado várias opiniões entre cientistas, tais como Elon Musk e Stephen Hawking, devido às consequências controversas que a IA poderá implicar. Além das opiniões dos especialistas, todos os que pertencem à sociedade irão ser afetados, quer positiva, quer negativamente.

O presente estudo visa a compreender quais as emoções espoletadas pela evolução da IA, e se essas emoções tomam um papel moderador no efeito da aceitação da IA na concordância ou discordância com a sua evolução, com ou sem reguladores.

Conduzimos um estudo com metodologia mista numa amostra de 205 participantes, aplicando um questionário *online*, onde avaliámos as emoções dos participantes relativamente aos estímulos apresentados em três tempos distintos. Posteriormente, realizámos uma entrevista semiestruturada.

Concluímos que, em oposição ao que é encontrado na literatura, as emoções negativas tendem a aumentar, à medida que o conhecimento relativo à IA é aprofundado. Simetricamente, as emoções não-negativas tendem a diminuir. As emoções negativas parecem funcionar como moderadoras da relação entre as variáveis supramencionadas.

Em adição, a visão dos participantes relativamente à evolução da IA parece ser favorável, no entanto, os participantes reconhecem a existência de reguladores como uma necessidade imperativa.

Palavras-chave: Inteligência Artificial; Emoções; Reguladores; Aceitação da IA

Abstract

Artificial Intelligence (AI) is one of the ultimate riddles of today's generations and its applications are increasing day by day. Without realising, we deal with AI in the subtlest ways. Its rapid development has triggered several opinions among scientists, such as Elon Musk and Stephen Hawking, due to the controversial consequences it may imply. Apart from experts' opinions, everyone belonging in society will be affected, whether positively or negatively.

The present study aims to understand what kind of emotions are triggered by AI evolution, and if those emotions play a moderator role on the effect of AI acceptance on the agreement on which AI should evolve or not, with or without regulators.

We ran a mixed methodology through a sample of 205 participants, applying an online survey, where we assessed the participants' emotions regarding AI stimuli across three times, and further conducting a semi-structured interview.

We concluded that, as opposite to what literature states, negative emotions tend to rise as the contact and knowledge regarding AI deepens. Symmetrically, non-negative emotions tend to decrease. Negative emotions seem to function as a moderator on the relationship between variables mentioned above.

Also, the participants' vision towards AI evolution seems to be hopeful, yet they recognise the need for regulators to be imperative.

Key-words: Artificial Intelligence; Emotions; Regulators; AI Acceptance

Index

List of Tables and Graphs

List of Figures

List of Acronyms

AI – Artificial Intelligence

TAM – Technology Acceptance Model

EAS – Emotion Assessment Scale

INTRODUCTION

Artificial Intelligence – or AI – is one of the ultimate riddles of today's generations. We have now entered an Era where the study of intelligence and its processes is no longer limited to psychologists and philosophers. Today, intelligence exists in the field of technology as a field of study, on which scientists and engineers' greatest ambition is to export intelligence that is characteristic to humans, to artificial agents, enabling them to perform any activities that are currently done by humans. AI exists in several forms and has applications in several fields. Basically, it is a set of complex algorithms build towards a specific purpose, and the main focus today is to improve those algorithms and build an AI that is capable of learning by itself. However, to do so, known algorithms so far are not enough, therefore there have been studies focusing on neural networks (Ertel, 2007; Luger, 2009), for example, in order to achieve more accurate and complex algorithms.

AI has several applications and opinions towards it being advantageous or disadvantageous are quite controversial. There is, in fact, a main focus on its advantages and how it may improve so many fields, for instance medicine, investigation, healthcare, and so many others. Notwithstanding, there are some identities with a special status towards AI that seem to be worried about its development and future consequences, such as Elon Musk and Stephen Hawking. These individuals reinforce the idea that the need for AI regulators is imperative, in order to avoid future catastrophes.

From this emerges the importance to look at society and understand people's knowledge and understanding towards AI, considering that if it really takes a relevant stand, society will be affected, whether positively or negatively. Therefore, we recognise that it is essential to assess people's acceptance towards AI and their vision regarding its evolution. It is central to measure what kind of emotions are triggered when people are exposed to the existent information regarding AI and its future perspectives, in order to comprehend if emotions play a decisive role when it comes to accepting AI and its evolution. Emotions are considered to be central in social psychology, therefore it only makes sense to assess them, since the development of such kind of technology will irrevocably affect society. These kinds of findings may end up to be considered essential, due to the fact that it is important that a certain evolution is well accepted by the ones involved, in order to avoid social disruptions.

The structure of this dissertation will focus on the aspects mentioned above, in order to provide a holistic knowledge regarding this theme and, hopefully, relevant conclusions.

We will begin contextualising the state of the art with Chapter I, highlighting the definition of AI, its application areas, and what really exists nowadays and its future perspectives. As explained before, emotions play a central role in social psychology, therefore, we will also clarify its definition and understand how AI acceptance and people's emotions are imperative when mentioning AI and its evolution.

Chapter II will describe the methodologies used in the present study, clarifying the variables studied as well as the procedure steps and used stimuli.

Results of the conducted analysis will be reported in Chapter III and further discussed in Chapter IV. This last chapter will also mention limitations occurred in the present study and provide suggestions for future studies.

CHAPTER I – Literature Review

1.1. Artificial Intelligence: Definition and Applications

Artificial Intelligence – or AI – is one of the ultimate riddles of today's generation. In fact, if we think properly, the "artificial" part of its name is truly recent. However, the study of "intelligence" itself has been done for years by philosophers and psychologists who try to understand memory, learning, thinking, seeing and reasoning processes (Russel & Norvig, 1995). The point is, while some are concerned about the study of intelligence processes, others are interested in building intelligence itself. And that is where AI enters. AI was formally introduced by John McCarthy – an American computer science pioneer and inventor, considered to be the father of AI – in a conference in Dartmouth in 1956 (Copeland, 1993; Gips, 1979; Russel & Norvig, 1995). Besides trying to understand it, engineers all over the world, following the previous work of the so named fathers of AI (Marvin Minsky, McCarthy, Allen Newell, Herbert Simon and Cliff Shaw) have been focusing on the development of agents that possess an artificial intelligence with abilities as closer as to the ones of human intelligence.

So, what is, in fact, AI?

It is considered to be a subdivision of computer science on which its main goal is to programme computers and create computing machines and systems that once would require human intelligence (Brent, 1988; Ertel, 2017; Gips, 1979). It uses symbolic reasoning and sophisticated knowledge structures and techniques so that its operations' performance can be analogous to human learning and decision-making (Atkinson, 2016; Brent, 1988; Hillman, 1985; Russel & Norvig, 1995).

1.1.1. AI's application areas

AI is an extremely complex subject – and we could not expect less, considering it works with such complicated processes that intelligence combines (Luger, 2009) – that embraces several fields (Brent, 1988; Gips, 1979; Ginsberg, 1993; Hillman, 1985; Russel & Norvig, 1995). Depending on the specifications of each field, it may be also named as machine learning, machine intelligence, deep learning, and cognitive computing (Atkinson, 2016). These specifications vary according to what each intelligent agent actually does. Please note that there are, in fact, AI application areas with different characteristics, however, we must not completely dissociate them from each other, considering most of AI systems are a merge of several AI fields. On the following topics we will expose some of the existent fields of AI and give a few examples of its applications within each field.

Game Playing:

As explained by Luger (2009), Ginsberg (1993) and Gipps (1979), this field begun its development with the use of board games such as chess, checkers and de 15-puzzle, due to the fact that this type of games has characteristics that make them ideal research models. This AI domain is dominated by the use of techniques that are called *heuristics* (i.e. problem-solving strategies), which constitute a major area of AI research (Luger, 2009). The success achieved with computer-based game playing lead this field to world championships (Gipps, 1979; Luger, 2009). One of most recent examples of these AI advances and their success is Google's *AphaGo AI* - a game-playing AI created by DeepMind (a company by Google) – who beat the world's best Go player last year in China (Mozur, 2017). This AI agent as recently been generalised so it may be able to learn other games as well. Now named *AlphaZero*, it beat the world champion chess program (*Stockfish 8*), after only four hours of learning the rules to chess (Gibbs, 2017).

Automated Reasoning and Theorem Proving:

The research regarding this field was responsible for much of the preliminary work in formalizing search algorithms and developing programming languages (Luger, 2009). Although these systems are not always perfectly accurate, they have shown importance in assisting several problems, for instance, the design and verification of logic circuits, control of complex systems, and verification of the correctness of computer programs (Luger, 2009). Furthermore, modern theorem provers behave as intelligent assistants to human activities (Luger, 2009), including medical diagnosis and information retrieval (Nilsson, 2014).

Expert Systems:

These systems work with a combination of a theoretical understanding of a certain problem and several heuristic problem-solving rules regarding that problem (Luger, 2009; Pannu, 2015). That is why expert systems are programs that work within a certain specialized domain (Copeland, 1993; Hillman, 1985; Luger, 2009; Nilsson, 2014), for example, medical diagnosis (Agha, Jarghon & Naser, 2017; Gipps, 1979; Hamet & Tremblay, 2017; Hillman, 1985; Luger, 2009; Nilsson, 2014; Rekhawi, Ayyad & Naser, 2017) , treatment prescription (Luger, 2009), study of molecules (Hamet & Tremblay, 2017; Luger, 2009; Nilsson, 2014) , computer system configuration (Hillman, 1985; Luger, 2009), geological information for oil prospecting (Hillman, 1985) and evaluating ore deposits (Luger, 2009; Nilsson, 2014), analysing the performance of electronic circuits (Stallman & Sussman, 1977), and preventing cyber assaults (Anwar & Hassan, 2017).

A very curious example of an expert system - showing the immense variety of this field - is *Watermelon Expert System* developed by Abu-Nasser and Abu-Naser (2018), which combines two different programming languages and helps farmers to detect watermelon diseases through a pleasant user interface, providing solutions for those problems.

Natural Language Understanding:

This field focuses on the development of programs that have the ability to read, understand, analyse, and generate human language despite the type of speech (formal, informal, colloquial or even slightly *slang*) (Gips, 1979; Hillman, 1985; Luger, 2009; Pannu, 2015). It relies on several disciplines other than AI, such as computational linguistics, philosophy, and cognitive psychology (Nilsson, 2014), and it is widely used in speech understanding, semantic information processing, question answering, information retrieval, language translation (Gips, 1979; Pannu, 2015). One very clear and known example of this type of systems is Apple's iOS *Siri*, "*OK Google*" in Google Now, and Microsoft's *Cortana*, all virtual assistants who use natural language understanding to answer questions, make recommendations, delegate user's requests to other apps, and a series of other functions using AI.

Modeling:

Whether we talk about human performance (Luger, 2009), natural systems (biological, sociological, economic or ecological), or problem-solving systems, modeling is referred by Pannu (2015) as "the ability to develop an internal representation and set of transformation rules which can be used to predict the behaviour and relationship between some set of real-world objects or entities" (p. 80).

Attached to this field is one of the most seductive approaches regarding AI to the ones who study it: neural networks. According to Luger (2009), this approach "seeks to build intelligent programs using models that parallel the structure of neurons in the human brain or the evolving patterns found in genetic algorithms and artificial life" (p. 29). Additionally, Ertel (2007) refers to this field as "the bionics branch within AI" (p. 245). We can illustrate this field by mentioning *Neuralink*, an American neurotechnology company founded by Elon Musk that claims to be developing implantable brain-computer interfaces, using AI and neural networks. The main goal of this company for now is to find ways to treat serious brain diseases, being that in the future it aims to achieve human enhancement (i.e. cognitive abilities) (Newitz, 2017).

Learning and Adaptive Systems:

These systems are able to adapt their behaviour based on previous experience, meaning they are capable of learning and building new algorithms in order to readjust themselves to a specific situation (Ertel, 2017; LeCun, Bengio & Hinton, 2015; Pannu, 2015). This happens, for exemple, in cybernetics, and concept formation (Pannu, 2015). Often embodied with perception techniques (Nilsson, 2014), they are also able to recognise patterns and analyse certain scenes (Pannu, 2015).

Robotics:

By last, the field of robotics is the most well-known area by population in general, since everyone has already seen a robot, whether it was complex or not. Robotics is (or might be, depending on each case) a combination of all of the fields above (Pannu, 2015), concerning the design and utility of intelligent agents (Hillman, 1985). It comprehends a spectrum of disciplines besides AI, such as mechanical engineering, industrial engineering, computer science, physics, materials science, manufacturing systems engineering, and control theory (Hillman, 1985). The difference from the other areas is that these systems – the robots – are capable of moving over terrain and manipulate objects (Gips, 1979; Luger, 2009; Nilsson, 2014; Pannu, 2015). There is a diversity of areas on which robotics has been applied: healthcare (Breazeal & Scassellati, 2000; Robinson, MacDonald & Broadbent, 2014), education (André, Baker, Hu, Rodrigo & Boulay, 2017), work (Nezhad, 2015), social companion (Dautenhahn, 2004; Konok, Korcsok, Miklósi & Gácsi, 2018), transportation/navigation (Pannu, 2015), industrial automation (Pannu, 2015), agriculture (Emamgholizadeh, Kashi, Marofpoor & Zalaghi, 2014; Pannu, 2015; Patrício & Rieder, 2018), rescue (Schneider, Wildermuth & Wolf, 2015), underwater operations (Riva, 2017), and so many others (Pannu, 2015). Another important example to give, also due to its controversial aspects, is robotics applied in military purposes with the use of autonomous robots, often known as *killer robots* or *killer drones*, aiming to replace soldiers in wars and battlefields (Coeckelbergh, 2011; Sauer & Schörnig, 2012). These weapons, once programmed, are capable of finding a target and operate according to its judgement of the situation, without human intervention or supervision (Coeckelbergh, 2018; Leveringhaus, 2016).

1.1.2. AI today and AI tomorrow

About five years ago we did not even have internet (3G mobile data) in our mobile phones. Today, our mobile phones show us adds about the thing we were just talking about to the person next to us.

It took just a few years for AI to evolve in a way that is actually unknown for many of us. In fact, it has already proven to exceed human performance in several fields (a few already mentioned above), such as image recognition, speech transcription and direct translation (Atkinson, 2016; Spiegeleire, Maas & Sweijs, 2017). We are almost entering some sort of wizarding world, considering we even have self-driving cars (we cannot really find these in everyday life yet, but there have been several competitions for autonomous robots on which a very common category are autonomous vehicles, and their development has evolved quite a lot). Despite all of these advances regarding AI, it has still not reached its peak and it is considered to be quite limited in terms of what it does right now and what it might be able to do in the future (Atkinson, 2016; Lu, Li, Chen, Kim & Serikawa, 2018).

As described by Spiegeleire, Maas and Sweijs (2017), literature commonly categorises AI in the three following generations:

- **Artificial Narrow Intelligence** (ANI or "narrow AI"), often paired with an infant, it is related to machine intelligence that is restricted to specific tasks, equalling or exceeding human intelligence. An example is Google's *AlphaGo* already mentioned;
- **Artificial General Intelligence** (AGI or "strong AI"), which is an advanced level of AI and it is considered to be paired with adults, meaning it equals human performance on any task;
- **Artificial Superintelligence** (ASI) is the pinnacle of AI, where it surpasses human intelligence and performance across all tasks and fields.

We can say that right now we stand somewhere between ANI and AGI, and the path just keeps flowing forward. There have been remarkable advances in neuroscience and computer science (Hinton & Salakhutdinov, 2006). AI is currently the nucleus of several enterprises' business model, such as Apple, Amazon, Google, Facebook, IBM, and Microsoft and there has been a punch in the automotive industry when it comes to the investment of AI systems and AI programs by Toyota, Ford Motors, Mercedes-Benz, and BMW (Spiegeleire, Maas & Sweijs, 2017). Furthermore, we can highlight the advances in natural language processing, cybersecurity, and face recognition algorithms - for example, Google's *GooLeNet*, Facebook's *DeepFace*, and Yahoo's *DeepDense* (Spiegeleire, Maas & Sweijs, 2017). All of these extraordinary advances culminate to the conclusion that we are currently in the *Cognitive Era* (Kelly, 2015), characterised by systems that do not need to be entirely programmed and are able to learn and function in an autonomous way (Spiegeleire, Maas & Sweijs, 2017).

So, what does this *Cognitive Era* provide us and what are AI's future perspectives?

As stated by Kelly (2015), "as with every revolutionary technology, our initial understanding will be limited – both by the world's complexity and by our own deeply ingrained biases and heuristics. However, for all these limitations, progress is imperative" (p. 10). But what exactly is this progress? According to Spiegeleire, Maas & Sweijs (2017), there are two possible scenarios regarding AI evolution. One is that AI will be continuously evolving and slowing down, meaning there will be a period of strong evolution and development, followed by a period of stagnation. The other scenario is that AI will evolve exponentially until there is an intelligence explosion where AI supersedes human intelligence across all fields (Spiegeleire, Maas & Sweijs, 2017). Spector (2006) asserts that “not only should the AI systems of the future grow and learn, but their developmental and learning processes should be crafted by the most powerful designer of adaptive complexity known to science: natural selection (p. 1253). But when will this happen? Actually, the timeline of AI evolution generates disagreement among experts (Bostrom, 2014). Armstrong and Sotala (2015) concluded that 50% of experts (from the sample they studied) consider this superintelligence will be achieved by 2040, whilst 90% assume it will happen by 2075.

Attached to this thematic of AI evolution, there is another important topic to mention and that is AI regulation.

Is AI evolution regulated somehow?

Right now, there is no entity concerned whatsoever with AI regulation, meaning their developers and users are the ones responsible for its control. However, as it has been said before, AI is evolving faster everyday and its final goal is to supress humans across all fields (Kelly, 2015; Spector, 2006; Spiegeleire, Maas & Sweijs, 2017). There is, somehow, a dichotomic vision towards this topic and, for that reason, we can say that there are several advantages and benefits linked to AI (mentioned in AI application areas), but there are also a number of potentially dangerous disadvantages (Armstrong, Bostrom & Shulman, 2016; Armstrong & Sotala, 2015). We are talking about labour displacement (or even suppression), negative economic impact (due to an increase in economic inequality between developed and underdeveloped countries), loss of privacy, societal disruptions, machine bias, automated surveillance, among others (Atkinson, 2016; Spiegeleire, Maas & Sweijs, 2017). This exponential development is somehow subject of concern and entails risks, especially when regarding governmental and national security aspects (Allen & Chan, 2017).

Furthermore, these concerns are exalted by illustrious identities in the field of technology, such as Elon Musk and Stephen Hawking:

“Success in creating effective AI, could be the biggest event in the history of our civilization. Or the worst. (...) Unless we learn how to prepare for, and avoid, the potential risks, AI could be the worst event in the history of our civilization. (...) It could bring great disruption to our economy. (...) I am an optimist and I believe that we can create AI for de good of the world.

We simply need to be aware of the dangers, identify them, employ the best possible practice and management, and prepare for its consequences well in advance."

(Stephen Hawking at the Web Summit 2017 technology conference in Lisbon, Portugal)

"I think we should be very careful about artificial intelligence. (…) Increasingly scientist 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. With artificial intelligence, we are summoning the demon."

(Elon Musk at the MIT Aeronautics and Astronautics department's Centennial Symposium, 2014)

Now we wonder… Why is there only a need for regulation in capital markets and not in technological markets? An example for this is the market of Crypto coins (the known *Bitcoins*), which is an unstable and volatile market that originated several parallel interests. Technology is ahead of regulation, instead of existing regulators to act in a preventive way in this market. Right now, we are dealing with a reactive mentality instead of a proactive one. From this, urges the need for control and regulation that focuses on defining rules and limits towards AI, controlling possible risks and preventing catastrophes (Ramamoorthy & Yampolskiy, 2018; Spiegeleire, Maas & Sweijs, 2017).

All these conclusions are taken within an experts' community, but it is well known that this evolution will affect society in general, either negatively or positively. And how do people really feel about AI and its evolution?

1.2. Emotions

Before we move on to the answer for the question stated above, it is important to clarify the concept of emotions, for further understandings.

Considered to be central to the field of psychology, emotions are more complex than we might think and its definition is equally complex and hard to stipulate (Arriaga, 2010; Reeve, 2009; Shiota & Kalat, 2007). They are considered to be multidimensional, existing as biological, subjective, purposive, and social phenomena (Izard, 1993; Reeve, 2009). However, emotion is not defined by any of these dimensions individually, but by the sum of its parts. Reeve (2009) declares that "emotion is that which choreographs the feeling, arousal, purposive, and expressive components into a coherent reaction to an eliciting event" (p. 301). In fact, the definition of emotion is often misleading by the definition of feelings and *vice-versa*. The truth is that they complement each other, but it is important to understand the differences. Damasio (2003) explains that, in a very raw conception, emotions come first and feelings come after. Feelings are perceptions and, somehow, the expression of emotions. They emerge when our body conforms to the characteristics of a certain emotion, let's say, when we are conscious of our emotions, enabling us to *feel* in a certain way (Damasio, 1994).

Furthermore, in an attempt to build an integrative and consensual definition of emotion, Kleinginna and Kleinginna (1981) came to the following conclusion: "emotion is a complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems, which can (a) give rise to affective experiences such as feelings of arousal, pleasure/displeasure; (b) generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labelling processes; (c) activate widespread physiological adjustments to the arousing conditions; and (d) lead to behaviour that is often, but not always, expressive, goal-directed, and adaptive" (p. 355).

And what really causes an emotion?

We know that what triggers an emotion is a significant stimulus event, however, it is not clear if they are primarily biological or cognitive phenomena (Reeve, 2009), yet we know both enter de equation. They work as an integrative system, reuniting social, cultural, and history of the individual, as well as evolutionary, phylogenetic, and history of the species (Buck, 1984; Levenson, 1994; Reeve, 2009). Also, Plutchik (1985) sees emotion as a process formed by a succession of events that converge into a complex feedback system.

And how do emotions manifest?

As explained by Arriaga (2010), there are several ways in which an emotion can manifest, such as: expressive behaviour (facial expressions, body language, vocalizations), physiologically (heart rate, breathing, blood pressure, muscle tension, electrodermal activity), and neurologically (evoked potentials).

We have by now given a holistic knowledge of what emotions are, but we have not yet mentioned how many emotions exist and which ones there are. Emotions can be classified as primary/basic (in a more biological perspective) or secondary (in a more cognitive perspective). Fear, anger, disgust, sadness, joy, and interest are considered to be the basic emotions (Reeve, 2009). Some authors also refer contempt, surprise, shame, and guilt (Ekman, 1992; Izard, 1991). Secondary emotions are the more complex ones, evolving a combination of primary emotions and other external factors. Examples of secondary emotions can be jealousy, envy, pride, love, gratitude, and so on (Lazarus, 1991; Reeve, 2009).

There are several models that help to understand emotions. We highlight Russel's Circumplex Model of Affect (Russel, 1980), which is commonly used to test emotional facial expressions, and affective states; Plutchik's Wheel of Emotions (Plutchik, 1980), a three-dimensional model that describes the relations among emotions, enabling us to understand how complex emotions interact and change over time; and a more recent one, Lövheim's Cube of Emotions (Lövheim, 2012), a three-dimensional model which explains eight basic emotions through combinations of neurotransmitters dopamine, noradrenaline, and serotonin.

According to Roseman, Antoniou and Jose (1996), emotions can be differentiated into categories: negative emotions, and positive (or non-negative) emotions. In negative emotions we find emotions such as fear, sadness, frustration, disgust, anger, contempt, shame, guilt, dislike, and regret. In positive emotions we find hope, joy, relief, liking, and pride. Surprise, however, stands in both negative and positive categories, due to the fact that it is caused by an unexpected circumstance.

1.2.1. Cognitive aspects of emotion

As previously mentioned, emotions do not only emerge from biological aspects, but they also emerge from information processing, social interaction, and cultural contexts (Reeve, 2009). Besides, complex emotions cannot be assessed by means of purely biological analysis, such as facial expressions, or endocrine systems activity. So, how can we understand these emotions?

From here emerges the concept of *appraisal*, which stands for an estimate of the personal significance of a certain event (Reeve, 2009). Several authors have focused on this thematic, and we will highlight two of them. The following figures illustrate these cognitive emotion's mechanism in a very understandable way, through Arnold's Appraisal Theory of Emotion (Arnold, 1960), and Lazaru's Complex Appraisals (Lazarus, 1991), respectively.

Figure 1.1 - Arnold's Appraisal Theory of Emotion (Reeve, 2009, p.345)

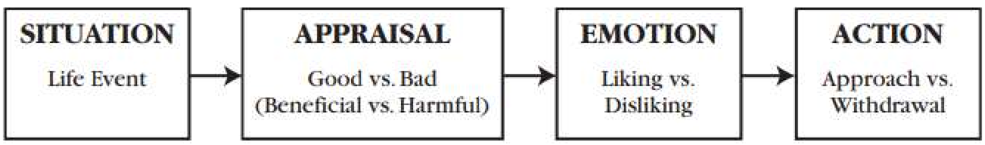
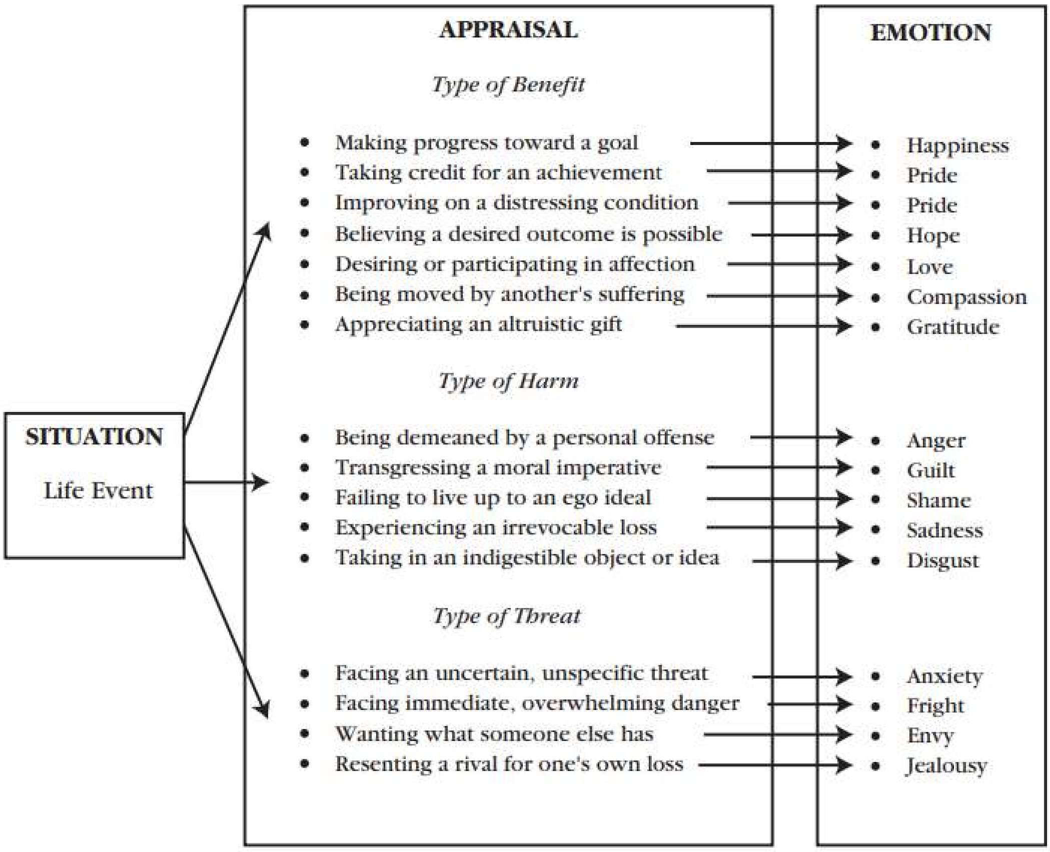


Figure 1.2 - Lazarus' Complex Appraisals: Types of Benefit, Harm, and Threat
(Reeve, 2009, p. 347)



1.3. Acceptance and Emotions towards AI and its evolution

Now that we have clarified the concept of emotions, we may resume the question previously asked and reflect about it: how do people really feel about AI and its evolution?

First, it is important to understand if people really are familiar with the concept of AI and how they accept it. Indeed, nowadays AI is present even in many smartphones, however, the public in general still has a very poor knowledge and understanding of technology and AI specifications (Atkinson, 2016). This acceptance towards AI may vary in two extremes, that is, people tend to have whether a very positive attitude towards AI, or a very negative one (Crowed & Friess, 2013), meaning people may recognise that AI might be extremely beneficial in several fields, as they may also hold onto the belief that AI will destroy humanity.

There are a few studies regarding technology acceptance, for example, Davis, Bagozzi & Warshaw (2016), that even proposed a Technology Acceptance Model (TAM). However, this study refers to technology in general, and does not mention AI specifically. On the other hand, there are, in fact, several studies that analyse people's acceptance towards robots (Breazeal & Scassellati, 2000; Cañamero & Fredslund, 2000; Dautenhahn, 2004; Hancock, Billings &

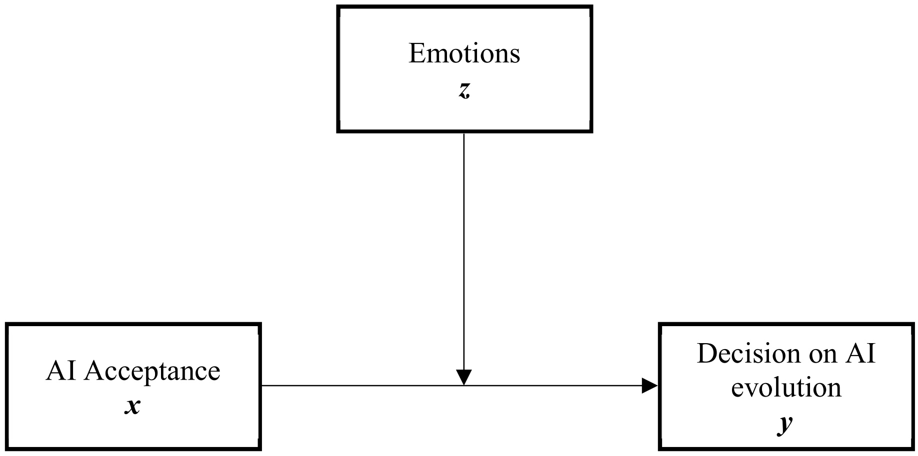
Schaefer, 2011; Kaplan, 2004; Nadel et al., 2006; Ray, Mondada & Siegwart, 2008; Nomura, Suzuki, Kanda & Kato, 2006; Partala & Surakka, 2004). All these studies concluded that individuals tend to accept robots in a very positive way and give them a very positive status. This, somehow, makes sense. If we look back at social psychology, we stand before several theories, and we highlight the Contact Hypothesis (Allport, 1954) which claims that under certain circumstances, as intergrupal contact increases, the preconceptions and prejudices among groups will decrease. This hypothesis is only valid when groups share similar status and characteristics. Having this said, we must not fully extend this theory when speaking of interactions between humans and AI. However, due to what is found in literature, it seems to apply in a holistic way, meaning that acceptance towards AI probably increases when the knowledge and contact with it also increases. However, as previously explained, we must remember that robots – the main focus of the existent studies (Breazeal & Scassellati, 2000; Cañamero & Fredslund, 2000; Dautenhahn, 2004; Hancock, Billings & Schaefer, 2011; Kaplan, 2004; Nadel et al., 2006; Ray, Mondada & Siegwart, 2008; Nomura, Suzuki, Kanda & Kato, 2006; Partala & Surakka, 2004) regarding this theme – belong to robotics, which is one of the fields of AI. From here we consider that studying only robots' acceptance is quite reductive, considering AI can assume so many diverse forms and act in so many different fields. Robotics is just a small piece of the puzzle and AI does not have to be confined into a human or animal-like figure. AI equals to algorithms combined to something grateful and capable of achieving so much more that public in general actually knows. These acceptance studies (Breazeal & Scassellati, 2000; Cañamero & Fredslund, 2000; Dautenhahn, 2004; Hancock, Billings & Schaefer, 2011; Kaplan, 2004; Nadel et al., 2006; Ray, Mondada & Siegwart, 2008; Nomura, Suzuki, Kanda & Kato, 2006; Partala & Surakka, 2004) focus mainly on whether people consider robots/technology to be positive or negative, but they do not focus on truly perceiving what kind of emotions are triggered by AI, meaning what people actually feel towards this theme. Instead, the existing studies that relate emotions and AI tend to be more specific and technical, often focusing on how to confer human emotions to artificial agents (Cañamero & Fredslund, 2000), while others only focus on studying people's anxieties towards humanoid robots (Nomura, Suzuki, Kanda & Kato, 2006; Szollozy, 2017).

And what about what people really feel and thing about this thematic? Should not we try to understand people's raw perceptions of AI and its evolution? Should not we provide society a holistic knowledge regarding AI and its future perspectives?

It is important to understand society's perspectives regarding AI today and AI in the future.

1.4. Research model and propositions

Figure 1.3 - Research Model



Based on literature review, our research model suggests that emotions moderate the effect of AI acceptance on the decision towards AI evolution, and we intend to understand how this relationship processes.

According Lazarus (1991) and Reeve (2009), negative emotions are a typical response to uncertainty and the unknown. Therefore:

Proposition 1 (P1): As the progressive contact with AI deepens (as enacted by exhibiting the stimuli throughout all three times), we expect negative emotions to diminish.

In addition, considering that society in general has a poor knowledge and understanding of AI (Atkinson, 2016):

Proposition 2 (P2): As the progressive contact with AI deepens (as enacted by exhibiting the stimuli throughout all three times), we expect non-negative emotions to increase.

Therefore, due to the progressive contact and knowledge regarding AI:

Proposition 3 (P3): We expect non-negative emotions to significantly and positively moderate the effect of AI acceptance on the decision towards AI evolution.

CHAPTER II – Method

2.1. Data Analysis Strategy

Considering this research combined both quantitative and qualitative methodologies, the nature of each variables will determine the data analysis specifications.

For quantitative variables, we will submit the variables "emotions" and "acceptance" to a confirmatory factor analysis (using *IBM SPSS Amos*) in order to assess their psychometric quality. According to Hair, Black, Babin and Anderson (2010), the following criteria should be applied so as to consider the factorial solutions as valid: 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. Also, PCFI is considered to be better when values are closer to 1, so we will report that value as well, in order to decide on the factorial solution parsimony. Additionally, if the previous criteria do not apply, we will use Lagrange Multipliers to identify possible biases from certain items, suggesting their removal (Hair, Black, Babin & Anderson, 2010).

In order to analyse the research models and propositions, our first thought was to conduct moderations using PROCESS v2.16 (Hayes, 2013), but that brought us a limitation due to the fact that we were not able to control other variables' effects (i.e.: "sex"). For that reason, we will conduct those moderations using the method proposed by Baron and Kenny (1986), so we can control the mentioned effects. The variable "sex" will then be controlled by means of a hierarchical regression so we can understand if it adds value to the initial moderation models. All of the analysis mentioned above will be achieved by using *Software IBM SPSS Statistics 24*. Furthermore, we will use software Modgraph (Jose, 2013) for a better understanding of the significative moderation models with significative interaction effects.

For qualitative variables, we will extract a global appreciation from the answers that participants gave to both questions asked, and we will categorise those answers in order to build a new set of variables, and further conduct frequencies analysis for each category.

2.2. Sample

The sample used in this study was made following a convenience and also snowball approach, considering we asked our friends and a few known people to answer the survey and those people recommended others. By doing it this way, we managed to gather the sample quickly, saving us a lot of time. The nature of this type of sampling advises caution when analysing data and interpreting findings and assuming them as externally valid. A total of 205 individuals composes the sample, of which 106 are male. The majority of the sample is single (92.2%), with ages ranging from 18 to 36 years old, averaging 23.7 years old. Table 2.1 shows the 11 categories in which the participants' area of activity (professional or academic) settles, and we can see that the larger group in the sample (27.3%) works/studies in the technology field.

Table 2.1 – Frequencies of participants' area of activity

	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	

2.3. 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 powered by Qualtrics, which started with acknowledging the nature of the study and giving the informed consent. Next, the participants were presented with an emotions scale to rate their baseline emotional status (e.g. sliding a bar from 0 to 100 according to each emotion intensity felt at that moment). This was the first moment that intended to establish the baseline measures.

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

The third moment started with video #2 offering examples of nowadays AI applications. Upon completion, participants registered how they felt emotionally while watching the video. For that purpose, they used the same scale as in moment one.

The fourth moment started with exhibiting video #3 that showed three testimonies from three renowned individualities with divergent opinions concerning AI. The first question after watching the video concerned asking whether each participant have already seen any of the testimonies (this is a control variable). After we asked each participant to state their degree of agreement in a six-point Likert scale (1 Total disagree to 6 Total agree) regarding the evolution of AI, the evolution of AI with regulation, and the evolution of AI without regulation. After participants took a stand, we asked them to fill in the emotions scale again always taken as reference the moment they were watching video #3. Lastly, participants gave some sociodemographic data for sample characterization.

The second phase of this study comprehended a qualitative approach, meaning the participants were submitted to a short semi structured interview, covering two main questions: 1 – "Considering all the questions you answered and the videos you saw, please justify your decision and clarify your vision regarding AI evolution."; 2 – "During the experiment and while watching the videos, was(were) there any emotion(s) that stood out?".

The reason why we opted for a mixed methodology (quantitative plus qualitative) settles in two main aspects. First, it could work as a *plan b* if the quantitative part failed for some reason (e.g.: assessing the emotions properly). The second aspect is that, considering that we are dealing with an innovative (and somehow controversial) theme, it became much more interesting to really listen to what the participants had to say, rather than limit their voice to quantitative measures. In addition, as stated in literature review, psychometric measures are not always enough to understand complex emotions, and this approach allows participants to express themselves freely. This approach adds much more richness to the study and provides interesting and valuable information.

2.3.1. Stimuli

The stimuli we used for this experiment were three videos. The following sections explain the contents of those videos and each video is illustrated with screenshots of itself.

Video #1:

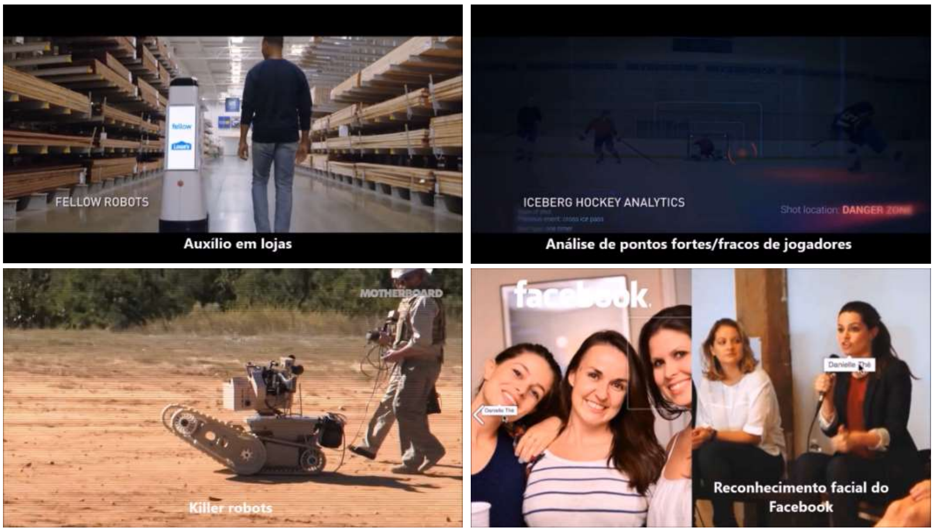
Video #1 has 1'11'' of duration and provides an AI definition, clarifying what it is and which kind of forms it may assume. The purpose of this video is to state an equal baseline of knowledge for all participants.

Figure 2.1 - Screenshot segments of Video #1

Video #2:

Video #2 lasts 1'30'' and illustrates a few examples regarding AI's utility and how it is used in today's days (e.g. in health, music, transportation, etc). This video makes the participants confront their previous knowledge with reality, in terms of AI's utility.

Figure 2.2 - Screenshot segments of Video #2



Video #3:

Video #3 is the longest (6’57’’ of duration) and shows three testimonies from three renowned individualities with divergent opinions concerning AI. 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). These testimonies were intended to face the participant with a critical positioning towards the future of AI.

Figure 2.3 - Screenshot segments of Video #3

2.5. Measures

The following topics describe the specifications of each variable under study - “emotions”, “AI acceptance”, and “decision” -, and how we will measure them.

2.5.1. Emotions

When searching for an instrument to measure the variable “emotions”, we struggled due to the number of instruments that, in fact, exist, but are quite extensive as to the number of items, for example, EAS - Emotion and Assessment Scale by Carlson, Collins, Stewart, Porzelius, Nitz and Lind (1989). Using a scale like this would overextend the time of the survey, what would encourage the participants’ withdrawal and also some difficulties in obtaining voluntary participations. Considering these aspects, we searched for a more objective scale that would not take too long for the participants to answer and so emotions were measured with eight items extracted from Lövheim's (2012) cube of emotions, following the example of Moyle, Moyle, Bec and Scott (2017). This is in fact an interesting instrument to use in this experiment, due to the fact that it has been applied in the development of AI and several attempts to reach the biologically-inspired artificial emotions (Hsu, Chen & Heh, 2014; Talanov & Toschev, 2014; Vallverdú, Talanov, Distefano, Mazzara, Tchitchigin & Nurgaliev, 2016). The items that we used were: shame/humiliation, distress/anguish, fear/terror, anger/rage, contempt/disgust, surprise, enjoyment/joy, interest/excitement and participants were asked to register on a Visual Analytic Scale (VAS) from 0 to 100 points (0 is equivalent to 'nothing') how much they were they feeling each emotion. These measurements took place three times: 1) as a baseline concerning the extent the participant was experiencing each emotion at the time of response, 2) after seeing video #1 and video #2 (what is AI and what are its applications) participants were requested to rate each emotion experienced while watching it, and 3) after seeing video #3 (personalities positioning concerning AI) and referring to subjective emotion experience while watching it. Psychometrically the emotions should organize around dimensions (as stated in literature review, one concerning positive/non-negative and the other negative emotions) which we can use for comparison purposes between moments if the factor structures hold across the three measurements. Table 2.2 reports findings for confirmatory factor analyses and the respective reliabilities

Table 2.2 - Confirmatory factor analysis of the variable Emotions in all three times

	T1	T2	T3
KMO	.752	.754	.743
Bartlett's X2	507.801, 28 gl, p<.001	663.804, 28 gl, p<.001	631.230, 28 gl, p<.001
Explained variance	60.8%	65.3%	64.2%
F1 Cronbach alpha	.801	.842	.815
F2 Cronbach alpha	.728	.714	.694
CFA	CMIN/DF=2.247, p<.01; CFI=.952, TLI=.929, PCFI=.646, RMSEA=.078.	CMIN/DF=3.835, p<.01; CFI=.921, TLI=.877, PCFI=.592, RMSEA=.118.	CMIN/DF=3.301, p<.01; CFI=.933, TLI=.895, PCFI=.600, RMSEA=.107.
Lagrange Multipliers	Suggest removal of "surprise"	Suggest removal of "surprise"	Suggest removal of "surprise"
CFA	CMIN/DF=2.004, p=.014; CFI=.969, TLI=.953, PCFI=.646, RMSEA=.070.	CMIN/DF=1.963, p=.020; CFI=.978, TLI=.965, PCFI=.606, RMSEA=.069.	CMIN/DF=1.220, p=.257; CFI=.995, TLI=.992, PCFI=.616, RMSEA=.033.
F1 Cronbach alpha	.801	.842	.815
F2 r_{SB}	.812	.804	.756

Table 2.3 - Factor loadings across all three measurement moments

Emotions	**T1f1**	**T1f2**	**T2f1**	**T2f2**	**T3f1**	**T3f2**
Anger/ Rage	,828	,048	,864	-,022	,856	-,003
Distress/Anguish	,817	,027	,848	,031	,769	-,043
Fear/Terror	,724	,225	,771	,036	,755	,007
Contempt/Disgust	,696	,067	,791	-,125	,792	-,061
Shame/Humiliation	,666	-,024	,652	,200	,734	,064
Interest/Excitement	-,081	,888	-,164	,870	-,181	,836
Enjoyment/Joy	,029	,878	-,025	,861	-,087	,858
Surprise	,242	,611	,341	,648	,382	,669

Extraction Method: Principal Component Analysis

Rotation Method: Varimax with Kaiser's Normalization

a. Rotation converged in 3 iterations

As the factor analyses are steady in all the three moments, we reason the emotions are comparable across time. Likewise, because "surprise" did not factorize, we will use it separately. So, emotions are measured on three aspects: negative emotions, non-negative emotions, and surprise.

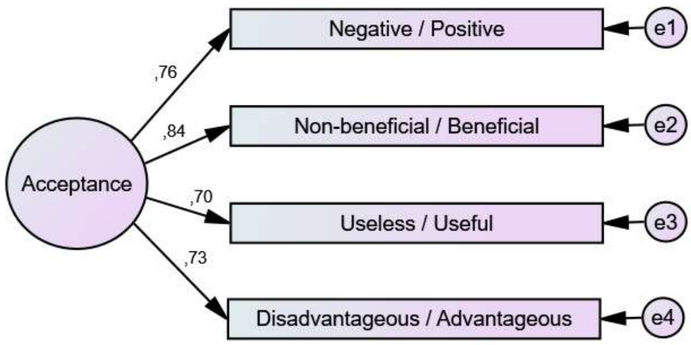
Besides using the scale mentioned above, we intended to assess the participants' emotions through their expressions while exposed to the stimuli, using FaceReader™ from software NOLDUS. For that, we recorded the participants' face and uploaded those recordings to the software mentioned, however, as we were not familiar enough we the software, we tested in ten participants before conducting that analysis through the whole sample. This method failed to work, so we decided to quit this option.

2.5.2. AI Acceptance

The instrument to measure the variable "acceptance" was based on the Fast Form of the Technology Acceptance Model (TAM) proposed by Chin, Johnson and Schwarz (2008), which was created from its Original Form introduced by Davis (1986). This model aimed to understand what determined the acceptance of a new system, especially computers, through the users' perceptions of it (Chin, Johnson & Schwarz, 2008). This could help to predict and identify if a certain new system would be acceptable or not for the population, and pursue with it (Davis, Bagozzi & Warshaw, 1989). Although this instrument refers to 'technology' in general, it gets somewhat specific to everyday gadgets used by people in general. However, when it comes to AI, it is impossible to reduce it to a simple day-to-day gadget and we struggled to dovetail the TAM instrument on its extent to the AI theme. Having that said, we decided to adapt a few items and build the scale by ourselves. "Acceptance" was then measured with six items through a six-point Likert scale (1-negative/6-positive, 1-non-beneficial/6-beneficial, 1-useless/6-useful, 1-ineffective/6-effective, 1-inefficient/6-efficient, 1-disadvantageous/6-advantageous). This scale was presented to the participants after video #1.

A confirmatory factor analysis showed two of the items (efficient and effective) had an unacceptable low communality and were excluded to have a single factor valid solution (KMO=809, X^2 Bartlett = 323.777, 6 gl, p=.000) explaining 68.1% variance after rotation. The CFA showed valid fit indices for this 4-item factor (CMIN/DF=1.237, p=.29, CFI=.999, TLI=.996, PCFI=.333, RMSEA=.034) which is also reliable (Cronbach's alpha=.842).

Figure 2.4 - CFA AI Acceptance



2.5.2. Decision on AI evolution

This variable will be named "decision" and is separated into three different variables, concerning three different decisions: "against/in favour of AI evolution", "against/in favour of AI evolution with regulators" and "against/in favour of AI evolution without regulators". Participants had to rate each of these variables with a 6-point Likert scale (1 – in favour; 6 – against). This variable took place after video #3, so it allows us to understand one's position regarding this subject, assuring that every participant got the exact same information throughout the experiment. In order to promote a clear visualization and consequent interpretation of the following results, these variables were recoded into: 1 – against and 6 – in favour.

CHAPTER III – Results

Chapter III will present all results obtained throughout the entire analysis. It will begin by presenting descriptive and bivariate statistics concerning quantitative variables, so we can understand the correlations among variables and their relevance for the study.

Further specific analysis regarding all quantitative and qualitative variables under study will be exposed.

3.1. Descriptive and Bivariate Analysis

Descriptive statistics show, for each variable, the sample number (N), minimum and maximum values registered on each response, mean and standard deviation. Bivariate statistics show the relationship between all variables under study, evidencing some relevant correlations. The variables covered in both descriptive and bivariate statistics are the socio-demographic variables ("sex"), "negative emotions", "non-negative emotions" and "surprise" in all three times of measurement (the items that compose these variables will also be reported separately and will be considered as separate variables for further analysis), "acceptance", and the three decisions regarding AI evolution ("decision against/in favour", "decision against/in favour with regulation", "decision against/in favour without regulation"). All descriptive analysis is shown in table 3.1 below.

Table 3.1 shows us clear changes through all three times. The average of negative emotions tends to rise from T1 to T2, and from T2 to T3. The reverse happens with non-negative emotions, on which the average tends to decrease throughout all three times. Surprise's average has its peak at T2 (where participants were exposed to existent AI applications) and then decreases in T3. The same applies to emotions individually in a very similar way, while negative emotions tend to rise throughout the times, and non-negative emotions tend to decrease. Interest, on the other hand, remains high through all three times (reaching its peak at T2, alike surprise). When looking at decisions on AI evolution, we can see that the decision on AI evolution is relatively high, being superior when adding regulators. The opposite happens when considering AI evolution without regulators, on which the average is sharply low.

Table 3.2 reports Spearman correlations, and we can see that there are several significant correlations. However, we must keep in mind that all variables are show, that is, we show emotions as composite variables (factors, as explained previously), and we also report them separately in that analysis, so it is only predictable that they tend to relate within each other.

Table 3.1 - Descriptive Statistics

Variable	N	Min-Max	Mean	Std. Deviation
Sex	205	-	-	-
T1 – Negative Emotions	205	0.00-59.80	6.78	10.60
T1 – Non-Negative Emotions	205	0.00-100	54.20	25.60
T1 – Surprise	205	0.00-100	17.80	24.23
T2 – Negative Emotions	205	0.00-69.00	9.19	13.52
T2 – Non-Negative Emotions	205	0.00-100	46.66	25.61
T2 – Surprise	205	0.00-100	38.48	29.25
T3 – Negative Emotions	205	0.00-90.80	13.20	14.70
T3 – Non-Negative Emotions	205	0.00-97.00	37.14	23.55
T3 – Surprise	205	0.00-97.00	25.68	23.89
T1 – Distress/Anguish	205	0.00-100	10.61	18.35
T1 – Fear/Terror	205	0.00-70.00	7.29	14.36
T1 – Anger/Rage	205	0.00-82.00	6.38	13.60
T1 – Contempt/Disgust	205	0.00-61.00	3.47	9.54
T1 – Enjoyment/Joy	205	0.00-100	51.53	29.07
T1 – Interest/Excitement	205	0.00-100	56.87	26.70
T1 – Shame/Humiliation	205	0.00-85.00	6.16	13.78
T2 – Distress/Anguish	205	0.00-90.00	11.35	18.22
T2 – Fear/Terror	205	0.00-100	17.05	22.01
T2 – Anger/Rage	205	0.00-84.00	6.22	15.51
T2 – Contempt/Disgust	205	0.00-96.00	6.59	16.35
T2 – Enjoyment/Joy	205	0.00-100	35.98	28.03
T2 – Interest/Excitement	205	0.00-100	57.34	27.99
T2 – Shame/Humiliation	205	0.00-70.00	4.84	12.90
T3 – Distress/Anguish	205	0.00-97.00	20.47	23.89
T3 – Fear/Terror	205	0.00-100	26.83	26.55
T3 – Anger/Rage	205	0.00-96.00	7.14	14.93
T3 – Contempt/Disgust	205	0.00-92.00	6.73	15.82
T3 – Enjoyment/Joy	205	0.00-97.00	25.68	23.89
T3 – Interest/Excitement	205	0.00-100	48.61	28.59
T3 – Shame/Humiliation	205	0.00-76.00	4.81	11.50
Acceptance	205	1-6	4.87	0.84
Against/In Favour	205	1-6	4.02	1.14
Against/In Favour With Regulators	205	1-6	5.02	1.12
Against/In Favour Without Regulators	205	1-6	1.79	1.18

Table 3.2 - Bivariate Spearman Correlations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Sex	1																
2. T1 – Negative Emotions	-.030	1															
3. T1 – Non-Negative Emotions	.068	.070	1														
4. T1 – Surprise	-.284**	.160*	.316*	1													
5. T2 – Negative Emotions	.096	.338**	.139*	.199**	1												
6. T2 – Non-Negative Emotions	-.231**	.071	.461**	.226**	-.015	1											
7. T2 – Surprise	.020	.278**	.262**	.185**	.263**	.398**	1										
8. T3 – Negative Emotions	.069	.364**	.220**	.092	.690**	.064	.228**	1									
9. T3 – Non-Negative Emotions	-.268**	-.005	.374**	.296**	-.070	.692**	.274**	-.059	1								
10. T3 – Surprise	-0.89	.081	.277**	.226**	.169*	.277**	.489**	.237**	.403**	1							
11. T1 – Distress/Anguish	-.015	.777**	.017	.090	.282**	.026	.177*	.315**	-.070	.032	1						
12. T1 – Fear/Terror	-.018	.728**	.153*	.167*	.348**	.104	.161*	.388**	-.009	.091	.520**	1					
13. T1 – Anger/Rage	-.102	.747**	-.054	.155*	.321**	-.054	.170*	.234**	-.058	-.016	.553**	.541**	1				
14. T1 – Contempt/Disgust	-.060	.543**	-.023	.243**	.336**	.023	.236**	291**	.013	.113	.443**	.434**	.588**	1			
15. T1 – Enjoyment/Joy	-.106	.112	.917**	.289**	.136	.416**	.265**	.239**	.334**	.241**	.053	.171**	.012	.041	1		
16. T1 – Interest/Excitement	.029	-.003	.891**	.276**	.128	.417**	.191**	.169*	.339**	.253**	-.038	.087	-.113	-.094	.656**	1	
17. T1 – Shame/Humiliation	-.009	.649**	-.023	.284**	.213**	.057	.179*	.154*	.043	.153*	.423**	.462**	.512**	.400**	-.034	-.027	1
18. T2 – Distress/Anguish	.104	.357**	.065	.156*	.794**	-.031	.266**	.613**	-.109	.183**	.350**	.367**	.330**	.353**	.074	.056	.236**
19. T2 – Fear/Terror	.113	.285*	.163*	.185**	.892**	.011	.243**	.641**	-.052	.135	.218**	.322**	.233**	.260**	.166**	.124	.127
20. T2 – Anger/Rage	-.037	.338**	.014	.129	.662**	-.032	.253**	.435**	-.108	.203**	.353**	.347**	.420**	.418**	.034	-.015	.273**
21. T2 – Contempt/Disgust	-.044	.294**	-.034	.140*	.612**	-.061	.161*	.424**	-.069	.193**	.269**	.352**	.323**	.452**	-.025	-.053	.344**
22. T2 – Enjoyment/Joy	-.222**	.153*	.435**	.196**	.050	.905**	.398**	.119	.620**	.264**	.055	.161*	.028	.066	.408**	.360**	.144*
23. T2 – Interest/Excitement	-.182**	-.007	.414**	.224**	-.046	.914**	.343**	.024	.635**	.259**	-.001	.041	-.114	-.005	.359**	.407**	-.019
24. T2 – Shame/Humiliation	.017	.231**	.065	.259**	.550**	.040	.311**	.283**	-.008	.273**	.192**	.325**	.291**	.390**	.033	.077	.364**
25. T3 – Distress/Anguish	.129	.243**	.205**	.039	.608**	.028	.192**	.841**	-.103	.220**	.286**	.261**	.117	.229**	.196**	.183**	.063
26. T3 – Fear/Terror	.121	.304**	.247**	.071	.575**	.088	.222**	.904**	-.035	.171*	.211**	.338**	.181**	.165*	.272**	.176*	.065
27. T3 – Anger/Rage	-.034	.400**	.054	.174*	.598**	-.060	.171*	.637**	-.108	.295**	.327**	.463**	.354**	.457**	.069	.023	.373**
28. T3 – Contempt/Disgust	-.079	.299**	-.059	.185**	.520**	-.084	.111	.522**	-.127	.188**	.316**	.345**	.256**	.487**	-.045	-.073	.388**
29. T3 – Enjoyment/Joy	-.191**	.046	.366**	.259**	.026	.598**	.324**	.009	.850**	.419**	-.035	.039	-.004	.072	.348**	.293**	.132
30. T3 – Interest/Excitement	-.282**	-.025	.295**	.274**	-.125	.620**	.193**	-.101	.920**	.330**	-.071	-.040	-.064	-.005	.249**	.295**	-.088
31. T3 – Shame/Humiliation	.016	.379**	-.016	.190**	.458**	-.019	.207**	.450**	-.078	.240**	.335**	.373**	.339**	.501**	.002	-.040	.479**
32. Acceptance	-.075	-.056	.026	.067	-.127	.225**	.062	-.086	.152*	.066	-.064	-.125	.001	-.094	.001	.071	.022
33. Against/In Favour	-.294**	-.060	.002	.094	-.309**	.295**	-.024	-.405**	.423**	.043	-.056	-.071	-.087	-.087	-.042	.048	.059
34. Against/In Favour With Regulators	-.114	-.020	.066	.065	-.134	.165*	.081	-.229**	.237**	-.002	-.037	-.048	.032	-.053	.070	.060	-.009
35. Against/In Favour Without Regulators	-.78	-.023	-.015	.114	.001	-.010	-.149*	-.126	.165*	.026	-.065	-.002	-.050	.036	-.037	.024	.084

* p<0.05; ** p<0.01

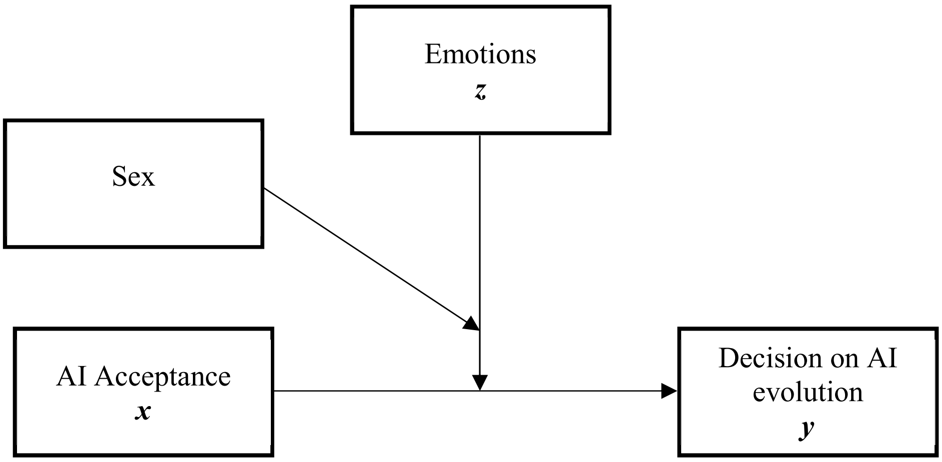
18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
1																	
.618**	1																
.570**	.545**	1															
.500**	.453**	.660**	1														
.018	.074	.075	.062	1													
-.052	-.029	-.115	-.141*	.666**	1												
.502**	.370**	.653**	.557**	.079	.019	1											
.641**	.519**	.354**	.356**	.052	.022	.269**	1										
.446**	.628**	.319**	.278**	.145*	.035	.161*	.647**	1									
.627**	.483**	.642**	.602**	.034	-.122	.491**	.512**	.448**	1								
.477**	.397**	.576**	.727**	.018	-.146*	.500**	.380**	.323**	.709**	1							
-.041	.044	-.001	.038	.644**	.450**	.100	-.049	.035	.001	.006	1						
-.130	-.114	-.154*	-.119	.482**	.640**	-.088	-.126	-.087	-.149*	-.188**	.592**	1					
.493**	.337**	.487**	.470**	.044	-.054	.552**	.353**	.278**	.634**	.621**	.034	-.132	1				
-.125	-.100	-.051	-.104	.183**	.222**	-.023	-.095	.012	-.088	-.106	.081	.162*	-.159*	1			
-.247**	-.339**	-.202**	-.183**	.215**	.307**	-.094	-.341**	-.415**	-.271**	-.172*	.309**	.424**	-.130	.307**	1		
-.144*	-.162*	-.047	-.096	.120	.171*	-.070	-.197**	-.183**	-.180**	-.123	.136	.256**	-.159*	.235**	.551**	1	
-.092	-.023	-.058	.048	-.034	.017	.025	-.122	-.181**	-.016	.094	.201**	.131	.052	.016	.239**	-.133	1

3.2. Quantitative Data

Considering variable "sex" had some correlations with other variables, we decided to incorporate it as a control variable in our research model. After running moderations with this variable, we realised that it added significant value to our initial model. For that reason, our research model now includes the variable sex as part of the relationship. We also tested a control variable concerning previous exposure to the stimuli (i.e. if the participants had already seen any of the videos) and it showed no relevance.

The updated research model is represented in figure 3.1. Please note that, as explained in Chapter I, this is the holistic model of our study, meaning all the analysis will comprehend all three emotion categories (negative, non-negative, and surprise), all three decisions on AI evolution (against/in favour, with regulators, and without regulators), through all three moments (T1, T2, and T3). For parsimony sake, we will only highlight and further discuss significative results.

Figure 3.1 - New Research Model



3.2.1. Analysis within the whole sample

The following tables (3.3, 3.4, and 3.5) report results obtained when conducting moderations with the whole sample, encompassing all variables mentioned above. The boxes in light grey mean the model is significative, whilst the boxes in dark grey mean the model is significative and that there is a significative interaction term.

Table 3.3 reports moderation findings regarding the moderator variable 'negative emotions', across all three times and all three decisions (dependent variables). As we can see, negative emotions seem to have a great impact on the relationship between AI acceptance and the decision towards its evolution. The strength of the model increases throughout the times, as we can see through the adjR^2. The models in T1 and T2 are significative with significative main effects (except when the dependent variable is 'decision without regulators'), however, it is only in T3 that the models show themselves as significative, with a significative interaction term.

Table 3.3 - Moderations with moderator 'Negative Emotions'

		T1	T2	T3
Negative Emotions	Against/In Favour	**Model:** adjR^2= .083 F(4,200)= 9.542, p<.001 **Sex:** B= -.631, t= -4.262, p<.001, 95%IC -.924, -.339 ***AI Acceptance*:** B= .352, t= 3.984, p<.001, 95%IC .178, .527 **Interaction term:** B= .010, t= 1.160, p=.247, 95%IC -.007, .026	**Model:** adjR^2= .257 F(4,200)= 18.600, p<.001 **Sex:** B= -.571, t= -4.117, p<.001, 95%IC -.844, -.297 ***AI Acceptance*:** B= .298, t= 3.618, p<.001, 95%IC .136, .460 **Interaction term:** B= -.014, t= -1.941, p=.054, 95%IC -.028, .000	**Model:** adjR^2= .363 F(4,200)= 30.044, p<.001 **Sex:** B= -.530, t= -4.137, p<.001, 95%IC -.783, -.277 ***AI Acceptance*:** B= .338, t= 4.443, p<.001, 95%IC .188, .488 **Interaction term:** B= -.016, t= -3.217, p=.002, 95%IC -.026, -.006
	With Regulators	**Model:** adjR^2= .049 F(4,200)= 3.617, p=.007 **Sex:** B= -.193, t= -1.254, p=.211, 95%IC -.496, .110 ***AI Acceptance*:** B= .309, t= 3.363, p=.001, 95%IC .128, .490 **Interaction term:** B= .003, t= .351, p=.726, 95%IC -.014, .020	**Model:** adjR^2= .082 F(4,200)= 5.582, p<.001 **Sex:** B= -.166, t= -1.094, p=.275, 95%IC -.465, .133 ***AI Acceptance*:** B= .274, t= 3.047, p=.003, 95%IC .097, .452 **Interaction term:** B= -.010, t= -1.297, p=.196, 95%IC -.025, .005	**Model:** adjR^2= .130 F(4,200)= 8.615, p<.001 **Sex:** B= -.150, t= -1.015, p=.311, 95%IC -.440, .141 ***AI Acceptance*:** B= .311, t= 3.549, p<.001, 95%IC .138, .483 **Interaction term:** B= -.019, t= -3.299, p=.001, 95%IC -.030, -.008
	Without Regulators	**Model:** adjR^2= .004 F(4,200)= 1.196, p=.314 **Sex:** B= -.190, t= -1.147, p=.253, 95%IC -.516, .137 ***AI Acceptance*:** B= -.037, t= -.375, p=.708, 95%IC -.232, .158 **Interaction term:** B= .014, t= 1.499, p=.136, 95%IC -.004, .032	**Model:** adjR^2= -.006 F(4,200)= .693, p=.597 **Sex:** B= -.170, t= -1.015, p=.312, 95%IC -.499, .160 ***AI Acceptance*:** B= -.058, t= -.585, p=.559, 95%IC -.254, .138 **Interaction term:** B= -.001, t= -.070, p=.944, 95%IC -.017, .016	**Model:** adjR^2= .029 F(4,200)= 2.511, p=.043 **Sex:** B= -.143, t= -.871, p=385, 95%IC -.466, .180 ***AI Acceptance*:** B= -.038, t= -.393, p=.695, 95%IC -.230, .154 **Interaction term:** B= -.015, t= -2.346, p=.020, 95%IC -.027, -.002

The moderation models using 'non-negative emotions' as the moderator are shown in the following table 3.4. The findings reported show that all models are significative with significative main effects (apart from the ones where the dependent variable is 'decision without regulators', which is only significative in T3), however, none of the models show a significative interaction term.

Table 3.4 - Moderations with moderator 'Non-Negative Emotions'

		T1	T2	T3
Non-Negative Emotions	**Against/In Favour**	**Model:** adjR^2= .138 F(4,200)= 9.193, p<.001 **Sex:** B= -.640, t= -4.277, p<.001, 95%IC -.935, -.345 ***AI Acceptance*:** B= .356, t= 3.973, p<.001, 95%IC .179, .533 **Interaction term:** B= -.002, t= -.715, p=.476, 95%IC -.009, .004	**Model:** adjR^2= .168 F(4,200)= 11.271, p<.001 **Sex:** B= -.527, t= -3.510, p=.001, 95%IC -.824, -.231 ***AI Acceptance*:** B= .297, t= 3.346, p=.001, 95%IC .122, .472 **Interaction term:** B= .001, t= .396, p=.693, 95%IC -.005, .007	**Model:** adjR^2= .236 F(4,200)= 16.724, p<.001 **Sex:** B= -.425, t= -2.930, p=.004, 95%IC -.711, -.139 ***AI Acceptance*:** B= .281, t= 3.325, p=.001, 95%IC .114, .447 **Interaction term:** B= .000, t= -.124, p=.901, 95%IC -.008, .007
	With Regulators	**Model:** adjR^2= .056 F(4,200)= 4.011, p=.004 **Sex:** B= -.210, t= -1.362, p<.175, 95%IC -.514, .094 ***AI Acceptance*:** B= .292, t= 3.155, p=.002, 95%IC .109, .474 **Interaction term:** B= .001, t= .433, p=.666, 95%IC -.005, .008	**Model:** adjR^2= .062 F(4,200)= 4.369, p=.002 **Sex:** B= -.129, t= -.822, p=.412, 95%IC -.439, .181 ***AI Acceptance*:** B= .284, t= 3.055, p=.003, 95%IC .101, .467 **Interaction term:** B= .004, t= 1.356, p=.177, 95%IC -.002, .011	**Model:** adjR^2= .084 F(4,200)= 5.689, p<.001 **Sex:** B= -.081, t= -.517, p=.606, 95%IC -.389, .227 ***AI Acceptance*:** B= .255, t= 2.799, p=.006, 95%IC .075, .434 **Interaction term:** B= -.003, t= -.870, p=.385, 95%IC -.011, .004
	Without Regulators	**Model:** adjR^2= -.004 F(4,200)= .818, p=.515 **Sex:** B= -.207, t= -1.239, p=.217, 95%IC -.537, .123 ***AI Acceptance*:** B= -.027, t= -.266, p=.791, 95%IC -.224, .171 **Interaction term:** B= -.004, t= --1.164, p=.246, 95%IC -.012, .003	**Model:** adjR^2= .005 F(4,200)= 1.231, p=.299 **Sex:** B= -.134, t= -.789, p=.431, 95%IC -.470, .201 ***AI Acceptance*:** B= -.038, t= -.378, p=.706, 95%IC -.237, .160 **Interaction term:** B= .007, t= 1.879, p=.062, 95%IC .000, .013	**Model:** adjR^2= .036 F(4,200)= 2.898, p=.023 **Sex:** B= -.055, t= -.324, p=.746, 95%IC -.388, .278 ***AI Acceptance*:** B= -.068, t= -.697, p=.487, 95%IC -.262, .125 **Interaction term:** B= .007, t= 1.680, p=.095, 95%IC -.001, .015

When conducting moderations using 'surprise' as the moderator (shown in table 3.5), all the models are significative, except when the dependent variable is 'decision without regulators'. However, only two models show a significative interaction term: in T3 when using 'decision against/in favour', and T2 when using 'decision with regulators'

Table 3.5 - Moderations with moderator 'Surprise'

		T1	T2	T3
Surprise	**Against/In Favour**	**Model:** adjR^2= .137 F(4,200)= 9.071, p<.001 **Sex:** B= -.638, t= -4.215, p<.001, 95%IC -.937, -.340 ***AI Acceptance*:** B= .345, t= 3.829, p<.001, 95%IC .168, .523 **Interaction term:** B= -.001, t= -.139, p=.890, 95%IC -.009, .008	**Model:** adjR^2= .138 F(4,200)= 9.183, p<.001 **Sex:** B= -.617, t= -4.140, p<.001, 95%IC -.912, -.323 ***AI Acceptance*:** B= .347, t= 3.771, p<.001, 95%IC .165, .528 **Interaction term:** B= .000, t= -.150, p=.881, 95%IC -.007, .006	**Model:** adjR^2= .162 F(4,200)= 10.850, p<.001 **Sex:** B= -.655, t= -4.450, p<.001, 95%IC -.946, -.365 ***AI Acceptance*:** B= .338, t= 3.881, p<.001, 95%IC .166, .510 **Interaction term:** B= -.009, t= -2.434, p=.016, 95%IC -.017, -.002
	With Regulators	**Model:** adjR^2= .044 F(4,200)= 3.363, p=.011 **Sex:** B= -.193, t= -1.227, p=.221, 95%IC -.502, .117 ***AI Acceptance*:** B= .303, t= 3.240, p=.001, 95%IC .118, .487 **Interaction term:** B= .001, t= .117, p=.860, 95%IC -.008, .010	**Model:** adjR^2= .074 F(4,200)= 5.076, p=.001 **Sex:** B= -.238, t= -1.563, p=.120, 95%IC -.538, -.062 ***AI Acceptance*:** B= .360, t= 3.840, p<.001, 95%IC .175, .545 **Interaction term:** B= .008, t= 2.471, p=.014, 95%IC .002, .015	**Model:** adjR^2= .045 F(4,200)= 3.389, p=.010 **Sex:** B= -.207, t= -1.337, p=.183, 95%IC -.512, .098 ***AI Acceptance*:** B= .301, t= 3.283, p=.001, 95%IC .120, .482 **Interaction term:** B= -.001, t= -.361, p=.719, 95%IC -.009, .006
	Without Regulators	**Model:** adjR^2= -.007 F(4,200)= .670, p=.614 **Sex:** B= -.182, t= -1.075, p=.284, 95%IC -.516, .152 ***AI Acceptance*:** B= -.024, t= -.239, p=.811, 95%IC -.223, .175 **Interaction term:** B= .006, t= 1.170, p=.243, 95%IC -.004, .015	**Model:** adjR^2= .005 F(4,200)= 1.245, p=.293 **Sex:** B= -.169, t= -1.019, p=.309, 95%IC -.497, .158 ***AI Acceptance*:** B= -.029, t= -.281, p=779, 95%IC -.231, .173 **Interaction term:** B= .001, t= .150, p=.881, 95%IC -.007, .008	**Model:** adjR^2= -.012 F(4,200)= .385, p=819 **Sex:** B= -.168, t= -1.004, p=317, 95%IC -.499, .162 ***AI Acceptance*:** B= -.048, t= -.480, p=.632, 95%IC -.244, .148 **Interaction term:** B= .002, t= .355, p=.723, 95%IC -.007, .010

3.1.2. Analysis with sample split by variable 'sex'

Due to the fact that the variable sex reported a significative impact in several moderation models, as depicted above, we have decided to split the data according to that variable, enabling us to conduct separated analysis and understand the existing differences. The models that prove to be significative, with a significative interaction term, will be illustrated using ModGraph.

The following tables report similar data to the previous ones, this time, separated by sex in order to compare male *versus* female.

As we can see through tables 3.6, 3.7, and 3.8, which correspond to all moderations across T1, some of the models are significant, yet none of the models show a significant interaction term.

*Table 3.6 - **T1**: Comparative moderations (by 'sex') with moderator 'Negative Emotions'*

	T1 – Negative Emotions		
	Against/In Favor	**With Regulators**	**Without Regulators**
Male	**Model:** adjR^2= .093 $F(3,102)$= 4.607, p=.005 ***AI Acceptance*:** B= .413, t= 3.152, p=.002, 95%IC .153, .672 **Interaction term:** B= .018, t= 1.040, p=.301, 95%IC -.017, .053	**Model:** adjR^2= -.010 $F(3,102)$= .669, p=.573 ***AI Acceptance*:** B= .133, t= .967, p=.336, 95%IC -.140, .406 **Interaction term:** B= -.013, t= -.717, p=.475, 95%IC -.050, .023	**Model:** adjR^2= .019 $F(3,102)$= 1.685, p=175 ***AI Acceptance*:** B= .300, t= 1.974, p=.051, 95%IC -.001, .602 **Interaction term:** B= .027, t= 1.316, p=.191, 95%IC -.014, .067
Female	**Model:** adjR^2= .050 $F(3,95)$= 2.708, p=.050 ***AI Acceptance*:** B= .308, t= 2.529, p=.013, 95%IC .066, .549 **Interaction term:** B= .012, t= 1.266, p=.209, 95%IC -.007, .032	**Model:** adjR^2= .091 $F(3,95)$= 4.262, p=.007 ***AI Acceptance*:** B= .431, t= 3.438, p=.001, 95%IC .182, .679 **Interaction term:** B= .007, t= .734, p=.465, 95%IC -.013, .028	**Model:** adjR^2= .050 $F(3,95)$= 2.719, p=.049 ***AI Acceptance*:** B= -.298, t= -2.344, p=.021, 95%IC -.550, -.046 **Interaction term:** B= .013 t= 1.230, p=.222, 95%IC -.008, .033

*Table 3.7 - **T1**: Comparative moderations (by 'sex') with moderator 'Non-Negative Emotions'*

	T1 – Non-Negative Emotions		
	Against/In Favor	**With Regulators**	**Without Regulators**
Male	**Model:** adjR^2= .063 F(3,102)= 3.337, p=.022 ***AI Acceptance*:** B= .389, t= 2.974, p=.004, 95%IC .130, .649 **Interaction term:** B= .001, t= .210, p=.834, 95%IC -.009, .011	**Model:** adjR^2= -.017 F(3,102)= 1.610, p=.192 ***AI Acceptance*:** B= .118, t= .884, p=.379, 95%IC -.147, .383 **Interaction term:** B= .001, t= .208, p=.836, 95%IC -.009, .011	**Model:** adjR^2= .006 F(3,102)= 1.222, p=306 ***AI Acceptance*:** B= .249, t= 1.656, p=.101, 95%IC -.049, .548 **Interaction term:** B= .002, t= .349, p=.728, 95%IC -.009, .013
Female	**Model:** adjR^2= .041 F(3,95)= 2.399, p=.073 ***AI Acceptance*:** B= .354, t= 2.664, p=.009, 95%IC .090, .617 **Interaction term:** B= -.005, t= -1.065, p=.289, 95%IC -.015, .004	**Model:** adjR^2= .083 F(3,95)= 3.967, p=.010 ***AI Acceptance*:** B= .446, t= 3.261, p=.002, 95%IC .174, .717 **Interaction term:** B= -.001, t= -.233, p=.816, 95%IC -.011, .009	**Model:** adjR^2= .039 F(3,95)= 2.3139, p=.081 ***AI Acceptance*:** B= -.241, t= -1.733, p=.086, 95%IC -.516, .035 **Interaction term:** B= -.006 t= -1.168, p=.246, 95%IC -.016, .004

*Table 3.8 - **T1**: Comparative moderations (by 'sex') with moderator 'Surprise'*

	T1 – Surprise		
	Against/In Favor	**With Regulators**	**Without Regulators**
Male	**Model:** adjR^2= .060 F(3,102)= 3.2317, p=.026 ***AI Acceptance*:** B= .403, t= 3.077, p=.003, 95%IC .143, .664 **Interaction term:** B= -.001, t= -.134, p=.893, 95%IC -.012, .011	**Model:** adjR^2= -.009 F(3,102)= .675, p=.569 ***AI Acceptance*:** B= .143, t= 1.059, p=.292, 95%IC -.125, .412 **Interaction term:** B= .001, t= .161, p=.872, 95%IC -.011, .013	**Model:** adjR^2= .001 F(3,102)= 1.043, p=377 ***AI Acceptance*:** B= .258, t= 1.706, p=.091, 95%IC -.042, .557 **Interaction term:** B= .000, t= .023, p=.982, 95%IC -.013, .013
Female	**Model:** adjR^2= .033 F(3,95)= 2.101, p=.105 ***AI Acceptance*:** B= .269, t= 1.785, p=.077, 95%IC -.030, .568 **Interaction term:** B= -.003, t= -.411, p=.682, 95%IC -.019, .013	**Model:** adjR^2= .092 F(3,95)= 4.298, p=.007 ***AI Acceptance*:** B= .521, t= 3.394, p=.001, 95%IC .216, .826 **Interaction term:** B= .008, t= 1.016, p=.312, 95%IC -.008, .025	**Model:** adjR^2= .025 F(3,95)= 1.853, p=.143 ***AI Acceptance*:** B= -.278, t= -1.764, p=.081, 95%IC -.592, .035 **Interaction term:** B= .002 t= .247, p=.805, 95%IC -.015, .019

Tables 3.9, 3.10, and 3.11 report findings correspondent to T2. Although there are some significative models, we can only find significative interaction terms within the female sample. When the dependent variable is ‘decision against/in favour’, the model explains 25.8% of the variation of the decision towards AI evolution (adjR^2=.258) and it is significative (F(3,95)= 12.365, p<.001). The effect of AI acceptance on that decision is positive and significative (B= .256, t= 2.364, p=.020, 95%IC .041, .471). The interaction term is negative (B=-.040), meaning that highest negative emotions decrease the effect of AI acceptance on the decision towards its evolution (t= -3.214, p=.002, 95%IC -.064, -.015).

The same applies to, in a similar way, when the dependent variable is ‘decision without regulators’.

These findings are illustrated in ModGraphs, in figures 3.2 and 3.3, respectively.

Table 3.9 - **T2**: Comparative moderations (by 'sex') with moderator ‘Negative Emotions’

	T2 – Negative Emotions		
	Against/In Favor	**With Regulators**	**Without Regulators**
Male	**Model:** adjR^2= .151 F(3,102)= 7.219, p<.001 ***AI Acceptance*:** B= .323, t= 2.585, p=.011, 95%IC .075, .571 **Interaction term:** B= -.001, t= -.172, p=.863, 95%IC -.019, .016	**Model:** adjR^2= -.008 F(3,102)= .718, p=.543 ***AI Acceptance*:** B= .128, t= .942, p=.348, 95%IC -.141, .397 **Interaction term:** B= .000, t= -.048, p=.961, 95%IC -.019, .018	**Model:** adjR^2= .017 F(3,102)= 1.600, p=194 ***AI Acceptance*:** B= .265, t= 1.762, p=.081, 95%IC -.033, .563 **Interaction term:** B= .013, t= 1.302, p=.196, 95%IC -.007, .034
Female	**Model:** adjR^2= .258 F(3,95)= 12.365, p<.001 ***AI Acceptance*:** B= .256, t= 2.364, p=.020, 95%IC .041, .471 **Interaction term:** B= -.040, t= -3.214, p=.002, 95%IC -.064, -.015	**Model:** adjR^2= .192 F(3,95)= 8.744, p<.001 ***AI Acceptance*:** B= .397, t= 3.338, p=.001, 95%IC .161, .634 **Interaction term:** B= -.025, t= -1.861, p=.066, 95%IC -.052, .002	**Model:** adjR^2= .100 F(3,95)= 4.611, p=.005 ***AI Acceptance*:** B= -.339, t= -2.716, p=.008, 95%IC -.586, -.091 **Interaction term:** B= -.033 t= -2.304, p=.023, 95%IC -.061, -.005

Figure 3.3 - ModGraph T2: Negative Emotion Against/In Favour (female)

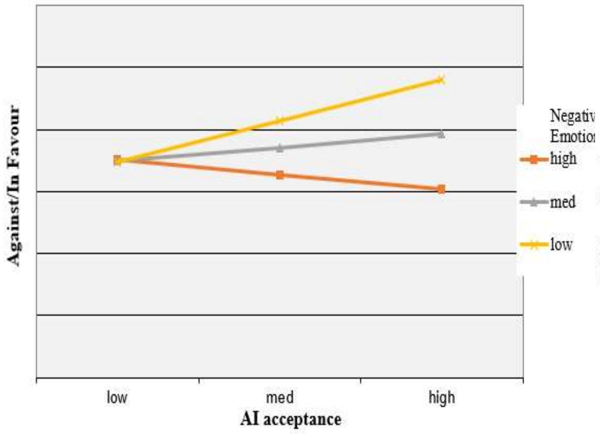


Figure 3.3 - ModGraph T2: Negative Emotions, Without Regulators (female)

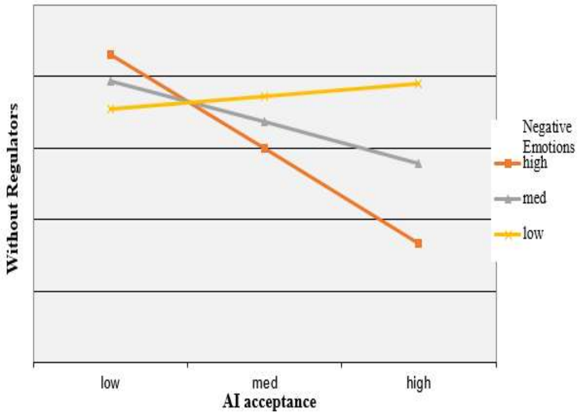


Table 3.10 reports analysis regarding 'non-negative emotions'. All models are significative within the female sample, but there only exists a significative interaction term when dependent variables are 'with regulators' and 'without regulators'. Unlike the previous results, the interaction terms in these models are positive, meaning that non-negative emotions tend to increase the effect of AI acceptance in the decision towards AI, with and without regulators. Also, when the dependent variable is 'without regulators', the main effect is negative, meaning that the more they accept AI, the less they agree on its evolution without regulators. These results are once again represented in ModGraphs, in figures 3.4 and 3.5, respectively.

Table 3.10 - **T2**: Comparative moderations (by 'sex') with moderator 'Non-Negative Emotions'

	T2 – Non-Negative Emotions		
	Against/In Favor	**With Regulators**	**Without Regulators**
Male	**Model:** adjR^2= .107 F(3,102)= 5.174, p=.002 ***AI Acceptance*:** B= .301, t= 2.2665, p=.026, 95%IC .038, .565 **Interaction term:** B= -.001, t= -.193, p=.847, 95%IC -.011, .009	**Model:** adjR^2= .025 F(3,102)= 1.912, p=.132 ***AI Acceptance*:** B= .096, t= .694, p=.489, 95%IC -.178, .371 **Interaction term:** B= -.008, t= -1.476, p=.143, 95%IC -.019, .003	**Model:** adjR^2= .003 F(3,102)= 1.091, p=356 ***AI Acceptance*:** B= .283, t= 1.804, p=.074, 95%IC -.028, .595 **Interaction term:** B= -.001, t= -.149, p=.882, 95%IC -.013, .011
Female	**Model:** adjR^2= .053 F(3,95)= 2.830, p=.043 ***AI Acceptance*:** B= .296, t= 2.376, p=.020, 95%IC .049, .543 **Interaction term:** B= .002, t= .475, p=.636, 95%IC -.006, .010	**Model:** adjR^2= .177 F(3,95)= 8.043, p<.001 ***AI Acceptance*:** B= .498, t= 4.074, p<001, 95%IC .255, .741 **Interaction term:** B= .013, t= 3.253, p=.002, 95%IC .005, .021	**Model:** adjR^2= .073 F(3,95)= 3.575, p=.017 ***AI Acceptance*:** B= -.247, t= -1.920, p=.058, 95%IC -.503, .008 **Interaction term:** B= .009 t= 2.193, p=.031, 95%IC .001, .018

Figure 3.5 - ModGraph T2: Non-Negative Emotions, With Regulators (female)

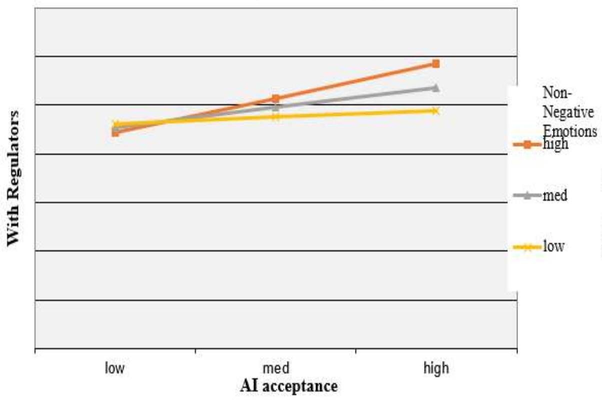
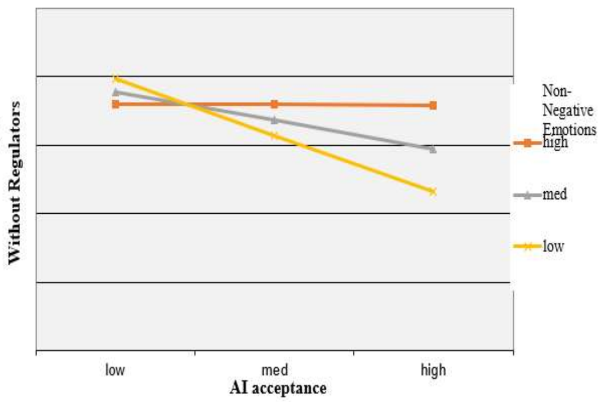


Figure 3.5 - ModGraph T2: Non-Negative Emotions, Without Regulators (female)

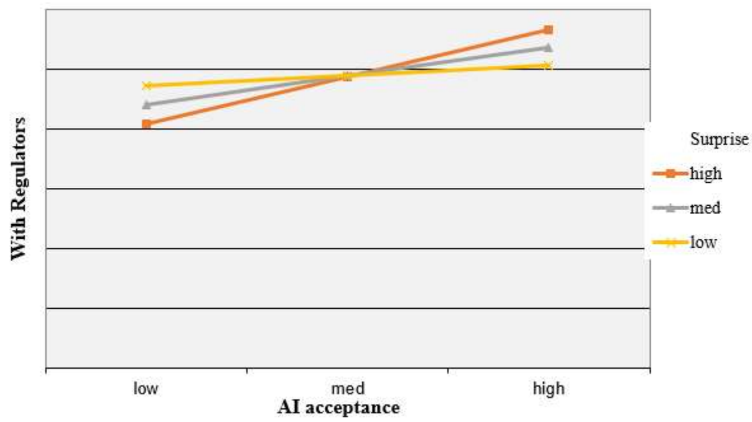


On table 3.11 we find moderation results regarding the moderator 'surprise'. There is a significative model with a significative interaction within females, when the dependent variable is 'with regulators'. The main effect is positive and significative, and the interaction term is also positive, meaning that the higher is surprise felt by female participants, the higher is the effect of AI acceptance in the decision towards AI evolution with regulators. This result is shown in table 3.6, through a ModGraph.

*Table 3.11 - **T2**: Comparative moderations (by 'sex') with moderator 'Surprise'*

		T2 – Surprise	
	Against/In Favor	**With Regulators**	**Without Regulators**
Male	**Model:** adjR^2= .062 F(3,102)= 3.304, p=.023 ***AI Acceptance*:** B= .309, t= 2.990, p=.004, 95%IC .131, .649 **Interaction term:** B= -.002, t= -.341, p=.734, 95%IC -.012, .009	**Model:** adjR^2= .002 F(3,102)= 1.061, p=.369 ***AI Acceptance*:** B= .169, t= 1.257, p=.212, 95%IC -.098, .435 **Interaction term:** B= .003, t= .647, p=.519, 95%IC -.007, .014	**Model:** adjR^2= .021 F(3,102)= 1.756, p=160 ***AI Acceptance*:** B= .232, t= 1.558, p=.122, 95%IC -.063, .528 **Interaction term:** B= -.006, t= -1.047, p=.297, 95%IC -.018, .006
Female	**Model:** adjR^2= .033 F(3,95)= 2.111, p=.104 ***AI Acceptance*:** B= .312, t= 2.347, p=.021, 95%IC .048, .576 **Interaction term:** B= .000, t= -.008, p=.994, 95%IC -.009, .009	**Model:** adjR^2= .153 F(3,95)= 6.893, p<.001 ***AI Acceptance*:** B= .566, t= 4.324, p<001, 95%IC .306, .826 **Interaction term:** B= .013, t= 2.841, p=.006, 95%IC .004, .022	**Model:** adjR^2= .046 F(3,95)= 2.590, p=.057 ***AI Acceptance*:** B= -.243, t= -1.759, p=.082, 95%IC -.516, .031 **Interaction term:** B= .003 t= .624, p=.534, 95%IC -.006, .012

Figure 3.6 - ModGraph T2: Surprise, With Regulators (female)



The results concerning moderations in T3 are shown in tables 3.12, 3.13, and 3.14.

As seen in table 3.12, regarding the moderator ‘negative emotions’, the analysis within the female sample reveals to be entirely significative, including models and interaction terms. All interaction terms are negative (and significative), meaning that the more female participants experience negative emotions, the lower is the effect of AI acceptance on the decision towards AI evolution, whether simply against/in favour, or with or without regulators. These results are represented with ModGraphs in figures 3.7, 3.8, and 3.9, respectively.

When it comes to the male sample, when the dependent variable is ‘with regulators’, the model itself proves to be quite weak and not significative (adjR^2= .043, F(3,102)= 2.565, p=.059). However, that model also reports a significative (and negative) interaction term (B= -.025, t= -2.251, p=.027, 95%IC -.047, -.003), meaning that although the studied variables are not strong enough to justify the relationships among them (there is not a significative main effect), negative emotions still play a significant role within these relationships, and the higher male participants experience negative emotions, the effect of AI acceptance on the decision with regulators decreases. Notwithstanding, the model does not prove to be sufficient in explaining these relationships.

Table 3.12 - **T3**: Comparative moderations (by 'sex') with moderator ‘Negative Emotions’

	T3 – Negative Emotions		
	Against/In Favor	**With Regulators**	**Without Regulators**
Male	**Model:** adjR^2= .234 F(3,102)= 11.690, p<.001 ***AI Acceptance*:** B= .330, t= 2.782, p=.006, 95%IC .095, .565 **Interaction term:** B= -.014, t= -1.367, p=.175, 95%IC -.034, .006	**Model:** adjR^2= .043 F(3,102)= 2.565, p=.059 ***AI Acceptance*:** B= .092, t= .700, p=.486, 95%IC -.169, .354 **Interaction term:** B= -.025, t= -2.251, p=.027, 95%IC -.047, -.003	**Model:** adjR^2= .005 F(3,102)= 1.164, p=327 ***AI Acceptance*:** B= .269, t= 1.782, p=.078, 95%IC -.030, .569 **Interaction term:** B= .007, t= .565, p=.573, 95%IC -.018, .033
Female	**Model:** adjR^2= .359 F(3,95)= 19.310, p<.001 ***AI Acceptance*:** B= .355, t= 3.486, p=.001, 95%IC .153, .557 **Interaction term:** B= -.017, t= -2.991, p=.004, 95%IC -.029, -.006	**Model:** adjR^2= .233 F(3,95)= 10.911, p<.001 ***AI Acceptance*:** B= .494, t= 4.207, p<001, 95%IC .261, .727 **Interaction term:** B= -.020, t= -3.058, p=.003, 95%IC -.033, -.007	**Model:** adjR^2= .137 F(3,95)= 6.203, p=.001 ***AI Acceptance*:** B= -.234, t= -1.894, p=.061, 95%IC -.479, .011 **Interaction term:** B= -.020 t= -2.834, p=.006, 95%IC -.034, -.006

Figure 3.8 - ModGraph T3: Negative Emotions, Against/In Favour (female)

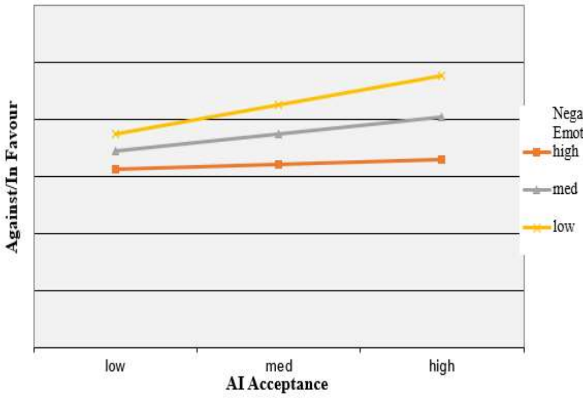


Figure 3.8 - ModGraph T3: Negative Emotions, With Regulators (female)

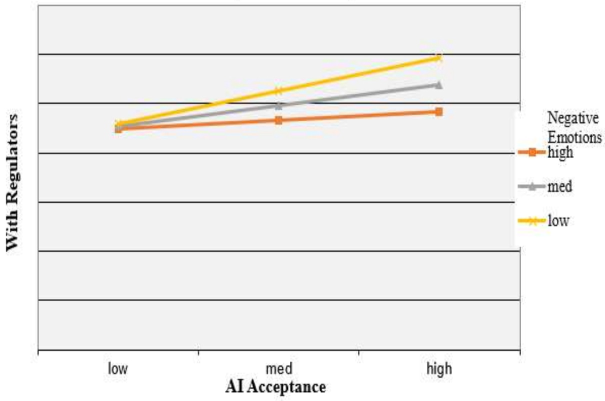
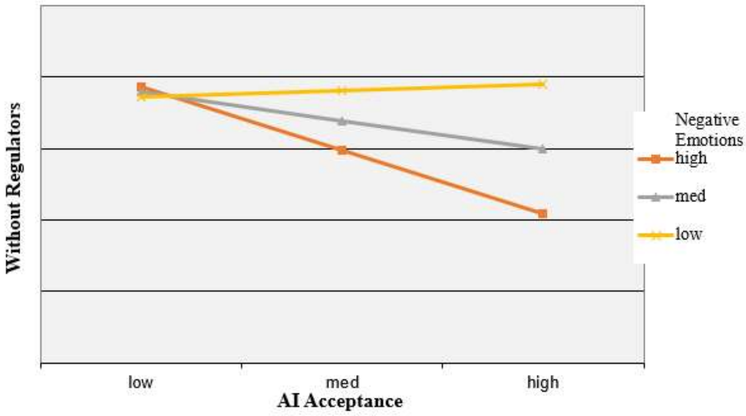


Figure 3.9 - ModGraph T3: Negative Emotions, Without Regulators (female)



As illustrated in table 3.13, all moderation models concerning the moderator 'non-negative emotions' prove themselves as significative. However, none of the models presents a significative interaction term. In those models, there is only a significative main effect when the dependent variable is 'with regulators' or 'without regulators'.

In the male sample, the moderation model is significative when the dependent variable is 'against/in favour', yet there is no significative interaction term.

When the moderator is 'surprise', as shown in table 3.14, there is only one significative model with a significative interaction term and that happens within the female sample, when the dependent variable is 'against/in favour'. This interaction term is negative, which means that the higher female participants experience surprise, the lowest will be the effect of AI acceptance on the decision towards its evolution. However, this main effect ($x \rightarrow y$) is not significative. This result is represented in figure 3.10, using a ModGraph.

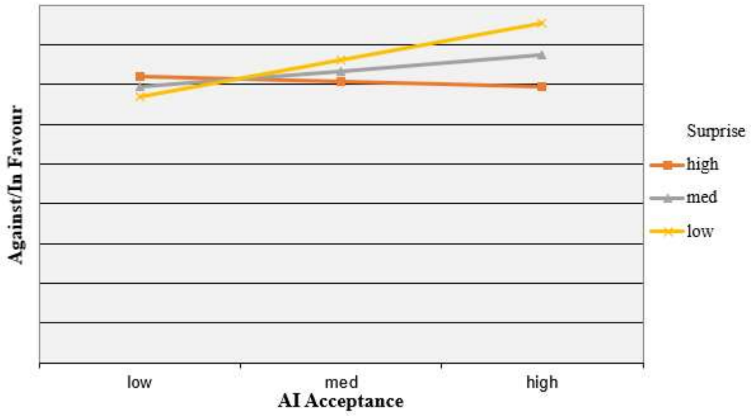
*Table 3.13 - **T3**: Comparative moderations (by 'sex') with moderator 'Non-Negative Emotions'*

	T3 – Non-Negative Emotions		
	Against/In Favor	**With Regulators**	**Without Regulators**
Male	**Model:** adjR^2= .231 F(3,102)= 11.484, p<.001 ***AI Acceptance*:** B= .274, t= 2.246, p=.027, 95%IC .032, .516 **Interaction term:** B= .006, t= 1.055, p=.294, 95%IC -.005, .016	**Model:** adjR^2= .023 F(3,102)= 1.821, p=.148 ***AI Acceptance*:** B= .109, t= .794, p=.429, 95%IC -.163, .380 **Interaction term:** B= -.001, t= -.146, p=.884, 95%IC -.012, .011	**Model:** adjR^2= .029 F(3,102)= 2.052, p=111 ***AI Acceptance*:** B= .200, t= 1.303, p=.195, 95%IC -.104, .503 **Interaction term:** B= .003, t= .494, p=.622, 95%IC -.010, .016
Female	**Model:** adjR^2= .109 F(3,95)= 4.990, p=.003 ***AI Acceptance*:** B= .195, t= 1.511, p=.134, 95%IC -.061, .451 **Interaction term:** B= -.008, t= -1.356, p=.178, 95%IC -.019,.004	**Model:** adjR^2= .121 F(3,95)= 5.478, p=.002 ***AI Acceptance*:** B= .376, t= 2.793, p=006, 95%IC .109, .644 **Interaction term:** B= -.002, t= -.326, p=.745, 95%IC -.014, .010	**Model:** adjR^2= .073 F(3,95)= 3.560, p=.017 ***AI Acceptance*:** B= -.294, t= -2.139, p=.035, 95%IC -.566, -.021 **Interaction term:** B= .001 t= .742, p=.460, 95%IC -.007, .016

*Table 3.14 - **T3**: Comparative moderations (by 'sex') with moderator 'Surprise'*

	T3 – Surprise		
	Against/In Favor	**With Regulators**	**Without Regulators**
Male	**Model:** adjR^2= .080 F(3,102)= 4.029, p=.009 ***AI Acceptance*:** B= .429, t= 3.270, p=.001, 95%IC .169, .689 **Interaction term:** B= -.009, t= -1.505, p=.135, 95%IC -.020, .003	**Model:** adjR^2= -.009 F(3,102)= .695, p=.557 ***AI Acceptance*:** B= .126, t= .924, p=.358, 95%IC -.145, .398 **Interaction term:** B= .005, t= .881, p=.381, 95%IC -.007, .017	**Model:** adjR^2= .013 F(3,102)= 1.459, p=230 ***AI Acceptance*:** B= .265, t= 1.747, p=.084, 95%IC -.036, .566 **Interaction term:** B= -.005, t= -.798, p=.427, 95%IC -.019, .008
Female	**Model:** adjR^2= .096 F(3,95)= 4.465, p=.006 ***AI Acceptance*:** B= .235, t= 1.924, p=.057, 95%IC -.008, .477 **Interaction term:** B= -.012, t= -2.265, p=.026, 95%IC -.022, -.001	**Model:** adjR^2= .088 F(3,95)= 4.148, p=.008 ***AI Acceptance*:** B= .399, t= 3.092, p=003, 95%IC .143, .655 **Interaction term:** B= -.005, t= -.864, p=.390, 95%IC -.016, .006	**Model:** adjR^2= .031 F(3,95)= 2.048, p=.112 ***AI Acceptance*:** B= -.291, t= -2.208, p=.030, 95%IC -.553, -.029 **Interaction term:** B= .002 t= .329, p=.743, 95%IC -.009, .013

Figure 3.10 - ModGraph T3: Surprise, Against/In Favour (female)



3.2.1. Qualitative data

As already mentioned, this study had a qualitative component, in which we conducted short semi-structured interviews to the participants (in this case, we were able to interview 204 out of 205 participants). In order to analyse the interviews and enrich the study will valuable information, we decided to categorise the participants' answers in order to build new variables and further conduct frequencies analysis.

As previously explained, we asked the participants two questions: 1 – "Considering all the questions you answered and the videos you saw, please justify your decision and clarify your vision regarding AI evolution."; 2 – "During the experiment and while watching the videos, was(were) there any emotion(s) that stood out?".

For question #1, we decided to transform it into the variable "Vision towards AI evolution", with four levels of response: 1 – completely unfavourable vision, 2 – unfavourable/apprehensive vision, 3 – favourable/balanced vision, 4 – favourable/hopeful vision.

For question #2, we decided to categorise each emotion as a different variable, according to the emotions uttered by the participants. In this case we have: fear, anguish, anger, surprise, joy, interest, excitement, contempt, and sadness.

The findings regarding these analyses are exposed in tables 3.15, and 3.16.

As we can see, the majority of the participants (68.6%) reported having a favourable/balanced vision towards AI evolution, meaning they understand it will bring advantages, but it must be developed cautiously and with the appropriate regulators. On the other hand, the following majority of responses (17.2%) goes to an unfavourable/apprehensive vision, that is the participants fear the consequences of AI development and some might even recognise advantages and benefits to it, however they feel apprehensive and reluctant towards some aspects of this evolution. Only 2.0% of the sample reported AI evolution as completely unfavourable, while the remaining participants seem to have a favourable and hopeful vision of this evolution.

When it comes to emotions, fear was the one participants mentioned the most (59.3%), followed by interest (57,8%), excitement (31.4%) and surprise (13.3%).

Table 3.15 - Frequencies of participants' vision on AI evolution

	Frequency	Percent	Valid Percent	Cumulative Percent
Completely unfavourable	4	2,0	2,0	2,0
Unfavourable/apprehensive	35	17.2	17,2	19,1
Favourable/balanced	140	68.6	6,.6	87,7
Favourable/hopeful	25	12,3	12,3	100,0
Total	204	100,0	100,0	

Table 3.16 - Frequencies of participants' experienced emotions

	Frequency	Percent	Valid Percent
Fear	121	59,3	59,3
Anguish	10	4,9	4,9
Anger	1	0,5	0,5
Contempt	1	0,5	0,5
Sadness	1	0,5	0,5
Interest	118	57,8	57,8
Joy	7	3,4	3,4
Excitement	64	31,4	31,4
Surprise	25	12,3	12,3

Note: There is no total in this table because each participant could mention more than one item.

CHAPTER IV – Discussion and Conclusions

We will begin this chapter by understanding whether or not our research model and our propositions have been confirmed and provide conclusions to those findings. Afterwards, we will explore additional data that has proven to be central to this theme. Throughout the entire discussion, we will illustrate findings using participants' citations from the qualitative phase of the study.

At the end of this chapter we will reveal which were the limitations to the present study and provide suggestions for future research.

4.1. Results discussion

As stated in our research model, we believe that emotions play a central role between AI acceptance and consequent decision towards its evolution, that is, emotions will most likely moderate the effect of AI acceptance on the decision towards AI evolution.

Having that said, our first proposition, P1, stated: as the progressive contact with AI deepens (as enacted by exhibiting the stimuli throughout all three times), we expect negative emotions to diminish. Looking back to the descriptive statistics, we may right away affirm that this proposition does not confirm. As the survey progressed, participants were exposed to stimuli regarding AI, as explained in chapter III, and associating this with what is stated in literature review, it would be expectable that this progressive contact and knowledge regarding AI would emphasize the agreement towards its evolution, alike all studies that mentioned human-robot interaction (Breazeal & Scassellati, 2000; Cañamero & Fredslund, 2000; Dautenhahn, 2004; Hancock, Billings & Schaefer, 2011; Kaplan, 2004; Nadel et al., 2006; Ray, Mondada & Siegwart, 2008; Nomura, Suzuki, Kanda & Kato, 2006; Partala & Surakka, 2004). However, when analysing the data, we realised that negative emotions actually rose across all three times, instead of decreasing and the last moment of the experiment, T3, was when the participants registered higher negative emotions

Symmetrically to P1, P2 stated: as the progressive contact with AI deepens (as enacted by exhibiting the stimuli throughout all three times), we expect non-negative emotions to increase. Similarly to what happened in P1, this proposition is also rejected, and non-negative emotions decrease as the survey progresses.

But why does this happen in both P1 and P2? According to the literature review, a progressive contact and knowledge before technology should evoke more positive emotions and camouflage negative ones. In fact, if we remember the Contact Hypothesis (Allport, 1954), progressive contact within groups reduces prejudice among them. Yes, that hypothesis also mentions that this is only valid when the groups share the same status and some common goals. Our purpose whatsoever is not to address this hypothesis into AI agents. What we intend to transmit is that, considering the final goal of AI is to flawlessly perform tasks that are currently only done by humans - and, therefore, be part of our society -, this suggests that it is important to reformulate social models and hypothesis that were not thought for this type of agents. This means that it is essential to re-educate society in terms of such evolution.

Another important fact to mention is that people must be clarified in terms of what really is AI, due to the fact that previous studies only focused on robots (Breazeal & Scassellati, 2000; Cañamero & Fredslund, 2000; Dautenhahn, 2004; Hancock, Billings & Schaefer, 2011; Kaplan, 2004; Nadel et al., 2006; Ray, Mondada & Siegwart, 2008; Nomura, Suzuki, Kanda & Kato, 2006; Partala & Surakka, 2004), and AI is much more than that.

«What I highlight is how people see AI, like those robots they do in Hollywood and, however, it is not what it is about at all. It is much more than that, and it is like it is an agglomerated of abstract things that are able of thinking for themselves. It does not have to necessarily be a robot; a robot is just a support for AI." – Participant 39.

«I was not that informed. When we think about AI, we imagine only robots and we do not remember the rest. » - Participant 40

These findings are somehow coherent with qualitative data, considering that the most mentioned emotion was fear. Interested was the second most mentioned emotion and, although quantitative data shows a decrease of non-negative emotions (in this case, interest) throughout the time, it also shows that this emotion remains high throughout the entire experiment.

The last proposition, P3, states: we expect non-negative emotions to significantly and positively moderate the effect of AI acceptance on the decision towards AI evolution. When looking at the results of our moderations with moderator 'non-negative emotions', we may claim that it is also rejected. However, we also noticed that variable sex plays a significative role in several analyses, that's why we split the sample and will discuss the results obtained separately, for both female and male participants.

Now still concerning P3, with the split sample we realise that this proposition is confirmed under certain circumstances. Only for female participants, non-negative emotions seem to play an important role in T2, when they are presented with AI applications. The more they feel non-negative emotions, the stronger is the effect of their AI acceptance on their agreement towards its evolution with and without regulators. Meaning that their level of AI acceptance influences whether they agree or not on its evolution, and this relation is emphasized when emotions such as interest and joy are present.

We decided to conduct further analyses in order to understand the relationships among all variables. As emotions were categorised in 3 factors, we decided to test the other two, applying the same analyses as to 'non-negative emotions'.

When testing 'negative emotions' as a moderator, we found that, in fact, it adds value to the model, once again, only in the female participants. These significative results happen in T2 and T3. These findings are somehow related to P1 and its rejection, but what do they tell us? Well, by T2 participants knew about AI applications and by T3 participants had already acquired a holistic knowledge regarding AI (its definition, applications that currently exist, and future perspectives), meaning that their contact with this theme became deeper. This means that, when fully aware of AI and its specifications, female participants tend to experience higher levels of negative emotions which, consequently, play an important role when they decide to be against or in favour of AI evolution, whether with or without regulators. In this case, the more negative participants feel about AI, the lower is the effect of their AI acceptance on their decision towards its evolution. That is, even if they accept AI, that is not enough for them to decide if they agree on its evolution or not. Instead, they are also driven by the negative emotions they are feeling, and not just by the rational facts. We must mention that the weakest models are when participants face the possibility of an evolution without regulators. This happens, most likely due to the fact that there is a need for control and regulation, as stated in literature review, and having to accept an evolution that comprehends so many implications as AI does, puts the participants in an uncontrollable situation and, as we know, we have a very great need for life stability (Ehrhardt, Saris & Veenhoven, 2000).

«Considering AI will be more or less like the internet, meaning it will be global, I think that if those regulators do not exist, what is going to happen is a conflict of interests and, last case scenario, it might be worse than a third World War. » - Participant 15

«I think there can never be an equality. I think humans should always be superior, there can never exist an equality, that is impossible. There has to be one on one side and one on the other side of the wall. » - Participant 12

As mentioned in literature, surprise is an emotion that can be either positive, or negative, considering the event that triggers it. That fact has been confirmed in our study, when conducting the CFA for the variable emotions. When testing that variable as a moderator, we concluded that, once again, there are only significative interaction terms within the female sample. In T2, after participants where presented with examples regarding AI's applications, the more surprised they felt, the higher they tend to accept AI and its evolution (with regulators). That might probably be due to the fact that participants feel overwhelmed by the amount of AI applications that exist and tend to perceive its evolution as beneficial, however, always with regulators.

«AI has a lot of potentialities and, when used in the right way, it can be an extremely impressive help towards humans. » - Participant 115

This finding changes in T3, when surprise plays an important role again, but this time, the opposite way. That is, when completely aware of AI's characteristics, the surprise they are feeling tends to diminish the effect of their acceptance towards AI on the decision of agreeing with its evolution.

Both of these last conclusions highlight the fact that surprise may be experience either in a positive way, or in a negative way, and that confers a completely different approach to reality.

«I felt surprised because I saw a few things that I did not know. When the scientists gave their opinions, there where a few things that I did not know, and that made me a bit scared of what this is becoming. » - Participant 52

One of the biggest question-mark these findings provide us is the fact that emotions only play a significative role as a moderator within female participants. The question is: why?

After some research, we came to the conclusion that, in general, women tend to experience emotions in a more intense way than men (Brebner, 2003). This is majorly due to hormonal aspects (Yeh, Lai, Lin, Lin & Sun, 2015). That is probably one of the explanations for our findings. The other explanation we found was that, after analysing the data base once again, we concluded that 40.6% of the male sample inserts in the area of technologies, whilst within the female sample, only 13.1% work/study in this field. That seems pretty obvious for us to assume as one of the most significant reasons why emotions did not play any significative role within the male sample. As stated by Conwell, Sharood and Els (2013), "those affected by a technology may provide more accurate predictions that the engineers involved in the development of the technology" (p. 14). Society in general tends to think more long-term, whereas technology developers tend to think in a more impulsive way.

«A lot of times, what happens is a complete disconnection between who develops AI and society's reality. That is, it becomes more and more necessary for universities to provide the logic of a global knowledge. Meaning, a technologies or engineering student must know how to understand and follow social processes and needs. That disconnection, in particular in the development of AI can be a threat and that is why regulators are extremely necessary. » - Participant 72

Looking at results within male sample, we realise that there are a few significative models, without a significative interaction term. What this means is that there are other factors which we did not reach in our study, that would provide better explanations and understandings of the models.

In terms of a global conclusion, highlighted by qualitative results, the majority of the participants recognise the benefits of AI and the several advantages it may provide. However, there is also a great reinforce regarding the need for regulation within AI evolution. In fact, if we look at the results, we realise that there are very few significative models when one of the variables is 'decision without regulators'. That is because this is understood as a very sensitive and controversial theme, and the need for regulation is starting to become somehow imperative. Conwell, Sharood and Els (2013) suggest that "we should look closely at where the technology is going and more importantly where we are sending it. (…) It is important to be careful, then, when developing any new technology especially controversial ones such as robotics. It is clear we will have to tread carefully when developing new regulations and when encouraging development of this technology" (p.55).

«I think AI must be used only for people's benefit. The line that separates beneficial from problematic is quite tenuous. » - Participant 62

« I believe that, in order to work in that field and regulate AI, it is crucial for that person to be submitted to a psychological assessment. » -Participant 115

As already mentioned, 'interest' was mentioned many times, however, 'fear' was mentioned even more. During the interviews, participants referred many times to the loss of jobs and economic differences, concerns that are also stated by several authors (Kelly, 2015; Spector, 2006; Spiegeleire, Maas & Sweijs, 2017). And the problem resides not only in labour displacement, but also, somehow, as an intelligence displacement.

« I'm scared of becoming inhuman. I think AI will stagnate human being's development and thinking, transforming us into vehicles that obey AI. I confess that I am scared of being replaced for something I am not quite sure what it is. I do not know what is going to happen and I think everyone should be afraid. » - Participant 47

«Maybe we are not that ready for what AI can bring. » - Participant 9

As a conclusion, we highlight the fact that the results obtained do not exactly follow what is stated in literature. Existent studies reveal that as the contact between technology deepens, people tend to highly accept its evolution and react towards it with more positive emotions (Breazeal & Scassellati, 2000; Cañamero & Fredslund, 2000; Dautenhahn, 2004; Hancock, Billings & Schaefer, 2011; Kaplan, 2004; Nadel et al., 2006; Ray, Mondada & Siegwart, 2008; Nomura, Suzuki, Kanda & Kato, 2006; Partala & Surakka, 2004). However, our study shows that, in fact, negative emotions tend to increase as participants' knowledge regarding AI deepens. This suggests that there should be a restructuring in social theories, when it comes to prepare society for certain changes, such as the ones concerning AI evolution. This is important because, as previously mentioned, AI's final goal will implicate its integration within society.

Furthermore, the lack of knowledge within participants reveals as a key point in their acceptance towards AI and its evolution. AI evolution, in general is considered to be positive and beneficial in several fields. However, the need for regulation is indeed a worry for the majority of participants, so much that AI evolution without those regulators reveals to be completely unthinkable.

4.2. Limitations and future research

The first limitation we would like to address concerns our methodology and the use of FaceReader™ by NOLDUS. When conducting tests to the software, we soon realised that, in order to make an adequate use of it and extract its potentialities and correct use, we would need to explore it in a way that was beyond our knowledge and time. That is, if we only used its simple functionalities, it would not work properly and would not provide us accurate results. This fact is well illustrated with the following situation that occurred: one of the participants recorded was extremely tired and presented a slight head down due to the fatigue. When analysing her expressions, FaceReader™ registered constant sadness when, in fact, the participant was not sad, but extremely tired. Most likely, there is a functionality in this software that allows us to create a baseline for each participant, however, doing it for 206 participants would require an immense amount of time and knowledge regarding the software. Another aspect we noticed was that the participants did not exteriorise their emotions in a very visible way and the software had a few troubles with detecting micro expressions. An explanation for that to happen is that the stimuli applied is mainly informative and steady, rather than capable of triggering an accentuated reaction (for example, chocking or very funny content). This causes a much more contained reaction, because the information processing is much more cognitive rather than emotional. For that reason, we decided to quit this option and advise caution in terms of data reliability for future studies that intend to use this software.

Another limitation is regarding the sample, meaning, participants were not all the same age or working in the same areas of activity. This means that their knowledge before AI is quite different. Although we tried to make sure all participants got the same information to work as a baseline of knowledge, we noticed that many of them came with a few preconceptions regarding this theme and the majority kept in mind the idea of Hollywood illustrated robots, which somehow limited their openness to the information provided by us. Associated with this comes the already mention fact that our male participants were majorly inserted in the technology field, whereas that did not happen among female participants. This reflected deeply in our results and for that reason we suggest that future studies always take into account the participants' area of activity in all analyses. Furthermore, it would be interesting to conduct a similar study with a sample only constituted by individuals in the technology field, and another sample with individuals from other fields, in order to obtain more accurate results and spot the different perspectives among samples.

For future studies, it would also be interesting to assess if people who are not familiar with AI somehow differentiate in terms of emotion standard.

We should also refer that some participants failed to understand a few of the quantitative questions. For instance, some thought that the scale of emotions was not to assess their own feelings, but to express what they thought a machine should feel. This might have created bias in the results. Another misconception had to do with the concept of regulators, which participants had some trouble in addressing what really were those regulators (a person?

Another machine?). From here we suggest that any future studies concerning these aspects must be extremely clear, avoiding to mislead the participants.

These last limitations were only perceived through the qualitative phase, where participants were able to express their opinion without barriers. This highlights the importance of qualitative studies, especially regarding such a controversial subject. Therefore, we suggest future studies to focus on qualitative methodologies, trying to understand society's readiness for AI evolution.

Last, but not least important, we highlight the scarce literature towards this thematic, considering that there are not studies emphasizing the importance of understanding how people really feel about AI specifically. These aspects are quite important when dealing with such controversial subject. From here urges a suggestion for future studies to focus on an understanding of society, instead of putting the main focus on AI.

REFERENCES

Abu-Nasser, B. S. & Abu-Naser, S. S. (2018). Cognitive system for helping farmers in diagnosing watermelon diseases. *International Journal of Academic Information Systems Research, 2(7),* 1-7.

Agha, M. E., Jarghon, A. & Naser, S. (2017). Polymyalgia rheumatic expert system. *International Journal of Engineering and Information Systems, 1(4),* 125-137.

Allan, G. & Chan, T. (2017). *Artificial intelligence and national security*. Cambridge, MA: Belfer Center for Science and International Affairs.

Allport, G. W. (1954). *The nature of prejudice.* Cambridge, Ma: Perseus Books.

André, E., Baker, R., Hu, X., Rodrigo, M. M. & Boulay, B. (2017). *Artificial intelligence in education: 18th International Conference, AIED 2017, Wuhan, China, June 28 – July 1, 2017, Proceedings* (Vol. 10331). Springer International Publishing AG.

Anwar, A. & Hassan, S. I. (2017). Applying artificial intelligence techniques to prevent cyber assaults. *International Journal of Computational Intelligence Research, 13(5),* 883-889.

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

Armstrong, S. & Sotala, K. (2015). How we're predicting AI – or failing to. In Romport, J., Zackova, E. & Kelemen, J. (Eds), *Beyond Artificial Intelligence – The Disappearing Human-Machine Divide* (pp. 11-29). Cham, Switzerland: Springer International Publishing AG.

Arnold, M. B. (1960). *Emotion and personality.* New York: Columbia University Press.

Arriaga, P. & Almeida, G. (2010). Fábrica de emoções: A eficácia da exposição a excertos de filmes na indução de emoções. *Laboratório de Psicologia, 8(1),* 63-80.

Arriaga, P., Franco, A. & Campos, P. (2010). Indução de emoções através de breves excertos musicais. *Laboratório de Psicologia, 8(1),* 3-20.

Atkinson, R. D. (2016). "It's going to kill us!" and other myths about the future of artificial intelligence. *Information Technology & Innovation Foundation*. Retrieved from http://www2.itif.org/2016-myths-machine-learning.pdf

Baron, R.M. & Kenny, D.A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51(6),* 1173-1182.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies.* Oxford, United Kindom: Oxford University Press.

Breazeal, C. & Scassellati, B. (2000). Infant-like social interactions between a robot and a human caregiver. *Adaptive Behavior, 8(1),* 49-74.

Brebner, J. (2003). Gender and emotions. *Personality and Individual Differences. 34(3),* 387-394.

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

Buck, R. (1994). *The communication of emotion.* New York: Guilford Press.

Cañamero, L. D. & Fredslund, J. (2000). How does it feel? Emotional interaction with a humanoid LEGO robot. *AAAI Technical Report FS-00-04. AAAI.*

Carlson, C.R., Collins, F.L., Stewart, J.F., Porzelius, J., Nitz, J.A. & Lind, C.O. (1989). The assessment of emotional reactivity: A scale development and validation study. *Journal of Psychopathology and Behavioural Assessment, 11(4),* 313-325.

Chin, W., Johnson, N. & Schwarz, A. (2008). A fast form approach to measuring technology acceptance and other constructs. *Management Systems Research Center, 32 (4),* 687-703.

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

Coeckelbergh, M. (2018). Should we ban fully autonomous weapons?. *Uni: View Magazin.*

Conwell, T., Sharood, T. & Els, N. V. (2013). *Fear the robots!: A study on the perceived future of robotics in modern society.* Worcester Polytechnique Institute.

Copeland, J. (1993). *Artificial intelligence: A philosophical introduction.* John Wiley-Blackwell.

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

Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain.* New York: Avon Books.

Damasio, A. (2003). *Looking for Spinoza: Joy, sorrow, and the feeling brain*. London, United Kingdom: William Heinemann.

Dautenhahn, K. (2004). Robots we like to live with?! – A developmental perspective on a personalized, life-long robot companion. *13th IEEE International Workshop on Robot and Human Interactive Communication (IEEE Catalog No.04TH8759),* 17-22.

Davis, F. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13(3),* 319-340.

Davis, F., Bagozzi, R. & Warshaw, P. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35(8),* 982-1003. http://dx.doi.org/10.1287/mnsc.35.8.982

Ehrhardt, J. J., Saris, W. E. & Veenhoven, R. (2000). Stability of life-satisfaction over time. *Journal of Happiness Studies, 1(2),* 177-205.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6,* 169-200.

Emamgholizadeh, S., Kashi, H., Marofpoor, I. & Zalaghi, E. (2014). Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *International Journal of Environmental Science and Technology, 11(3),* 645-656. http://dx.doi.org/10.1007/s13762-013-0378-x

Ertel, W. (2017). *Introduction to artificial intelligence* (2nd ed.). Cham, Switzerland: Springer International Publishing AG.

Gibbs, S. (2017). AlphaZero AI beats champion chess program after teaching itself in four hours. *The Guardian.* Retrieved from https://www.theguardian.com/technology/2017/dec/07/alphazero-google-deepmind-ai-beats-champion-program-teaching-itself-to-play-four-hours

Ginsberg, M. (1993). *Essentials of artificial intelligence.* San Francisco, California: Morgan Kaufmann Publishers, Inc.

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

Hair, J.F., Black, W.C., Babin, B.J. & Anderson, R.E. (2010). *Multivariate data analysis*. Upper Saddle River, New Jersey: Prentice-Hall.

Hamet, P. & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism Clinical and Experimental, 69,* S36-S40. http://dx.doi.org/10.1016/j.metabol.2017.01.011

Hancock, P. A., Billings, D. R. & Schaefer, K. E. (2011). Can you trust your robot?. *Ergonomics in Design, 19(3),* 24-29.

Hayes, A.F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: The Guilford Press.

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

Hinton, G. E. & Salakhutdinov, R. R. (2017). Reducing the dimensionality of data with neural networks. *Science, 313(5786),* 504-507.

Hsu, C.M., Chen, T.T. & Heh, J.S. (2014). Emotional and conditional model for pet robot based on neurol network. 7th *International Conference on Ubi-Media Computing and Workshops,* 305-308. http://dx.doi.org/10.1109/U-MEDIA.2014.33

Izard, C. E. (1991). *The psychology of emotions.* New York: Plenum Press.

Izard, C. E. (1993). Four systems for emotion activation: Cognitive and noncognitive processes. *Psychological Review, 100,* 68-90.

Jose, P.E. (2013). *ModGraph-I: A programme to compute cell means for the graphical display of moderational analysis: The internet version, Version 3.0.* Victoria University of Wellington, Wellington, New Zealand. Retrieved from https://psychology.victoria.ac.nz/modgraph/

Kaplan, F. (2004). Who is afraid of the humanoid? Investigating culturan differences in the acceptance of robots. *International Journal of Humanoid Robotics, 1(3),* 465-480.

Kelly, J. E. (2015). Computing, cognition and the future of knowing. *Whitepaper, IBM Research, 2.*

Kleinginna, P. R. & Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion, 5(4),* 345-379.

Konok, V., Korcsok, B., Miklósi, Á. & Gácsi, M. (2018). Should we love robots? – The most liked qualities of companion dogs and how they can be implemented in social robots. *Computers in Human Behavior, 80,* 132-142.

Lazarus, R. S. (1991). *Emotion and adaptation.* New York: Oxford University Press.

LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep Learning. *Nature, 521,* 436-444.

Levenson, R. W. (1994). Human emotion: A functional view. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 123-126). New York: Oxford University Press

Leveringhaus, A. (2016). What's so bad about killer drones?. *Journal of Applied Philosophy, 35(2),* 341-358.

Lövheim, H. (2012). A new three-dimensional model for emotions and monoamine neurotransmitters. *Medical Hypotheses, 78,* 341-348. http://dx.doi.org/10.1016/j.mehy.2011.11.016

Luger, G. F. (2009). *Artificial intelligence: Structures and strategies for complex problem solving (6th Edition).* Boston, United States of America: Pearson Education, Inc.

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.

Moyle, B., Moyle, C., Bec, A. & Scott, N. (2017). The next frontier in tourism emotion research. *Current Issues in Tourism.* http://dx.doi.org/10.1080/13683500.2017.1388770

Mozur, P. (2017). Google's AlphaGo defeats Chinese Go master in win for A.I.. *The New York Times.* Retrieved from https://www.nytimes.com/2017/05/23/business/google-deepmind-alphago-go-champion-defeat.html

Nadel, J., Simon, M., Canet, P., Soussignan, R., Blancard, P., Cañamero, L. & Gaussier, P. (2006). Human responses to an expressive robot. In *Procs of the Sixth International Workshop on Epigenetic Robotics.* Lund University.

Newitz, A. (2017). Elon Musk is setting up a company that will link brains and computers. *ARS Technica.* Retrieved from https://arstechnica.com/information-technology/2017/03/elon-musk-is-setting-up-a-company-that-will-link-brains-and-computers/

Nezhad, H. R. (2015). Cognitive assistance at work. In *AAAI Fall Symposium Series. AAAI Publications.*

Nilsson, N. J. (1980). *Principles of artificial intelligence.* Morgan Kaufmann Publishers, Inc.

Nomura, T., Suzuki, T., Kanda, T. & Kato, K. (2006). Measurement of anxiety toward robots. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on (pp. 372-377). IEEE.*

Pannu, A. (2015). Artificial intelligence and its application in different areas. *International Journal of Engineering and Innovative Technology, 4(10),* 79-84.

Partala, T. & Surakka, V. (2004). The effects of affective interventions in human – computer interaction. *Interacting With Computers, 16,* 295-309.

Patrício, D. I. & Rieder, R. (2018). Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Computers and Electronics in Agriculture, 153,* 69-81.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience: Vol. 1. Theories of emotion* (pp.3-33). New York: Academic.

Plutchik, R. (1985). On emotion: The chicken-and-egg problem revisited. *Motivation and Emotion, 9,* 197-200.

Ramamoorthy, A. & Yampolskiy, R. (2018). Beyond mad?: The race for artificial general intelligence. *ITU Journal: ICT Discoveries, 1(2),* 1-8.

Ray, C., Mondada, F. & Siegwart, R. (2008). What do people expect from robots?. In *Proceedings of the IEEE/RSJ 2008 International Conference on Intelligent Robots and Systems. IEEE Press.*

Reeve, J. (2009). *Understanding motivation and emotion* (5th ed.). United States of America: John Wiley & Sons, Inc.

Rekhawi, H., Ayyad, A. & Naser, S. (2017). Rickets expert system diagnoses and treatment. *International Journal of Engineering and Information Systems, 1(4),* 149-159.

Riva, G. (2017). SubCULTron: Robotic exploration of unconventional environmental niches. *Cyberpsychology, Behavior, and Social Networking, 20(12),* 775.

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.

Roseman, I. J., Antoniou, A. A. & Jose, P. E. (1996). Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory. *Cognition and Emotion, 10,* 241-277.

Russel, S. & Norvig, P. (1995). *Artificial intelligence: A modern approach.* Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Sauer, F. & Schörnig, N. (2012). Killer drones: The ‘silver bullet’ of democratic warfare?. *Security Dialogue, 43(4),* 363-380.

Schneider, F. E., Wildermuth, D. & Wolf, H. L. (2015). ELROB and EURATHLON: Improving search & rescue robotics through real-world robot competitions. In *Robot Motion and Control (RoMoCo), 2015 10th International Workshop on* (pp. 118-123). IEEE.

Shiota, M. N. & Kalat, J. W. (2007). *Emotion* (2nd ed.). Belmont: Wadsworth.

Spector, L. (2006). Evolution of artificial intelligence. *Artificial Intelligence, 170(18),* 1251-1253.

Spiegeleire, S. D., Maas, M. & Sweijs, T. (2017). *Artificial intelligence and the future of defense: Strategic implications for small – and medium-sized force providers.* The Hague, The Nederlands: The Hague Centre for Strategic Studies.

Stallman, R. M. & Sussman, G. J. (1977). Forward reasoning and dependency-directed backtracking in a system for computer-aided circuit analysis. *Artificial Intelligence, 9(2),* 135-196.

Szollosy, M. (2017). Freud, Frankenstein and our fear of robots: projection in our cultural perception of technology. *AI & Society, 32(3),* 433-439.

Talanov, M. & Toschev, A. (2014). Computational emotional thinking and virtual neurotransmitters. *International Journal of Synthetic Emotions, 5(1),* 1-8.

Vallverdú, J., Talanov, M., Distefano, S., Mazzara, M., Tchitchigin, A. & Nurgaliev, I. (2016). A cognitive architecture for the implementation of emotions in computing systems. *Biologically Inspired Cognitive Architectures, 15,* 34-40. http://dx.doi.org/10.1016/j.bica.2015.11.002

Yeh, Y., Lai, G., Lin, C. F., Lin, C. & Sun, H. (2015). How stress influences creativity in game-based situations: Analysis of stress hormones, negative emotions, and working memory. *Computers & Education, 81,* 143-153.