FACULTY OF LAW
Lund University

Tany Calixto Bonfim

Criminal liability of artificial intelligent machines: eyeing into AI's mind

JAMM07 Master Thesis

International Human Rights Law
30 higher education credits

Supervisor: Valentin Jeutner

Term: Spring 2022

CONTENTS

Summary

> *"Laws of robotics: First: A robot may not injure a human being or, through inaction, allow a human being to come to harm. Second: A robot must obey orders given it by human beings except where such orders would conflict with the First Law. Third: A robot must protect its own existence as long as such protection does not conflict with the First or Second Law. Fourth: A robot may not harm humanity or, by inaction, allow humanity to come to harm"*[1] *Isaac Asimov*

The concept of criminal responsibility in modern legal systems is grounded on the notion of agency, which encompasses concepts such as autonomy, intentionality, and individual accountability. However, with the ever-accelerating progress of technology and consequently increasing use of artificial intelligence in our daily lives, important tasks are entrusted to AI-driven systems. As artificial intelligence entities learn from experience, they acquire additional data and improve their capacity to write their own algorithms. Consequently, AI will progressively perform autonomously from humans. The more AI and sophisticated robots become "smart", the more likely they can react to impulses. As a result, we are gradually handling with agents rather than tools and thus, advanced artificial intelligence systems deployed carelessly will be the source of future concerns, as AI systems are already autonomously engaging in activities that would be deemed illegal for a human. Hence, there may be no one to blame for the negative impacts of their actions – mostly when one considers a self-learning machine. Since the accountability loophole is identified, the present thesis examines the feasibility of ascribing *mens rea* (criminal mind) to artificial intelligence entities, specifically autonomous systems such as robots. Apart from being a prerequisite for criminal accountability, the ability to be responsible is crucial under modern criminal law.

However, artificial intelligence entities do not possess personhood – since they are machines and lack consciousness – and therefore, they are not capable of falling under criminal liability. Furthermore, considering that AI entities can commit crimes, the question regarding whether criminal law can harbour autonomous machine behaviors rises. The issue is overly complex, since we are dealing with algorithms that we do not properly comprehend, and thus this dissertation will require analysis through the machine perspective aligned with a philosophical approach. By assuming that AI entities have their own degree of consciousness, the likelihood of seeing them as holders of a guilty mind leave the inconceivable to the prospective.

[1] Isaac Asimov, " Runaround", in *I, Robot* (London: Harper Voyager 2013), 31. Runaround was originally published in *Astounding Science Fiction* (New York: Street & Smith, March 1942). Owing to the potential weaknesses in his first three laws, Asimov later added the Fourth or Zeroth law. See Isaac Asimov, "The Evitable Conflict", *Astounding Science Fiction* (New York: Street & Smith, 1950).

The conclusion of this thesis highlights the fact that it is strongly necessary creation of a regulation at national and international levels to deal with the subject and protect present and future generations.
Since AI systems are constantly evolving, the threshold and consequence response of criminal law approaches.

With that in mind, it is important to mention that corporate legal doctrine has evolved to allow for penal law to impute guilty mind states to legal persons, even if they are not able to fulfil the *mens rea* requirement officially.
Hence, the same type of legal construction could open the floor for AI entities to be considered criminally liable.
Ultimately, when considering the analogy made to accommodate corporate unlawful behaviors under criminal law, it is plausible to reach a legal framework for the criminal accountability of AI systems.
There are proposals envisaging redefining the legal status of robots and granting them the status of "electronic persons" that are being discussed in the European Parliament in an effort to regulate the matter.

The issue is highly problematic due to, among other challenges, the inverse proportion of the speed at which AI technologies have been advancing and its own regulation.

Keywords: *Mens rea* – AI – Artificial Intelligence – Criminal Liability – Consciousness - E-personhood – Robots – Autonomous systems

Acknowledgements

Initially, I would like to thank my supervisor, Professor Valentin Jeutner. My eternal gratitude for his readiness, feedback, and invaluable patience and wisdom, along with his expertise and kind support during the tortuous development of this complex topic.

I am also extremely grateful to my classmates, especially the lovely friends that I made in Lund: Lui, Tsiko, Cristina, Julia, and Larissa, for their moral support and friendship, besides many useful and thoughtful discussions during the reading group. Thanks, must also go to Lena Olsson, the librarian at Raoul Wallenberg Institute of Human Rights and Humanitarian Law, who has been always pleasing and supportive in providing any book whenever I could not find it at the library – not to mention plenty of snacks, treats and cinnamon buns gently made by herself.

Further, I am grateful to my former boss, Kele, and my current boss, Marcio for having always supported me throughout the period that I decided to pursue a master's program overseas.

I also could not have embarked on this journey without my parents, especially my mom, who has always backed me up in every moment and situation. Thank you for not having let me give up on everything throughout the difficult moments, in the middle of a pandemic and a terrible war nearby. I would also like to thank my cats, Veridiana and Brisa for all the entertainment and emotional support. I could have not even moved abroad without them.

Helsingborg, Sweden
25 May 2022

With all my heart and gratitude,
Tany

Abbreviations

AI	Artificial Intelligence

Introduction

A) Background

Humans have had an aversion to releasing something they do not control since the dawn of time. This fear was encapsulated in Greek mythology by a figure named Pandora.[2]

In 1818 Mary Shelley gave the world Frankenstein[3] (or The Modern Prometheus) and his monster. It is about that composite image of the scientific creator and his unruly creation that is a central figure of modern mythology: the hubris of the scientist striving for divinity, followed by a vengeful monster.
It did not take long for the association with the monster and the fast development of artificial intelligence over the past decade to find an echo: artificial intelligence is growing stronger, faster, smarter, and more dangerous than its clever programmers. As with the creature of Frankenstein, artificial intelligence is not born, but it is still made by circumstances. [4]
Moving beyond the frightening dystopian narrative about AI, technology is here, and it does not show any sign that it will go away – on the contrary. Hence, we cannot afford to be Frankenstein watching as this technology unfolds before our eyes; rather we must be Prometheus.

With a few exceptions, the criminal liability resulting from artificial intelligence activities has only been addressed in connection with humanitarian law and autonomous weapons. In this thesis, I seek to consider the likelihood to ascribe *mens rea* to artificial intelligence entities and consequently their criminal liabilities for their conduct by covering cases in which AI systems autonomously perform acts which would be regarded as offenses if they were performed by humans. There is a possibility that an AI system engages in criminal activities for which no human is aware of the mental state needed; thus, no human planned, foresaw or directed such a crime. Therefore, a gap in criminal liability is raised here since a crime is committed for which no one can be held criminally responsible. Although AI systems are far from replicating the complexity of human psychology, I argue that with an appropriate level of abstraction and a machine approach taken

[2] 'Pandora | Myth & Box | Britannica' <https://www.britannica.com/topic/Pandora-Greek-mythology> accessed 5 April 2022. Greek mythology depicts human behavior most lively in the myth of Pandora´s box. It served both not only as a form of education, but also as a justification for a variety of human tragedies for ancient Greeks, since Prometheus fetched fire (technology) to humanity and, therefore, was severely punished by Zeus.
[3] 'Frankenstein; or The Modern Prometheus.' (*Library of Congress, Washington, D.C. 20540 USA*) <https://www.loc.gov/item/53051218/> accessed 5 April 2022.
[4] 'What Frankenstein's Creature Can Really Tell Us about AI | Aeon Essays' (*Aeon*) <https://aeon.co/essays/what-frankensteins-creature-can-really-tell-us-about-ai> accessed 18 May 2022.

into consideration, AI systems can be viewed as having cognitive attitudes relevant to the realization of *mens rea*.[5]

B) Purpose and the Research Questions

The purpose of this dissertation is to explore the feasibility of ascribing *mens rea* to artificial intelligence entities, specifically autonomous systems such as robots. The ability to be responsible is pivotal under modern criminal law, besides of being a prerequisite for criminal accountability. Regarding the *mens rea* requirement in criminal law, two fundamental questions are addressed in this dissertation:

1. Is it possible to ascribe *mens rea* to artificial-intelligent entities?
2. If yes, is it possible for criminal law to accommodate the autonomous behavior of machines?

For the purpose of this work, it will be demonstrated that robots are able to commit crimes and it will be assumed that they have their own consciousness degree in order to be criminally liable. This assumption is regarded to be extremely problematic since robots are neither alive nor legal entities endowed with personhood. However, the issue is extremely important due to the skyrocketing speed with which artificial intelligence has been developing, and yet, its actions and activities are not regulated. The probability of a robot engaging in offenses is no longer restricted to sci-fi movies and hence the liability gap arises – mostly when it comes to machine learning systems. This research aims to shed a light on this complex theme and perhaps propose some feasible alternatives to the matter.

C) Delimitations

The whole discourse on artificial intelligence is quite complex. Hence, for the purposes of this work, some subjects have been excluded.

To keep the thesis concise and to determine the jurisdiction, the *mens rea* element focused on is the one that stems from Common Law, given that the first computers appeared in England. In addition, just three types of *mens rea* will be addressed – intention, recklessness, and negligence. Besides the fact that they are the most common in the majority of the legal systems, apart from intention, the first two types of *mens rea* are less challenging to identify for the purposes of artificial intelligence regulation.

[5] Francesca Lagioia and Giovanni Sartor, 'AI Systems Under Criminal Law: A Legal Analysis and a Regulatory Perspective' (2020) 33 Philosophy & Technology 433, at 434.

Moreover, the study is circumscribed to the internal element of criminal liability and excludes the *actus reus* - for concision reasons.

The repercussions on tort law will be left aside since the focus of this analysis will be narrowed to criminal accountability in order to identify the specific legal challenges and consequences posed by artificial intelligence technology.

Regarding the numerous types of artificial intelligence, I will restrict the discussion to issues surrounding machines displaying intelligent behavior - particularly robots. That is because autonomous machines hold a certain degree of autonomy.

The thesis does not also deal with issues related to the criminal responsibility of the person behind the robot, the manufacturing company, or the computer developer. The focus is the investigation of criminal accountability of machine-autonomous systems themselves due to the thesis limitation.

Lastly, the issue regarding punishment is massively vast. Thus, the author does not go into detail but provides a brief overview of the matter.

Due to time constraints and the scarcity of research resources, this study is only for academic purposes.

D) Methodology and Material

This thesis applies the legal/doctrinal research method in conjunction with academic discussions and empirical research in order to comprehend the law that governs a particular area. Whenever a problem is identified, the aim of the research is tailored and reduced to specific research questions. Following that, relevant data from legal doctrine, academic commentaries and empirical studies are collected and analyzed from various perspectives adopting scientific, psychologic, and sociologic evidence in order to find a more holistic view to answer the research question. Finally, the findings are discussed, and potential reforms are proposed.[6]

Hence, the primary law to be considered is criminal law from the Common law system. The interpretation regarding the applicability of criminal law to AI systems will be drawn upon scholarly articles, journals, and doctrines. When discussing the relevant rules, the standard legal methodology will be employed.

Lastly, due to the fact that the aim of this work is to try to answer an under-researched topic, this thesis will use empirical findings regarding science – particularly the results of psychological studies and robotics experiments, nonetheless sometimes also including some neuroscience research – targeting

[6] Ishawara Bhat, P. (2019) Idea and Methods of Legal Research. Oxford University Press, pp. 145-161

to investigate whether machines possess a kind of consciousness or not and hence, if they are capable of having a guilty mind. By investigating the findings, one can be able to formulate an alternative approach to liability. The dissertation is largely based on experimental research, though it also includes some theoretical and philosophical reflections.

E) Outline

The structure of this thesis follows its aims – providing analysis regarding the possibility of ascribing *mens rea* to artificial-intelligence entities.

Hence, the work is divided into three chapters, wherein Chapter 1 introduces the mental element of the crime and its grounds, as well as its principal types envisaging to explain the framework of this thesis.

The core of this thesis and perhaps the main questions – the feasibility of ascribing *mens rea* on artificial-intelligence entities as well as the assumptions for criminal law to host autonomous machines behaviors - are discussed in Chapter 2. In the first part of this chapter, the author highlights the arisen of artificial intelligence and its challenges presented according to its evolution, along with the discussion regarding the mental element of artificial intelligence systems and whether they are capable of committing criminal offenses.

Chapter 3 is reserved for the discussion about the future of artificial-intelligence systems and the lack of a legal framework to deal with the urgent need to regulate the matter in conjunction with alternatives and perspectives.

After summarizing all the work in this thesis, the author offers some general remarks on the responsibility gap triggered by the technological advancements in the field of artificial intelligence.

CHAPTER 1. *MENS REA*

> *"Every mind has a horizon in respect to its present intellectual capacity but not in respect to its future intellectual capacity."*[7] *Wilhelm Leibniz*

1.1 GENERAL INTRODUCTION

This first chapter will introduce the overall concept of the mental element in criminal liability. A brief historical background is also presented to address the elements of the crime, but the emphasis remains on *mens rea* rather than the *actus reus*. In addition, the main types of *mens rea* and their main categories in Common law systems will be succinctly explained. The context provided is necessary as preparation for the discussion in the following chapter. Comprehending the origins and meaning of *mens rea*, will enable one to predict the challenges that artificial intelligence machines will face in the near future.

1.1.1 The elements of crime

'All rational beings are expected to know certain elementary truths about harmful behaviour, such as that it is morally wrong to kill innocent people'.[8]

This idea of what is considered universally harmful or awry has been developed throughout history as communities started to form cities and later States. Over the centuries, criminal law become vulnerable and mutable in national and international legal systems. Although '[t]here is no serious possibility of developing a value-free, quasi-scientific language of criminal law that could claim universal understanding',[9] some core criminal law notions are applicable across cultures.

A cohesive criminal justice system requires concepts of human action and personal guilt.

> Comparisons between the systems of penal law developed in the western European countries, and those having their historical origins in the English common law must be stated cautiously. Substantial variations exist even among the

[7] 'Gottfried Leibniz Quote' (*A-Z Quotes*) <https://www.azquotes.com/quote/900273> accessed 25 May 2022.
[8] George P. Fletcher, *The Grammar of Criminal Law: American, Comparative, and International,* Vol. 1: Foundations, Oxford University Press, 2007, p. 9-10 ISBN 978-0-19-510310-6, Citing Aristotle, *The Nichomacheon Ethics of Aristotle* (Sir David Ross Trans.), London: Oxford University Press, 1925.
[9] Gerhard OW Mueller, 'On Common Law Mens Rea' (1957) 42 Minnesota Law Review 1043, 1060.

> nations that adhere generally to the Anglo-American system or to the law derived from the French, Italian, and German codes. In many respects, however, the similarities of the criminal law in all states are more important than the differences.[10]

All criminal systems demand an element of criminal intent for most crimes. However, every language develops its own word in an effort to encapsulate the depths of guilt and punishment. Some terms, such as "intention", and "causation", are readily translated across languages, but others are unique to certain linguistic cultures and criminal law languages are profoundly rooted in particularistic cultures of guilt and blaming.[11]

The expression *mens rea*[12] initially originated in civil law systems with the introduction of Roman Law, often known as The Law of the Twelve Tables[13], however, this kind of *mens rea* is derived from written regulations.[14] On the other hand, Anglo-American systems have the substantive law grounded in cases and legislation. Both systems have different approaches to definition, grading and punishment attached to each type of *mens rea.*[15]
The word *mens rea* quickly become popular in the common-law system, and many countries still use the same Latin terminology or a word equivalent to "guilty mind" today. In certain countries, such as France, the expression "mental element" is used, which implies the same thing but is spelled in the local language.[16]

In Common law legal systems, two elements of crime are necessary to convict a person for having committed a crime: *actus reus*[17] (the external element) and *mens rea*[18] (the internal one).

For the sake of brevity, this thesis focuses on the *mens rea* element in Anglo-American systems. Moreover, the analysis of the *mens rea* coming from countries with an Anglo-Saxon legal tradition is more appropriate when considering the context of artificial intelligence – since the first computers were built in England.[19] Furthermore, the pioneer of artificial intelligence,

---

[10] 'Criminal Law | Definition, Types, Examples, & Facts | Britannica' (14 March 2022) <https://www.britannica.com/topic/criminal-law> accessed 14 March 2022.
[11] Fletcher (n 8) at 117-118.
[12] This expression is deployed throughout the thesis aiming to set up the mental element required to someone be convicted for a crime.
[13] 'Law of the Twelve Tables | Roman Law | Britannica' (14 March 2022) <https://www.britannica.com/topic/Law-of-the-Twelve-Tables> accessed 14 March 2022.
[14] Khalid Saleh Al-Shamari, 'The Emergence Of Mens Rea In Common Law And Civil Law Systems' [2019] 95 مجلة كلية القانون الكويتية العالمية.
[15] ibid at 94.
[16] ibid at 106.
[17] Jonathan Law (ed), *A Dictionary of Law* (Oxford University Press 2018) <https://www.oxfordreference.com/view/10.1093/acref/9780198802525.001.0001/acref-9780198802525> accessed 14 March 2022.
[18] Ibid
[19] 'Belford, Geneva G. and Tucker, Allen. "Computer Science". Encyclopedia Britannica, 9 Nov. 2021, Https://Www.Britannica.c'. https://www.britannica.com/science/computer-science

Alan Turing[20], proposed the Turing test[21] – a criterion to determine whether an artificial computer can "think".
On the other hand, the analysis and conclusions of the thesis can be adaptable to other legal systems.

1.2 THE MENTAL ELEMENT

One of the most complex issues when it comes to definition is related to the term *mens rea.* It stems from the Latin maxim *actus reus non facit reum nisi mens sit rea* i.e, 'An act is not necessarily a guilty act unless the accused has the necessary state of mind required for that offence'.[22] The phrase was taken from a sermon by Saint Augustine[23] against the crime of perjury. According to his maxim, acts depend on the guilty of the mind.[24]

> [T]he maxim is indelibly fixed in the common law through the centuries to the present day as a declaration of the principles of common law as to criminal responsibility. In the application of this maxim, no man can be convicted of crime, unless the two requirements which the maxim contemplates are fulfilled, namely, that there be both the physical element of actus reus, and the mental element of mens rea.[25]

Indeed, the closer we look at the language of criminal law, the clearer it seems that law and theology are inextricably linked.[26] The teaching of the penitential books that punishment should be based on moral guilt gave a powerful impetus to this evolution under the pervasive influence of the Church because moral guilt has a mental component at its core.[27]

Albeit the meaning of *mens rea* has fluctuated throughout the course of twelve centuries, by the twentieth century the idea had become a "universal"

---

[20] Andrew Hodges, 'Alan Turing' in Edward N Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Winter 2019, Metaphysics Research Lab, Stanford University 2019) <https://plato.stanford.edu/archives/win2019/entriesuring/> accessed 15 March 2022.
[21] 'Turing Test | Definition & Facts | Britannica' <https://www.britannica.com/technology/Turing-test> accessed 15 March 2022.
[22] Jonathan Law, 'Actus Reus' in Jonathan Law (ed), *A Dictionary of Law* (Oxford University Press 2018) <https://www.oxfordreference.com/view/10.1093/acref/9780198802525.001.0001/acref-9780198802525-e-79> accessed 15 March 2022.
[23] Augustine, Sermons (148-183) (John Rotelle ed, Edmund Hill trans, New York City Press, 1992) vol 5, at 315
[24] 'What makes the difference is how the word comes forth from the mind. The only thing that makes a guilty tongue is a guilty mind'.
[25] Brendon Murphy, 'The Technology of Guilt' (2019) 44 Australasian Journal of Legal Philosophy 64., p.72, citing [1979] 2 NSWLR 1, 24
[26] Fletcher (n 8) at 118, citing Gottfried Wilhelm Freiherr von Leibniz, *Nova Methodus Discendae Docendaeque Iurisprudentiale* (1748) (Gashuttenim Taunus: D. Auvermann, 1974).
[27] Francis Bowes Sayre, 'Mens Rea' (1932) 45 Harvard Law Review 974, 988.

requirement of modern legal systems[28]. *Mens rea* has had a significant impact on the concept of criminal liability; over time, the idea has matured and developed, and now it plays a critical part in determining punishment.[29] Nonetheless, the goal of modern criminal justice is shifting away from enforcing adequate punishment for moral wrongdoing and toward safeguarding social and public interests.[30]

The importance of *mens rea* is further emphasized by neuroscience findings, which provides increasingly sophisticated and specific knowledge about how the human mind works.[31]

Although most legal systems recognize the importance of the guilty mentality or *mens rea*, statutes have not always defined this term precisely.[32] Accordingly, 'It is lamentable that, after more than a thousand years of continuous legal development, English law should still lack clear and consistent definitions of words expressing its basic concepts'.[33]
Therefore, the concept of *mens rea* will depend on the interpretation of legal scholars. In the common-law system, *mens rea* is generally defined as the mental state that a defendant possesses when he commits a crime, whether it is a general intent to commit the conduct or specific intent to cause the criminal result.[34] Alan Norrie describes *mens rea* as 'a shorthand term denoting the existence of either *intention* to commit a crime, or *recklessness* (running a risk) as to whether a crime will occur as a result of one´s actions'.[35]

Ultimately, it can be concluded that the ability to make choices is the ground for criminal law liability.[36]

1.3 TYPES OF *MENS REA*

[28] Deborah W Denno, 'Concocting Criminal Intent' (Social Science Research Network 2017) SSRN Scholarly Paper ID 2909005 323–378 <https://papers.ssrn.com/abstract=2909005> accessed 14 March 2022. citing Guyora Binder, Criminal Law ch. 5 (2016); Gerhard O.W. Mueller, On Common Law Mens Rea, 42 MINN. L. REV. 1043 (1958); Francis Bowes Sayre, Mens Rea, 45 HARV. L. REV. 974 (1932); J.W.C. Turner, The Mental Element in Crimes at Common Law, 6 Cambridge L.J. 31 (1936), for a broad overview of the history and purpose of *mens rea*, see generally; and Morissette v. United States, 342 U.S (1952), at. 250
[29] Al-Shamari (n 14) at 97.
[30] Sayre (n 27) at 1017.
[31] Deborah W Denno, 'The Place for Neuroscience in Criminal Law' (Social Science Research Network 2016) SSRN Scholarly Paper ID 2806641 at 328, citing Deborah W. Denno, The Place for Neuroscience in Criminal Law, in Philosophical Foundations Of Law and Neuroscience 69, 77-80 (Dennis Patterson & Michael S. Pardo eds., 2016); Joseph R. Simpson, Introduction to Neuroimaging in Forensic Psychiatry: From the Clinic To the Courtroom XV, xv-xvii (Joseph R. Simpson ed., 2012) <https://papers.ssrn.com/abstract=2806641> accessed 14 March 2022.
[32] 'Criminal Law | Definition, Types, Examples, & Facts | Britannica' (n 10).
[33] G Williams, *Textbook on Criminal Law* (2nd edition) (London: Steven & Sons, 1983, at 73
[34] David C. Carson LL.B. & Alan R. Felthous M.D., "Introduction to This Issue: Mens Rea", 21 Behavioral Sciences and the Law 559, 559 (2003).
[35] Alan Norrie. *Crime, Reason and History*. A Critical Introduction to Criminal Law, 2nd edition, Butterworths, at 35. ISBN 0 406 93246 8
[36] Fletcher (n 8) at 266.

There are several possibilities of *mens rea* states including intention, purpose, recklessness, knowledge, wilfulness, fraudulence, dishonesty, negligence, etc. The discussion, however, will be limited to the most prevalent and significant types of fault elements. Further, it is crucial to determine what type of culpability is appropriate for a crime in order to determine liability, punishment, and legal consequences.[37]

A) Intention

Motives and intentions are the building blocks of human agency. Human beings generate intentions and carry out activities to accomplish desired objectives through a complicated psychological and sociological process (the production of motives for actions).[38] English law features intent as the highest level of *mens rea*, being expressed by the intent to achieve a particular result through a certain act.[39] It describes a criminal's ambition to attain a criminal outcome.

When we talk about motive, we sometimes mean an emotion like jealousy or greed, and other times we mean a type of intention.[40]As a result, intent differs from motive, which has little legal significance in determining criminal responsibility.[41]

B) Recklessness

Recklessness usually entails taking intentional and irrational risks, either in the hope of avoiding a particular unfavourable occurrence or in the hope of preventing some evil from occurring. The reckless person takes a risk on purpose which involves an offender being irresponsible, as they are aware of the risk of harm but disregard it. [42]

Glanville Williams provided the following exposition of recklessness:

> We learn as a result of experience and instruction, and our learning brings awareness of the dangers of life. We can guess at the probable present even when we cannot directly perceive it and can protect ourselves into the future by foreseeing the probable consequences of our acts. Our

[37] Roman Dremliuga and Natalia Prisekina, 'The Concept of Culpability in Criminal Law and AI Systems' (2020) 13 Journal of Politics and Law 256, at 257.
[38] Williams (n 33), at 36
[39] Mohamed Elewa Badar, *Studies in International and Comparative Criminal Law: The Concept of Mens Rea in International Criminal Law* (1st ed. 2013), p.33, ISBN 978-1-84113-760-5. citing Glanville Williams, *Text Book of Criminal Law*, 2nd edn (London: Stevens & Sons, 1983) 71
[40] Norrie (n 35), at 37 citing (Smith and Hogan, 1999, 78; cf Wasik, 1979)
[41] Al-Shamari (n 14) at 109.
[42] Badar (n 39), at 51, citing Glanville Williams, *Text Book of Criminal Law*, 2nd edn (London: Stevens & Sons, 1983), at 96

> memory works forwards. This is the foundation of the notion of recklessness.[43]

There are two types of recklessness: objective and subjective. In the first one, recklessness entails taking unjustifiable risks knowingly. By contrast, in its subjective form, it becomes a word for negligence.[44]

While intent supposes a direct desire for a particular result, recklessness depends on the subjective estimation of matters and choices.[45]

C) Negligence

Negligence does not imply any specific state of mind on the part of the defendant – differently from the previous *mens rea* aforementioned. Here, 'there is no requirement that the defendant foresees the risk that the *actus reus* might occur', since the idea of carelessness and thoughtlessness is captured by negligence.[46] Negligence does not incriminate the blind person for not seeing, but only those who have the ability to see but do not use it.[47] In other words, negligence is the failure to act as a reasonable person would have acted in circumstances where the law requires such an act.[48] That is, even if the defendant was unaware of the danger on the occasion in question, he would have been aware of it if he had taken proper precautions.[49]
As a result, negligence differs from other forms of *mens rea* because the defendant is punished according to his/her negative state of mind.[50]

Thus, it is important to note that if the defendant was unaware of the risk, but he/she should have been aware of it, there will be negligence. On the other hand, if the defendant was aware of the risk and still took it, there will be recklessness.[51]

1.4 CONCLUSION

Over the centuries, criminal law shaped the element of criminal intent for most crimes in different legal systems across the world. Despite the fact that each country has its own definition and understanding regarding the mental

[43] Ibid
[44] Al-Shamari (n 14) at 112
[45] Ibid
[46] Graham Virgo, 'Criminal Law: Theory and Doctrine. By A.P. Simester and G.R. Sullivan [Oxford: Hart Publishing. 2000 Lxix, 651, (Bibliography) 19 and (Index) 23. Paperback. £22.50 Net. ISBN 1 901362–60–4.]' (2002) 61 The Cambridge Law Journal 719, at 144-145.
[47] Gabriel Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law* (UPNE 2013) at 87.
[48] Badar (n 39), at at 66, citing Glanville Williams*, Text Book of Criminal Law*, above (n 26)
[49] Ibid, at 67
[50] Ibid
[51] Ibid

element of the crime, there is an overall consensus regarding the "guilty mind" - one of the required elements to establish the criminal liability of defendants. This mental state is known as *mens rea* in Anglo-American legal frameworks and the Latin maxim *actus reus non facit reum nisi mens sit rea* became well ingrained in the common law, whereas in Roman law *dolus* ascended to be the common denominator in ascertaining the guilt of the perpetrator.[52]

Albeit there are various debates regarding the meaning of *mens rea* among scholars all over the world, the crux or substance of *mens rea* at common law is the awareness of evil, the sense of doing something one should not.[53]

This part of the thesis analysed the mental element necessary to set the criminal liability, known as *mens rea*. The study is limited to the internal element of criminal liability and excludes the *actus reus*. In addition, the highlighted common types of blame are intention, recklessness, and negligence and they entail distinct degrees of 'fault' in common law jurisdictions.[54]

To conclude this chapter, the *mens rea* requirement of criminal liability was introduced to investigate whether artificial intelligence can fulfil that element in its aforementioned sub-forms. It is crucial to keep in mind the concept of *mens rea* in order to gain a thorough understanding of the present work, as it will point the way to the subsequent parts of the thesis.

[52] Ibid, at 30
[53] Mueller (n 9), at 1060.
[54] Williams (n 33), at 36.

CHAPTER 2. *MENS REA* IN THE TECH CONTEXT – THE ISSUE OF CRIMINAL LIABILITY

"Everyone takes the limits of his own vision for the limits of the world."[55] *Arthur Schopenhauer*

The previous chapter introduced the concept of *mens rea*, which is one of the necessary elements required in common law countries to establish criminal responsibility. This second chapter will discuss and analyze the issues raised by the mental element when determining whether artificial intelligence-driven entities are capable of perpetrating wrongdoings or not.
Before I go on to further consider the issue, it is useful previously explore the questions regarding machine consciousness in order to evaluate if it is possible to establish a mental element in AI-based systems. Sequentially, the autonomous systems along with machine learning topics will also be addressed. Following that, I will finally approach the issues regarding criminal liability and the personhood of autonomous agents. Chapter 2 will be ended by looking at the attribution of moral insights to robots, as well as its issues of personification; importantly, the criminal responsibility of corporations and the analyze of *mens rea* in crimes committed by AI entities will also be investigated.

2.1 GENERAL INTRODUCTION

When most of us think of Artificial Intelligence (AI), we immediately picture robots and sci-fi thrillers in which machines take over the world. However, AI is already present among us — in our smartphones, air-traffic control systems, driverless cars, personal robots, and so on.

The world is changing even faster as people, devices, and information become more interconnected.[56] The artificial intelligence technology of the twenty-first century is now able to do very many things that were considered science fiction in the past.[57] Machines are now programmed to 'think' like humans and act in human-like ways, like performing skills of surgeons. The ideal feature of AI, which distinguishes it from traditional software programs, is its

[55] Philosiblog, 'Every Person Takes the Limits of Their Own Field of Vision for the Limits of the World.' (*philosiblog*, 19 April 2012) <https://philosiblog.com/2012/04/19/every-person-takes-the-limits-of-their-own-field-of-vision-for-the-limits-of-the-world/> accessed 25 May 2022.
[56] Stephen Hawking, *Brief Answers to the Big Questions* (John Murray 2018) at 194.
[57] Gabriel Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems.* (Springer International Pu 2016) at 23.

ability to learn and rationalize on its own, and then, when necessary, take actions that have the best chance of achieving a particular goal.[58]

However, all the advantages arising from the countless benefits of AI also convey its downsides – as the current powers of AI entities are both striking and worrisome. The growing dependence of humans on machines appears to have grown their fear of them.

In addition, the assumption that humans will create intelligent machines to replace us is even held by well-known academics. The famous physicist Dr. Stephen Hawking said, 'The development of full artificial intelligence could spell the end of the human race'.[59]
The most urgent warning issued by Hawking concerns the rise of artificial intelligence: It will either be the best or worst thing that has ever happened to us[60]. If we are not cautious, it may be the last one. This may sound like science fiction, but Hawking says it is not. 'It would be a mistake, and potentially our worst mistake ever'.[61]

Humans are clumsy in comparison to robots. It takes generations to iterate because of the slow pace of evolution. Robots, on the other hand, can improve on their own design much faster, and will likely be able to do so without our assistance in the near future. According to Hawking, this will result in an "intelligence explosion" in which machines' intelligence will surpass ours "by more than ours surpasses that of snails".[62] Essentially, AI will be very good at achieving its objectives; if humans get in the way, we may be in trouble.[63]

Hawking also pointed out:

> You're probably not an evil ant-hater who steps on ants out of malice, but if you're in charge of a hydroelectric green-energy project and there's an anthill in the region to be flooded, too bad for the ants. Let's not place humanity in the position of those ants.[64]

For those who are still not convinced, he proposes a different metaphor. "Why are we so worried about AI? Surely humans are always able to pull the plug?" a fictitious person asks him. Hawking answers: "People asked a computer, 'Is there a God?' And the computer said, 'There is now,' and fused the plug".[65]

---

[58] 'Are We Prepared for the Rise of AI? | Ccier' <https://cuts-ccier.org/are-we-prepared-for-the-rise-of-ai/> accessed 18 March 2022.
[59] Kelsey Piper, 'The Case for Taking AI Seriously as a Threat to Humanity' (*Vox*, 21 December 2018) <https://www.vox.com/future-perfect/2018/12/21/18126576/ai-artificial-intelligence-machine-learning-safety-alignment> accessed 18 March 2022.
[60] Hawking (n 56) at 184.
[61] Ibid, at 184
[62] Ibid
[63] Ibid at 188
[64] Ibid
[65] Ibid at 193

Setting aside the mythic approach of the "Frankenstein complex" [66] we ought to bear in mind that the issue is supposed to be informative in the sense of drawing attention rather than alarmist – given that the AI we have today is still in its primitive stages. Nonetheless, experts worry about what will happen when that intelligence outpaces us. Or, as Hawking puts it, 'Whereas the short-term impact of AI depends on who controls it, the long-term impact depends on whether it can be controlled at all'.[67]

Since it is impossible for human beings to distinguish the truly impossible from the simply fantastic *a priori*, all possibilities must be considered.[68] Aside from that, all future possibilities are fiction before they become reality.

In one way or another, robot freedom might lead to some harmful behaviors, even if well-intended. In part, the literal-mindedness of a computer is responsible for this, which may carry out an order *ad absurdum* because it "is logical but not reasonable".[69] Tay Microsoft's chatbot[70] is an example of an AI acting erratically by performing tasks that its original programmers might not have intended. The actions of Tay were entirely digital and limited to the Twittersphere; but artificial intelligence can cause physical effects when incorporated within or controlled by hardware, like a robot.[71] As a result, nobody is potentially culpable.

Furthermore, there have already been reported robot-provoked fatalities. The first one occurred in 1979 in Michigan, USA, when a Ford assembly worker Robert William was struck in the head by a robot arm. Also, the autonomous vehicle (AV) operated by Uber in Arizona, killed a pedestrian in 2018. In its self-driving software, Uber detected the victim but failed to stop in time.[72] Unfortunately, none of the examples aforementioned is science-fiction.

The more AI systems and smart robots become better at "sensing", "thinking", and "acting" - at least in the engineering sense - the more likely they can

---

[66] The expression "Frankenstein complex" was coined by science writer Isaac Asimov (1920-2002) to describe men's fear of machines rebelling against their creators, a clear allusion to Mary W. Shelley's novel's legendary monster. Asimov then devised the famous Three Laws of Robotics in order to counteract the Frankenstein complex. In a sense, these laws served as a sort of moral code for robots, preventing them from rebelling against their creators. 'Isaac Asimov | Biography & Facts | Britannica' <https://www.britannica.com/biography/Isaac-Asimov> accessed 5 April 2022.

[67] Ibid at 188

[68] Sam N Lehman-Wilzig, 'Frankenstein Unbound: Towards a Legal Definition of Artificial Intelligence' (1981) 13 Futures 442, at 444.

[69] Ibid, at 445, citing I. Asimov, *The Naked Sun* (London, Granada Publishing, 1975), p. 143

[70] Tay, a chatbot, repeatedly made racist and rude comments on Twitter before it was shutdown.'In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation' (*IEEE Spectrum*, 25 November 2019) <https://spectrum.ieee.org/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation> accessed 29 March 2022.

[71] Ryan Abbott, *The Reasonable Robot: Artificial Intelligence and the Law* (Cambridge University Press 2020) at 33.

[72] Summer, 'Four Crazy Real Cases of Humans Killed by Robots' (*Medium*, 29 March 2022) <https://historyofyesterday.com/four-crazy-real-cases-of-humans-killed-by-robots-7ab9bc0a9e38> accessed 29 March 2022.

respond to impulses. As a result, we are increasingly dealing with agents, rather than simple tools of human interaction.[73]
Therefore, there will be more and more connections between regular people and increasingly competent - and mobile - machines as robots leave the factories and move into homes and offices.

Having said that, when an unmanned vehicle causes a car accident, a surgical robot makes a surgical error or a trading algorithm commits fraud, and so on, the issue of criminal liability arises. After robots are programmed to react based on external circumstances, they could begin committing crimes completely without human assistance. As a result, how do you punish a robot? Is reprogramming sufficient? And who should be prosecuted for these offenses: the manufacturer, the programmer, the user, or the AI entity itself?[74]
Ultimately: is it possible to ascribe *mens rea* to artificial intelligence machines?

This second chapter seeks to assess what can be understood by consciousness – since "to be conscious" is a state of mind that is seen as a prerequisite to have it in order to be held accountable for some action/decision. In short, whether machines are able to think and to have awareness, the possibility of considering artificial intelligent entities as holders of a guilty mind might leave from the unthinkable to the plausible.

In addition, understanding AI is critical for considering how it should be regulated and how it may challenge existing legal systems.[75]
The rate at which AI has been developing is skyrocketing and policymakers along with lawmakers are being urged to regulate the matter at both national and international levels. The world has always changed, and so has the legislation. On the other hand, because of technological advancements, the world is moving at a faster pace. Thus, the first step is to recognize the need for a legal framework to protect current and future generations.

Our society is faced with a gap of ever-widening responsibility that, if not addressed, poses a threat to the moral framework of society as well as the legal concept of liability.[76]

Within ten to twenty years, the biggest challenge will not be preventing artificial intelligence from destroying humanity, but how to live alongside it.[77]

In the next subchapter and subsections, I will briefly expose the arising of artificial intelligence in order to provide a backdrop. Aiming to understand

---

[73] Michael McGuire and Thomas J Holt (eds), *The Routledge Handbook of Technology, Crime and Justice* (Routledge 2020) at 643., citing Pagallo Ugo (2013) *The Laws of Robots: Crimes, Contracts, and Torts.* Dordrecht: Springer
[74] Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems.* (n 57) at XV.
[75] Abbott (n 71) at 32.
[76] Andreas Matthias, 'The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata' (2004) 6 Ethics and Information Technology 175, at 176.
[77] Jacob Turner and SpringerLink (Online service), *Robot Rules. Regulating Artificial Intelligence* (1st ed. 2019., Springer International Publishing 2019) at 37.

the present and perhaps predict the future, it is important to have a look into the past. As will be discussed further, the mental element scan and the issue of machine awareness are essential to subject them to criminal liability. In parallel, the potential for self-learning AI systems to engage in unlawful actions aligned with the issue of dealing with algorithms that we do not properly understand will also be explored. Finally, I will move towards criminal responsibility and the topic of attributing personhood to autonomous systems, where moral perceptions are assigned to robots along with their personification, besides the criminal liability of legal persons and the guilty mind analysis of crimes perpetrated by AI systems. All these topics will be covered in subsections.

2.2 HOW DID ARTIFICIAL INTELLIGENCE EMERGE?

> *"Viewed narrowly, there seem to be almost as many definitions of intelligence as there were experts asked to define it."*[78] *Robert. J. Sternberg*

Throughout history, humankind has always sought tools to simplify daily living. The track record of artificial intelligence can be traced back to ancient times when philosophers pondered the possibility that artificial beings, mechanical humans, and other automatons might have existed or could exist in some form.[79]
As humans have developed systematic methods of rationality, the concept of thinking machines has evolved with them.[80] Artificial intelligence became more tangible through early thinkers in the 1700s and beyond. Philosophers pondered how intelligent non-human machines could artificially mechanize and manipulate humans. The thought processes that fueled interest in AI began when classical philosophers, mathematicians, and logicians considered the mechanical manipulation of symbols, eventually leading to the 1940s invention of the programmable digital computer, the Atanasoff Berry Computer (ABC)[81]. This particular invention sparked the interest of scientists in developing an "electronic brain"[82] or an artificially intelligent being.[83]

The concept of creating a machine that could be considered 'intelligent' is quite abstract. Intelligence is a subjective concept, considering a pragmatic perspective. This brings up an important point that has been present since the

[78] Cindy Wigglesworth, *SQ21: The Twenty-One Skills of Spiritual Intelligence* (SelectBooks, Inc 2014).
[79] 'A Complete History of Artificial Intelligence' (*G2*), by Rebeca Reynoso, May 25, 2021 <https://www.g2.com/articles/history-of-artificial-intelligence> accessed 14 March 2022.
[80] Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems.* (n 57) at 1-2.
[81] 'Atanasoff-Berry Computer | Britannica' <https://www.britannica.com/technology/Atanasoff-Berry-Computer> accessed 31 March 2022.
[82] Neuralink, which is baked by serial entrepreneur Elon Musk is developing a "neural lace" for artificial processors to communicate with human brain tissue. See https://neuralink.com/applications/
[83] Supra note 68

beginning of the Enlightenment: we do not know everything. This is significant in the field of AI because there is still a lot that we do not know.[84]

There are currently over 70 different definitions of "intelligence" according to Legg and Hutter[85]. On the other hand, the lack of agreement over what intelligence is does not prevent researchers from discussing Artificial Intelligence.
Artificial intelligence systems that think like humans are difficult to detect unless human thinking is first defined. However, artificial intelligence technologies designed as general problem solvers have been shown to make decisions that are very similar to human decisions when presented with the same information.[86]

Although the universal lack of consensus regarding a single definition of artificial intelligence, the discussion that follows is better placed in some kind of context by drawing a rough boundary around the concept.

Any system that takes advantage of its environment to achieve a goal can be defined as having general intelligence. The biological goal is to maintain autonomy and reproduce, or in other words, to survive; whereas the goal of machines is to solve a specific task or problem using both internal and external resources. Intelligent beings, as well as robots and computers, fall under this general definition. Hence, intelligence is an umbrella term that covers different levels of intelligence, contextual influences, and various types of systems with different degrees of intelligence.[87]
Another school of thought on artificial intelligence holds that intelligence necessitates comprehension. Machines, regardless of what they can do, do not qualify as intelligent in this view because they do not understand what they are doing. Even for a super-intelligent AI, action without comprehension simply stimulates intelligence.[88]

The following definition has been proposed by the European Commission´s High-Level Expert Group on Artificial Intelligence in 2019:

> Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans[89] that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information,

[84] 'Where Did AI Come From?' <https://www.rs-online.com/designspark/where-did-ai-come-from> accessed 14 March 2022.
[85] Ben Goertzel and Pei Wang, *Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms: Proceedings of the AGI Workshop 2006* (IOS Press 2007) 18–22.
[86] Hallevy (n 51) at 7 citing Masoud Yazdani and Ajit NArayanan, Artificial Intelligence: Human effects.
[87] Camilo Miguel Signorelli, 'Can Computers Become Conscious and Overcome Humans?' (2018) 5 Frontiers in Robotics and AI <https://www.frontiersin.org/article/10.3389/frobt.2018.00121> accessed 11 April 2022.
[88] Abbott (n 62) at 25
[89] Humans design AI systems directly, but they may also use AI techniques to optimize their design.

> derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions. As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems).[90]

In recent AI definitions, logic is emphasized rather than the link to humanity. When an AI system thinks rationally, it is driven by goals and reasons. To act rationally, a machine must perform in a way that can be described as goal-directed.[91]
In contrast, this rationalist definition might collide with more advanced AI entities, such as unsupervised machine learning (since they develop new goals) as it will be demonstrated in the next subchapter.

With this brief notion of AI in mind, the next section will delve further into the subject of consciousness, addressing the question of whether machines can be conscious or not. Besides being puzzling and controversial, the issue is meant to be evaluated from a philosophical perspective. In addition, I will restrict discussion primarily to issues surrounding machines displaying intelligent behavior. This is important because if machines are/will be able to have/develop a type of awareness, the possibility of ascribing the mental element of the crime to them will arise.

2.3 ARTIFICIAL INTELLIGENCE AND MENTAL ELEMENT: ARE A.I DRIVEN ENTITIES ABLE TO COMMIT CRIMES?

> *"The feeling of intelligence is a mirage, if you achieve it, it ceases to make you feel so. As somebody has competently put it - AI is Artificial intelligence until it is achieved; after which the acronym reduces to Already Implemented."*[92]

The goal of this subchapter is to present an overview regarding machine consciousness, with an eye toward answering the question concerning the

[90] Eric LEMONNE, 'Ethics Guidelines for Trustworthy AI' (*FUTURIUM - European Commission*, 17 December 2018) <https://ec.europa.eu/futurium/en/ai-alliance-consultation> accessed 25 May 2022.
[91] Turner and SpringerLink (Online service) (n 77) at 13.
[92] Mohit Thakkar, *Artificial Intelligence: A Theoretical Guide* (Mohit Thakkar 2018) at 2.

probability of AI systems possessing consciousness and consequently whether AI entities fall under criminal responsibility. Starting with a philosophical perspective on machine consciousness to frame the work, this subchapter firstly focuses on machine awareness, and then moves towards machine-autonomous systems – where the issue of self-learning machines and their potential engagement in criminal activities will be investigated. Following that, the next subsection will elaborate on the criminal liability and the personhood of autonomous agents, taking into account the moral perceptions attributed to robots and their personification. Additionally, the criminal responsibility of legal persons will be discussed. The chapter concludes by discussing the guilty mind in criminal offenses perpetrated by AI entities, highlighting negligence.

2.3.1) Machine consciousness

For the purpose of answering questions about artificial consciousness, one must consider the philosophical reflections on consciousness that emphasize human (and animal) consciousness.[93]

The query of whether robots can have intelligence has prompted philosophers, psychologists, and cognitive scientists to ponder questions such as "What does intelligence really mean?" and "Can machines think?" As humanity accepts robot intelligence, it is critical to investigate the meaning and significance of consciousness, not only because consciousness is a fundamental aspect of thought, but also because consciousness is the cornerstone of thought and intelligence.[94]

There are as many senses of consciousness as there are uses of the word "conscious". The act of perceiving necessarily involves consciousness, since perception is a form of consciousness, much in the way that walking is a form of exercise. Likewise, a person is conscious whenever he/she perceives, imagines, dreams, and so forth.[95] "Consciousness' is in fact primarily a clinical term, and "conscious" is usually used as a synonym for "deliberate"; thus, both terms are limited in their common applications to human affairs'.[96] We have fairly firm beliefs about what things other than men are conscious. We believe that dogs, cats, and horses, as well as rabbits, in some sense, very likely are. Some people also tend to think that catching fish is cruel; however, the fisherman does not demonstrate scruples towards the worm. For most reflective people, however, all of this is conjecture; even if we had a vast understanding of dogs and cats, we may be hesitant to assert it. In fact, the main reason we do not know whether cats and dogs are conscious or not is

[93] Elisabeth Hildt, 'Artificial Intelligence: Does Consciousness Matter?' (2019) 10 Frontiers in Psychology <https://www.frontiersin.org/article/10.3389/fpsyg.2019.01535> accessed 11 April 2022.
[94] David Levy, *Robots Unlimited: Life in a Virtual Age*, 2005, CRC Press, Taylor & Francis Group
[95] Kenneth M Sayre, *Consciousness: A Philosophic Study of Minds and Machines* (Random House 1969) at 73.
[96] Ibid, at 5

that we do not have enough knowledge about consciousness itself. In the same way, we are not certain what to say about machines. Convinced that machines are completely inanimate, we are nonetheless impressed by claims that with a little ingenuity, machines can do almost anything humans can. It follows that machines being able to do everything humans can do should enable them to perform the things humans can do that indicate their consciousness. If that is true, why would we not consider the possibility that machines might also be conscious?[97]

That being said, it is 'important to distinguish between human intelligence and artificial intelligence, as it is clear that the two at least operate on a different basis'.[98]

The concept of consciousness is a daunting one, and the goal of this thesis is not to answer that question, but rather to establish a link between awareness and the guilty mind in artificial intelligent entities - thus enabling the possibility of, in theory, their criminal responsibility.

Our current inability to define consciousness poses a challenge to conceiving supposedly intelligent and conscious artificial intelligence entities. Intelligence and human consciousness are dynamic. As babies, we have a fairly weak consciousness and intelligence. The capacity of our consciousness and intelligence can increase rapidly as we grow older, but we may lose consciousness if our brain is injured during surgery or if anaesthesia is applied. Besides that, as for comatose people, it is unknown whether the vegetative patient's consciousness has been lost completely.[99]
Yet, 'our ordinary conception of consciousness leads us to doubt whether fish and worms are conscious but does not assure us that they are not'.[100]

Accordingly, 'AI does not think the way a person does. AI is not conscious or self-awareness in the same sense a person is,'[101] and in this sense, there is a challenge in arguing that machines cannot think since the concept of human thought remains largely unrecognized and only implicitly detectable via introspection.[102] Likewise, as humans claim to experience the transcendental, it is difficult to prove it. Therefore, how can be possible to demonstrate this in machines?[103]

---

[97] Ibid, at 73
[98] Ryan Browne, 'Science Fiction and Artificial Intelligence: Dissecting the Cultural Fear of Robots and Androids' at 24 <https://www.academia.edu/30977210/Science_Fiction_and_Artificial_Intelligence_Dissecting_the_Cultural_Fear_of_Robots_and_Androids> accessed 6 April 2022.
[99] Deyi Li, Wen He and Yike Guo, 'Why AI Still Doesn't Have Consciousness?' (2021) 6 CAAI Transactions on Intelligence Technology 175.
[100] Sayre (n 95) at 5.
[101] Abbott (n 71) at 27.
[102] ibid at 26.
[103] Steven Schkolne, 'Machines Demonstrate Transcendental Self-Consciousness' (*Medium*, 6 November 2020) <https://schkolne.medium.com/machines-demonstrate-transcendental-self-consciousness-1e1340ad7d58> accessed 14 April 2022.

Several journals and conference papers have proposed theories and architectures regarding the possibility of conscious machines and robots.[104] Furthermore, various AI enthusiasts argue that not only do robots have the theoretical possibility of "life," but they will almost certainly be perceived as such.[105] We perceive artificial intelligence as familiar and unfamiliar because of its ubiquitous presence in modern life and uncanny resemblance to humans. As a result, humans subordinate AI purely on the basis that it is not human.[106]In fact, the problem with robots and humans is not their differences, but instead our humancentrism - our insistence on seeing the world strictly from a human perspective, as evidenced by the structural bias in our language - that prevents us from seeing robots as human-like now and in the future. Of course, there are numerous arguments against this perspective. Man has a soul which is directly gifted by God, according to traditional western religions; robots, on the other hand, have no soul and are therefore dead and without rights. From a humanistic perspective, it can be argued that robots are alive only by comparing our brains to these machines and by other reductionist arguments. Blood is the life force that keeps flesh and bones alive. Therefore, robots remain complex dead machines, capable of acting and looking like humans, but remain robots, not humans.[107]
Consequently, the question of whether machines can be made conscious is a question about our understanding of what it means to be a conscious entity, and it will not be resolved by any amount of knowledge about actual and potential machines.[108]

Thus, a software program like Siri[109] can answer questions in the same conversational way as if it were a real person. However, what makes Siri self-aware is not the technology of speech synthesis and voice recognition, but rather the underlying understanding of self. When seeking machines' self-awareness, one should avoid attempting to compare them directly with human experience.[110] Consciousness is a mind property, and if all other mind properties result from the brain, and if the brain can be modelled on a computer, then perhaps consciousness can be reproduced by artificial intelligence.[111]

---

[104]There have been several international workshops on the subject (see www.machineconsciousness.org); there are regular sessions at consciousness conferences - as several ongoing projects; the first books have started to appear; and there are funded project. Owen Holland, 'The Future of Embodied Artificial Intelligence: Machine Consciousness?' in Fumiya Iida and others (eds), *Embodied Artificial Intelligence: International Seminar, Dagstuhl Castle, Germany, July 7-11, 2003. Revised Papers* (Springer 2004) at 51 <https://doi.org/10.1007/978-3-540-27833-7_3> accessed 12 April 2022.

[105] Phil McNally and Sohail Inayatullah, 'The Rights of Robots: Technology, Culture and Law in the 21st Century' (1988) 20 Futures 119.

[106] Browne (n 98) at 7.

[107] McNally and Inayatullah (n 105).

[108] Ryan Abbott and Alex Sarch, 'Punishing Artificial Intelligence: Legal Fiction or Science Fiction' 53 62, at 9.

[109] 'Siri | Computer Application | Britannica' <https://www.britannica.com/topic/Siri> accessed 15 April 2022.

[110] Steven Schkolne, 'Machines Demonstrate Self-Awareness' (*Medium*, 6 November 2020) <https://becominghuman.ai/machines-demonstrate-self-awareness-8bd08ceb1694> accessed 4 April 2022.

[111] Lawrence B Solum, 'Legal Personhood for Artificial Intelligences Essay' (1991) 70 North Carolina Law Review 1231, at 265.

In building robots like Qbo[112] that pass the mirror test, researchers are making a major step in history. Nonetheless, these anthropomorphic feats reveal more about the ability of machines to mimic a human than they do about consciousness. By taking a closer look at the machine's external sense of self, through the lens of what is important to the machine, then it will be viable to comprehend it at a far deeper (and more accurate) level. In order to truly understand the machine, one must examine its external sense of self through the eyes of what matters to the machine. For instance, an IP address appears everywhere on any device connected to the internet. Through this, machines identify their actions, and that can be deemed as external self-awareness.[113] Even toy pet robots today display some self-awareness. One can consider a robot whose operation is governed by both its "desire" to play with its human owner and its desire not to drain its battery. Hence, when the battery is low, it will return to its charging station, even if the human wishes to continue playing with it.[114]

On the other hand, internal self-awareness is regarded as the pinnacle of consciousness. This self-reflexivity - or the ability to introspect – plays a pivotal role in conscious experience.[115] While currently, robots are unconscious (according to the in-force notion about consciousness), they possess a degree of intelligence. They are able to learn from examples and follow procedures, as well as help solve some intelligent problems.[116]
Bringsjord emphasizes this by claiming that as a result of a logico-mathematical analysis, the structure and form of self-consciousness can be clarified and specified. These specifications can then be rendered computationally in a way that will support clear tests of mental ability and skill. [117]
In contrast, tests for machines seem to make sense only when compared with human intelligence – as the Turin test itself. If a machine becomes conscious it may also develop a non-anthropocentric morality and even provide new answers to many moral dilemmas.[118] Then, it may not be appropriate to apply human moral standards to artificial intelligence since AI does not work in the same way as the human mind.[119]

Associating cognitive states with artificial systems, and therefore applying legal qualifications, contradicts the belief that mentalistic concepts are

---

[112] 'Qbo Robot Passes Mirror Test, Is Therefore Self-Aware' (*IEEE Spectrum*, 6 December 2011) <https://spectrum.ieee.org/qbo-passes-mirror-test-is-therefore-selfaware> accessed 15 April 2022.
[113] Schkolne (n 110).
[114]Don Norman, '7: THE FUTURE OF ROBOTS' at 4 <https://www.academia.edu/2849702/7_THE_FUTURE_OF_ROBOTS> accessed 27 April 2022.
[115] Schkolne (n 110).
[116] Li, He and Guo (n 99).
[117] Selmer Bringsjord and others, 'Real Robots That Pass Human Tests of Self-Consciousness', *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (2015) at 499.
[118] Signorelli (n 87).
[119] *Robot Rules*, p at 66 <https://link.springer.com/book/10.1007/978-3-319-96235-1> accessed 16 May 2022.

exclusive to humans.[120] Hence, "cognitive concepts can be interpreted in a flexible and neutral way so that they are applicable also to some artificial entities".[121] Similarly, one may say that an entity has the purpose of realizing a certain result if there is an internal state of the entity and while it has that internal state, it will tend to achieve that goal. As the intention is a state of mind, in order to determine whether an AI system possesses such an internal state it is necessary to look into its internal structure and, more specifically, its functioning.[122] It is possible that transistors or their kin are simply too slow to generate what we would recognize as consciousness and state of mind.[123] Ultimately, one must take a machine-oriented approach to the problem, since humans are blinded by the opacity of their own perception at first – which makes it hard to witness machine self-awareness.[124]

Following the above discussion, we can conclude that consciousness and intelligence are not clearly correlated. However, that does not mean intelligence can only exist when consciousness is present. The history of artificial intelligence shows that intelligence can exist without consciousness.[125]

The next sub-chapter will explore the issue of machine learning[126] in autonomous systems, as criminal responsibility gaps are arising due to their development. Subsequently, I will analyze whether is possible to ascribe *mens rea* to artificial intelligence entities. Further, in order to narrow the dissertation down on artificial intelligence machines, the discussion will be restricted to autonomous machines – such as robots[127] - as they possess a certain degree of autonomy.

2.3.2) Autonomous system and machine learning: the issue of dealing with algorithms that we do not utterly understand and the potential for self-learning programs to engage in unlawful actions

[120] Giovanni Sartor, 'Cognitive Automata and the Law: Electronic Contracting and the Intentionality of Software Agents' (2009) 17 Artificial Intelligence and Law 253, at 254.
[121] ibid.
[122] Lagioia and Sartor (n 5) at 447.
[123] Solum (n 111) at 1265 citing DANIEL C. DENNETT, THE INTENTIONAL STANCE 327-28 (1987).
[124] Schkolne (n 110).
[125] Li, He and Guo (n 99).
[126]Machine learning is described as an inductive method of learning in which the AI system analyses various given data and identifies patterns for future use without being explicitly programmed for that purpose. In this sense, the computer's behavior is not predictable by either the operator–owner or the original programmers. 'Machine Learning | Artificial Intelligence | Britannica' <https://www.britannica.com/technology/machine-learning> accessed 18 April 2022.
[127] Examples of robots include robotic manipulators, autonomous vehicles (e.g., cars, drones,), humanoid robots, robotic vacuum cleaners, etc.

Artificial intelligence doesn't always have an explanation for its actions. Generally, we can determine what a machine has done, but not how or why it made certain choices.[128] The behavior of artificial intelligence may exhibit high levels of autonomy and irreducibility. A machine that has some autonomy is capable of receiving sensory input, setting goals, assessing outcomes against criteria, making decisions, and adjusting behavior to ensure success without being actively controlled by a human. As a consequence, a new situation emerges since the manufacturer/operator of a machine is, in theory, not able to predict the future behavior of the machine.[129]
Thus, they cannot be held liable for it and consequently, there is a responsibility gap when it comes to ascribing criminal liability to AI entities.

There are currently machines in development or in use that can decide on a course of action and act without human intervention. The rules by which they act are not fixed during the manufacturing process - but can be changed by the machine itself during operation. This is referred to as machine learning.[130] In this context, the term autonomy is usually used to describe the ability of agents to make certain decisions while performing a task without the need for constant monitoring and intervention on the part of the user. Recognizing the need for initiative implies an acknowledgement that some outcomes of agent activity may be difficult for the user to predict. Occasionally, they may even contradict what the user perceives as their interests or desires[131]. Further, the process by which the software of an autonomous system determines how a robot should act in any given scenario is extremely complex – mostly in the case of "deep learning" systems.[132] A concrete example previously aforesaid is the chatbot "Tay" designed to reply to questions people ask on social media and become increasingly 'smarter' over time. Consequently, this will lead to highly negative social effects.

In this manner, the more artificially intelligent systems are controlled by algorithms that were not written by humans, the more likely it is that they will behave in ways that are not only unexpected by humans but also totally unexpected. As a result, once the deep learning machine is able to make its own decisions, any interference from humans is removed.[133] Thus, the machine can use this learned programming process to acquire sufficient knowledge to craft and execute its own algorithms. After this, deep learning is no longer under the control of humans because only the machine

[128] Abbott (n 71) at 113.
[129] Andreas Matthias, 'The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata' (2004) 6 Ethics and Information Technology 175, at 175.
[130] ibid at 177.
[131] Carolyn Dowling Australian, 'Intelligent Agents: Some Ethical Issues and Dilemmas' at 3.
[132] Singapore Academy of Law and others, *Report on Criminal Liability, Robotics and AI Systems* (2021) at 32. Deep learning is a subset of machine learning that employs 'artificial neural networks' to model and derive insights from complex data structures and relationships. Artificial neural networks, in a nutshell, are a collection of 'layered' algorithms that analyze, classify, learn from, and interpret input data. The values from one layer are fed into the next layer, resulting in increasingly refined insights.
[133] Priya Persaud, 'PROTECTING AGAINST ULTRON: EXPLORING THE POTENTIAL CRIMINAL LIABILITY OF SELF-PROGRAMMING DEEP LEARNING MACHINES' (2019) 72 28, at 587, citing Jo Best, IBM Watson: The inside story of how the Jeopardy-winning supercomputer was born, and what it wants to do next, TECH REPUBLIC (Sept. 9, 2013).

understands what its algorithms are supposed to do.[134] This is significant in terms of the law because negligence is determined by foreseeability. Thus, if an AI involves in conducts that was unforeseen, a byproduct of its capacity to "think" and plan its own course of actions, who then is liable for its actions?[135] In situations where an AI system truly has the ability and freedom to make its own decisions, it may be unjust to blame its owner, user, or programmer since these individuals lack the intent, knowledge and control over the AI after it has been implemented.[136] In other words: let us assume an artificial intelligence-controlled machine is performing tasks in a unique and unpredictable way to humans. Who should be held liable if harm to a person or damage to property occurs – the AI program that generated the damage but is not a person or the human who lacks knowledge of how the machine functions or even its intent? [137]
In this sense, considering that an autonomous system has intent, can it be assumed that it also possesses a guilty mind?

If the answer is positive, this means that the artificial intelligence system itself is a "gun" with its own cognition and will. A gun of that sophistication can be taught and developed by itself, and thus, it can shoot at different targets based on input signals or experience.[138]

These questions are far from easy to be answered considering the technological developments. 'Since Al is not one thing but is constantly evolving, the answer - and with it, criminal law's response - will hugely depend upon the individual facts of the case at hand'.[139] Such occurrences are unavoidable, necessitating a re-examination of both community and legal understandings of liability.[140] The more autonomous the system, the more challenging it is to establish effective rules governing liability for harmful acts.[141] With such autonomous decision-making processes growing stronger, the threshold for criminal liability for entities is rapidly approaching.[142]
We must therefore assess how the mental states requirements for the criminal law field may change over time in comparison to the decisions made by some AI machines.[143]

---

[134] ibid, citing Aaron Mak, Google Taught A.I. How to Program More A.I., SLATE.
[135] Woodrow Barfield and Ugo Pagallo (eds), *Research Handbook on the Law of Artificial Intelligence* (TJ International Ltd 2018) at 4-24.
[136] Michael T Stuart and Markus Kneer, 'Guilty Artificial Minds: Folk Attributions of Mens Rea and Culpability to Artificially Intelligent Agents' (2021) 5 Proceedings of the ACM on Human-Computer Interaction 363:1, at 363:2.
[137] Barfield and Pagallo (n 135) at 4.
[138] Dremliuga and Prisekina (n 37) at 260.
[139] Dafni Lima, 'Could AI Agents Be Held Criminally Liable: Artificial Intelligence and the Challenges for Criminal Law' (2017) 69 South Carolina Law Review 677, at 681.
[140] Carolyn Dowling, 'Intelligent Agents: Some Ethical Issues and Dilemmas', *Selected papers from the second Australian Institute conference on Computer ethics* (Australian Computer Society, Inc 2000) at 31.
[141] Nora Osmani, 'The Complexity of Criminal Liability of AI Systems' (2020) 14 Masaryk University Journal of Law and Technology 53, at 75.
[142] Persaud (n 133) at 604.
[143] Barfield and Pagallo (n 135) at 400.

2.3.3) Moving towards Criminal Liability and Personhood of autonomous agents

Before going on to further consider the basis on which criminal liability may be imposed in such cases, it is useful to explore the issues regarding the attribution of mental states to robots as well as the legal corporations and their criminal accountability. Besides, the personification of robots will also be briefly addressed. These topics, and their implications for the imposition of criminal liability, will be investigated in the following sections. The mentioned subjects are crucial for understanding the possibility of ascribing *mens rea* to AI entities, where the matter will be finally debated. Additionally, due to the thesis limitation, this subchapter is devoted only to questions about the potential criminal liability of the robot itself, instead of the criminal responsibility of the person behind the robot, the manufacturing company, or the computer programmer.

A) Moral perceptions – assigning mental states to robots

Currently, robots cannot be held criminally responsible, in line with existing traditional accounts. The theory of criminal responsibility has traditionally taken free will as its starting point.[144] This approach is considered to be an expression of the principle of autonomy: Individuals are considered to be autonomous persons who have the ability to choose between alternative actions in general.[145]According to traditional notions of criminal responsibility, the concept of 'guilt' or 'responsibility' of robots contained in this first impulse must be categorically rejected regardless of the future of technology.[146] Robots do not have free will. As such, if the concept of criminal responsibility truly relies on free will - as traditional theories of blame posit - this concept faces the same potential instability as it has been facing due to technological advances in robotics and artificial intelligence.[147]

The concept of free will as the foundation of criminal responsibility is indeed problematic, and it may be unnecessary. However, the fact that robots are not physical entities does not exclude their criminal responsibility, since the issue is not a biophysical or metaphysical one, but a sociological or psychological one.[148]

[144] Monika Simmler and Nora Markwalder, 'Guilty Robots? – Rethinking the Nature of Culpability and Legal Personhood in an Age of Artificial Intelligence' (2019) 30 Criminal Law Forum 1, at 10.
[145] ibid at 11.
[146] ibid at 11, citing Gless and Weigend, 'Intelligente Agenten und das Strafrecht', ZStW 126(3) (2014), at; Wohlers, 'Individualverkehr im 21. Jahrhundert: das Strafrecht vor neuen Herausforderungen', Basler Juristische Mitteilungen 3 (2016), at 123–124.
[147] ibid at 11.
[148] ibid at 15.

According to Gless, Silverman, and Weigend, this is because it is assumed they are not morally responsible agents and therefore they cannot be the recipients of punishment, or, more explicitly, because they do not have the capacity to understand what punishment is.[149]
Moral accountability is indeed too complex for machines and humans alike. Generally speaking, there is no definition of morality that is valid for all societies and individuals Yet, although artificial intelligence technologies do exist and are present in our everyday lives, both in the private and industrial sectors, they cause harm from time to time, whether they are morally responsible or not.[150]

The idea of punishing AI from the start may seem confusing to sceptics – similar to hitting a computer that crashes. A machine ruled by artificial intelligence would surely lack the elements of criminal law such as culpability - a "guilty mind," which is characterized by a disregard for legally protected values - and the requirement of voluntary action. It is important to mention that, as seen elsewhere, there is not that much controversy regarding the attribution of mental states to animals, children, or people with mental disorders. However, claiming them to computers and corporations – and consequently being capable of committing offences - is highly controversial.[151] Still, the law already punishes artificial persons in the form of corporations, even though they do not have mental states and do not engage in voluntary activities.[152] Thus, although it seems to be inconceivable to attribute criminal responsibility to AI entities, abstaining from its establishment would also be unfeasible – even though the absence of a regulatory body to deal with it.

While the absurdity of blaming robots persists, holding someone or something blameworthy presupposes the moral agency of the blame. In this sense recent empirical research demonstrated that 'people are rather willing to ascribe *blame* to AI-driven systems and robots, to hold them *morally* responsible, and to deem their actions morally *wrong*'.[153]A number of studies examine the willingness of people to treat artificial agents, such as robots, as moral agents: Malle and colleagues observed that about 60-70% of participants felt comfortable blaming artificial agents for violating moral standards across multiple experiments.[154] In their remarks, 'a good number of ordinary people are ready to apply moral concepts and cognition to the actions

[149] ibid at 4, citing Gless, Silverman and Weigend, 'If Robots Cause Harm, Who Is to Blame? Self-Driving Cars and Criminal Liability', New Criminal Law Review 19(3) (2016), at 412.
[150] Hallevy (n 51) at 20-21.
[151] Stuart and Kneer (n 136) at 363:8.
[152] Abbott (n 71) at 112.
[153] Stuart and Kneer (n 118) at 363:2.
[154] ibid, citing B. F. Malle, S. T. Magar, and M. Scheutz, "AI in the Sky: How People Morally Evaluate Human and Machine Decisions in a Lethal Strike Dilemma," in *Robotics and Well-Being*, M. I. Aldinhas Ferreira, J. Silva Sequeira, G. Singh Virk, M. O. Tokhi, and E. E. Kadar, Eds. Cham: Springer International Publishing, 2019, pp. 111–133. doi: 10.1007/978-3-030-12524–0_11. [27] B. F. Malle, M. Scheutz, J. Forlizzi, and J. Voiklis, "Which Robot Am I Thinking About? The Impact of Action and Appearance on People's Evaluations of a Moral Robot," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, Christchurch, New Zealand, Mar. 2016, 125–132.

of artificial agents'.[155] Moreover, and contrary to expectations, participants are willing to judge artificial agents as harshly, if not harsher, than a human performing the same task.[156] In particular, participants justified high levels of blame towards either an artificial agent or a human based on the same things: the agent's thoughts, intentions, and ability to make choices.[157] Although artificial agents are not considered to be persons, they possess this autonomy and so it may be argued that they have something resembling a "will" or, at the very least, what is called "intentional states".[158]
A certain number of human-robot interactions tend to use the intentional stance to understand, explain, and predict the actions of robots. Moral judgement rests on a set of mental states, so the intentional stance leads to the appearance of moral agency in robots.[159]

Technological breakthroughs may lead to a situation in which social robots not only engage in simple interactions with humans but also perform highly complex daily tasks. The more complex and advanced a task is, the greater the chance for failure.[160] Then it will not be long before humans recognize the robot's autonomy for what it is and attribute to it the corresponding "capabilities".[161]

Whenever the concept of artificial intelligence social responsibility is discussed, the debate will always revolve around the concept of machines becoming more and more human-like. Criminal law is the primary social tool used in human society to deal with such situations since it stipulates the punishment for those who harm or endanger society.[162]

However, considering that an artificial intelligence system has its own, in a sense, designed cognition and will, courts cannot easily apply the traditional concept of culpability in intentional crimes, where the intent is determined by the actions of the offender.[163]

Conversely, rather than "baulking" the possibility of holding robots responsible, several studies – though not all – indicate that robots are held to the same levels of blame and responsibility as humans in otherwise identical situations. A possible explanation could be this: Across major moral psychology theories, inculpability is predominantly related to inculpating

[155] ibid, citing M. Fricker, "What's the Point of Blame? A Paradigm Based Explanation," Noûs, vol. 50, no. 1, pp. 165–183, 2016, doi: 10.1111/nous.12067.
[156] ibid at 363:3. Researchers conducted a study which consisted of participants being given different versions of the trolley problem. In the trolley problem, people are asked to divert a runaway trolley heading to kill many people to another track where it would only kill one. A switch of tracks and killing one person was deemed less wrong for artificial agents, who were blamed more than humans if they failed to save the vast majority of people.
[157] Stuart and Kneer (n 118).
[158] Pedro Miguel Freitas, Francisco Andrade and Paulo Novais, 'Criminal Liability of Autonomous Agents: From the Unthinkable to the Plausible' in Pompeu Casanovas and others (eds), *AI Approaches to the Complexity of Legal Systems* (Springer 2014) at 146, citing Sartor, G.: Cognitive automata and the law: electronic contracting and the intentionality of software agents. Artificial Intelligence and Law 17, 253–290 (2009).
[159] Stuart and Kneer (n 118) at 363:7.
[160] Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems.* (n 57) at 21.
[161] Simmler and Markwalder (n 144) at 15.
[162] Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems.* (n 57) at 22.
[163] Dremliuga and Prisekina (n 37) at 256.

mental states (or "*mens rea*", e,g., intention, knowledge, or recklessness). [164] Similarly, the way how human mental states shape our legal reactions to human crimes, the cognitive states of AI systems could determine our legal reaction to their harmful behavior: intentional or reckless harm caused by AI systems might require different responses from the "inculpable" harm caused by the same system.[165] Perhaps, if more people were willing to ascribe mental states to artificial agents, then their tendency to hold them morally liable might seem less bizarre.[166]

Ryan Abbot counterargues by positing that AI lacks culpability since it simply executes its programming, even if it behaves in ways that are intuitively blamed by society.[167] Nonetheless, by focusing solely on the criminal liability of humans instead of AI machines, courts fail to conduct a systematic analysis of liability. Considering machine learning's decision-making process can function without human involvement, criminal liability should not be limited to human beings. [168]

It is not in doubt that criminal responsibility has been designed to apply to humans rather than to other creatures or to the capabilities of other creatures. Human consciousness, soul, and mind constitute the mental element requirement. In light of this question, it is inevitable to ask whether artificial intelligence technologies can be compared with human standards of spirit, soul, and mind. The greater question, but not a legal one, is how criminal liability can be imposed upon spiritless and soulless entities, based on these insights.[169]

In a view of evaluating that, the following section will address the criminal liability of legal persons alongside the analysis of its application to AI entities.

B) Criminal responsibility of corporations

The term "AI crime" refers to cases in which AI would be criminally liable if a human did the same thing. Although machines have been causing harm since ancient times, and robots have been causing fatalities since at least the 1970s, the majority of machine-caused harm is regarded as an accident. In other words, in cases involving machines, criminal law is applied to individuals or companies instead of the machines themselves.[170]

Nonetheless, one cannot overlook the fact that we live in a rapidly developing society, characterized by the discourse of the global risk society, entailing

[164] Stuart and Kneer (n 136) at 363:4.
[165] Lagioia and Sartor (n 5) at 434.
[166] Stuart and Kneer (n 136) at 363:4.
[167] Abbott (n 71) at 117.
[168] Persaud (n 133) at 591, citing Claudia Geib, Lawmakers Want You to Be Able to Sue Robots, FUTURISM (Apr. 13, 2018), https://futurism.com/robots-rights-eu-personhood (discussing the point of view of European lawmakers who believe "electronic personalities" should be granted legal personhood such that it can be held accountable for its conduct).
[169] Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems.* (n 57) at 39.
[170] Abbott (n 71) at 113.

important paradigm shifts in our community's cultural, economic, sociological, and technological dimensions, as well as significant changes in the way criminality manifests. Criminal activity has become increasingly complex and organized in recent years, often involving corporations, societies, and associations as key players.[171] In the nineteenth century, courts recognized the need for large corporations to be held criminally liable because of the substantial impact they have had on the economy, especially the railroad industry. This resulted in corporations being punished for actions committed by individuals.[172] Over time, the traditional principle of *societas delinquere non potes*t[173] gradually begun to fade away. Due to this, most western national jurisdictions recognize corporations as legal persons and have enacted domestic laws regulating the criminal liability of corporations[174].
In essence, this type of criminal liability is legitimated by a material analogy between the behavior of natural persons and legal persons. With infants, one limits and removes blameworthiness – even though they have the capacity to act. Therefore, they do not fall under criminal liability. Parallel to that, it would not be completely unreasonable to punish artificial persons despite their incapacity to act on a physical or anthropological level.[175]
Since the human agency is no longer an absolute and insurmountable criterion: legal entities are now criminally liable for certain offenses, thus opening the door for AI-based entities to also be criminally liable.[176]

Corporate legal doctrine has evolved to allow for penal law to impute guilty mental states to corporations, even if corporations cannot formally satisfy the mental state (*mens rea)* requirements. In a corporate context, *respondeat superior* is one of the most important doctrinal instruments. It allows mental states possessed by an agent of a corporation to be attributed to the corporation if the agent was acting within the scope of their employment and in furtherance of the corporation's interests. *Respondeat superior* renders corporations capable of being convicted of crimes without violating the principle of legality since imputation principles of this kind are well understood and legally accepted. The same type of legal construction of mental states could be used to make AI punishable since corporations can be put within the ambit of proper punishment through this mechanism and avoid the eligibility challenge.[177]

Ultimately, criminal liability for legal persons is an innovative advance in criminal law and the models that supported that advance could offer us

[171] Freitas, Andrade and Novais (n 158) at 148, citing Dias, F.: Pressupostos da Punição e Causas que Excluem a Ilicitude e a Culpa. In: Centro de Estudos Judiciários (org.): Jornadas de Direito Criminal I, 41-83 (1983).
[172] Osmani (n 141) at 70-71, citing Dragatsi, H. (2011) Criminal Liability of Canadian Corporations for International Crimes. Canada: Thomson Reuters, 147.
[173] Literally means "corporations cannot commit crimes" and postulates that legal entities do not bear moral and criminal responsibility. See: Stoitchkova, D. (2010) *Towards Corporate Liability in International Criminal Law.* Utrecht: Utrecht University, 7.
[174] Osmani, 'The Complexity of Criminal Liability of AI Systems' (n 141) at 71.
[175] Freitas, Andrade and Novais (n 158) at 148, citing Costa, F.: Responsabilidade jurídico-penal da empresa e dos seus órgãos (ou uma reflexão sobre a alteridade nas pessoas colectivas, à luz do direito penal). Revista Portuguesa de Ciência Criminal 2(4), 537-559 (1992).
[176] Freitas, Andrade and Novais (n 158) at 153.
[177] Abbott (n 71) at 119.

immense insights into a plausible dogmatic framework for the criminal responsibility of artificial entities. Additionally, it demonstrates that criminal law is able to be flexible when criminal policy requires it.[178] Nonetheless, the entailing outcomings will probably enhance the anthropomorphization of robots, as will be discussed in the sequence.

C) Robots and the issue of personification

Artificial intelligence is fundamentally different from both animals and legal persons. Artificial intelligence entities traditionally do not qualify as legal persons. It is said that they are mere objects, and that is perhaps the crux of the question.[179] A robot is not alive, but it is also not a mere fiction, like corporations. Nevertheless, it has the potential to exist (at least after its initial creation) independently and without the involvement of humans, as it is conceivable that it can reason, which sets it apart from both legal persons and animals.[180]

Robots, according to some legal scholars, are objects created by humans and, as such, are perceived as items or property, which, by default, do not have any legal rights to any sort of control or command. Instead, they are simply seen as tools in the hands of their producers or owners. Conversely, others believe that AI systems differ from other objects because they actively intervene in human relations.[181]
While corporations already have a significant impact on our daily lives and social interactions, and thus have become agents in the social system, this is not yet true for robots. But the prospect of this change in the future is not out of the question.[182] Besides, robots may even be programmed to display emotions[183] and to react emphatically in the real world, like an embodied counterpart – making it easier to consider them as social actors than fictitious entities.[184] Incidentally, robotic rights have already gained prominence on a global scale. Sophia, a humanoid robot, was granted citizenship by Saudi Arabia in 2017.[185] Also, according to a British study, humans avoid lying to humanoid robots in order to avoid "hurting their feelings".[186]

---

[178] Freitas, Andrade and Novais (n 158) at 154.
[179] Ibid at 149.
[180] Lima (n 139) at 688.
[181] Osmani, 'The Complexity of Criminal Liability of AI Systems' (n 141) at 59.
[182] Simmler and Markwalder (n 144) at 19.
[183] Alok Jha, 'First Robot Able to Develop and Show Emotions Is Unveiled' *The Guardian* (8 August 2010) <https://www.theguardian.com/technology/2010/aug/09/nao-robot-develop-display-emotions> accessed 12 May 2022. See also Schwiegershausen, 'The World's First Robot with Feelings is a Big Hit', N.Y. Magazine (22 June 2015), available at: <https://www.thecut.com/2015/06/worlds-first-robot-with-feelings-is-a-big-hit.html>
[184] Susanne Beck, 'Intelligent Agents and Criminal Law—Negligence, Diffusion of Liability and Electronic Personhood' (2016) 86 Robotics and Autonomous Systems 138, at 142.
[185] 'Saudi Arabia, Which Denies Women Equal Rights, Makes A Robot A Citizen' (*NDTV.com*) <https://www.ndtv.com/world-news/saudi-arabia-which-denies-women-equal-rights-makes-a-robot-a-citizen-1768666> accessed 5 April 2022.
[186] Adriana Hamacher and others, 'BiB - Believing in BERT' See on the experiment with the robot 'Bert': Hamacher, Bianchi-Berthouze, Pipe and Eder, Believing in BERT: Using expressive

In contrast, we are faced with a dilemma today: if we define a robot as a kind of person, then the moral responsibility for its actions has to be reckoned with as well.[187] Humanity has encountered numerous problems in the past when trying to include women, slaves, and superior mammals in the circle of persons. Similarly, the slave gradually gained a more "human" legal character with rights and duties associated with freemen, so too the humanoid may gradually acquire a quasi-human view, as their abilities towards morality, aesthetics, creativity, and logic approach those of humans. [188] Likewise, by taking a Kelsenian perspective into consideration, it is possible to state that entities are holders of duties and rights, since a person is 'a totality of rights and obligations which have the behaviour of a human being as its content and thus form a unity'[189].
On the flip side, we might have to concede some moral rights to the robot, such as their right not to be switched off. Assuming artificial intelligence machines are capable of making autonomous decisions similar to what humans do, one must bear in mind that consciousness is closely related to the concept of personhood. In the event that an entity can have subjective experiences, and eventually suffer, then this entity should be treated as an individual. A sceptical perspective on consciousness shows that despite assuming that others have rights based on their capacity to suffer, we may not know what others are really feeling. In short, it seems that we protect the rights of others based on what we feel rather than what they are actually feeling.[190] In this context, the studies on robotic consciousness may force us to rethink our definition of the concept of person.[191] Consequently, the crimes of intent, negligence, strict liability, etc., will be profoundly altered. With AI and robots becoming more advanced and more integrated into our society, our notions of moral rights will also be forced to change. By allowing robots the same protections as other creatures, the question could naturally switch from "why should we allow robots to have rights?" to asking, "why should we continue to deny them?"[192]

In any case, the ground-breaking and thus contentious question is whether robots can be criminals themselves, i.e., whether they can be encountered the guilty mind element in criminal behaviors carried out by them, and this will be tackled in the next sections.

communication to enhance trust and counteract operational error in physical Human-Robot Interaction' (2016) <http://data.bris.ac.uk/data/dataset/1xkj9m7z4vi6l137gihezyd9o3/> accessed 3 May 2022.
[187] Riccardo Manzotti and Antonio Chella, 'Good Old-Fashioned Artificial Consciousness and the Intermediate Level Fallacy' (2018) 5 Frontiers in Robotics and AI <https://www.frontiersin.org/article/10.3389/frobt.2018.00039> accessed 1 April 2022.
[188] Lehman-Wilzig (n 68) at 447, citing Wiener, "The brain and the machine", in Hook. p.115-116.
[189] Hans Kelsen, The pure theory of law. University of California Press, 1967
[190] Turner and SpringerLink (Online service) (n 77) at 155.
[191] Manzotti and Chella (n 187).
[192] Turner and SpringerLink (Online service) (n 77) at 171.

D) *Mens rea* in criminal offenses perpetrated by AI systems

As already discussed in chapter one, under the general theory of criminal law, the main mental states requirements for criminal responsibility are knowledge, intent, negligence, etc. If an offender fulfils the *actus reus* and *mens rea* requirements, then liability arises.

By m*ens rea* being reported generally as a criterion by which individual culpability is assessed, it can be said that an act of causing harm with more *mens rea*, such as intent, would generally be considered more culpable than an act of recklessness or negligence with less *mens rea*.[193] Thus, how should we deal with the question of *mens rea* for AI?

First of all, recognizing *mens rea* of AI entities can be challenging. The first step is determining the specific developmental level of the AI entity. It should be remembered that not all AI entities possess the same abilities, such as cognitive ability, and this should be taken into consideration when determining if *mens rea* may be attributed to such entities. However, as aforesaid, autonomous systems such as robots will be considered in the analysis.

Furthermore, a different state of mind for each crime must be attributed to the accused.[194] Some writers[195] argue that the only mental requirement needed to impose criminal liability is knowledge, intent, and negligence, among others, and affirm that both knowledge and specific intent can be attributable to AI agents when these agents receive sensory data, which is analyzed by the AI entity. If an AI entity has sensors that provide it with data that could be processed internally, can we say that the entity comprehends or understands what information is being processed? [196]

To prove that the offender was fully aware of the crime beyond any reasonable doubt, as required by criminal law, is a difficult task. Awareness is an internal state of the mind that is not necessarily expressed externally. Because of this, criminal law has developed evidential substitutes. Substitutes like these are presumptions (such as the willful blindness presumption and awareness presumption), which, in certain situations, presuppose awareness

---

[193] '"Should the Model Penal Code's Mens Rea Provisions Be Amended?" By Kenneth W. Simons' at 181 <https://scholarship.law.uci.edu/faculty_scholarship/797/> accessed 14 May 2022.

[194] Freitas, Andrade and Novais (n 158) at 153.

[195] Gabriel Hallevy, 'The Criminal Liability of Artificial Intelligence Entities - From Science Fiction to Legal Social Control' (2010) 4 Akron Intellectual Property Journal 171, 171–201.
He argues that it cannot be denied that the structure of criminal liability has been designed for humans and their capabilities, rather than for other creatures or their capabilities. The requirement for mental elements is based on the human spirit, soul, and mind. At this point, the inevitable question is whether artificial intelligence technologies can be evaluated using human standards of spirit, soul, and mind. The deeper, but not legal, question is how criminal liability can be imposed on spiritless and soulless entities based on these insights. It is important to remember that criminal liability is not dependent on these terms of deep psychological significance. Criminal liability may be imposed, with or without spirit or soul, if an offender meets both the factual and mental element requirements of the specific offense- such as corporations.

[196] Freitas, Andrade and Novais (n 158) at 153.

and replace it.[197] Nevertheless, awareness here should not be acknowledged in the same sense as in psychology and philosophy, since awareness now is analyzed from the machine perspective.

An action taken by a machine is considered rash when it does not weigh one relevant factor as significant enough when deciding to act in a certain way. AI-technology decision making is generally complicated due to a large number of factors to be considered. The human mind in such situations tends sometimes to miscalculate the weight of some factors. The difference is: people are driven by hope and belief, computers are not.[198]

Let us assume a deep learning machine makes weighted decisions after performing data mining that is, ideally, loosely related to the original code that it received from its programmer. In other words, due to the extensive amount of data that deep machine learning has at its disposal, it is much more likely that it will make a decision based on its own analysis than if it were to blindly follow its source code.[199] For instance, consider the 2014 news article where an AI was programmed to purchase products on the black market. Initially, the AI was given instructions to buy items over the Internet but was not specifically directed to purchase illegal items. As the AI grew accustomed to purchasing items, it started to explore the depths of the Internet and eventually began purchasing illegal items, such as Ecstasy pills.[200] This agency would certainly count as a crime – a commissive one - if it were executed by humans. Here, the programmers that developed the bot will be liable if they did not set up proper restrictions regarding the type of goods that the web robot could purchase or the websites it could go for it. However, if it is proven that the machine developers did not act either recklessly or negligently, then it can be assumed that the web robot committed the crime. On the other hand, it will not be subject to prosecution, since it is not a legal person under criminal law.[201]
An analysis of the AI's code could provide valuable insight into its *mens rea* and whether it acquired that behavior on its own or whether it was instructed to do so.[202] Therefore, it can be concluded that if the AI was not programmed to acquire illegal items, but nonetheless did so, then it would satisfy both the *mens rea* and *actus reus* requirements to be criminally liable.[203]

Lastly, there is another aspect for determining criminal responsibility: punishment[204]. Its retributive feature imposes a constraint, grounded in justice, aiming to promote social welfare.[205] Thus, it is frequently argued that the punishment of robots that would have the same purpose as punishment of

[197] Hallevy, *When Robots Kill* (n 47) at 51-52.
[198] Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems.* (n 57) at 101.
[199] Persaud (n 133) at 599, citing Geib.
[200] Mike Power, 'What Happens When a Software Bot Goes on a Darknet Shopping Spree?' *The Guardian* (5 December 2014) <https://www.theguardian.com/technology/2014/dec/05/software-bot-darknet-shopping-spree-random-shopper> accessed 28 April 2022.
[201] Lagioia and Sartor (n 5) at 456.
[202] Persaud (n 133) at 598.
[203] ibid at 600.
[204] This subject is so wide that it cannot be adequately summarized in this space.
[205] Abbott (n 71) at 115.

humans is not feasible today.[206] Hence, the following questions will arise: How should artificial intelligence be punished? With prison time or monetary sanctions? Should the source code be removed, or should the robot be destroyed? Is the owner of a robot liable for the fine even if he cannot control it? Analogously, one could deduce it is the same as for damage caused by an animal, so criminal responsibility would be borne by the person who did not secure the animal.[207] However, it is challenging to apply to AI the laws on liability for animals. First, there is a difference between wild and domesticated animals; second, animals are limited by their own natural faculties. A variety of tasks can be taught to animals, depending on their species. For instance, a dog can be taught to retrieve a ball, but it cannot be taught to fly an airplane or perform brain surgery. Third, the ways in which an animal attains a goal are generally predictable and tend to be influenced by evolution rather than individual choice.[208]

Another argument to respond to those questions is that aside from the fact that corporate entities would hardly be 'punishable' in this sense (and thus the punishment of such legal persons, which is practiced in many legal orders, would be pointless), it is highly questionable whether this 'punishability' should really be a requirement for criminal responsibility. Mostly considering that it can be assumed that punishment is primarily defined by its symbolic force as a reaction to the disappointment of expectations.[209] Responsibility as attribution is a social operation, not a real substrate within a person. To call someone "guilty" simply means that we attribute a flaw, the disappointment of a normative expectation, to them.[210] Moreover, while punishing AIs is not possible conceptually, applying criminal law to them, in order to convict them of crimes, might be. Society may still benefit from AI convictions while avoiding the conceptual confusion that arises when AI may be punished.[211]

Since negligence is deemed the most probable common kind of *mens rea* to happen in the tech context regarding robots, this, and its implications for the imposition of criminal liability will be explored in the last subsection.

E) Negligence

'[N]egligence does not incriminate persons who are incapable of forming awareness, but only those who failed to use their existing capabilities to form awareness.'[212] As a general rule, negligence consists of cognitive instead of

[206] Simmler and Markwalder (n 126) at 27, citin Gless and Weigend 'Intelligente Agenten und das Strafrecht', ZStW 126(3) (2014), at 578.
[207] Karel Nedbálek, 'The Future Inclusion of Criminal Liability of the Robots and the Artificial Intelligence in the Czech Republic' (2018) 1 Expert: Paradigm of Law and Public Administration 86, at 93.
[208] Turner and SpringerLink (Online service) (n 77) at 56.
[209] Simmler and Markwalder (n 144) at 27.
[210] ibid 25–26, citing Gunther, 'Freiheit und Schuld in den Theorien der positiven Generalpra¨vention', in Schunemann, von Hirsch and Jareborg (eds), Positive Generalpravention (1998), page 157.
[211] Abbott (n 71) at 124.
[212] Hallevy, *Liability for Crimes Involving Artificial Intelligence Systems.* (n 57) at 122.

volitional aspects. Considering that volition is underpinned by cognition, and negligence does not require awareness, volition cannot be part of negligence.[213] Thus, even if the offender was unaware of their actions posing a risk, they can be held criminally liable under negligence.

It is important to note that one cannot disregard the possibility for robots to satisfy the fault element of certain offenses under criminal law – which is rooted in a fault-based framework.[214]Through the development of autonomous robots that can learn, we are building machines capable of taking over responsibilities even at the decision-making stage, as aforementioned in subchapter 2.3.2. Robots with adaptive and learning capabilities could react in unpredictable ways to new inputs if they were left to interact with humans in a non-supervised environment. [215] Robots, even though they are programmable to be very precise, can produce inaccurate results and be subjected to high rates of failure when they are placed in social situations – meaning that AI risks are never totally mitigated in practice at a single point in time. The reason for this is a simple fact that most social situations have subjective interpretations while robot tasks tend to be objectively defined. [216] Then, if the robot caused damage as a result of these reactions, it is unlikely that this was the fault of the programmer, producer, or user.[217]

The case in which an unmanned vehicle killed a pedestrian while crossing the road in the wrong place during testing is a classic example of criminal negligence. As a result, the driverless vehicle failed to anticipate the potential repercussions of its activities, through inaction, caused on society. Taking the appropriate precautions, foresight should have been able to predict these outcomes. In this event, it can be said that the erroneous activity of the artificial intelligence system is most likely the imperfection of algorithms at the present stage of its development. As a result, the main focus of blame should still be shifted to the program's creators if they could objectively and subjectively predict the appearance of errors and take preventative measures with the necessary care in the future. However, the feature of artificial intelligence is its self-learning ability, which can lead to inappropriate actions(inaction) resulting in damage to the user regardless of the actions taken by the manufacturer or the user. If and when the machine makes the decision on its own, it would not be fair to attribute the blame to another person. [218]

Therefore, the “black box” issue becomes indeed a challenge when it comes to assessing its liability regarding negligence. Since it can be difficult or even impossible to predict how an autonomous system will react, the matter of

[213] ibid at 123.
[214] Singapore Academy of Law and others (n 132) at 29.
[215] Beck (n 184) at 140.
[216]Sandeep Nagar, ‘Human-Robot Interactions: A Psychological Perspective’ at 6 <https://www.academia.edu/22879983/Human_robot_interactions_A_psychological_perspective> accessed 5 May 2022.
[217] Beck (n 184) at 140.
[218] Dr Atif Ali, ‘Artificial Intelligence and Criminal Liability: Challenges in Articulation of Legal Aspects for Counter-Productive Actions of Machine Learning’ (2021) 2 International Journal of Instructional Technology and Educational Studies 13, at 21.

what a reasonable person should (or should not) do to prevent the harm from arising might never be addressed. Liability for criminal negligence rests ultimately on its uncertainty. [219] Jurors using the reasonable person test are asked to put themselves in the perspective of a reasonable person and decide what that person would have done if they were in their place. The jury may have difficulty following that reasoning in the case of a reasonable robot, but it is a far less nebulous and fictional concept than the reasonable person.[220] The human mind is perhaps more like a black box of algorithms than artificial intelligence. When questioned appropriately, AI is more transparent about its internal rules, which can also be explicitly overwritten.[221] Accordingly, the courts would have the discretion to determine what a reasonable person/robot would or would not do in each case harm was caused by autonomous machines. By doing so, the courts can adapt or apply existing criminal negligence standards or establish new ones when there is no precedent. It is likely that any other negligence-based offense created specifically to cover damages caused by AI systems will be similarly broad and widely applicable. As a result, the law can be tailored to various circumstances. The downside, however, is that some autonomous systems may generate types of conduct for which existing precedents are inappropriate. Alternatives could include specifying the nature and extent of the relevant standards of conduct in legislation specific to a particular sector or technology, rather than leaving it to the courts to establish them over time.[222]

Thus, instead of focusing on whether an AI was negligently designed or marketed, the negligence test should evaluate the actions of the AI. This would apply the negligence paradigm to AI as an individual rather than a product, and the AI would be treated as a person as opposed to a product. The way an AI acted is what matters to an accident victim, not what it was thinking.[223]

2.4.CONCLUSION

This chapter discussed the likelihood of ascribing *mens rea* to AI entities, as well as if they are able to commit crimes and consequently the issue of their criminal responsibility.

Given that the essence of robots is their autonomous decision-making function, they are qualitatively different from existing technologies in that they must make independent decisions at times. This poses a challenge to established legal systems because, for the first time, a piece of technology is interposing itself between humans and a possible outcome.[224] Additionally,

[219] Singapore Academy of Law and others (n 132) at 31.
[220] Abbott (n 71) at 67.
[221] ibid at 138.
[222] Singapore Academy of Law and others (n 132) at 31.
[223] Abbott (n 71) at 62.
[224] Turner and SpringerLink (Online service) (n 77) at 64-65.

some conducts performed by robots indeed qualify as crimes if they were practiced by human beings, as was demonstrated.

In order to reach a proper conclusion, several aspects required for the purposes of convicting someone have been considered. Seen in this light, the potential moral justifications for granting AI personhood as well as its implications have been addressed. The main argument used to support the thesis is the analogous application of criminal responsibility to non-physical persons, such as corporations.

In view of the aforesaid and answering the questions posed in this thesis, it can be concluded that: on the one hand it is possible to ascribe *mens rea* to artificial intelligence entities if the humancentric approach when dealing with AI crimes is put aside. On the other hand, they will not be subject to prosecution since they lack legal personality under criminal law; thus, ruling criminal law cannot accommodate autonomous machines' behaviors.

Ultimately, the 'criminal responsibility of robots can thus be excluded today, due to the sociological fact that they have not yet acquired personhood, but it cannot be excluded for the future'.[225]

The third and final chapter will approach the lack of regulatory bodies to deal with the issue as well as possible solutions to the identified questions.

[225] Simmler and Markwalder (n 144) at 27.

CHAPTER 3 – THE LEGAL FUTURE OF AI-BASED SYSTEMS: ALTERNATIVES AND PERSPECTIVES

"What you can do now would be artificial intelligence fifty years ago. What we can do fifty years from now will not be artificial intelligence."[226] *Kevin Kelly*

3.1 GENERAL INTRODUCTION

In the preceding chapter, it has been analyzed the arising of AI entities along with their development into machine learning systems. It has also been argued the mental element issue of AI entities by exploring the machine consciousness and the possibility for self-learning programs to perform activities that would constitute criminal activities if human beings had carried them out. To examine the criminal liability of robots, it was necessary to assess if they are able to fulfil the requirements of *mens rea* offenses. It has also shed a light on relevant matters such as the criminal accountability of corporations and the personification of robots. Issues such as guilty mind and particularly negligence have been examined and finally, it has been demonstrated that AI systems can perform activities that would amount to crimes if they were executed by humans. The next step is to determine how the law may respond to AI crimes and also to look at solutions to the liability issue.[227] Prior to looking at possible alternatives to the liability issue, it is vital to argue that the absence of a legal framework for AI is the main issue to be dealt with.

3.2 THE LACK OF REGULATORY STRUCTURES TO ACCOUNT AI ENTITIES AND THEIR OBSTACLES

[226] Kevin Kelly, *The inevitable: Understanding the 12 Technological Forces that Will Shape our Future, Viking Press, 2016*
[227] Lagioia and Sartor (n 5) at 456.

In the early days of the Internet, there were no clear legal rules. It is known that most problems can be avoided at the beginning of technological development and introduction. Exceptions aside, legal rules follow along with the development of technology as technology creates powerful economic stakes. In this way, the law follows innovation but is not prepared for future challenges[228] – particularly when it comes to self-learning AI based-systems ones and their unpredictable behavior. Thus, as AI evolves and becomes smarter, more legal issues will arise.
New technological development can have an impact on society and its norms, which is why new regulations are necessary to deal with new dangers such as robots.[229]

Artificial intelligence systems can be regulated in a variety of ways - from requiring explainability to placing restrictions on how certain AI systems can be used. However, lawmakers and regulators still have not reached a broad consensus on what artificial intelligence is as a concept, an essential prerequisite to building a common standard to govern it. As an example, some definitions are drafted so narrowly that they only apply to sophisticated uses of machine learning. Other definitions (such as the one from the recent European Union proposal[230]) seem to apply to almost any software involved in decision-making, and these definitions would apply to decades-old systems. Different definitions of artificial intelligence are only one of many indications that global efforts to regulate AI are still in their infancy.[231]

At present, some[232] specific legal provision on AI exists – such as the European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).[233]Nevertheless, it is not in itself a legislative initiative, but instead a set of recommendations and it provides solely for civil liability. Although it is a significant step forward in the advancement of AI framework regulation, if only civil liability governed AI crimes, it would not be able to

[228] Dremliuga and Prisekina (n 37) at 257.
[229] Beck (n 184) at 138.
[230] 'Proposal for a Regulation Laying down Harmonised Rules on Artificial Intelligence | Shaping Europe's Digital Future' <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence> accessed 13 May 2022.
[231] Andrew Burt, 'New AI Regulations Are Coming. Is Your Organization Ready?' [2021] *Harvard Business Review* <https://hbr.org/2021/04/new-ai-regulations-are-coming-is-your-organization-ready> accessed 12 May 2022.
[232] AI regulation approach in the world: The United States of America (USA) has to date taken a rather hands-off approach towards AI regulation. The National Artificial Intelligence Initiative Act of 2020 was passed primarily with the objective to foster investments and research and development (R&D) in AI, and the US Federal Trade Commission believes that, at this stage, the existing US legal framework sufficiently addresses the risk of biases and discrimination associated with the growing use of AI systems. The United Kingdom (UK) published its National AI Strategy in September 2021, setting out how the UK will invest in AI applications and plans to present its AI regulation in 2022. At the international level, the Organization for Economic Co-operation and Development (OECD) has adopted a (non-binding) Recommendation on AI, while the Council of Europe is currently working on a legal framework for the development, design and application of AI. Furthermore, in the context of the newly established EU-US tech partnership (the Trade and Technology Council), the EU and the USA seek to develop a mutual understanding of the principles underlining trustworthy and responsible AI.
[233] European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) 2017.

respond to cases in which AI systems attempt to engage in criminal activity and then fail to do so.[234]

Of a conservative nature, the law must always cite precedents to justify its decisions. In the beginning, legal issues are analyzed according to concepts and principles of current law.[235] Furthermore, in common-law systems, judges strive to resolve specific cases in a way that best fits the legal landscape.[236]

In general, criminal law is retrospective and individual-centric and incapable of directing the development of technologies.[237]In contrast, technology is progressing at a pace that allows us to test the theory of criminal responsibility while revealing its foundations and sociological implications,[238] as the side effects of technological development will be more and more severely clashed into criminal law negligence regimes.

To further complicate the issue, in terms of robotics, the traditional orientation of criminal law - towards individual responsibility - is challenged.[239] Additionally, as already posed in the previous chapter, criminal rules apply only to human beings (and sometimes to corporations acting on behalf of humans). Because humans possess mental states, have the capacity to make decisions, and can be influenced by criminal laws, they qualify to be submitted to the law. Conversely, a non-human entity cannot understand what norms and sanctions mean, nor the social ramifications of criminal behavior, therefore they cannot be subject to criminal law and norms.[240]

Although all the issues mentioned above, the role of AI in society, as well as its relationship with individuals, will need to be regulated no matter what form it takes. It is unlikely that humanity will have to wait centuries to see the massive consequences of AI. According to McKinsey, compared to the Industrial Revolution 'this change is happening ten times faster and at 300 times the scale, or roughly 3,000 times the impact'.[241]
Likewise, as Broersen pointed out, 'our tendency to delegate responsibilities to artificially intelligent systems will become a serious problem for our society and for our legal systems globally'[242].
Hence, the regulatory approach should be international (at least to set out basic rules) since AI will not be limited within the boundaries of a particular jurisdiction.[243]

---

[234] Lagioia and Sartor (n 5) at 458.
[235] Barfield and Pagallo (n 135) at foreword xxi.
[236] Solum (n 111) at 1233.
[237] Beck (n 184) at 141.
[238] Simmler and Markwalder (n 144) at 1.
[239] Beck (n 184) at 138.
[240] Lagioia and Sartor (n 5) at 437.
[241] 'No Ordinary Disruption: The Four Global Forces Breaking All the Trends | McKinsey Global Institute | McKinsey & Company' <https://www.mckinsey.com/mgi/no-ordinary-disruption> accessed 16 May 2022.
[242] Jan Broersen, 'Responsible Intelligent Systems' (2014) 28 KI - Künstliche Intelligenz 209.
[243] Barfield and Pagallo (n 135) at preface xxv.

Robots and artificial intelligence must be regulated, but not because they could take decision-making powers away. Rather, the primary objective of regulation should be to ensure transparency and accountability over who is responsible for what – either civilly or criminally. The aim should be to promote an improved interaction between makers and workers/users, whose interaction is closely entwined.[244]

Ultimately, as this author sees it, crimes committed by AI systems demand a specific legal response, because they are particularly dangerous; not only might there be a liability gap, but they may also have extremely serious social repercussions. Consider, for instance, two cases in which a patient died as a result of therapy delivered by a medical robot. In the first case, the therapy was administered according to established protocols, but the patient died as a result of an allergy unknown to the robot; in the second case, the robot was aware of the allergy, was aware that a drug would cause death under the relevant conditions, but chose to administer the drug to kill the patient, possibly to save money on expensive treatment. Should the second case be managed differently because it would have been considered homicide if a human had been involved?[245] Thus, it may not be unimaginable or inconceivable for a robot to become guilty in the future, even as fictional as that may appear today. [246]

To date, developing minimal regulations towards criminal behavior would help allocate responsibility and liability for situations in which an artificial agent has 'learning and teaching' capabilities and is able to exercise unintended outcomes.[247]

Since AI systems lack a statutory framework governing their legal status, their development into legal persons remains merely theoretical.[248] Surprisingly, nonetheless, the European Parliament in the 2017 European Parliament Resolution on Civil Law Rules on Robotics aforesaid investigates whether robots should be granted legal rights as corporations enjoy. This will be discussed in the following subsection.

3.3 THE INTRODUCTION OF E-PERSONHOOD?

[244] 'A Law on Robotics and Artificial Intelligence in the EU?' (*etui*) at 9 <https://www.etui.org/publications/foresight-briefs/a-law-on-robotics-and-artificial-intelligence-in-the-eu> accessed 13 May 2022.
[245] Lagioia and Sartor (n 5) at 438.
[246] Simmler and Markwalder (n 144) at 1.
[247] ibid at 10, citing Grodzinsky F., Miller K. W. and Wolf M.J. (2008) The ethics of designing artificial agents, Ethics and Information Technology, 10 (2-3), 115-121. DOI 10.1007/s10676-008-9163-9; Vanderelst D.; Winfield A. (2016) An architecture for ethical robots, arXiv:1609.02931v1 [cs.RO] 9 September 2016. https://arxiv.org/pdf/1609.02931v1.pdf and Winfield A. (2012) Robotics. A very short introduction, Oxford, Oxford University Press.
[248] Osmani (n 141) at 56.

A solution to some of these challenges could be to impose liability directly on the AI system itself. However, in order to make that feasible, AI systems would require a separate legal personality – given that AI entities are not conscious as humans are according to the criminal law in force - like the way companies are given legal personalities currently.[249] Hence, the creation of a detached legal personality for AI might be an elegant and pragmatic solution to the issues raised.[250]
In the discussion of robot personhood, some researchers have advanced the concept of 'electronic personhood' or 'e-person.[251] This legal status intends to ensure rights and duties for the most capable AI agents. This would provide robots with the status of a legal subject.[252]

Legal personality for AI is no longer merely a topic for academic discussion. Many proposals have been made over the years to grant AI legal personality. In response to the advancements in AI and robotics, the European Parliament Resolution of 16 February 2017 mentioned in the previous subchapter further requested the commission to consider the possibility to:

> 'creating a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently'.[253]

Despite the vagueness of the concept of e-personhood - the Committee does not elaborate on the definition of electronic personhood - the report underlines that it could reflect a legal status similar to corporate personhood. It would establish specific legal rights and responsibilities for AI agents, but it is not intended to grant robots human rights.
European Union oversight of legal personhood for autonomous systems could serve as a starting point for allocating rights and responsibilities to Al systems; however, not all scholars accept this insight.[254]

As shown in subchapter 2.3.3 in the section regarding criminal responsibility in chapter 2, granting legal personality to artificial agents is complex. On the one hand, when robots or other autonomous artificial agents are attributed personhood, they become subjects (as opposed to things or objects) and enter

[249] Singapore Academy of Law and others, *Report on Criminal Liability, Robotics and AI Systems* (2021) at 36.
[250] Turner and SpringerLink (Online service) (n 77) at 81.
[251] Simmler and Markwalder (n 144) at 19, citing Beck, ÔU¨ber Sinn und Unsinn von Statusfragen – zu Vor-und Nachteilen der Einfu¨hrung einer elektronischen Person', in Hilgendorf and Gu¨nther (eds), Robotik und Gesetzgebung (2013); Gruber, ÔRechtssubjekte und Teilrechtssubjekte des elektronischen Gescha¨ftsverkehrs', Beck (ed), Jenseits von Mensch und Maschine (2012), page 150.
[252] ibid at 19, citing Muller, 'Roboter und Recht. Eine Einfuhrung', Aktuelle Juristische Praxis 5 (2014), at 604.
[253] European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) (n 233) at paragraph 59.
[254] Osmani, 'The Complexity of Criminal Liability of AI Systems' (n 141) at 60.

the realm of legal persons.[255] Nonetheless, the concept of corporate personhood is a legal fiction, a tool of convenience to make matters simpler. Hence, a regulatory "toolbox" for AI may also be an alternative to deal with the lack of personhood of sophisticated autonomous robots.

Beck considers a pivotal distinction between legal personhood for corporations and electronic personhood: electronic personhood may develop emphatic abilities and relate to humans more than legal persons. She argues that a novel type of personhood is required and should be designed particularly for electronic entities in order to bridge the gap that presently exists between traditional legal personality and the issues that arise from the actions of electronic entities that do not fall under the conventional personality.[256]

Moreover, when deciding whether to grant AI legal personality, the point is not whether the potential legal person understands the meaning of its actions. Indeed, we recognize the legal personality of humans who are unaware that they have it, such as young children and people in permanent comas. Despite the fact that children and those with diminished faculties are usually only able to act through other representatives, they are still legal persons. In this light, there is no magic to bestowing legal personality on AI. We do not declare it to be alive.[257] Making something an "electronic person" is more of a legal fiction than a philosophical assertion.[258]

Thus, the creation of a legal personality for a computer system does not mean that it will be treated as a human, nor will it serve as an excuse for people to blame computers for their actions.[259] Creating a new legal status for robots might be feasible to bring them into the current system of civil liability at least. For instance, let us assume that a disabled person relies on a home robot to help them around the house. This robot monitors the diet of that person to ensure they are receiving enough of the right foods. If the robot notices that the vegetable supplies are running low and orders some more, it is vital to ensure that a valid contract exists between the robot and the shop. It should not be expected for the grocer to claim that the contract is null and void because it was negotiated with a machine. Consequently, and in a purely technical sense, this makes it easier to grant the robot a legal personality.[260]

---

[255] 'A Law on Robotics and Artificial Intelligence in the EU?' (*etui*) at 7 citing Chopra S. (2010) Rights for autonomous artificial agents?, Communications of the ACM, 53 (8), 38-40; Chopra S. and White L. F. (2011) A legal theory for autonomous artificial agents, Ann Arbor, The University of Michigan Press.<https://www.etui.org/publications/foresight-briefs/a-law-on-robotics-and-artificial-intelligence-in-the-eu> accessed 13 May 2022.

[256] Beck (n 184) at 141-142.

[257] Turner and SpringerLink (Online service) (n 77) at 190 citing avid J. Calverley, "Imagining a Non-biological Machine as a Legal Person", AI&Society, Vol. 22 (2008), 523–537, 526.

[258] 'Giving Robots "Personhood" Is Actually about Making Corporations Accountable - The Verge' <https://www.theverge.com/2017/1/19/14322334/robot-electronic-persons-eu-report-liability-civil-suits> accessed 23 May 2022.

[259] Turner (n 71) at 205.

[260] 'Giving Robots "Personhood" Is Actually about Making Corporations Accountable - The Verge' (n 258).

In addition, a personality for AI could give it the motivation to adhere to certain rules that otherwise it might abandon or ignore due to a conflict with its own interests. By assuming artificial intelligence is trained to value its assets, giving it personality would give it a sense of ownership.
Although an AI entity may not be influenced by the psychological and emotional aspects of wanting to be seen by its peers as acting lawfully, it appears easier to imagine an AI system acting rationally to avoid asset depletion.[261]

Technology should not threaten our world – it should serve us. Therefore, one must have a realistic perspective on what is possible.[262]

3.4 CONCLUSION

This final chapter touched upon the urgent need to structure a legal body to deal with AI systems, as well as explained why AI presents a unique difficulty for legal regulation – mostly due to the fact that criminal law, in practice, 'is not always dictated by the anachronistic whims of society'.[263] However, in the absence of specific legislation for now it is essential to seek legal alternatives. The chapter has begun by setting out arguments which have been raised against making major legal changes to accommodate AI – although the existence of some legal provisions that have been heading by some praised initiatives. After that, it has been analyzed the creation of electronic personhood designed for AI systems in an attempt to fit them into established legal structures.

As AI becomes more self-sufficient, traditional theories in both criminal and private law face increasing difficulty in assigning culpability to recognized legal persons. Former Director of the Bureau of Consumer Protection at the US Federal Trade Commission, David Vladeck, stated the following:[264]

> So long as we can conceive of these machines as 'agents' of some legal person (individual or virtual), our current system of products liability will be able to address the legal issues surrounding their introduction without significant modification. But the law is not necessarily equipped to address the legal issues that will start to arise when the inevitable occurs and these machines cause injury, but when there is no 'principal' directing the actions of the machine. How the law chooses to treat machines without principals will be the central legal question that accompanies the introduction of truly autonomous

[261] Turner and SpringerLink (Online service) (n 77) at 188-189.
[262] 'Giving Robots "Personhood" Is Actually about Making Corporations Accountable - The Verge' (n 258).
[263] Turner and SpringerLink (Online service) (n 77) at 39.
[264] ibid at 185.

> machines, and at some point, the law will need to have an answer to that question.[265]

The purpose of this closing chapter is to advocate for a viable solution to the responsibility gap assessed throughout this thesis. This author acknowledges that no single approach is likely to be adequate to address the full range of AI system applications and potential harms. Therefore, an attempt to stress the benefits and drawbacks of those approaches has been set.

Importantly, we cannot lose sight of that by creating machines which make decisions for us and subsequently providing them with legal personality, we give away part of our (social) identity – or, maybe more properly, we reconstruct our identity in order to include them since we have previously decided to use them for a specific part of our autonomy. In reducing human decision-making potential in certain situations, and potentially making machines liable for these decisions, we change our understanding of autonomy, personhood, and responsibility.[266] And therefore, robots and other autonomous machines might be able to think outside their algorithm.

[265] David Vladeck, 'Machines Without Principals: Liability Rules and Artificial Intelligence' (2014) 89 Washington Law Review 117, at 150.

[266] Beck (n 184) at 142, citing S. Beck, B, Zabel, Person, Persönlichkeit, Autonomie –Juristische Perspektiven, in: O. Friedrich, M. Zichy (Eds.), Persönlichkeit –Neurowissenschaftliche und neurophilosophische Fragestellungen, Mentis, 2014, pp. 49–82.

CONCLUDING REMARKS

In conclusion, this thesis argues that artificial intelligence is unique compared to other technology created by humanity since it is capable of autonomous decision-making, i.e, taking independent decisions that were neither planned nor predictable by its designers.

However, huge problems will emerge coming most from a combination of two aspects: The first one is that AI entities are becoming more and more integrated into our lives. Second, is the lack of a holistic regulation that serves most jurisdictions, because technological advances are worldwide.

The thesis has identified two problems of a singular impact that will become more worrisome as technology evolves: the fact that AI entities are behaving and acting criminally and, consequently, their impunity stemming from their activities.

A legal principle rarely changes dramatically overnight but rather evolves over an extended period of time and this might be the biggest challenge to deal with the inverse proportion to the speed with which the AI technologies have been happening.

The thesis argues that, in theory, it is possible to ascribe *mens rea* to AI autonomous systems if the issue is considered by taking a machine approach. In contrast, this work concludes that they will not be subject to prosecution since they lack legal personality under criminal law; ultimately, ruling criminal law cannot accommodate autonomous machines' behaviors.

Importantly, my interest in arguing the thesis that some machines - robots - might be agents is not to show its absolute and irrefutable truth but rather to defend its plausibility. Of course, this work has more questions than answers, but the author hopes that it prompts future studies

Since autonomous systems and machines are capable of learning by themselves, one cannot rule out the chance of offenses being practised by these entities. Therefore, there is a responsibility gap detected in this work, as AI systems are not endowed with legal personality in order to fall under criminal liability.

In view of the challenges that have been arising in seeking to apply existing criminal laws and principles in relation to the actions of increasingly autonomous AI systems, it has also been considered ways in which those laws or principles might need to be adapted, or new laws or approaches adopted – such as the "e-personality" proposal.

What remains clear is that AI technologies will continue to involve new forms of harm, thereby continuously challenging legal and regulatory frameworks, and requiring legislators to respond with agility to emerging risks.[267]

One day, it is possible that the criminal law framework will be fully applicable to AI, however, the current capabilities of the systems do not meet the legal standard of awareness and volition that the criminal law requires.[268]

[267] Singapore Academy of Law and others (n 249) at 45.
[268] Rachel Charney, 'Can Androids Plead Automatism - A Review of When Robots Kill: Artificial Intelligence under the Criminal Law by Gabriel Hallevy Book Review' (2015) 73 University of Toronto Faculty of Law Review 1, at 70-71.

BIBLIOGRAPHY

Books and Books Chapters/Sections

Asimov, Isaac. *I, robot*. Harper Voyager, 2013.

Abbott, Ryan. *The Reasonable Robot: Artificial Intelligence and the Law*. Cambridge University Press, 2020.

Badar, Mohamed Elewa. *The concept of mens rea in international criminal law: the case for a unified approach*. Studies in international and comparative criminal law: 12. Hart, 2013. (1st ed. 2013), p.33, ISBN 978-1-84113-760-5

Barfield, Woodrow, e Ugo Pagallo, orgs. *Research Handbook on the Law of Artificial Intelligence*. Storbritannien: TJ International Ltd, 2018.

Bhat, P. Ishwara. "Objectivity, Value Neutrality, Originality, and Ethics in Legal Research". Em *Idea and Methods of Legal Research*. Delhi: Oxford University Press, 2020. https://doi.org/10.1093/oso/9780199493098.003.0003.

Casanovas, Pompeu, Ugo Pagallo, Monica Palmirani, e Giovanni Sartor. *AI Approaches to the Complexity of Legal Systems: AICOL 2013 International Workshops, AICOL-IV@IVR, Belo Horizonte, Brazil, July 21-27, 2013 and AICOL-V@SINTELNET-JURIX, Bologna, Italy, December 11, 2013, Revised Selected Papers*. Springer, 2014.

Fletcher, George P. *The Grammar of Criminal Law: American, Comparative, and International Volume One: Foundations*. Oxford University Press, USA, 2007. ISBN 978-0-19-510310-6

Freitas, Pedro Miguel, Francisco Andrade, e Paulo Novais. "Criminal Liability of Autonomous Agents: From the Unthinkable to the Plausible". Em *AI Approaches to the Complexity of Legal Systems*, organizado por Pompeu Casanovas, Ugo Pagallo, Monica Palmirani, e Giovanni Sartor, 145–56. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2014. https://doi.org/10.1007/978-3-662-45960-7_11

Goertzel, Ben, e Pei Wang. *Advances in Artificial General Intelligence: Concepts, Architectures and Algorithms : Proceedings of the AGI Workshop 2006*. IOS Press, 2007.

Hallevy, Gabriel. *Liability for crimes involving artificial intelligence systems.* Springer International Pu, 2016.

Hallevy, Gabriel. *When Robots Kill: Artificial Intelligence Under Criminal Law*. UPNE, 2013.

Hawking, Stephen. *Brief Answers to the Big Questions*. London: John Murray, 2018.

Hodges, Andrew. "Alan Turing". Em *The Stanford Encyclopedia of Philosophy*, organizado por Edward N. Zalta, Winter 2019. Metaphysics Research Lab, Stanford University, 2019. https://plato.stanford.edu/archives/win2019/entriesuring/.

Holland, Owen. "The Future of Embodied Artificial Intelligence: Machine Consciousness?" Em *Embodied Artificial Intelligence: International Seminar, Dagstuhl Castle, Germany, July 7-11, 2003. Revised Papers*, organizado por Fumiya Iida, Rolf Pfeifer, Luc Steels, e Yasuo Kuniyoshi, 37–53. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2004. https://doi.org/10.1007/978-3-540-27833-7_3

Kelly, Kevin. *The Inevitable: Understanding the 12 Technological Forces That Will Shape Our Future*. Reprint edition. New York, New York: Penguin Books, 2017.

Kelsen, Hans.The pure theory of law. University of California Press, 1967.

Levy, David. *Robots Unlimited: Life in a Virtual Age*. CRC Press, 2005.

McGuire, Michael, e Thomas J. Holt, orgs. *The Routledge Handbook of Technology, Crime and Justice*. Routledge International Handbooks. London: Routledge, 2020.

Norrie, Alan. *Crime, Reason and History: A Critical Introduction to Criminal Law*. Cambridge University Press, 2001. ISBN 0 406 93246 8

Rotelle, John ed, Edmund Hill. Augustine, Sermons (148-183) trans, New York City Press, 1992) vol 5, 315 -

Sayre, Kenneth M. *Consciousness: A Philosophic Study of Minds and Machines*. Random House Studies in Philosophy. New York: Random House, 1969.

Shelley, Mary. "Frankenstein; or, The Modern Prometheus." Library of Congress, Washington, D.C. 20540 USA Image. Acessado 5 de abril de 2022. https://www.loc.gov/item/53051218/.

Thakkar, Mohit. *Artificial Intelligence: A Theoretical Guide*. Mohit Thakkar, 2018.

Turner, Jacob e SpringerLink (Online service). *Robot Rules. Regulating Artificial Intelligence*. 1st ed. 2019. Springer International Publishing, 2019.

Wigglesworth, Cindy. *SQ21: The Twenty-One Skills of Spiritual Intelligence*. SelectBooks, Inc., 2014

Williams, Glanville. *Textbook of Criminal Law*. London: Stevens, 1983.

Scholarly Articles

Dremliuga, Roman, e Natalia Prisekina. "The Concept of Culpability in Criminal Law and AI Systems". *Journal of Politics and Law* 13, nº 3 (2020): 256–62.

Murphy, Brendon. "The Technology of Guilt". *Australasian Journal of Legal Philosophy* 44 (2019): 64–99.

Sayre, Francis Bowes. "Mens Rea". *Harvard Law Review* 45, nº 6 (1932): 974–1026. https://doi.org/10.2307/1332142.

Reports and Proposals

Singapore Academy of Law, Law Reform Committee, Kannan Ramesh, Charles Aeng Cheng Lim, Siyuan Chen Chew, Desmond, Josh Kok Thong Lee, Gilbert Leong, et al. *Report on Criminal Liability, Robotics and AI Systems*, 2021.

European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)) (2017). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52017IP0051.

Conferences

Bringsjord, Selmer, John Licato, Naveen Sundar Govindarajulu, Rikhiya Ghosh, e Atriya Sen. "Real robots that pass human tests of self-consciousness". Em *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 498–504, 2015. https://doi.org/10.1109/ROMAN.2015.7333698.

Dowling, Carolyn. "Intelligent agents: some ethical issues and dilemmas". Em *Selected papers from the second Australian Institute conference on Computer ethics*, 28–32. CRPIT '00. AUS: Australian Computer Society, Inc., 2000.

Web materials

A-Z Quotes. "Gottfried Leibniz Quote". Acessado 25 de maio de 2022. https://www.azquotes.com/quote/900273.

"Applications - Neuralink". Acessado 25 de maio de 2022. https://neuralink.com/applications/.
"Are We Prepared for the Rise of AI? | Ccier". Acessado 18 de março de 2022. https://cuts-ccier.org/are-we-prepared-for-the-rise-of-ai/.
"Atanasoff-Berry Computer | Britannica". Acessado 31 de março de 2022. https://www.britannica.com/technology/Atanasoff-Berry-Computer.
"Computer Science | Definition, Types, & Facts | Britannica". Acessado 14 de março de 2022. https://www.britannica.com/science/computer-science.
"criminal law | Definition, Types, Examples, & Facts | Britannica". Acessado 14 de março de 2022. https://www.britannica.com/topic/criminal-law.
"Giving robots 'personhood' is actually about making corporations accountable - The Verge". Acessado 23 de maio de 2022. https://www.theverge.com/2017/1/19/14322334/robot-electronic-persons-eu-report-liability-civil-suits.
"Law of the Twelve Tables | Roman Law | Britannica", 14 de março de 2022. https://www.britannica.com/topic/Law-of-the-Twelve-Tables.
"Machine Learning | Artificial Intelligence | Britannica". Acessado 18 de abril de 2022. https://www.britannica.com/technology/machine-learning.
"No Ordinary Disruption: The Four Global Forces Breaking All the Trends | McKinsey Global Institute | McKinsey & Company". Acessado 16 de maio de 2022. https://www.mckinsey.com/mgi/no-ordinary-disruption.
"Proposal for a Regulation Laying down Harmonised Rules on Artificial Intelligence | Shaping Europe's Digital Future". Acessado 13 de maio de 2022. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence.
"Siri | Computer Application | Britannica". Acessado 15 de abril de 2022. https://www.britannica.com/topic/Siri.
"Turing Test | Definition & Facts | Britannica". Acessado 15 de março de 2022. https://www.britannica.com/technology/Turing-test.
"Where did AI come from?" Acessado 14 de março de 2022. https://www.rs-online.com/designspark/where-did-ai-come-from.
Aeon. "What Frankenstein's Creature Can Really Tell Us about AI | Aeon Essays". Acessado 18 de maio de 2022. https://aeon.co/essays/what-frankensteins-creature-can-really-tell-us-about-ai.
"A Law on Robotics and Artificial Intelligence in the EU?" Acessado 13 de maio de 2022. https://www.etui.org/publications/foresight-briefs/a-law-on-robotics-and-artificial-intelligence-in-the-eu.
"A Complete History of Artificial Intelligence". Acessado 14 de março de 2022. https://www.g2.com/articles/history-of-artificial-intelligence.
IEEE Spectrum. "Qbo Robot Passes Mirror Test, Is Therefore Self-Aware", 6 de dezembro de 2011. https://spectrum.ieee.org/qbo-passes-mirror-test-is-therefore-selfaware.
IEEE Spectrum. "In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation", 25 de novembro de 2019. https://spectrum.ieee.org/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation.
Jha, Alok. "First Robot Able to Develop and Show Emotions Is Unveiled". *The Guardian*, 8 de agosto de 2010, seç. Science.

https://www.theguardian.com/technology/2010/aug/09/nao-robot-develop-display-emotions.
Law, Jonathan LawJonathan. "Actus Reus". Em *A Dictionary of Law*, organizado por Jonathan Law. Oxford University Press, 2018. https://www.oxfordreference.com/view/10.1093/acref/9780198802525.001.0001/acref-9780198802525-e-79.
LEMONNE, Eric. "Ethics Guidelines for Trustworthy AI". Text. FUTURIUM - European Commission, 17 de dezembro de 2018. https://ec.europa.eu/futurium/en/ai-alliance-consultation.
Matthias, Andreas. "The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata". *Ethics and Information Technology* 6, nº 3 (1º de setembro de 2004): 175–83. https://doi.org/10.1007/s10676-004-3422-1.
NDTV.com. "Saudi Arabia, Which Denies Women Equal Rights, Makes A Robot A Citizen". Acessado 5 de abril de 2022. https://www.ndtv.com/world-news/saudi-arabia-which-denies-women-equal-rights-makes-a-robot-a-citizen-1768666.
Norman, Don. "7: THE FUTURE OF ROBOTS", 1º de janeiro de 2003. https://www.academia.edu/2849702/7_THE_FUTURE_OF_ROBOTS.
Philosiblog. "Every Person Takes the Limits of Their Own Field of Vision for the Limits of the World." *Philosiblog* (blog), 19 de abril de 2012. https://philosiblog.com/2012/04/19/every-person-takes-the-limits-of-their-own-field-of-vision-for-the-limits-of-the-world/.
Piper, Kelsey. "The Case for Taking AI Seriously as a Threat to Humanity". Vox, 21 de dezembro de 2018. https://www.vox.com/future-perfect/2018/12/21/18126576/ai-artificial-intelligence-machine-learning-safety-alignment.
Power, Mike. "What Happens When a Software Bot Goes on a Darknet Shopping Spree?" *The Guardian*, 5 de dezembro de 2014, seç. Technology. https://www.theguardian.com/technology/2014/dec/05/software-bot-darknet-shopping-spree-random-shopper.
Schkolne, Steven. "Machines Demonstrate Self-Awareness". Medium, 6 de novembro de 2020. https://becominghuman.ai/machines-demonstrate-self-awareness-8bd08ceb1694.
Schwiegershausen, Erica. "The World's First Robot With Feelings Is a Big Hit". The Cut. Acessado 12 de maio de 2022. https://www.thecut.com/2015/06/worlds-first-robot-with-feelings-is-a-big-hit.html.
Signorelli, Camilo Miguel. "Can Computers Become Conscious and Overcome Humans?" *Frontiers in Robotics and AI* 5 (2018). https://www.frontiersin.org/article/10.3389/frobt.2018.00121.
Summer. "Four Crazy Real Cases of Humans Killed by Robots". Medium, 29 de março de 2022. https://historyofyesterday.com/four-crazy-real-cases-of-humans-killed-by-robots-7ab9bc0a9e38