Artificial Intelligence and the External Element of the Crime
An Analysis of the Liability Problem

Matilda Claussén-Karlsson

Spring 2017
Abstract

The rise of artificial intelligence ("AI") raises questions about liability for crimes an AI commits, mainly because the AI acts autonomously and with limited control from humans. The purpose of this thesis is to enquire this liability problem concerning AI, with focus on the external element of liability, the actus reus element. By using the doctrinal study of law as method, the thesis examines the liability problem by interpreting general theories and doctrines in criminal law, common to most legal systems. The analysis aims to define AI for legal purposes and to analyse whom to hold liable when an AI commits a crime de lege lata. With regard to the conclusions from the de lege lata analysis, the thesis then normatively analyses possible solutions to the liability problem de lege ferenda.

When the AI acts autonomously and the defendant omits to intervene there must be a legal duty to act for the defendant. It is not possible to state that the launch or use of an AI always constitutes a serious risk for harm. Depending on the situation, the actor may have a duty to act based on an assumption of a particular responsibility over the AI. Self-created dangerous situation where there is a serious risk that the AI causes harm can also impose a duty to act for the defendant. Limited foreseeability and unpredictability of the AI’s actions will however constrain criminal liability. The defendant can never be expected to avoid harms that were unpredictable from his or her position. Neither can the defendant be held liable for harms he or she did not cause. Besides the duty to act, the thesis acknowledges that an AI could be used as an innocent agent in order to perpetrate a crime, if the actor has possibilities to instruct or directly influence the AI’s behaviour. Still, the problematic features of AI persist.

The liability problem could supposedly be solved partially; either by introducing a civil law supervisory duty for the owner of the AI or by granting legal personhood for AI’s and thus create AI criminal liability. None of these solutions are sufficiently correcting the liability problem, though. But, a supervisory duty for the owner would be the most suitable solution of these two. It has the possibility to qualify the defendant’s behaviour as wrong when he or she breaches the civil law duty and the AI as a consequence causes (foreseeable) harm. The analysis draws the conclusion that criminal law may not be the best branch of law to solve these problems, and the liability problem with AI in criminal law remains challenging.
Preface

This thesis marks the end of my studies in law at Örebro University and thus symbols the beginning of a new era in my life. My choice of subject this thesis reveals the ‘nerdy’ side of me, the part of my brain that loves science, mathematics and the development of artificial intelligence.

AI and the rapid technological development will indeed affect the law, not only criminal law but law of all kinds. Hopefully, lawyers and jurists have a few more decades of work before we are replaced by artificially intelligent attorneys or robot judges. Maybe, we will one day be forced to write law in code instead of words and sentences in the future. Until that day, we must consider how to handle AI in the law of today. This thesis is a first attempt to handle AI in criminal law.

I would like to thank my parents Christel and Åke, for always giving me your support when life is not turning out as expected. I would also like to thank my supervisor Jacob Öberg for his important comments and advices regarding this thesis.

Matilda Claussén-Karlsson
### Table of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AI</strong></td>
<td>artificial intelligence.</td>
</tr>
<tr>
<td><strong>art/arts</strong></td>
<td>article/articles.</td>
</tr>
<tr>
<td><strong>BGH</strong></td>
<td>Bundesgerichtshof; the German Federal Supreme Court.</td>
</tr>
<tr>
<td><strong>c/cs</strong></td>
<td>chapter/chapters (of statutes).</td>
</tr>
<tr>
<td><strong>cf</strong></td>
<td>confer; compare.</td>
</tr>
<tr>
<td><strong>ECHCR</strong></td>
<td>European Convention of Human Rights.</td>
</tr>
<tr>
<td><strong>ed/eds</strong></td>
<td>editor/editors.</td>
</tr>
<tr>
<td><strong>edn</strong></td>
<td>edition.</td>
</tr>
<tr>
<td><strong>eg</strong></td>
<td>exempli gratia; for example.</td>
</tr>
<tr>
<td><strong>HovR</strong></td>
<td>Hovräten; Swedish Court of Appeal.</td>
</tr>
<tr>
<td><strong>ibid</strong></td>
<td>ibidem; in the same place.</td>
</tr>
<tr>
<td><strong>ie</strong></td>
<td>that is.</td>
</tr>
<tr>
<td><strong>n</strong></td>
<td>footnote.</td>
</tr>
<tr>
<td><strong>NJA</strong></td>
<td>Nytt Juridiskt Arkiv.</td>
</tr>
<tr>
<td><strong>OUP</strong></td>
<td>Oxford University Press.</td>
</tr>
<tr>
<td><strong>s/ss</strong></td>
<td>section/sections.</td>
</tr>
<tr>
<td><strong>SFS</strong></td>
<td>svensk författningssamling.</td>
</tr>
</tbody>
</table>
Table of Contents

ABSTRACT 2

PREFACE 3

TABLE OF ABBREVIATIONS 4

CHAPTER 1. INTRODUCTION 7
  1.1 ARTIFICIAL INTELLIGENCE AND THE LIABILITY PROBLEM 7
  1.2 PURPOSE AND DELIMITATION 8
  1.3 METHODOLOGICAL CONSIDERATIONS AND MATERIAL 9
  1.4 ETHICAL CONSIDERATIONS 11
  1.5 OUTLINE 12

CHAPTER 2. WHAT IS ARTIFICIAL INTELLIGENCE? 14
  2.1 GENERAL INTRODUCTION 14
  2.2 DEFINING THE TERMS 14
  2.3 SCIENTIFIC ARTIFICIAL INTELLIGENCE 15
  2.4 EXAMPLES OF AIs 17
    2.4.1 BOTS 17
    2.4.2 DRONES AND AUTONOMOUS CARS 18
    2.4.3 HIGH FREQUENCY TRADING AIs 19
    2.4.4 AUTONOMOUS WEAPON SYSTEMS AND MILITARY ROBOTICS 19
    2.4.5 AIs IN HEALTH AND MEDICAL SERVICES 20
  2.5 CONCLUSION 20

CHAPTER 3. AI AND THE ACTUS REUS 22
  3.1 GENERAL INTRODUCTION 22
  3.2 CRIMES AND CRIMINAL ACTS 24
  3.3 OMISSIONS 24
    3.3.1 OMISSIONS IN GENERAL 24
    3.3.2 A SERIOUS RISK FOR HARM? 26
    3.3.3 A DUTY TO ACT ASSUMING A PARTICULAR RESPONSIBILITY? 27
    3.3.4 A DUTY TO ACT BECAUSE OF A SPECIAL RELATIONSHIP TO THE HARM? 29
  3.4 CAUSATION AND AI 33
    3.4.1 CAUSATION IN GENERAL 33
    3.4.2 CAUSATION AND AI 34
  3.5 THE INNOCENT AGENT DOCTRINE 35
  3.6 CONCLUSION 36

CHAPTER 4. TWO POSSIBLE SOLUTIONS? 38
  4.1 GENERAL INTRODUCTION 38
  4.2 A SUPERVISORY DUTY? 39
    4.2.1 WHAT KIND OF LEGAL MEASURE? 39
    4.2.2 POSSIBLE EFFECTS 41
  4.3 AI CRIMINAL LIABILITY? 42
  4.4 CONCLUSION 43

CHAPTER 5. CONCLUDING REMARKS 45

REFERENCES 47
  LITERATURE 47
  LEGISLATION 50
I. Background

“There are two kinds of creation myths: those where life arises out of the mud, and those where life falls from the sky. In this creation myth, computers arose from the mud, and code fell from the sky.”

George Dyson

Chapter 1. Introduction

1.1 Artificial Intelligence and the Liability Problem

Humankind has for millenniums dreamt of creating an artificial being that thinks and acts humanly, in fiction as well as philosophy. This dream is about to come true, perhaps already within this century. The rapid technological change in our society has increased the quantity of technology influencing our lives. For example, some law firms have already hired their first artificially intelligent attorney. Our phones have artificially intelligent assistants that learn which applications we use the most and where we are heading when starting the car’s engine. Robotic nurses and surgeons are not fiction anymore. We are living in the age of artificial intelligence (“AI”).

AI can briefly be described as the science of making machines intelligent, to be able to perform tasks that generally require human intelligence. Driving a car, trading stocks at the stock exchange and defining a military target in war are examples of tasks that afore required human intelligence. Today, there are AIs able to perform the exact same tasks without a human involved.

AI technology is often using methods of reinforcement or machine learning to process big amounts of data. The AI learn its task gradually to be more efficient and to become better; just as humans – without further programming. Big technology companies are using AI together with reinforcement learning methods, to give the user ‘unique, personalized experiences’. In existing fields AIs have demonstrated a surprising ability to take unforeseeable decisions. Numerous AIs have also been involved in accidents of deadly character, where the

---

5 cf Bostrom (n 3) 14.
9 Bostrom (n 3) 16 and 138-39.
contribution from the AIs themselves are questionable. This has given rise to a public concern that there will be crimes committed without any possibility to hold a human liable.\textsuperscript{10}

Criminal law aims to prevent the occurrence of harm, embedded in communicating the wrongfulness and moral blame of the conduct that the crimes proscribe.\textsuperscript{11} The moral directions that criminal law gives us humans somewhat require the potential offender to be morally attributable and to be deterred by the threat of penal sanctions.\textsuperscript{12} The race towards creating a super-intelligent artificial being challenges criminal law, as human control is one of the essential keys when holding a person liable for a crime.\textsuperscript{13} When an AI acts autonomously, the human’s limited control over the AI seems problematic already when examining the guilty act of the crime. The characteristics of AI will collide with the requirements for establishing liability, obviously.

Due to the absence of guidance concerning responsibility over the AI’s behaviour from legislation and cases, criminal law and its principles will be the utmost constraint limiting how far we can stretch human responsibility over the AI. Previous research in law and AI is limited in general, to be almost absent in criminal law and AI, except Hallevy’s \textit{Criminal Liability for Artificial Intelligence Systems}\textsuperscript{14} and a few scholarly articles of which Karnow’s\textsuperscript{15} and Grimm, Smart and Hartzog’s\textsuperscript{16} are very valuable for this thesis. But, the research is not as comprehensive as to state there is nothing else to explore within this branch of law. Most scholars seem to focus their research on civil liability such as product and tort liability, although the problems that AI gives rise to are even worse in criminal law.\textsuperscript{17} Hopefully, this thesis can provide with useful guidelines on how to solve the liability problem the first part of the \textit{actus reus}, i.e. concerning whom the defendant is or should be and why, when an AI causes harm.\textsuperscript{18}

1.2 Purpose and delimitation

The main purpose of this thesis is to examine and enquire the legal position in the external \textit{actus reus} element of criminal liability for crimes an AI commits. The enquiry includes defining AI legally and assess when a human can and cannot be held liable for an AI’s crime. The author aims to examine these questions from a general perspective based on the doctrines of the general part of criminal law, common to most legal systems.\textsuperscript{19} In addition, the author aims to examine several solutions that possibly can meet the future of AI technology and solve the liability problem. Subsequently, the following sub-questions must be addressed:

- What is AI and how can AI be defined for legal purposes?

\textsuperscript{10} Stone and others (n 7) 46-47.


\textsuperscript{12} Simester and von Hirsch (n 11) 18.

\textsuperscript{13} George P Fletcher, \textit{Basic Concepts of Criminal Law} (OUP 1998) 44.


\textsuperscript{17} Stone and others (n 7) 46-47.

\textsuperscript{18} Bostrom (n 3).

\textsuperscript{19} eg Fletcher (n 13).
• Who can be ascribed criminal liability for an AI’s crimes?
• De lege ferenda, who should be ascribed criminal liability for crimes an AI commits?

The thesis is not limited to any particular criminal legal system. Instead, the fundamentals of the thesis are doctrines of the general part of criminal law, common to most legal systems. The study is confined to the external element of criminal liability (actus reus) when an AI commits a crime, i.e. to offences where an AI is the de facto offender. Accordingly, the internal element of the crime (mens rea) will not be assessed other than briefly, as the actus reus may exist without mens rea, but mens rea will not exist without a conduct to blame. Elements of specific crimes will only be examined merely and shall be seen as illustrations. The de lege ferenda analysis enquires civil liability and legal personhood for AI, and consequently the thesis leaves the criminal law partially. Still, this ‘detour’ is only an attempt to solve the liability problem for the criminal law, and the thesis do not analyse civil liability otherwise.

Moreover, secondary participation in forms of providing the offender with aid, abet, procurement or counsel will not be analysed. Possible defences such as justifications and excuses are out of the scope of the thesis, and will not be examined at all. In addition, civil liability will only be mentioned briefly in the de lege ferenda analysis. Other aspects of AI in law will only be mentioned in passing.

1.3 Methodological Considerations and Material

Among the most debated issues within legal scholarship is what legal method jurists use in their research, and under what label the method should be categorised. The debate demands methodological considerations concerning the legal reasoning in this thesis. To begin with, the thesis encompasses legal reasoning de lege lata and de lege ferenda, as it is based on a legal issue that needs a legal solution rather than only confined to interpret a certain law or set of rules. This thesis does not only include doctrinal studies de lege ferenda, but also the philosophy of law and legal theory, to strengthen the legal analysis within. In order to recognise and examine general concepts of the actus reus element in the general part of criminal law, the thesis takes advantage of analysing different sources of law from different legal systems. The study of the sources of law to establish what the law is, is generally labelled as legal dogmatics or the doctrinal study of law inter alia. Many jurists are criticising the use of the term legal dogmatics for excluding legal reasoning de lege ferenda, while others claim it includes such reasoning, indeed. To avoid misconceptions, the thesis will refer to the doctrinal study of law as its method. The aim of the doctrinal study of law is to interpret and analyse the law as it is and what it should be. It allows the analysis to remain critical and reflect over the law as it is, to draw conclusions about how the law should be.

20 Andrew P Simester and others, Simester and Sullivan's Criminal Law (6th edn, Hart Publishing 2016) 72, about the fact that thoughts alone ought not to be punished.
21 cf Simester and others (n 20) 218-24.
23 Stig Strömholm, Rätt, rättskällor och rätstillämpning (Nordstedts Juridik 1996) 41.
25 cf Peczenik (n 24) 313.
26 ibid.
Within this method, the sources of law including legislation, case law, legal doctrine, general principles of law and preparatory works are used in a dynamic hierarchy. The hierarchy varies between legal systems, accordingly. The analysis uses legal doctrine in forms of literature and articles as sources of law, together with cases to a greater extent than what is usual in the doctrinal study of criminal law. Normally, legislation is the primary and only binding source of law with respect to the rule of law. In absence of cases and legislation that targets AI directly, the analysis uses the existing sources of law to reason normatively about how these sources should be interpreted in cases concerning AI in criminal law.

A reasoning de lege ferenda loses most of its dignity when it appears to be contrary to general principles of law. The general principles recognised in criminal law explain the rules and the motive behind them. Even though reasoning de lege ferenda can be disengaged from previous or existing norms, general principles will be the ultimate boundaries limiting the reasoning together with prevailing higher norms. General principles of law can be derived from a higher norm or from methodological analysis of other sources of law. Legal doctrine is formidable when it comes to systematising the law to extract the recognised general principles that have support in the law. In a difficult case, principles can be of useful help for judge to reach a solution.

Regarding material, this thesis relies on a large amount of material and only the most important material will be pointed out hereafter. The traditional sources of law will be used and analysed dynamically, to distinguish and apply the basic concepts of criminal law to an AI’s crimes. Legal doctrine in forms of literature and articles are crucial for the analysis, especially concerning legal systems of which the author is unexperienced. In order to uphold a high quality of the analysis throughout the whole thesis, the leading textbooks on criminal law and criminal legal theory provide the thesis with great insight in every analysed legal system.

Theories of the general part of criminal law provide the thesis with the fundamentals of the analysis. Hence, legal doctrine in forms of standard companions such as Simester and Sullivan’s Criminal Law, Ashworth’s Principles of Criminal law and Kriminalrättens grunder, coupled with other legal doctrine, are used to strengthen the basic assertion of the similarities in the general parts of different criminal legal systems. In absence of authorised translations of penal codes into English, textbooks written directly in English concerning the doctrines and principles of the general part of criminal law are extremely useful to avoid any discrepancy that

27 Legal doctrine means both the study of law, and the result of the doctrinal study of law. Here, it is used as the result of the doctrinal study of law. Peczenik (n 24) 260.
28 Peczenik (n 24) 241.
29 Simester and others (n 20) 48.
30 ibid 99.
31 Strömholm (n 23) 323.
32 Peczenik (n 24) 189.
33 ibid 308-09.
36 cf Peczenik (n 24) 260.
37 Simester and others (n 20).
39 Asp, Ulväng and Jareborg (n 22).
otherwise risks occurring. For example, Bohlander’s Principles of German Criminal Law\(^{40}\) provides with valuable theoretical background of the German criminal legal system. In chapter three the thesis derives general doctrines of the actus reus from these authoritative works together with the Penal Codes, in order to interpret the doctrines in relation to AI in criminal law.

Current research in AI and law will be cited as sources. But, to examine criminal liability for crimes an AI commits, there is a pressing need for use of material from other sciences; mainly computer science, mathematics and philosophy to define AI legally and to understand the relationship between AI and law. Nevertheless, the material is used mostly as informative sources about AI and its current phase in development. Among the sources are Norvig and Russell’s authoritative work Artificial Intelligence, a Modern Approach\(^{41}\) and Nilsson’s The Quest for Artificial Intelligence\(^{42}\) among the most frequently used. These works provide the thesis with the mathematical and technological foundations of AI, as well as a helpful guide when attempting to find a definition of AI coherent with the general idea of AI among scientists in mathematics and computer science. In addition to these works, the thesis relies on recently published interdisciplinary research reports such as "Artificial Intelligence and Life in 2030"\(^{43}\) from Stanford University. Where there is legal doctrine in the intersection between law and AI, these are used and cited, although almost all available sources on law and AI analyses civil liability. By analysing these works in chapter two, the thesis aims to define AI in a way that is suitable both for the law and for AI as a science. Besides, in chapter two the thesis uses articles from reliable newspapers to exemplify what an AI can be used for in society and to illustrate how easy things can go wrong. The simple reason is the impossibility to receive updated information about occurrences involving AI otherwise. This examples then serve as a basis for the fictive examples that chapter three analyses.

To summarise these methodological considerations, the thesis uses the doctrinal study of law as method, since it is the most appropriate method in pursuance of the thesis’ purpose. The thesis studies these materials beside the sources of law in order to (1) find information about and systematise the doctrines of the general part of criminal law and (2) normatively apply these doctrines to AI in criminal law.\(^{44}\)

1.4 Ethical considerations

The research in this thesis comprises a lot of ethical considerations, mainly divided in three parts. Firstly, one must consider the use and handling of data concerning natural and legal persons within the thesis. In criminal law, the state is wielding its power against the individual defendant, who is in a very vulnerable position, exposed as a suspected criminal. It is inevitable, though, that the cases include personal data, both apropos the defendant and the victim of the crime. In tradition, many legal systems generally name cases with the surname of one or both of the parties, e.g. Claussén v. State or something similar. When the thesis refers to these cases

\(^{40}\) Michael Bohlander, Principles of German Criminal Law, (Hart Publishing 2009).
\(^{41}\) Norvig and Russell (n 6).
\(^{42}\) Nilsson (n 2).
\(^{43}\) Stone and others (n 7).
\(^{44}\) Aarnio (n 34) 19.
to mention the name of the case is unavoidable. Subsequently, the choice is to refrain use of any names or initials either than in the name of the case and instead simply refer to the ‘the defendant’. In contrast, when the analysis is discussing real occurrences and existing AIs the real names are referred to. The main reason is that such phenomena have been frequently featured in media, which often mention the nicknames of the AIs initially in the headlines. The AIs and their producers serve primarily as illustrative examples of existing AIs and therefore these are mentioned by their real names.

Secondly, it is necessary to consider interpretation and implementation of the sources of the law from an ethical point of view. When lawyers in general and judges in specific interprets the law in order to administer justice, the solution is reached through ethical considerations of what is right and what is wrong on the basis of the sources of law. As unique as every human being is, as unique is every criminal case that reaches the courts to get a solution; a judgment over what the defendant did, or not did, wrong. In criminal law, justice relates to justice by the rule of law, and not justice according to existing morals in society. Consequently, some forms of moral wrongdoing will fall outside the punishable area of wrongdoing. This is the price to pay for having a criminal legal system that is acceptable irrespective whom the defendant is. The defendant will always be someone’s family member, even though he or she is not a member of your family. This is of the essence when the defendant is to be attributed as a principal of the crime, when an agent has committed the crime de facto.

The last ethical issue concerns AI as an occurrence. AI is a part of our society and some AIs are already taking part in issues that require ethical considerations, at least as a part in the process to solve a problem. According to the principle of nullum crimen sine lege, everything not explicitly forbidden in law is allowed and can never constitute a crime. Likely, there will be AIs and activity concerning AIs that are legal, yet constitute what society deems as immoral and unethical. This thesis concerns the liability problem of AI, which is an opaque ethical issue in criminal law. Is it ethically appropriate to hold a person liable in certain situations? All necessary ethical considerations concerning AI cannot be examined within this thesis, due to the thesis’ scope. It is for the legislators and policymakers to consider and decide the ethics of AI and its future.

1.5 Outline

This thesis consists of three general parts. Part I, the Background, comprises this introductory chapter one, and chapter two. Chapter two analyses the first sub-question of this thesis and aims to define artificial intelligence in a way that is acceptable in law as well as among AI scientists and entrepreneurs. Thereafter follows a brief description of existing AIs that fulfil the stated definition, in order to illustrate for the reader what AI is, what its current capabilities are and the impact AI has on our lives already.

45 Peczenik (n 24) 146.
46 Ibid 89-91.
47 Ashworth (n 38) 70.
48 For a more comprehensive overview over the ethics of AI, see Norvig and Russell (n 6) 1034-40.
Part II, *AI and the Liability Problem*, consists of chapter three and contains the deeper analysis of AI in criminal law *de lege lata*. Chapter three analyses whom to hold liable, i.e. which actors that can be considered when discussing liability for crimes an AI commits, and why these actors should be considered. The analysis focuses on actors that may have an influence or impact over the AI’s decision-making. The chapter examines AI and the general part of the *actus reus*, i.e. criminal liability from an external perspective, to clarify when an actor can be held liable, and when no one is liable for the harm an AI creates. In addition, chapter three demonstrates what theory to use when, by providing the reader with fictive examples inspired by existing AIs.

Part III, *AI and the Legal Future* includes chapter four and five. Chapter four analyses AI and the liability problem from a *de lege ferenda* perspective, due to the conclusions drawn in the previous chapters. The chapter analyses a supervisory duty and AI criminal liability as solutions to the issues regarding criminal liability that chapter four illustrates, based on the principles for criminalisation of conduct. Chapter five is the final chapter and comprises a conclusive summary of the analysis and its conclusions, as well as a brief prognosis about the future of AI and its challenges for the criminal law.
Chapter 2. What is Artificial Intelligence?

“Intelligence is the ability to adapt to change.”

Stephen Hawking

2.1 General Introduction

What is artificial intelligence? It is a great yet difficult question to answer. Many scientists and journalists refer to the term artificial intelligence, AI, without a notice of what this phenomenon is or what it aims to be. One could believe AI is something that is so axiomatic that a clear definition is unnecessary. In contrast, it is rather a consequence of the fact that scientists over the world have not yet reached consensus about the definition of AI.

No one has provided the law with a legal definition of AI yet, since the legislators obviously tend to regulate occasions that have already occurred, rather than to look at present or future phenomena. This analysis aims to enquire what AI is and how it can be defined for legal purposes.

2.2 Defining the Terms

It is suitable to begin with an attempt to find the lexical meaning of the words artificial and intelligence. The English word artificial is synonym with words like factitious, synthetic and unnatural. A thing that is artificial is man-made or constructed by humans, usually to appear like a thing that is natural. The Swedish word for artificial, artificiell, and the French artificiel have equivalent synonyms. The Latin precedent artificialis origins from artificium, meaning handicraft or theory. In relation to law, artificial is used as in artificial person (i.e. legal person) and artificial insemination (i.e. human assisted reproduction). Artificial is thus used in the same manner irrespective the branch of law.

The word intelligence is more difficult to define. In English, as well as in Swedish and in French, the word has many meanings. Intelligence is explained as the ‘faculty of understanding’, ‘the action or fact of mentally apprehending something’ or simply as ‘intellect’. Hawking’s definition of intelligence stated above is also useful, but like most of the other definitions, it is vague. How does one adapt to change; by simply accepting the change or by learning how to handle the change, for instance? Accordingly, intelligence must

---

50 Stone and others (n 7) 12.
53 For example, see the British Agriculture (Artificial Insemination) Act 1946 (c. 29) or NJA 1983 s. 320 about artificial insemination.
55 Above n 49.
be further explained, since the meaning of the word appears to be vague. Intelligence and what it is de facto, is contested among psychologists, and has been for a long time.\textsuperscript{56} As a result, there is not any standard definition of intelligence hitherto.

Another issue concerning the different approaches to intelligence is that most of them relate to the human intellect. Intellect, as a synonym to mental abilities, can be considered as limited to the cognitive brain.\textsuperscript{57} An intellectual person is generally considered as a person with high intelligence and a great ability to comprehend complex problems in its environment.\textsuperscript{58} In law, the word ‘person’ encompasses natural persons like humans, together with legal persons in forms of corporations etcetera.\textsuperscript{59} The philosopher John Locke once defined persons as ‘agents capable of a law, and happiness and misery’.\textsuperscript{60} Expectantly, even an artificially intelligent agent will be capable of law, as happiness and misery relates to cognition and emotions rather than rules.\textsuperscript{61}

Conclusively, artificial intelligence might lexically be understood as an unnatural or synthetic intellect. Yet, AI represents more than this literal explanation. Words, as trivial parts of a sentence, give the sentence a practical meaning.\textsuperscript{62} It is therefore necessary to examine AI in a broader and more scientific context in order to find the practical meaning of artificial intelligence.

2.3 Scientific Artificial Intelligence

In 1955, the computer scientist McCarthy described the goal of AI as “to develop machines that behave as though they were intelligent”.\textsuperscript{63} As of today, we know that this definition is unsatisfactory with the requirements of modern technology. The leading AI computer scientists Norvig and Russell are providing us with more comprehensive thoughts about what AI is. They arrange the different scientific definitions of AI in four different categories of thought processes and human behaviour; (1) thinking humanly, (2) thinking rationally, (3) acting humanly and (4) acting rationally.\textsuperscript{64}

1. \textit{Thinking humanly} relates to the cognitive abilities a human brain has. That includes abilities such as making decisions, solving problems and to learn by mistakes and experiences. Imagine a machine with a mind.\textsuperscript{65}

2. \textit{Thinking rationally} can be explained as logical and deductive reasoning. For example, Aristotle’s syllogism that generates correct conclusions drawn from correct premises

\begin{itemize}
\item \textsuperscript{57} ibid.
\item \textsuperscript{58} ibid.
\item \textsuperscript{59} Ashworth (n 38) 114.
\item \textsuperscript{60} John Locke, An Essay Concerning Human Understanding (Book II 27th edn, T. Tegg and Son 1836) 234.
\item \textsuperscript{63} Wolfgang Ertel, Introduction to Artificial Intelligence, (Springer-Verlag 2011) 1.
\item \textsuperscript{64} Norvig and Russell (n 6) 2-8.
\item \textsuperscript{65} ibid 3.
are demonstrating rational thinking. ‘All humans can think; Steve is a human; therefore, Steve can think.’

3. **Acting humanly** is the objective approach to AI taken by Alan Turing in 1950, when he proposed the *Turing Test* that involved what he called the *Imitation Game*. The test is still relevant today, and involves mostly of the disciplines concerning AI. If a human interrogator cannot decide if a human or computer has answered a couple of written questions, the computer has passed the test.

4. **Acting rationally** is another objective approach to AI, where the AI is acting for the best result possible always, like a rational agent. For a computer to act rationally, it must be able to act humanly (3). This is the most developed form of AI.

Although Norvig and Russell’s taxonomy is broad, it serves with different attributes of a potential AI, rather than a clear definition. The taxonomy is also targeting extremely developed and advanced general AI, which do not exist yet. It will consequently leave narrow AI systems that already exist out of the scope. Narrow AI systems are intelligent when solving a specific problem, but would not pass general intelligence tests such as the Turing Test, for instance.

In order to develop a general theory of liability for crimes involving all kinds of AIs, AI must be defined more broadly.

The computer scientist Nils J. Nilsson provided the debate with a broad and important definition of AI a few years ago:

‘Artificial intelligence is that activity devoted to make machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment.’

According to this definition, both virtual assistants and a human brain are intelligent. The definition includes both narrow and general AI technology. A virtual assistant functions appropriately and with foresight in its environment, within the mobile phone. The assistant is performing tasks on command, and the assistant tells the user when it cannot perform the task or assist the human otherwise. A virtual assistant is not construed to walk away; nor would that be an appropriate function in its environment.

---

66 ibid 4.
67 Alan M Turing, ‘Computing Machinery and Intelligence’ (1950) 59 MIND 433.
68 ibid 435.
69 Norvig and Russell (n 6) 4-5.
71 ibid 7. Virtual assistants, the software in an autonomous car or algorithms used in high frequency trading at the financial markets are examples of narrow AI.
72 Or any other test of general intelligence. eg Shane Legg and Marcus Hutter, ‘Universal Intelligence: A Definition of Machine Intelligence’ (2007) 17:4 Minds and Machines 391.
73 Nilsson (n 2) 13.
74 Stone and others (n 7) 12-13.
Legg and Hutter, computer scientists as well, presented a similar definition of intelligence in 2007, which they call universal intelligence: ‘Intelligence measures an agent’s ability to achieve goals in a wide range of environments.’\(^75\)

The definition of universal intelligence, as well as Nilsson’s definition, is independent from definitions of human intelligence. Furthermore, both these definitions are neutral and evolution resistant. The definitions thus include both the AIs of today (i.e. like SIRI, autonomous cars, IBM’s Watson or search-engine algorithms) and tomorrow (human-like robots and autonomous weapon systems). ‘An agent’s ability to achieve goals’ might imply that all agents have a goal. Philosophically, this can be discussed forever. What is a goal? What is the purpose of that goal?

It is almost as difficult to define goal as to define intelligence. For regulatory purposes, Legg and Hutter’s definition are too vague.\(^76\) Therefore, Nilsson’s definition is more suitable from a regulatory perspective. This definition is also unlimited in relation to different kinds of technology. As we all know, modern technology evolves considerably faster than law and regulation. Nilsson’s definition is the broadest existing that could be found. The definition involves both the science of AI and the meaning of AI itself, but AI will be the term I use when I refer to AI as technology.\(^77\) It is also the most common definition that the bigger research projects concerning AI refer to when trying to define AI.\(^78\) The broadness signifies that the definition includes everything from calculators and simple algorithms to virtual assistant systems in smartphones and autonomous weapons systems. Conclusively, AI is an entity enabled to ‘function appropriately and with foresight in its environment’.\(^79\)

2.4 Examples of AIs

Before examining criminal liability, it is suitable with a few, selected, illustrative examples of already existing, real AIs that may cause problems with regard to criminal liability. The examples provided below are narrow AI systems, i.e. systems developed to perform a certain task. These examples will be recalled as examples in the de lege lata analysis.

2.4.1 Bots

Bots operating on the internet and Darknet are on the rise. Darknet is a totally anonymous part of the web, built upon TOR (“The Onion Router”) or similar technology.\(^80\) A few years ago, the Random Darknet Shopper, a bot programmed to go shopping on Darknet for an art exhibition, went wild and bought ecstasy pills among other illegal items despite any such instruction from the programmers, who had given the bot a 100 dollars in bitcoin to shop for every week.\(^81\) Comparably, a bot designed to compose tweets from words written in previous tweets, composed the death threat ‘I seriously want to kill people’, and published the threat.

---

\(^{75}\) Legg and Hutter (n 72) 12.


\(^{77}\) The researchers in the project ‘Artificial Intelligence and Life 2030’ (Stone and others (n 7)) are using Nilsson’s definition of AI too.

\(^{78}\) cf Stone and others (n 7) 12-13.

\(^{79}\) Nilsson, (n. 2) xiii.


using its owner’s Twitter account.82 The owner was investigated by the police, who claimed the threat should be seen as the owner’s own words, since it was published in his name.83

Another bot is Tay, Microsoft’s chatterbot developed with a millennial mind, who was autopsied within 24 hours after launch. Tay had machine learning technology and the thought behind Tay was that she would learn how millennials, i.e. older teenagers, act on Twitter, and then produce tweets and respond to questions sent to her. However, a few hours after launch, Tay’s Twitter account had to be shut down, since she began to show abusive, sexist and racist behaviour. She stated assertions like ‘Hitler was right’ and that ‘feminists should … burn in hell’ as well as ‘Taylor Swift rapes us daily’.84 Even though this behaviour was a result of interactions with other people online, it is an illustrative example of AI systems with a good intent from the producer and developer, but despite that the AI did act unforeseeable and not as intended.85

2.4.2 Drones and Autonomous Cars

Concerning cars, there are automated features that have been in use for over a decade already, e.g. parking assist systems and systems for cruise control.86 By time, nevertheless, autonomy has increased. Today, it is possible to buy a car with autopilot functions, that can change lanes without human assistance or drive itself to you in a parking lot. Several accidents leading to death involving autopilot systems have recently brought up a debate about the safety of these systems.87

Similar problems have arisen concerning domestic drones.88 Drones are used both for military and civil purposes, with their wide range of functions stretching from surveillance and deliverance of equipment to being used simply as toys. Police reports confirm that the use of drones in criminal activity are on the rise as well, from use in burglaries to acts of violence and vicinity against children.89

83 ibid. Another futuristic and philosophical question is if the freedom of speech and other human rights should be extended to cover AIs in the future.
86 Stone and others (n 7) 18-19.
2.4.3 High Frequency Trading AIs

Algorithmic trading and high frequency trading (“HFT”) algorithms are nowadays common at stock exchanges all over the world.\textsuperscript{90} By using machine learning technology, the algorithms can trade stocks and other financial instruments rapidly without human intervention. Consequently, incidents like the Flash Crash of 2010 are getting more and more common.\textsuperscript{91} An HFT algorithm triggered the Dow Jones Industrial Average Index by issuing a large order to sell Mini Futures for a total value of 4 billion dollars. An algorithm performed the order in error and HFT algorithms followed its steps, causing a great downturn in the index for a few minutes.\textsuperscript{92}

EU’s Market Abuse Regulation (No 596/2014)\textsuperscript{93} (“MAR”) and Market Abuse Directive (2014/57/EU)\textsuperscript{94} (“MAD”) address some of these issues, by forbidding certain identified strategies of algorithmic trading and HFT that give rise to market manipulation.\textsuperscript{95} The American Securities Exchange Commission has also adopted several regulations following the rise of computerised trading, but is focusing more on transparency.\textsuperscript{96} MAR imposes administrative sanctions on legal and natural persons for insider dealing and market manipulation, among other forbidden activities.\textsuperscript{97} MAD stipulates ‘minimum rules for criminal sanctions’ for insider dealing, unlawful disclosure of inside information and market manipulation.\textsuperscript{98}

There are examples of trading strategies from algorithms and HFT traders, which are claimed to manipulate the markets. An illustrative example is the Timber Hill case\textsuperscript{99} from the Norwegian Supreme Court (Høyesterett). Two investors were charged with market manipulation, when they had found out the market strategy of an algorithm and used their knowledge to make money. However, the court squashed the charges and stated that it was not proved whether it was the investors or the algorithm that had manipulated the market.

2.4.4 Autonomous Weapon Systems and Military Robotics

AI is currently used in many different ways in the military, e.g. in military robotics and semi-autonomous weapon systems. With regard to the increasing autonomy in weapons, several AI researchers and entrepreneurs among many others signed an open letter proposing a proscription of autonomous weapon systems (“AWS”), to avoid an AI arms race.\textsuperscript{100} Such

\textsuperscript{91} Jerry W Markham, Law Enforcement and the History of Financial Market Manipulation (ME Sharpe 2014) 318-19.
\textsuperscript{92} ibid 321.
\textsuperscript{95} cf art 12 and 15 MAR.
\textsuperscript{96} Markham (n 80) 323-25.
\textsuperscript{97} cf art 14-15 MAR.
\textsuperscript{98} Art 1.1 MAD.
\textsuperscript{99} Rt 2012 s 686.
weapon systems can acquire targets and initiate force without human intervention or supervision. But, whom to blame when an AWS acquires the wrong target and fires?

2.4.5 AIs in Health and Medical Services

Automation has not stayed away from healthcare either. AI is used for a wide spectrum of tasks within healthcare nowadays, from robotic surgeons to health analytics and diagnostics. Still, there are many legal obstacles to use of some kind of AI technology in healthcare though. For instance, if an AI gives a cancer diagnosis after analysing some test results, is the AI legally obligated to have a medical degree itself or is the AI just an assistant to the registered medical practitioner, i.e. the human? Or, whom to blame for maltreatment in healthcare, the practitioner or the AI helping the practitioner?

2.5 Conclusion

Evidently, there are many different kinds of AI that already exist or will exist in the near future. Common to all kinds is that most of the AIs were created with a good intention, yet some of them went rogue by buying illegal drugs or threatening other people. Accidents of deadly character with AIs involved have raised public concern about human liability in these cases.

To conclude this section, one could highlight some key features common to all kinds of AI.

- **Autonomy.** Humans are only limitedly involved or in the future not involved at all in the AIs decision-making. The autonomy varies between the different fields of AI, from autopilot mode in autonomous cars where the driver is required to stay in charge of the car, to the high frequency trading algorithms that function without humans engaging in their activity.

- **Unpredictability.** Like us humans, you can never know for sure how anyone else than yourself will react to something. An AI lacks cognition and may react totally different than a human facing exactly the same situation. In addition, most AIs discussed here are self-learning, i.e. they learn from mistakes and by processing a large amount of data. The outcome of the AI’s conduct is unpredictable, when the conduct is not a result of an instruction from the programmer, but a self-learned strategy.

- **Unaccountability.** As long as AIs lacks legal personality, they can behave in a way that for a human would have given legal consequences. The situation with accountability is comparable with the ones concerning animals. For example, a dog can

---

102 Stone and others (n 7) 26-29.
103 cf c 5, s 1 of the Swedish Patientsäkerhetslag (SFS 2010:659) (Patient Safety Act), which stipulates that other than licensed health and medical services staff are prohibited to examine patients professionally.
104 Stone and others (n 7) 46-47.
106 Sherer (n 76) 363.
107 Calo (n 105) 538.
bite a human to death, but will never be held legally accountable for their actions.\textsuperscript{109} As a consequence, we need to find a principal for the crime to hold liable. Increased autonomy equals decreased human control. Still, criminal law regulates human conduct. This take us back to a main question of criminal law: is it fair to be punished for an act you cannot control? Concerning criminal liability, already a narrow AI may cause issues when trying to find a liable person. But why? Criminal law in general targets humans and human behaviour. The general basis for criminal liability is usually the act requirement. However, only human acts be a ground for imposing a punishment.\textsuperscript{110} The observant reader may at this point realise there can be serious problems concerning liability for crimes involving an AI system, when the technology itself acts given the potential number of actors involved. Thus an AI’s crime must be possible to ascribe to a human, that can fulfil actus reus, the guilty act.\textsuperscript{111} This thesis focuses on the actus reus requirement, since without actus reus, there is no need to assess mens rea.\textsuperscript{112} No one ought to be punished for thoughts alone.\textsuperscript{113}

\begin{itemize}
\item \textsuperscript{109} Rather would the owner of the dog be held liable for damages. cf the British Animals Act 1971 (c 22) or the Swedish lag om tillsyn över hundar och katter (SFS 2007:1150) that imposes strict civil liability in tort for owners of cats and dogs.
\item \textsuperscript{110} Fletcher (n 13) 44; cf Asp, Ulväng and Jareborg (n 22) 71.
\item \textsuperscript{111} A Dictionary of Law (8th edn, 2015) 15.
\item \textsuperscript{112} ibid 395-96.
\item \textsuperscript{113} Simester and others (n 20) 72.
\end{itemize}
II. AI and the Liability Problem

“One can imagine such technology outsmarting financial markets, out-inventing human researchers, out-manipulating human leaders, and developing weapons we cannot even understand. 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.”

Stephen Hawking, Stuart Russell, Max Tegmark and Frank Wilczek

Chapter 3. AI and the Actus Reus

3.1 General Introduction

The general basis for criminal liability is usually the act requirement. Only human acts be a ground for imposing a punishment. An AI’s crime must be possible to ascribe to a human, that can fulfill the elements for criminal liability, actus reus and mens rea. This thesis focuses on the actus reus requirement.

In order to analyse the actus reus element, it is necessary to identify the actors involved in the AI and its decision-making. The first obvious actor is the user. The user is the person who launches the AI in the first place and instructs it about its tasks, and is benefitted by the AI’s work. The tendency thus far is that the user, together with the supervisor has been targeted in criminal investigations concerning Al’s behaviour. The next possible actor is the supervisor, who oversees the AI and has the possibility to intervene in the AI’s decision-making if necessary.

The producer produces the AI and may be responsible for everything that concerns the production of the AI, such as hardware, software and other features. The producer also knows the technology behind the decision-making process in the AI, at least in its state when introducing the AI to the markets. The producer is also the only actor that may affect the other actors’ expectation of what the AI de facto is capable of. The software developer counts as part of the producer in this analysis, even though the developer might be a third party acting on behalf of the producer.

The owner will in almost every case coincide with the user or the supervisor, and before sold the owner coincides with the producer. Despite that, it is necessary to mention the owner, as this role will be of importance for the de lege ferenda analysis. At last, the outsider is a third

---

115 Fletcher (n 13) 44; cf Asp, Ulväng and Jareborg (n 22) 71.
117 ibid 395-96.
118 Simester and others (n 20) 72.
119 Weaver (n 70) 21.
120 See s 2.4.
121 cf Bostrom (n 3) 167, about tripwires.
122 See s 2.4 about AIs that ‘learns wrong’ despite a good intention from the producers.
party who does not have a relationship to the AI, but despite that has the ability to affect the AI’s behaviour or interacts with it somehow. \(^{124}\) That can be a human such as a hacker or someone similar, or in hard cases, a malware or virus of some kind.

Whilst impact and ability will be the main determinants of responsibility as matters for causation of the consequences of the crime, they may also be important for criminal liability. Since the discussion here is confined to liability for crimes an AI commits, and we know that an AI is not legally accountable for its conduct, we need to trace the criminal behaviour back to a human behind the AI. That human must be in a position where he or she has a possibility to influence the AI and its conduct in one way or another. Seemingly this will be determined through consideration of the specific circumstances of each alleged crime.

However, it is suggested that the user’s and supervisor’s responsibilities are primarily linked to the use of the AI, i.e. when the AI is performing something. These actors may impact the AI by remotely control it, by giving exact instructions or by omitting to intervene and override the AI’s decisions. A user can, for instance, remotely control a drone and intentionally fly it into an aeroplane, or give the drone the exact instructions for how to fly while up in the air. From a liability perspective, the first of the aforementioned examples are not that difficult to solve, if you consider the drone as a simple tool used to injure the aeroplane.\(^{125}\) But, the question of liability gets more complicated as soon as the drone acts autonomously and ignores the instructions from the user and causes serious damage. Does the user or supervisor have a duty to override such behaviour by the drone?

The producer’s responsibility is primarily linked to the hardware and software of the AI, including all from mechanical elements to the code and algorithms within, and the education and training of the AI.\(^{126}\) The producer may influence the AI in any area, since the code is the AI’s brain, i.e. its core and the key to everything the AI is capable to do.\(^{127}\) A malfunction that is a consequence of a producer who is at fault, will probably be traced back to the producer and its deputies. But, what if the AI causes wrongs by default, not foreseen by the producers when they wrote the code? Does the user or supervisor have a duty to override such malfunctioning?

Even though the owner coincides with the user or supervisor primarily, it is ahead of time to rule out the owner as a possible defendant already at this stage in the analysis. In fact, the owner will be one of the indispensable actors for the future of liability.

The outsider who influences the AI in its decision-making somehow or other, is more or less self-evident in criminal liability. The outsider can be a hacker who remotes the AI or changes its code to act in a certain way, or a third party who presents an idea to the AI.\(^{128}\) One could also think of an outsider who is fooling the AI to misunderstand its environment in a way that results in wrongdoing, for instance.

\(^{124}\) Weaver (n 70) 23.

\(^{125}\) cf AI as an innocent agent, which will be further discussed below. See Ashworth (n 38) 437-38.

\(^{126}\) Grimm, Smart and Hartzog (n 16).

\(^{127}\) Bostrom (n 3) 35-37.

\(^{128}\) Secondary participation should be considered for the outsider as well.
To conclude this, there are many actors involved, with a different responsibility over the AI depending on the specific situation. It is therefore necessary to examine their responsibility in criminal law, with focus on the external element of liability, *actus reus.*

3.2 Crimes and Criminal Acts

As suggested above, there must be a human to whom the AI’s crime can be attributed, as it is generally considered that only humans can commit a criminal act\(^{129}\) The *actus reus,* the guilty act, is the act that forms the external parts of the offence, i.e. not referring to the mental state of the defendant.\(^{130}\) Specific crimes may vary among different states, but many concepts and theories of these legal systems are surprisingly similar.\(^{131}\) The elements of the *actus reus* for any criminal offence can be divided into three parts; *behaviour, circumstances* and *consequences.*\(^{132}\) The legally relevant behaviour includes both acts and omissions, which must have caused the consequences of the crime.

An act requires certain human control to be a criminal offence.\(^{133}\) Generally, the legally relevant behaviour must be a positive act *prima facie.*\(^{134}\) A positive act is easily distinguished, and in most cases, the defendant has control over the act that causes a result expectable from the defendant’s position.\(^{135}\) The *act* in criminal law is thus an act of will, an act in human control.

But, we know that part of the liability problem with AI is the lack of human control by the time the crime is committed. Increased autonomy equals decreased human control, when it comes to AI.\(^{136}\) The AI is thus performing the positive act in a legal sense, while the human, in contrast, *omits* to act or is at least passive when the AI acts. Positive acts from humans are therefore of minor importance when discussing liability for crimes an AI commits. Fortunately, there are exceptions to the positive act requirement.

3.3 Omissions

3.3.1 Omissions in General

In most legal systems, the *actus reus* requirement also includes failures to act, i.e. *omissions.*\(^{137}\) In such cases, the defendant behaves in a way that causes a *duty to act* to avoid harm, yet fails to act.\(^{138}\) The failure to act is generally considered to be an omission. Dubber and Hörnle are using the example of a car driver who falls asleep while driving, and harms a pedestrian walking on a footpath along the street.\(^{139}\) The legally relevant behaviour could be either that the driver began to drive the car even though he or she was tired and should have avoided driving, or

\(^{129}\) eg c 1, s 1 of the brottsbalken (SFS 1962:700) (Swedish Penal Code).

\(^{130}\) cf objektiver Tatbestand in German criminal law and brottsbeskrivningselnighet in Swedish criminal law. Accord Bohlander (n 40) 16; Asp, Ulväng and Jareborg (n 22) 69.

\(^{131}\) Fletcher (n 13) 3-5.

\(^{132}\) Simester and others (n 20) 71.

\(^{133}\) Asp, Ulväng and Jareborg (n 22) 71.

\(^{134}\) Simester and others (n 20) 72-73.


\(^{136}\) cf Bostrom (n 3).

\(^{137}\) Bohlander (n 40) 36; Asp, Ulväng and Jareborg (n 22) 72; Simester and others (n 20) 72.

\(^{138}\) See *R v Stone and Dobinson* [1977] QB 354.

\(^{139}\) Dubber and Hörnle (n 135) 195.
simply the failure to stop the car when feeling tired. But, would there be a duty to act for a passenger of an autonomous car, when the car heads for the footpath where pedestrians are walking?

Concerning crimes an AI commits, the possibility for a human to commit a crime by omission is fundamental for punishment. When an AI commits a crime, there are a few situations in which such a duty to act may arise for any of the identified actors.\(^{142, 143}\)

(I) A duty to act assuming that the defendant has a particular responsibility over a risk.\(^{144}\)

a) This duty corresponds with a certain role that the defendant takes in order to fulfil a civil law obligation that arises from either a contractual obligation or through law or customs imposing a certain responsibility.\(^{145}\)

(II) A duty to act because of special relationship to the the harm.\(^{146}\)

a) Duties arising through self-creation of serious risk for harm, which gives the defendant a duty to act to avoid the harm.\(^{147}\)

b) Even an unintentional act can create a legal duty to prevent harm, when the risks of a harm are imminent, i.e. when a person creates a harm by accident, but omits to prevent risks of that harm.\(^{148}\)

A variation of the latter is the continuing act doctrine, where the judge instead stretches the notion of the act to include both the actus reus and the mens rea, even though mens rea occurred after the actus reus.\(^{149}\) Obviously, a mere omission alone cannot create the actus reus without expressed liability for omissions in law, unless the crime concerns a state of affairs, which does not require an actus reus.\(^{150}\) Consequently, actus reus is not always involving a behavioural element, even though that is the general view taken in most criminal legal systems.

According to the general principle lex non cogit and impossiblia, it must have been possible for the defendant to act in conformity with the law.\(^{151}\) Consequently, only omissions where the

---

140 eg People vs Decina (1956) 2 NY 2d 133, where the defendant was indicted and charged for having killed four schoolgirls during an epileptic convulsion while driving on a public motorway, which the defendant demurred.


142 Fletcher (n 13) p 47; Bohlander (n 40) 40; Simester and others (n 20) 75.

143 Note the German division between Garantenstellung (duty of care) and Garantenpflicht (the duty to act de facto). See Bohlander (n 40) 41.

144 eg the Swedish case NJA 2005 s. 372, where a company had a civil law obligation to shovel snow and remove ice from a building’s ceiling, but neglected to do so. A representative for the company was held liable for the death of a pedestrian, who got hit by a big chunk of ice falling from the ceiling. See also Bohlander (n 40) 41; Simester and others (n 20) 77.

145 Bohlander (n 40) 43.

146 eg the Swedish Supreme Court Case NJA 2013 s. 588, where a step-parent was charged with assault for having failed to take the child to hospital, when the child had burnt his hands. The step-parent did not have that special relationship to the child, and did not have a duty to act according to the court. cf R v Stone and Dobinson (1977) QB 354.

147 Note that in Swedish and German criminal law the behaviour that creates a duty to act has to be dangerous in relation to the legal interest at risk. Legal behaviour cannot give rise to a commission by omission. A mere moral duty to act is not enough. Bohlander (n 40) 44-45; Asp, Ulväng and Jareborg (n 22) 122-23.

148 eg Miller [1983]2 AC 161, where a house caught fire after the defendant had set his mattress on fire with a cigarette, but instead of trying to extinguish the fire, the defendant simply moved to another room and continued sleeping.

149 For example, see Fagan v. Metropolitan Police Commissioner [1969] 1 QB 439, where the defendant accidently drove onto a policeman’s foot, but then refused to move his car. The refusal to move the car was a continuation of the first positive act of battery. Without the mens rea from the refusal, the first positive act would not have been an assault itself.

150 Simester and others (n 20) 85.

151 R v Bamber [1843] 5 QB 279.
defendant de facto could act, but failed, ought to be prosecuted. For AI, the result is that the defendant must have had a possibility to influence the AI and its decision-making. Hence when an AI causes the consequences of a crime, there might not be a defendant to whom the court can ascribe criminal liability.

3.3.2 A Serious Risk for Harm?
Before discussing whom of the actors that may have a duty to act and when that duty occurs, it is necessary to discuss AI as a risk of harm. John Stuart Mill’s Harm Principle is a liberal principle examining when a state is allowed to regulate a certain conduct of the individual. The harm principle is a fundamental basis for criminalisation, yet without defining harm.

According to Feinberg, harm is a ‘thwarting, setting back, or defeating of an interest’. Criminalised behaviour and their crimes, accordingly aims to protect the legal interests in question from harm. Thus harm is the consequence that criminalisation aims to prevent, and subsequently the use of the AI must cause a danger for violating the interest the crime is designed to protect, otherwise the AI cannot be considered as causing a serious risk for harm when launched. The premise for the discussion is that as long as the AI in its nature is not explicitly forbidden, the mere use or launch of an AI is not enough to count as a serious risk for harm. Nevertheless, some forms of AIs are, or will be, dangerous in nature.

Compare a potential AI in a weapon with an AI chatterbot. Although a chatterbot can cause minor risks for harm, the harm is most likely not of lethal nature. But, is the launch of the chatterbot itself a creating a serious risk for harm? The answer is probably no. The risk for the chatterbot to cause harm is minor, as its goal is not to cause harm. Weapons, by contrast, are dangerous in nature, although the harm is reactive and does not occur until someone pulls the trigger. The harm occurring from the use of a weapon is not incidental, like the harm from a chatterbot is.

For example, many legal systems require that animal owners have sufficient supervision over their beloved animals. If you walk your dog without a leash, there is a risk the dog runs away from you while ignoring your commands. Notwithstanding that risk, it is socially accepted to walk with your dog loose in most circumstances. Is it still accepted to walk a dog that has previously shown aggressive behaviour without a leash, then? The risk that the aggressive dog hurts other animals or humans, is potentially higher. While there is a risk for harm in the first example, the risk for harm is more serious with the aggressive dog in the last example.

---

153 Simester and others (n 20) 660.
155 Simester and others (n 20) 663.
156 Bohlander (n 40) 44. Asp, Ulväng and Jareborg (n 22) 109; Simester and others (n 20) 664.
157 This is consistent with the principle of legality, namely nullum crimen sine lege, ‘no crime without law’. All forms of conduct are thus allowed unless explicitly forbidden. See Jerome Hall, General Principles of Criminal Law (2nd edn first published 1960, Lawbook Exchange 2005) 27- 45; Simester and others (n 20) 22.
158 Simester and von Hirsch (n 11) 83-84.
159 cf ibid 84, about the difference between the reactive harm from a car compared with the reactive harm from a weapon.
160 Asp, Ulväng and Jareborg (n 22) 115.
161 eg the Swedish Court of Appeal case, HovR B 2127-13, where a dog owner was convicted and charged with fines for (negligently) causing bodily injuries when the owner’s three German Shepherds ran loose and bit several victims.
162 The owner has a duty to act to avoid harm. cf Asp, Ulväng and Jareborg (n 22) 164-65.
The launch of an AI that has previously shown harmful behaviour may constitute a serious risk for harm, similar to the aggressive dog. Crucial for the harm assessment is also what kind of limitations the AI has. If the producer clearly states what the AI is designed for and what limitations it has, it must be considered as causing a serious risk for harm when the AI is used outside the environment it is designed for.\textsuperscript{163} For example, to use an autonomous car in snowy weather, even though the producer clearly has stated the car is not constructed to operate in snow might be seen as causing a serious risk for harm. Accidents are more probable in an environment for which the car is not constructed. But, this risk is foreseeable from the defendant’s position. Unforeseeable risks cannot be taken into consideration when assessing.

Another aspect necessary to discuss is what effect \textit{machine learning} and similar methods causes for the harm assessment. Most AIs are designed to learn their environment through processing a large amount of data.\textsuperscript{164} With such a technology, the AI makes decisions based on what it has learned before. This makes the AI unpredictable, i.e. the AI may invent strategies or take decisions that are unforeseeable.\textsuperscript{165} There will almost certainly be situations where all humans in the chain from production to use of the AI have done everything ‘right’, yet the AI learns wrong nonetheless.\textsuperscript{166} One could argue that the unpredictability that machine learning within the AI gives rise to should result in more precautions regarding expectable events, but this is not enough to state that the launch in itself is causing a \textit{serious} risk for harm.\textsuperscript{167}

Subsequently, it is not possible to conclude \textit{de lege lata} that every launch or use of an AI, irrespective type of AI, causes a serious risk of harm.\textsuperscript{168} The conclusion is therefore that an AI as such, does not constitute a serious risk for harm although certain situations may contain such a serious risk.

3.3.3 A Duty to Act Assuming a Particular Responsibility?

As mentioned, not all forms of AI give rise to harm just because someone uses them. In absence duties of care stipulated explicitly in law for anyone in relation to the harm an AI may cause, the question is where and when, such a duty to act is created. Even though a human nowadays can have a close relationship to an AI, the relationship between a human and an AI is not a responsibility of welfare and of care, at least not yet.\textsuperscript{169}

To begin with, the responsibility the defendant has over an AI cannot be as extensive and unspecific as the duty of welfare a parent has to his or her child. It is therefore necessary to examine if it can be assumed that the defendant has a particular responsibility to avoid risks occurring from the AI in certain situations. This kind of duty to act is based on the defendant’s

\textsuperscript{163} Simester and others (n 20) 663.
\textsuperscript{164} Stone and others (n 7) 14-15.
\textsuperscript{165} Grimm, Smart and Hartzog (n 16) 13.
\textsuperscript{166} eg s 2.4, for several examples of AIs that learned wrong. cf Turing (n 67) 459 who claimed that ‘\textit{processes that are learnt do not produce a hundred per cent certainty of result; if they did they could not be unlearnt’}.
\textsuperscript{167} cf Simester and others (n 20) 665.
\textsuperscript{168} cf ibid 665.
\textsuperscript{169} cf \textit{R v Stone and Dobinson} [1977] QB 354. One could discuss if there would be such a responsibility when a \textit{general AI} is present, ie a general duty of care towards one’s own general AI, similar to the duty a parent has towards his or her child.
role and does only occur in specific situations. Duties arising because of civil law obligations and customs can give a person a duty to act when in charge of the AI.

The first obvious actor to assess liability for is the supervisor. The supervisor is in a position close to the AI when it performs its tasks, with a possibility to override the AI’s decisions. From an external perspective it appears as the supervisor has the de facto responsibility to override the AI’s wrongful decisions correspondingly. The possibility to override the AI’s decision-making is the vital element in a supervisor’s duty to act. The situation is illustrated in this example:

**Fictive example A.** A is responsible to supervise drones delivering packages directly to customers’ homes. The drones ‘know’ where to fly and act autonomously, while A supervises the drones’ routes through a screen. A has a remote from which he can take over the control over the drone. While watching a drone on its way to customer B’s home, he sees that the route is going straight over an airport’s busy runways that lays adjacent to B’s home. There is a serious risk that the drone causes an accident with both aeroplanes as well as humans involved. A has the possibility to change the drone’s route to avoid the harm, yet A neglects to intervene. The drone flies in over the runway and crashes into an aeroplane on its way to departure, causing several human injuries and a lot of damage.

A has a duty of supervision in this example, where the drones are flying autonomously but A can intervene if necessary. A had the possibility to avoid the crash, yet he omitted to act. The responsibility A has as a supervisor arises from the civil law obligation A has undertaken. The obligation formed a legal duty for A to act in that specific situation. Whilst his behaviour is a breach of his obligation to supervise the AI, it is also a criminal offence since A breaches a duty to act that is created by his voluntary undertaking to supervise the drones.

Concerning other actors, it is not possible to constantly assume a responsibility over the AI to the same extent vis-à-vis the supervisor’s responsibility. The supervisor is obligated to supervise the AI and can be forced to perform the duties in accordance with the contract, which no other actor can. In contrast with the supervisor, the user is not always a person with a de facto responsibility over the AI. Yet the user is likewise close to the AI:

**Fictive example B.** B uses a semi-autonomous car on his way to work. B drives in autopilot mode on the motorway, but is still expected to take over the control over the car whenever it is needed. Despite that, he watches a movie on his phone. The car miscalculates the speed of the car in front and does not slow down, which causes a crash.

---

170 Simester and others (n 20) 77-78.
171 cf R v Pittwood (1902) 19 TLR 37; NJA 2005 s. 372. See also Bohlander (n 40) 43.
172 cf NJA 1938 s. 69. See also Asp, Ulväng and Jareborg (n 22) 113; Dubber and Hörnle (n 135) 220-21.
173 cf the German Überwachungsgaranten and the Swedish övervakningsgarant; Bohlander (n 40) 41; Asp, Ulväng and Jareborg (n 22) 115.
175 cf Simester and others (n 20) 75-76, about parents’ duties toward their children.
In this example, B has the possibility and ability to take over the control of the car, which is a requisite for liability. The fact that he is expected to take over the control of the car when necessary, similar to a civil law obligation. When B activates the autopilot mode of the car, it is possible to assume he has similar responsibility as a supervisor to avoid serious risk for harm that the car may give rise to. Thus only a person in a supervising position can be considered to have particular responsibility and thus a duty to act to avoid harm caused by an AI, when in charge of the AI. Actors in a situation where one can assume a particular responsibility over the AI has a duty to act irrespective self-created danger or not.

3.3.4 A Duty to Act Because of a Special Relationship to the Harm?
Generally, the defendant can create a legal duty to act because of a special relationship to the harm, when behaving in a way that causes a serious risk for harm. However, the conditions for the behaviour to create such a duty to act varies among legal systems. The division mainly regards whether the creation of the serious risk of harm requires a fault or not. Countries with an Anglo-Saxon legal tradition allow unintended creations of a duty to act, in contrast with Germanic legal systems, which require a fault from the defendant to give him or her a duty to act. Asp, Ulväng and Jareborg asserts that actors has a duty to act irrespective self-created danger or not if in a situation where particular responsibility over the AI can be assumed. Likewise, the self-creation of the dangerous situation requires a wrongful conduct from the defendant.

What causes the harm is thus undeniably important. To require wrongdoing from the defendant causes issues when an AI acts autonomously. Only when the defendant de facto has done something wrong will his or her behaviour create a serious risk for harm. Some risks are permissible and socially accepted to take, like the use of an AI that may be remotely harmful after launch.

As suggested above, the user and the supervisor in charge of the AI have a particular responsibility from which it is possible to assume a duty to act. These actors will generally have a duty to act based on their role, either because of an obligation or because of customs. One could on the one hand state that these actors create a dangerous situation when disregarding their obligations. On the other hand, it is not necessary, since the obligation itself already

---

176 cf the principle of lex non cogit and impossiblia. It must be possible for the user to override the wrongful decision. See Asp, Ulväng and Jareborg (n 22) 48.
177 cf Bohlander (n 40) 44. See also Asp, Ulväng and Jareborg (n 22) 115 about the Swedish equivalent övervakningsgarant.
178 Simester and others (n 20) 78.
179 Bohlander (n 40) 44-45; Asp, Ulväng and Jareborg (n 22) 122-23.
180 See Ashworth (n 38) 113-14; Bohlander (n 40) 44-45; Asp, Ulväng and Jareborg (n 22) 122-24; Simester and others (n 20) 78-80.
181 cf Asp, Ulväng and Jareborg (n 22) 123.
183 Asp, Ulväng and Jareborg (n 22) 125. Contra the German case BGH, NJW 1999, 3132, where the court stated that the ‘hostile motivation’ of the defendant made the behaviour wrongful, not the behaviour objectively. The case indicates that the intending mind may qualify the passive behaviour as a criminal behaviour, without any prior wrongdoing.
184 cf Asp, Ulväng and Jareborg (n 22) 123.
imposes a duty to act to avoid harm. For example, to start a car, is not constituting a creation of a serious risk of harm, but to activate autopilot mode might well be.\footnote{185}

Think of the aforementioned trading bot that bought illegal drugs Darknet. The bot was instructed to buy items on Darknet, where both legal and illegal objects are sold. The owners did not de facto create a dangerous situation with a serious risk of harm by launching the bot, though there behaviour was legal.\footnote{186} Imaginably one could argue that there in fact was a risk the bot would buy something illegal at Darknet. But is that risk serious? It depends on to what extent such objects are sold on Darknet; a question with an ambiguous answer. Plausibility, probability and the defendant’s ability to approximate what the bot can do will be decisive for liability.\footnote{187} Possibly, if there was an expectable risk that the bot would acquire an illegal object, the defendants did create such a situation.

_producers_ of AIs, normally lack the capacity and opportunity to intervene in the AI’s decision-making once the AI is sold. Regarding producers, the question is whether a producer can create a serious risk of harm and consequently a duty for the producer to act to avoid the risk.\footnote{188} As long as the AI itself is legal and the production follows the specific rules for that kind of AI, the producer has not created a serious risk for harm merely by vending the AI.\footnote{189} There is undeniably a risk that some of the AIs misbehave or do not perform their tasks as expected. Yet taking that risk must be considered as acceptable, at least insofar the AI is a narrow AI developed for a specific task.\footnote{190}

Hence, can a duty to act be created anyway? Updates of the programme within the AI are the only way for the producer to influence the AI’s behaviour, once sold.\footnote{191} As illustrated in the following example, the producer may have a duty to act to avoid the serious risk of harm that the programme within the AI creates: \footnote{192}

**Fictive example C.** Producer C produces AI healthcare robots, mainly robotic surgeon assistants that can assist surgeons when replacing knee joints and hips on humans.\footnote{193} To C’s disappointment, reports have disclosed that the robotic surgeon assistants are malfunctioning because of a defective algorithm in their code. The assistants have been cutting the surgeons in their wrists, despite instructions to remain motionless during the surgery. C omits to rectify the malfunction through a new update of the programme in the AI, notwithstanding the fact that an update of the programme is the only solution.

C is informed of the malfunction in the robotic surgeon assistants. It is immaterial for C’s duty to act whether the malfunction was an unintentional production error or not.\footnote{194} As soon as C gets information about the malfunction and its effects, there is a duty to act. C created a situation with serious risk for harm by not updating the software even though he had the ability, which

\footnotesize
\begin{itemize}
  \item As long as the user can be considered to be the driver of the car, the user has an obligation to stay focused and drive carefully with respect of the ambiance, even though the car functions autonomously, in addition. cf Dubber and Hörnle (n 135) 195 and eg c 2 s 1 trafikförordningen (SFS 1998:1276) (Road Traffic Regulation).
  \item cf Simester and others (n 20) 79.
  \item cf Karnow (n 15) 180.
  \item cf Bohlander (n 40) 45.
  \item cf Ashworth (n 38) 112-13.
  \item cf Asp, Ulväng and Jareborg (n 22) 124.
  \item Stone and others (n 7) 20.
  \item cf Stone and others (n 7) 28.
  \item It is possible to argue the producer has a particular responsibility to rectify the erroneous behaviour.
  \item cf Asp, Ulväng and Jareborg (n 22) 122-23.
\end{itemize}

30
he failed to avoid. If C would have updated the programme, any future injuries on surgeons could have been avoided. It is also possible to argue that the producer has a duty to act and to avoid the serious risk for harm in this case, if he is in a position where he or she can act to avoid the risk and where other actors, from an external perspective, expects the producer to act.

Furthermore, the producer controls the education of the AI, e.g. all possible training of the AI before vending it. As a matter of taking the right precautions in the production, the producer has to identify all possible conditions under which the AI may be expected to operate, or at least make a respectable approximation. But, is the producer at fault when not identifying all possible conditions? Imagine this situation:

**Fictive example D.** Producer D produces autonomous cars. All of D’s cars are tested and trained to function in even the most extreme kinds of weather. Unfortunately, D has not recognised that some of the cars may be used in the deserts. D’s cars are not tested in any desert. There is a risk the sensors of the cars will be affected by all the dust and sand. Thus the cars risk misjudging their environment, with accidents as a result. Has D created any duty to act to avoid serious risk of harm in this case?

This situation is rather vague in comparison with the previous situation in example C. One could argue that a producer like D have a duty, if not legal, at least a moral duty, to inform the users under what conditions the cars (or AIs in general) are tested and trained. But generally, it is impossible to inform another party about an unknown circumstance. In this case, D should be excluded from liability already because of the absence of an actus reus. D might have caused a chain of events that might result in harm, by breaching a moral duty. D is very far away from the harm in this case. For the harm to occur in such a case, one can imagine a few causes that must eventuate:

(I) a user must decide to drive the car in the desert,
(II) the sensors must be affected by the sand and dust de facto, and
(III) these circumstances must produce an accident.

Consequently, it is the user who is using the car in the desert that creates the risk of harm, not D. D should be excluded from criminal liability already at the external level of the crime, on several reasons. To claim that D begins the chain of events by not informing about an unknown circumstance, i.e. how the car operates in the desert, would be inappropriate in criminal law. D’s contribution is a factual cause, but not a legal cause. D could not act in conformity with the law, which would be even more clear when assessing C’s mens rea. Testing the car or the AI in a certain environment makes the AI more predictable for the producer, indeed. But, as unpredictable the world is, as unpredictable are all situations an AI may be forced to confront.

---

195 Bohlander (n 40) 45.
196 See Bohlander (n 40) 43; Asp, Ulväng and Jareborg (n 22) 115.
197 Grimm, Smart and Hartzog (n 16) 24.
198 cf ibid 16.
199 cf Bohlander (n 40) p 44; Asp, Ulväng and Jareborg (n 22) 122.
200 cf Simester and others (n 20) 79.
201 Dubber (n 182) 102-03.
202 cf Asp, Ulväng and Jareborg (n 22) 48.
203 Grimm, Smart and Hartzog (n 16) 12.
Nevertheless, the scope for punishing a producer is thus limited. When an AI is learning how to act and behave through machine learning, for instance, it is possible the producer cannot be held liable at all, once the AI is sold. The more the AI learns, the more unpredictable it becomes. In the example above, the AI was not equipped with any self-learning method. A learning entity has the ability to learn ‘wrong’, to the same extent as it has the ability to learn ‘right’. 204 Uncertain is what impact ‘learning wrong’ has on the environment surrounding the AI. Recent research in AI and deep neural networks shows that mere adjustments in the environment, impossible for humans to recognise, can make the AI to comprehend the environment as totally different. 205 For example, a sticker on a traffic sign can make the sign totally unrecognisable for the AI, even though a human brain would apprehend the sign and its meaning despite the sticker. If a conduct like that is not foreseeable from the producer’s position, the producer has not behaved in a punishable way. 206

Conclusively, this raises other general issues. If the launch of an AI is legal and does not create any serious risk for harm, even though there is a chance that harm occurs remotely in the future, when is a potential duty to act created then? 207 The plausibility of the risks for harm will be undoubtedly important. 208 But, think of a situation where an AI invents its own behaviour, which constitutes a crime. Can we really argue that the defendant has created the dangerous situation? Does the defendant create such a situation by not controlling the AI, or maybe by not correctly approximating the AI’s behaviour? Arguably, this situation is not constituting a self-created dangerous situation at all, as long as there is no obligation to control the AI. 209 If we can conclude that the launch of the AI is not harmful, the launch cannot give rise to a special relationship to the harm the AI.

But, is it possible to argue that the legal behaviour of the defendant ceases after a while; that the defendant cannot argue that the launch was legal and then stay passive? 210 When the harm the AI gives rise to comes to the defendant’s knowledge, there should be some form of duty to act in order to interrupt the AI’s behaviour, at least. 211 If the defendant simply lets the AI commit a crime of which the defendant benefits, the defendant should also have created a dangerous situation with a duty to act to avoid harm. 212 Complex problems like this jeopardises the division between the guilty act and the guilty mind, as the guilt of the behaviour from an external perspective may be judged from an internal perspective instead. 213

---

204 cf Turing (n 67) 459 about learning machines.
206 cf Karnow (n 15) 191.
207 cf Simester and others (n 20) 666.
208 Asp, Ulväng and Jareborg (n 22) 96-97.
209 eg art 17 MiFID II, which obligates investment firms to take sufficient measures of control over algorithmic trading. In such a situation, an assumption of responsibility over the AI would put the defendant in a situation with a duty to act.
210 cf Bohlander (n 40) 45.
211 cf BGH, NJW 1999, 3132. See Bohlander (n 40) 50.
212 A different approach would be that defendant is using the AI as an innocent agent instead. Simester and others (n 20) 215.
213 cf Bohlander (n 40) 29; see also Asp, Ulväng and Jareborg (n 22) 69 and 123, about dolus subsequens and culpa subsequens.
3.4 Causation and AI

3.4.1 Causation in General

A consequence is the result of the defendant’s behaviour, i.e. the behaviour must cause the harm. This is in other words the ‘but for’-test, conditio sine qua non.\textsuperscript{214} The ‘but for’-test is very extensive and may distinguish the factual cause of the consequence, but not the legal cause.\textsuperscript{215} For omissions, the causal ‘but for’-test relates to what would have happened if the defendant had acted de facto, i.e. an act would have altered the outcome of the crime.\textsuperscript{216} Consider the following example:

\textit{Fictive example E.} E is an engineer working for a company selling drones. E has contributed with the production of the drones’ engines, making it possible for the drones to fly. One of the customer who buys a drone, later intentionally uses the drone to crash into an aeroplane.

The engineer E has caused the accident by making it able for the drones to fly but is excluded from liability already at the external level. E’s contribution is a factual cause, but not a legal cause. E’s contribution in this case, had an insignificant role in the cause of death albeit the fact that E fulfils the ‘but for’-test.\textsuperscript{217} The causation must be relevant to the result to be considered a legal cause.\textsuperscript{218} If the defendant punches the victim, and the young and healthy victim dies of an unforeseen heart attack, the punch is not the cause of death.\textsuperscript{219} The heart attack was not foreseeable for the defendant.

Sometimes the actus reus can only occur in combination with other causes, such as additional conditions or events, or acts from the victim itself or a third party. There may also be intervening causes in forms of natural events and knowing or unknowing interventions from other beings than the defendant or the victim. Intentional interventions will in general break that chain of causation, according to the principle of novus actus interveniens.\textsuperscript{220} In the future, when general AI exists, there will almost certainly be discussions if an AI can break the chain of causation as a novus actus interveniens. When an AI acts and thinks humanly,\textsuperscript{221} it would probably have the ability to deliberately intervene and thus break the chain of causation of a crime.\textsuperscript{222}

\textsuperscript{214} Bohlander (n 40) 46.
\textsuperscript{215} Bohlander (n 40) 45; Asp, Ulväng and Jareborg (n 22) 164-165; Dubber and Hörnle (n 135) 298.
\textsuperscript{216} Simester and others (n 20) 111.
\textsuperscript{217} For the meaning of an insignificant role in the chain of events, consider the English case \textit{R v White} [1910] 2 KB 124, where the defendant gave poison to the victim, who, before the poison took effect died of another cause. The defendant was not guilty of murder but attempted murder, since the poison was not the cause of death.
\textsuperscript{218} Bohlander (n 40) 47, about the German \textit{Adäquanztheorie}, the theory of adequate causation. cf Asp, Ulväng and Jareborg (n 22) 172; Simester and others (n 20) 90-91.
\textsuperscript{219} \textit{R v Carey & Ors} [2006] EWCA Crim 17, where a teenager died of a severe heart disease that the defendants’ violent behaviour against her had triggered. The violence was not considered to be the cause of the death.
\textsuperscript{220} Bohlander (n 40) 45; Simester and others (n 20) 97-101.
\textsuperscript{221} cf Norvig and Russell (n 6) 2-4.
\textsuperscript{222} cf Bohlander (n 40) 48; Bostrom (n 3) 129.
3.4.2 Causation and AI

The analysis regarding causation and AI focuses on causation from a more theoretical perspective, as causation in the end must be examined case by case. This section illustrates potential issues in causation when an AI commits a crime, and is not a comprehensive analysis of all matters regarding causation.223

As pointed out before, when the defendant has a duty to act the failure to fulfil that duty must have caused the harm in a legal sense.224 The big issue concerning causation and AI is how to examine causation and culpability depending on different sources of the harm within the AI. Concerning narrow AIs, harm may occur either because of a fault in the chain of production or by default. The question of causation can only be assessed with regard who the defendant is, what position the defendant has in relation to the AI and what the specific AI is educated for.225

In addition, are all foreseeable actions of the AI, actions that give the defendant a duty to act? Given the requirements of the actions’ relevance for the result, not all foreseeable actions will give the defendant a duty to act even though he or she has an obligation to act.226 Foreseeable yet improbable actions cannot give rise to liability for the defendant.227

When an AI requires sufficient supervising, it is normally culpable behaviour when leaving a supervising position for a longer period.228 Accordingly, a prudent person in the defendant’s position must take the serious risk for harm when leaving the position into consideration.229

The following example can illustrate a clear case of causation of harm when one can assume the defendant has a responsibility over the AI that causes harm:

Fictive example F. F is the supervisor over an AI that trades securities at the financial market. As a human on the loop, F is responsible for overseeing the algorithm’s decision-making and has the possibility to override manipulative behaviour from the algorithm. F has the de facto responsibility over the AI. F leaves his supervising position to get a cup of coffee, but he ends up talking with his colleague for three hours. The algorithm is mistakenly issuing a very large sell order in a certain security, consequently causing the trading price fall abruptly when other algorithms follows the activity of F’s algorithm. As a result, the financial market in question are manipulated.230

F’s decision to leave his position is voluntary, and he has control over his choice to walk away from the supervising position. F has an duty to act to avoid harm with the AI as its source, based on a contractual obligation to supervise the trading algorithm.231 If F had stayed in his position as expected of him, he could have overridden the wrongful decision that the AI took.232 The act

224 Bohlander (n 40) 45; Simester and others (n 20) 90-91.
225 cf Grimm, Smart and Hartzog (n 16) 16.
226 cf Karnow (n 15) 189.
227 Following Karnow’s arguments, even foreseeable actions from an AI can be too far away to be a legal cause. See ibid 180.
228 Asp, Ulväng and Jareborg (n 22) 164-65.
229 ibid.
230 Note that I refrain from analysing the crime market manipulation further than this. Even though market manipulation is criminalised both within the EU as well as in the US and other states, the particular elements of the crime might differ. The elements of the example are similar to the Flash Crash of 2010, that resulted in criminal investigations against an investor and his wrongly programmed trading algorithm. cf Bostrom (n 3) 20-21.
231 Asp, Ulväng and Jareborg (n 22) 115.
232 cf Hart and Honoré (n 223) 389; Asp, Ulväng and Jareborg (n 22) 139; Simester and others (n 20) 91.
of the AI was thus something F could foresee and expect when he left his position as a supervisor for such a long time.\textsuperscript{233} But, does it matter how the AI manipulates the markets?

It does, indeed. The defendant must be able to expect the chain of events to relevantly cause the consequences of the crime.\textsuperscript{234} Consider the situation where F leaves his AI in the exact same situation as example F. But when F returns, the AI has gone wild and acquired an AK47 illegally at Darknet. If F had stayed in his position, the AI would not have acquired the AK47. F fulfils the ‘\textit{but for}’-test.\textsuperscript{235} F’s behaviour can at least be seen as the factual cause of the crime in this case. F’s omission can be claimed to have \textit{caused} the illegal acquisition. Even though there might be a minor risk that the AI does something else than trading stocks, this was, however, not a chain of events that F should have expected from the AI.\textsuperscript{236}

Concerning self-created dangerous situations, causation is rather a question if the defendant has caused a dangerous situation at all, i.e. caused a situation with a serious risk for harm.\textsuperscript{237} If the AI itself can be considered as harmful, already the launch of the AI can put the defendant in a situation with a duty to act to avoid harm.\textsuperscript{238} Like in example F, the result of the chain of events that the defendant has created must be expected from the defendant’s position.\textsuperscript{239} Yet, the required duty to act limits the potential defendants, which makes the defendant’s omission relevant in relation to the caused consequences.\textsuperscript{240} But, when the defendant has acted in accordance with the law yet the AI acts in an unpredictable way, it will be very difficult to claim causality, irrespective whom the defendant is.

3.5 The Innocent Agent Doctrine

There are many ways a person can participate in a crime without being the offender \textit{de facto}. The defendant can encourage the AI’s crime through aid, abet, counsel or procurement, without committing the crime oneself.\textsuperscript{241} When the defendant is the \textit{de facto} offender but via an innocent agent, he or she is a secondary participant but is treated as the principal of the crime.\textsuperscript{242} For example, the grownup defendant A, uses his 10 year old brother B to commit a burglar. A tells B to break into their neighbours’ house and take the computer standing on the neighbours’ kitchen table.

An interesting thought regarding AI is the possibilities of remote control and use of AI as an innocent agent in order to perpetrate a criminal offence. Whilst the AI is unaccountable and lacks criminal capacity for its actions, it is possible to state that a defendant uses the AI as an instrument or tool to perpetrate a crime.\textsuperscript{243} Any of the identified actors that have the possibility

\textsuperscript{233} cf Asp, Ulväng and Jareborg (n 22) 164.
\textsuperscript{234} ibid.
\textsuperscript{235} cf Ashworth (n 38) 124-25.
\textsuperscript{236} cf § 2.03 Model Penal Code as cited in Dubber (n 182) 101.
\textsuperscript{237} Ashworth (n 38) 112; Bohlander (n 40) 44; Dubber (n 182) 31.
\textsuperscript{238} cf Asp, Ulväng and Jareborg (n 22) 167.
\textsuperscript{239} Asp, Ulväng and Jareborg (n 22) 164; cf Grimm, Smart and Hartzog (n 16) 12.
\textsuperscript{240} Asp, Ulväng and Jareborg (n 22) 167.
\textsuperscript{241} Bohlander (n 40) 153; Simester and others (n 20) 213.
\textsuperscript{242} Simester and others (n 20) 216.
\textsuperscript{243} Bohlander (n 40) 154-55; Simester and others (n 20) 216-17.
to influence the AI’s decision-making can be taken into consideration as principals of the crime. This theory can easily be illustrated with this example:

**Fictive example G.** G is disappointed over his tax return this year and wants to take revenge against the Tax Agency. G programmes his AI to make a hacker attack, in order to sabotage the Tax Agency’s web servers. The AI makes the attack with a devastating result that causes immense disturbances for several days. At the time for the attack, G enjoys a cold Daiquiri at a sunny beach in St Tropez.

Here G, an outsider, performs the *actus reus* through the AI. He deliberately programmes and instructs the AI from the beginning how to behave, whilst he does nothing more. The AI in this case is a simple tool to perpetrate the offence. But, what if the AI spontaneously causes harm? Is it possible for the defendant to just *accept* the AIs illegal behaviour ex post and thus the AI consequently is legally considered as a tool? Again, there is a risk for *dolus subsequens*, i.e. the intention appears after the crime, as it is impossible to withdraw or avoid harm that have already occurred.

Even though such cases are not common at the moment, they will certainly be more common in the future. Conclusively, an AI cannot be considered as an innocent agent acting on behalf of the defendant *de lege lata* if the AI itself has come up with the act, e.g. through machine learning or reinforcement learning. In such conditions, finding a duty to act for the defendant is the only possibility to hold a person liable.

3.6 Conclusion

In essence, there are several possibilities to hold some of the identified actors liable for their omissions. The analysis observes that a lawful *narrow* AI can never be considered as a serious risk of harm in its mere existence, despite when its *goal* or *effects* of the AI is fulfilling the requirements of harm. Furthermore, the analysis recognises a few situations where omission liability is possible for some of the actors, i.e. when it is possible to assume an actor has a particular responsibility over the AI, and when one of the identified actors is capable of creating a dangerous situation with a serious risk of harm. It is also possible for any of the actors to use the AI as an innocent agent, in order to perpetrate a crime either remotely or through instructions.

In conclusion, it is not possible to state that there is, or is not, a general duty to act to avoid harm that an AI causes. It is additionally impossible to state that one of the actors is always responsible for the AI’s conduct *de lege lata*. The actor’s responsibility over the AI depends substantially on the actor’s degree of control over the AI. The closer the actor is, the easier is it

---

244 Note that sabotage as a criminal offence and its elements may vary between legal systems. cf c 13, s 4 Swedish Penal Code.
245 Hallevy (n 14) 113.
246 cf Turing (n 67) 456, who discussed learning machines as ‘child machines’.
247 cf Asp, Ulväng and Jareborg (n 22) 82-83. cf c 23, s 6 Swedish Penal Code about a duty to interrupt certain crimes in progress.
248 cf Simester and others (n 20) 115.
249 Machine learning is used in self-driving cars and HFT algorithms already, but will likely be more common in the future. See Stone and others (n 7) 20; art 4.1.39-40 MiFID II.
250 cf Simester and others (n 20) 662.
251 cf Hallevy (n 14) 110.
to argue that the actor is responsible for the harm the AI causes. The responsibility will alter depending on situation, whom the actor is and what kind of AI that causes harm. The user and the supervisor are the most probable defendants, since they are close to the AI when it acts. Concerning the actors’ special relationship to the harm, spontaneous events that the AI causes cannot be considered to create a duty to act for the defendant. In such situations, the defendant must have a duty to act based on an assumption of the defendant’s responsibility instead.

The legal position when an AI uses machine learning technology and its criminal behaviour emerges spontaneously is unclear. As the analysis acknowledges, already narrow AIs who commit crimes are problematical for criminal law. The foundation of the actus reus in criminal law rests on the fundamental concept of voluntariness regarding criminal behaviour. Erroneous behaviour that the AI gives rise to without human involvement, raises the difficult question concerning whom to hold liable.

When there is no duty to act for the defendant, the next question regards if the AI can be considered to be an innocent agent acting on behalf of the defendant. If the AI is not acting on instructions or is controlled by the human, the mandatory question is whether the AI’s behaviour was expectable or not from the defendant’s position. If the harm was expectable, it is possible to argue that the defendant’s omission was voluntary and that the AI thus acts as an agent. Otherwise, the claim would be contrary to the act requirement. The defendant’s limited control over the AI will increase the importance of probability and approximation instead. In many cases, the defendant will not be the cause of the harm, and should then not be liable in accordance with any theory in criminal law.

Consequently, not all crimes an AI commits have an offender to prosecute. The law as it is, is thus very vague. The de lege ferenda analysis will therefore examine two different solutions to this liability problem.

---

252 cf Karnow (n 15) 180-81.
253 David Ormerod, Smith and Hogan’s Criminal Law (13th edn, OUP 2011) 55.
255 cf Simester and others (n 20) 667 about remote harms.
256 cf Hart, Punishment and Responsibility (n 254) 104-05, about involuntary acts that are not ‘subordinated the [defendant’s] conscious plans of action’.
257 cf Karnow (n 15) 180.
III. AI and Its Legal Future

‘The greater the freedom of a machine, the more it will need moral standards.’

Rosalind Picard

Chapter 4. Two Possible Solutions?

4.1 General Introduction

The previous analysis of AI in criminal law _de lege lata_ concludes that AI will undoubtedly cause problems when attempting to find a liable actor for crimes an AI commits. Imagine a scenario where there are a lot of actors involved, where all of them have contributed to the harm alongside the AI. Transparency within the AI and division of responsibilities between the actors will be decisive for holding the right defendant liable. The former is rather a question for the producers of AI than for the law to decide. Instead, we focus on how the law _should_ be construed to solve the problem of ascribing an AI’s criminal offence to a responsible defendant.

Before analysing suitable solutions for the liability problem, it is important to recall the causes of these problems. First of all, all forms of narrow AI are not considered to be harmful, even though some types of AIs will be harmful in nature. A launch of an AI is thus not always a risk relevant enough to create a duty to act for the defendant. As a result the defendant has the role as a simple bystander, who can watch the chain of events without risking liability for the crime. Furthermore, the unpredictability feature in the AI reduces the relevant behaviour for which someone else than the AI itself can be responsible. It is unclear when and where the defendant is expected to act and not. Such an order is not appropriate for criminal law, as it risks violating the rule of law.

There are those who may find the existing theories of _actus reus_ and criminal liability as sufficient to satisfy even these afore exemplified situations. The problem with AI is that there will almost certainly be crimes committed without a (legal) offender, especially when AIs are more autonomous than today. Will we accept that a crime has an agent but not a principal to blame? This is an analysis of two possible solutions of the liability problem.

---

259 cf Simester and others (n 20) 187.
260 cf Ormerod (n 253) 70.
261 Karmow (n 15) 189.
262 cf Simester and others (n 20) 21-22. The principle of the rule of law comprises a proscription against vague and unclear statutory regulations.
263 cf Ryan Calo (n 105) 544.
4.2 A Supervisory Duty?  

4.2.1 What Kind of Legal Measure?  

Scientists and entrepreneurs working with AI claim AI can be the most devastating invention ever created, as well as an invention that benefits humankind to a great extent. If we envisage a situation where a general AI might be dangerous, but at the same time creates the utmost utility for mankind, should such technology be prohibited? From a utilitarian perspective the answer is definitely no. If the production and development of a general AI generates the greatest happiness and maximum utility for humankind, such development should not be prohibited. The utility outweighs the potential risks for harm. A straightforward prohibition of dangerous AI or a similar legislative measure is not a suitable solution for the liability problem.  

Similar to human error there will be AI error, by design or by default. In accordance with John Stuart Mill’s Harm Principle the state is justified to regulate conduct of individuals that risk causing harm to others. Legal moralists argue that the state may use criminal law to prohibit immoral wrongs, even in absence of harm. Morals are sometimes said to be a non-legal factor that affects the law. If an AI can benefit mankind, it is important not to limit the beneficial possibilities that lay ahead. Still there are risks for harm to humans. In fact, the different identified actors may in many situations be considered to behave morally wrong when not intervening in an AI’s commission of a crime. Yet when the behaviour is neither harmful prima facie or remotely, a moral wrong should not be invoked as a satisfying justification for the state to intervene in the individual’s autonomy.  

The harm from the actors’ behaviour does not occur immediately, but it may occur in the future when the AI acts. The launch of a narrow AI will in most cases not be harmful in itself, but mediating actions from the AI may cause harm in the future. The actors’ behaviour is legal and do not constitute any wrongdoing without the AI’s contribution. It should also be emphasised that this is not a suggestion to hold any of the actor liable for the launch of an AI. But, the launch or use of an AI somewhat presupposes a duty of control and supervision over the AI and its actions. Taking a cynical view of the law, this presupposition would be enough for the

---

264 The structure of this section follows the Standard Harms Analysis in part, as proclaimed by Simester and von Hirsch, to examine whether certain behaviour fulfils Mill’s Harm principle. See Simester and von Hirsch (n 11) 54-55.
266 cf John Stuart Mill, Utilitarianism (2nd edn, Parker, Son, and Bourn 1863) 16.
267 ibid 22-23.
269 Feinberg (n 154) 27.
270 Penner and Melissaris (n 35) 125.
271 See Bohlander (n 40) 44; Simester and others (n 20) 78-79.
272 cf Simester and von Hirsch, (n 10) 35.
273 See Simester and others (n 20) 672. See also Peter Cane, Responsibility in Law and Morality (Hart Publishing 2002) 12, ‘morality is not subject to criticism in legal terms’.
274 Simester and von Hirsch (n 11) 57-59.
275 ibid 48.
276 ibid 19.
court to impose criminal liability.\textsuperscript{277} But, a behavioural analysis of what the court would do in fact gives us a pretentious answer.

When launching, using or instructing an AI to do something, to just ‘fire and forget’ is precarious per se. Although the risk for harmful behaviour from the AI is minor, there \textit{is} a risk, which may not be \textit{serious} at the moment.\textsuperscript{278} That risk intensifies coherently with the AI’s rising intelligence.\textsuperscript{279} Another important notion is why someone would use an AI in the first place – to act in one’s interest, like an agent that acts for the human principal.\textsuperscript{280}

Before our knowledge about AI and its future gets better, a legal responsibility to supervise an AI has its benefits. The unpredictability of an AI’s actions should result in precautions. If the supervisory duty could involve control and supervision over the AI, that would be enough to create a duty to act to avoid harm for the defendant predominantly based on the responsibility over the AI’s actions.\textsuperscript{281} Simester and von Hirsch states that the question to be asked is ‘\textit{what if this act were always permitted?’} when considering measures for reactive harm such as the harm that an AI may give rise to.\textsuperscript{282} What if it were always permitted to use an AI without any supervision?

At issue is whether a supervisory duty would clarify the responsibility over the AI and avoid harmful consequences of its behaviour. Such a duty would thus be a form of quasi criminalisation concerning the \textit{secondary, reactive harm} that the actors’ behaviour may give rise to.\textsuperscript{283} Due to the problems with unpredictability and foreseeability, criminal law and criminalisation is only partially solving the liability problem. Instead, the supervisory duty should be a civil law duty, analogous with the strict liability that is sometimes invoked on owner of animals or the duty that enters into force next year on the EU’s financial markets.\textsuperscript{284} The reason is simply that AIs in general are not harmful enough to justify the state to intervene, as concluded above. Yet the supervisory duty will have \textit{effect} in criminal law, which is what the analysis is seeking.\textsuperscript{285}

Article 17 MiFID II imposes a duty of control and supervision for investment firms engaged in algorithmic trading, and additionally obligates the firms to prevent erroneous orders and behaviour that may violate MAR or trading platform’s rules. Even though article 17 MiFID II does not impose criminal sanctions, it creates a \textit{duty to act} for the investment firms and their principals, to avoid serious risks for harm that the trading algorithms may cause.\textsuperscript{286} A similar duty for all kinds of AIs would not be a de facto criminalisation of the behaviour, but in its effect it extends the group of potential defendants.

\textsuperscript{277} cf Penner and Melissaris (n 35) 125.
\textsuperscript{278} cf the serious risk that is required for the defendant to create a dangerous situation. The risk for harm must be imminent and the harm a plausible consequence of the behaviour. Bohlander (n 40) 44-45; Asp, Ulväng and Jareborg (n 22) 174; Simester and others (n 20) 78-79.
\textsuperscript{279} Karmow (n 15) 193.
\textsuperscript{280} Bostrom (n 3) 157.
\textsuperscript{281} cf Bohlander (n 40) 41; Asp, Ulväng and Jareborg (n 22) 115-16; Dubber (n 182) 30; Simester and others (n 20) 77.
\textsuperscript{282} cf Simester and von Hirsch (n 11) 48.
\textsuperscript{283} ibid 48-49.
\textsuperscript{284} cf above n 109.
\textsuperscript{285} cf Simester and von Hirsch (n 11) 47.
\textsuperscript{286} cf Asp, Ulväng and Jareborg (n 22) 115-16 about \textit{övervakningsgarant} (duty of supervision).
Who should be responsible for the supervision then? It is reasonable to consider the user or the supervisor as potential targets for the proposed supervisory duty. Imaginably, one could argue that creating a supervisory duty for the user would make the causal chain from any wrongdoing to the harm shorter, since the user is closer to the AI than all other actors. Yet there are arguments against giving the user such responsibility. Generally, the user and the supervisor are already in positions with a duty to act, either because their role imposes a particular responsibility over the AI or because of a dangerous situation that the actors may create.

The producer’s relationship with the AI is not causally relevant enough when the AI commits a crime, in general. Asp, Ulväng and Jareborg are using the example of a car manufacturer, who is not criminally responsible for all accidents involving the cars just because the manufacturer produced the cars and accidents involving the cars are expected. A similar conclusion must be drawn regarding producers of AIs. In addition, it is possible to take other measures to force producers to take their responsibility in the chain of production of AI.

The supervisory duty must be clear, to minimise potential misconceptions of responsibility over the AI. It is therefore of importance to refer to a legal notion of which any potential defendant is fairly familiar; ownership. Hence, the possible legal solution is to impose a supervisory duty on the owner to either supervise the AI oneself, or being responsible to contract a supervisor one way or another. That would create a role based duty to act for the defendant, independent of the defendant’s own creation of a dangerous situation. With a clear duty for the owner, this model is accordingly favourable with regard to the rule of law.

4.2.2 Possible Effects

A civil law supervisory duty for the owner would have effect in criminal law as well. The civil liability will create a basis for the assumption of the owner’s responsibility over the AI. Albeit the proposed supervisory duty over the AI, the unpredictability feature remains problematic. The rule of law requires the law in itself to be predictable, besides a predictable interpretation and adjudication of the law.

---

287 ie the law would not risk that the defendant’s causal relationship to the harm is too distant. See Asp, Ulväng and Jareborg (n 22) 134-35; Simester and others (n 20) 92-93.
288 cf above 4.4.
289 cf Asp, Ulväng and Jareborg (n 22) 134-36.
291 See generally Fletcher (n 13) 67-69. Naturally, it is possible to hold a producer liable for mistake of his or her own, as illustrated in examples C and D.
292 See Bohlander (n 40) 43-44; Asp, Ulväng and Jareborg (n 22) 115-17; Dubber (n 182) 32.
293 cf the indicating effect civil law duties and customs can have regarding the notion of wrongdoing in criminal law. Asp, Ulväng and Jareborg (n 22) 115.
294 cf Peczenik (n 24) 684, about the qualification theory of norms.
295 cf Bohlander (n 40) 41; Simester and others (n 20) 77.
296 Peczenik (n 24) 63.
the stated solution. When adjudicating such cases, the court must decide what the defendant should have foreseen from his position. Harms that are not a plausible consequence, foreseeable from the defendant’s position, will and should not be punished.\footnote{cf Karnow (n 15) 189.} But, the decision concerning foreseeable harm risks not to give the supervisory duty any effect at all. A supervisor cannot be expected to override decisions he or she could not predict, as he or she would not have the \textit{ability} to override the decision.\footnote{cf Bostrom (n 3) 23. AI researchers believe general AI will reach human level intelligence approximately around year 2075.} If unexpected harm occurs, that would probably qualify the harm a foreseeable harm in the future, which would impose a duty on the supervisory to break the chain of events next time.

This suggested solution should be criticised for being unsustainable in a longer perspective. The aim of most AI research is to develop general \textit{AI}, a super-intelligent artificial being that can act like us humans, but better and faster.\footnote{cf Bostrom (n 3) 157.} This will almost certainly not impede the development of beneficial AI. But, it will probably thwart the development of general AI.\footnote{cf John Chipman Gray, \textit{The Nature and the Sources of Law} (Cambridge University Press 1909) 27-28.} As long as the AI has a human supervisor it \textit{has to be possible} to override the AIs decisions or control it by some means. Consequently, the AI will not be able to function totally autonomous from humans. Thus producers cannot manufacture AIs that once launched leave humans out of the loop. At least, there has to be a kill switch of some sort, to disconnect the AI from its decision-making. Conclusively, creating a supervisory duty to act would not rectify all causes of the problem with holding a defendant liable, but it would guide the actors involved in the AI’s decision-making and clarify who the liable actor is.

4.3 AI Criminal Liability?

From the law we can derive our legal rights and duties. To follow the law is thus to perform duties and to be granted rights.\footnote{cf John Chipman Gray, \textit{The Nature and the Sources of Law} (Cambridge University Press 1909) 27-28.} Legal personhood for AI is consequently a question whether AIs should have rights and duties in accordance with the law. The solution may be futuristic and progressive, yet a sufficient analysis should involve a brief discussion about legal personhood for AIs, as it would make the AIs accountable for their own actions.\footnote{cf Hildebrandt (n 108), who designates her article to criminal liability for smart environments.}

AI criminal liability requires legal personhood for the AIs, and would be similar to corporate criminal liability that some legal systems are recognising. Corporate criminal liability is deemed to be a fiction; a construed form of liability where the corporation is \textit{attributed} with its employee’s acts.\footnote{eg generally Hildebrandt (n 108); Weaver (n 70); Calo (n 105).} In contrast with corporations, AIs would be accountable for their own behaviour, not attributed with anyone else’s.\footnote{eg generally Hildebrandt (n 108); Weaver (n 70); Calo (n 105).} Even though it seems to be a simple solution not violating the rule of law requirements, it requires a more comprehensive assessment than this.

Entities that are already granted legal personhood are still in human control, within the daily decision-making or at shareholder level, for instance. The shareholders’ economic liability is
normally limited, yet the responsible humans are affected of the entities wrongdoing in other means. Humans are nearly absent in autonomous AIs, which subsequently indicates that there is no one to blame within the AI when an AI causes harm. In other words, there are no humans to deter and prevent from wrongdoing.

To make this solution function appropriately, we need to figure out how to deter an AI from commit a crime and how to punish the AI after the crime when physical imprisonment is not possible. An alteration of the fundamentals of criminal law is thus necessary. Can we compare an AI’s behaviour in a specific situation with the behaviour of a prudent person in the exact same situation when the AI lacks common sense?

Hart stresses that holding a person responsible for his or her actions presupposes the person possesses particular capacities, such as understanding what a legal rule imply, to understand and to reason, among others. If persons really should be ‘agents capable of law, happiness and misery’ as Locke stated, AI cannot be considered a person yet. Existing AIs do not possess all these capacities, although general AI will do that in the future. Probably, enforcing AI criminal liability will require a system where fulfilling the actus reus on an external level is enough for criminal liability. Is it possible to state that an entity without cognition can perform an act of will? The ‘unconscious’ state of the AI should be analogous with involuntary acts that excludes criminal liability. As long as AIs do not think and act humanly or possess the capacities that imposing responsibility for acts in general requires, direct criminal liability for AIs are not a solution good enough to solve the liability problem of AI.

4.4 Conclusion

The suggested solutions have their shortcomings and flaws, like many other legal instruments. The civil law supervisory duty would direct humans who launches or uses an AI, to take sufficient measures to control and prevent the AI from wrongdoing. Yet it does not counteract the issues with the unpredictability of an AI’s actions. A probable consequence of that is that the defendant’s omission will still not be considered as the cause of the event, though the defendant does not have the ability to avoid an event that is unknown. But, similar to the supervisory duty that article 17 MiFID II imposes on investment firms engaging in high frequency trading, a supervisory duty concerning AI’s would prevent ‘fire and forget’ scenarios.

Unfortunately, the supervisory duty would do a minor difference in criminal law, but clarifies the division of responsibility marinally. However, it would establish the causal chain between the defendant’s omission and the AI’s act, as a duty of supervision can be invoked irrespective

---

305 cf c 1 s 3 aktiebolagslagen (SFS 2005:551) (Swedish Companies Act); s 3 Companies Act 2006 (c 46).
307 Stone and others (n 7) 46-47.
308 cf Karnow (n 15) 203.
309 Hart, Punishment and Responsibility (n 254) 227.
310 Locke (n 60) 234.
311 cf section 2.4.
312 Hildebrandt (n 108) 525.
313 See Hart, Punishment and Responsibility (n 254) 98. cf Hildebrandt (n 108) 523, about a form of morality that is mindless.
314 Norvig and Russell (n 6) 3.
315 Ashworth (n 38) 133.
of the defendant’s own causation. It is thus possible to assume a duty to act based on the civil law supervisory duty the defendant has, even in those cases the AI itself has caused the foreseeable harm. As it targets humans and not the AI itself, the AI’s degree of intelligence is of minor importance for the duty to act.\textsuperscript{316} The degree of intelligence will however be a significant factor when assessing the potential risks of the AI at issue.

AI criminal liability may be the solution of the future, as many scholars suggest.\textsuperscript{317} But, if we are not to change the entirety of our criminal legal systems, we should not resort to AI criminal liability before more general AIs arrive. An AI that for example passes the Turing Test, will have the capacity it takes to be held liable for its own actions.\textsuperscript{318} Another disadvantage of AI criminal liability is that it lacks the deterrent and preventive functions, among other functions of criminal law, that directs the human conduct behind.\textsuperscript{319}

The conclusion of this analysis is that the most efficient countermeasure against the liability problem is to impose a supervisory duty in civil law with a duty to oversee an AI’s actions in order to prevent the AI from wrongdoing. Yet the solution is not enough to counteract the liability problem totally, as liability still depends on what is foreseeable and expectable from the defendant’s position. Only those who \textit{should} have expected that specific harm will be held liable for the AI’s actions.\textsuperscript{320}

\footnotesize
316 Bohlander (n 40) 41; Asp, Ulväng & Jareborg (n 22) 115; Dubber (n 182) 30; Simester and others (n 20) 77.
317 eg Karnow (n 15); Hildebrandt (n 108).
318 cf Hart and Honoré (n 223) 227-28; Hildebrandt (n 108) 531. See section 2.3 about the Turing Test.
319 cf Simester and others (n 20) 7.
320 cf Karnow (n 15) 188.
Chapter 5. Concluding Remarks

In this thesis, we have taken an evanescent glance into how the future will challenge the basic concepts of criminal law. As concluded in chapter two:

“Artificial intelligence is that activity devoted to make machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment”.321

Consequently, there are many different kinds of AIs but they all share a few common features; unaccountability, unpredictability and autonomy. These characteristics are also the primary reasons behind the liability problem. Unpredictability together with autonomy limit the potential defendants to humans who have a duty to act, and as a result liability can in some cases not be established for actors who should be liable. The primary cause of that issue is the lack of relevant causation when the AI acts autonomously without involving any human. The rule of law restrains the possible criminal behaviour for humans to controlled acts and omissions, which are voluntary.322 An act that is not willed, are not voluntary.323 If the AI acts autonomously, there is no established causal chain between the defendant and the AI, unless the launch or use of AI alone is harmful.

This thesis discusses the guilty act of the crime; the *actus reus* element, and illustrates the difficulties when the court struggles to find a liable defendant for crimes an AI commits. Not surprisingly, these legal challenges will increase with the decreased control that the defendant has. The *mens rea* element is still left to analyse, but if finding a liable actor is challenging already at the external level of the crime, it will be even more demanding to prove the required *mens rea* of the defendant.324

The seemingly best of the two analysed solutions, to curtail these difficulties, is to impose a supervisory duty on the owner, i.e. create a civil law duty to supervise the AI and to intervene to avoid risk for harm, even when the action was not foreseeable for the supervisor.325 With such a solution, the law would not hinder the creation of AI beneficial to mankind, but reduces possible hazardous risk taking from actors that know their contribution or passiveness will almost certainly not qualify for criminal liability.326 The main issue with that solution is that the unpredictability of the AI’s actions remains, and thus only foreseeable actions from the AI will be punishable even in the future.

AI criminal liability would solve the liability problem, since the AI itself is then always liable for its own actions. But before that, the AI must possess certain capacities, which in the current

---

321 Nilsson, (n 2) xiii.
322 Simester and others (n 20) 22 and 73.
324 cf Stone and others (n 7) 47.
325 cf Bohlander (n 40) 43-44; Asp, Ulväng and Jareborg (n 22) 115; Simester and others (n 20) 76.
326 cf Bostrom (n 3) 184-95.
state of art are still absent.\textsuperscript{327} In the future, there is a good chance an AI can fulfil the requirements for criminal liability.\textsuperscript{328} Until then, the liability problem persists.

The juristic profession is primarily engaged in advising others, by simply telling what is right and what is wrong according to the law.\textsuperscript{329} Despite that, this thesis concludes that the liability problem concerning crimes an AI commits may be the biggest challenge criminal law has ever faced, and there might not be an easy solution to propose. The solutions provided for in this thesis are imperfect and partially unsatisfactory options, but are ‘the best’ we can do about the liability problem for the moment.

Criminal law targets humans, and if we want to maintain the retributive and deterrent functions of punishment in criminal law, we need to direct the law at humans with the possibility of moral accountability, i.e. the humans behind and not the AI itself. The supervisory duty is de facto directed at the humans behind, yet is not the perfect way to solve the liability problem. Certainly, it may be to ‘bark up the wrong tree’ when trying to solve the liability problem with AI, in criminal law. At the moment, the AI and its principals levitate in an empty space without a clear notion of what is right and what is wrong in criminal law.

***

\textsuperscript{327} cf Bostrom (n 3) 24.
\textsuperscript{328} cf Hallevy (n 14) 229.
\textsuperscript{329} ‘People want to know under what circumstances and how far they will run the risk of coming against what is so much stronger than themselves, and hence it becomes a business to find out when this danger is to be feared. The object of our study, then, is prediction, the prediction of the incidence of the public force through the instrumentality of the courts.’ Oliver Wendell Holmes, Jr ‘The Path of the Law’, (1897)10 Harvard Law Review 457, s 1.
References

Literature

Articles and Books


Mill J S, Utilitarianism, (2nd edn, Parker, Son, and Bourn 1863).


Ormerod D, Smith and Hogan’s Criminal Law (13th edn, OUP 2011).


Peczenik A, Vad är rätt? (Fritzes Förlag, 1995).


The Oxford Companion to the Mind (2nd edn, OUP 2006).

Turing A M, 'Computing Machinery and Intelligence’ (1950) 59 MIND 433.


Newspaper Articles


Legislation

EU


Germany

Sweden
(SFS 2007:1150) lag om tillsyn över hundar och katter
(SFS 2005:551) aktiebolagslagen (Companies Act);
(SFS 2010:659) patientsäkerhetslagen (Patient Safety Act)
(SFS 1962:700) brottsbalken (Swedish Penal Code)
(SFS 1998:1276) trafikförordning (Road Traffic Regulation)

UK
Agriculture (Artificial Insemination) Act 1946 (c 29).
Animals Act 1971 (c 22)
Companies Act 2006 (c 46).

Cases
Germany
BGH, NJW 1999, 3132

Norway
Rt 2012 s. 686.

Sweden
NJA 2013 s. 588
NJA 2005 s. 372
NJA 1983 s. 320

United Kingdom
Fagan v Metropolitan Police Commissioner [1969] 1 QB 439
R v Bamber [1843] 5 QB 279
R v Carey & Ors [2006] EWCA Crim 17

NJA 2005 s. 372
R v Evans [2009] EWCA Crim 650
R v Miller [1983] 2 AC 161
R v Pittwood [1902] 19 TLR 37
R v Stone and Dobinson [1977] 1 QB 354
R v White [1910] 2 KB 124.

USA
People vs Decina (1956) 2 NY 2d 133.

Other Cited Materials
Websites and Blogs


Correspondence