Hochschule Ruhr West

Business Administration: Energy and Water Economics (M.Sc.)

Campus Mülheim an der Ruhr

HRW
HOCHSCHULE RUHR WEST
UNIVERSITY OF APPLIED SCIENCES

Master thesis

The ethical implications of AI-based mass surveillance tools

Submitted to Prof. Dr. Christian Weiß and Martin-Sebastian Abel

Marvin Wiesenthal
10008386
Ulmenallee 32a
45478 Mülheim an der Ruhr
Business Administration: Energy and Water Economics (M.Sc.)
Semester: 4

Contents

1 Introduction

Artificial intelligence (AI) is one of the most auspicious yet controversial technologies with virtually unlimited potential to solve almost all of the existential problems humanity is facing today.[1] Huge resources are poured into the development, testing and application of AI that is supposed to be utilized in almost all areas of everyday life.[2] It may be used to combat genetically inherited diseases, to revolutionize the economy, to bring prosperity and equality to everyone and to counter the effects of climate change.[3] With AI as the enabling technology humanity may experience a better future. Today, AI capabilities can already drastically improve analytic processing tasks and algorithmic systems and have beaten humans in games such as chess.[4] Yet, AI and all of its applications bring about a myriad of ethical challenges. Revolutionary weapon systems that achieve autonomy via AI and genome-editing powered by AI are just some specific examples.[5] An omnipotent AI will be either the greatest or the vilest thing that has happened to humanity in its brief existence.[6] However, even today more and more computational devices are connected to each other, spurring a huge increase in global data streams that can be used to further train and enhance AI systems.

The prowess of AI for executing analytic tasks paves the way for the use of AI in more and more applications. One of these applications, that shows great promise, is the use of AI in surveillance applications.[7] AI surveillance applications are proliferating at a fast rate, with a number of applications already being in use today.[8] These applications are aimed at accomplishing a number of policy objectives, some are in accordance with basic human laws, some are definitely not and some

[1] Cf. Hawking (2018). P. 183ff

[2] Cf. Hawking (2018). P. 183ff.

[3] Cf. Hawking (2018). P. 183ff.

[4] Cf. Burton (2015). P. 1ff.

[5] Cf. Hawking (2018). P. 183ff.

[6] Cf. Hawking (2018). P. 183ff.

[7] Cf. Feldstein (2019). P. 1.

[8] Cf. Feldstein (2019). P. 1.

belong in the nebulous area in between lawful and unlawful.[9] But what are lawful and unlawful uses of AI surveillance systems and what are their ethical implications?

This thesis will examine the ethical implications of AI based mass surveillance systems and try to **answer the first central question, if it is possible to use AI based mass surveillance applications in an ethical way.** Furthermore, the thesis will attempt to answer **the second central question and find out how the ethical use of AI based mass surveillance systems, if this ethical use is possible, materialize.** Governmental agencies will be in the focus of this discussion, as their use of the technology may have bigger ethical challenges. Yet private companies will play a part as well. In an attempt to accomplish these two aims, the thesis will inspect the basics of ethics and possible ethical theories that can be utilized to answer the questions. Normative ethics will be studied first with a focus on consequentialism and utilitarianism. To gain a deeper understanding of utilitarianism, act and rule utilitarianism will be compared. Afterwards, deontological theories will be the focus of the discussion with a concentration on deontological pluralism. Next, the mentioned theories will be evaluated, discussing advantages and weak spots of the theories, to assess which theory may serve as the ethical framework of this thesis and the subsequent answer to the two main questions.

The next step will be the establishment of the AI framework. This contains the definition of AI and a distinction of terms that are commonly used in the its environment such as automation and autonomy. The importance of data for AI will be discussed. Afterwards, the technological basis of AI will be outlined, discussing key concepts such as machine learning and deep learning. Additionally, it will be examined how an AI learns. The possible uses of AI in general will be outlined in a brief fashion, blazing the trail to discussing the moral challenges of AI. Afterwards, the current pace of AI development will be studied.

In the chapter that follows, the use of AI in surveillance technology is going to be highlighted. The possible ways of how AI can be used for surveillance purposes are reviewed here, discussing facial

[9] Cf. Feldstein (2019). P. 1.

and behavioral recognition systems, smart cities, smart policing, communications/data driven surveillance and their enabling technologies. Then, the global proliferation of AI surveillance systems is going to be outlined.

Subsequently, the accordance of AI surveillance with basic human laws and rights, such as the right to privacy, will be checked to find out if the law and the international framework of human rights allow for AI surveillance or at least have restrictions that would greenlight the use of AI surveillance technology. All the aspects of the thesis, especially including the selected ethical framework, will be combined in this last section in order to enable the adaptation of a framework that allows to find out, if AI surveillance systems can be ethically permissible while also creating insights how this ethical AI surveillance system must be engineered. To finish, the thesis will end with a conclusion.

2 The ethical framework of the thesis

The following chapter is aims to lay out the ethical groundwork for this thesis. This ethical framework will be used to assess the two guiding questions of this thesis. Different ethical concepts, consequentialism and deontology and their expressions, will be examined. These ethical concepts are discussed because they are used to assess the morality of actions, policies and more based on different premises. Afterwards, the concepts will be critically reflected, paving the way to selecting the guiding concept of this thesis.

2.1 The terminology of the thesis

To understand the basic concepts and ideas of ethics, the environment of the aforementioned ethical concepts and their terminology must be understood. Consequentialism and deontology are both concepts belonging to the branch of normative ethics.[10] Normative ethics deal with theories that try to determine what must be regarded as right or obligatory, in order to find out the what must be done.[11] To find the definition of rightness, it is mandatory to first sharpen the understanding of moral standing, moral rights and intrinsic value.[12] Furthermore, key concepts such as obligation, rightness and supererogatory actions must be defined.[13]

Moral standing describes entities that should matter in the decision-making process and the respective assessment of the morality. [14] Therefore, entities that have moral standing will henceforth be referred to as moral entities. For example, a person possesses moral standing if the treatment of the person makes a moral difference and therefore changes our assessment whether something is morally acceptable or not.[15] Moral standing typically attaches to all living creatures, future and present.[16] Attfield further describes that this could lead to the conclusion, that every living creature has moral rights. According to him, this consideration bears weight, because moral rights create a

[10] Cf. Attfield (2012). P. 71ff.

[11] Cf. Attfield (2012). P. 71ff.

[12] Cf. Attfield (2012). P. 71ff.

[13] Cf. Attfield (2012). P. 71ff.

[14] Cf. Attfield (2012). P. 71ff.

[15] Cf. Andre and Velasquez (1991). P. 1f.

[16] Cf. Attfield (2012). P. 72ff.

strong case against harming the holder of the moral rights and therefore, avoiding harm to the holder of the moral rights becomes obligatory. However, the range of morality is wider than the reality of rights and not all obligations are in a causal relationship with moral rights. [17]

Agency describes the ability of a being to act.[18] Additionally, it is vital to also understand the moral agent. A moral agent is a moral entity that possesses the ability to differentiate right from wrong and therefore may be held accountable for its actions.[19] Explicitly with assigning people with moral agency we can hold them accountable for the harm that they may cause to other moral entities.[20] Having discussed the moral entity and the moral agent, the next important concept of ethics that must be understood is obligation.

To respect and uphold rights we must take a step back and define what a right is. Several important historical documents use this term such as the American Declaration of Independence from 1776 and the United Nations Universal Declaration of Human Rights from 1948.[21] A right may be defined as a justifiable claim on others.[22] This justification stems from a standard that is generally acknowledged and accepted by society.[23] If only the claimant sees a justification for his right and society does not, no moral standard to base a right on exists.[24] The justification can be as tangible as codified law but is not limited to it.[25]

In order to understand obligation, understanding intrinsic value is key. Intrinsic value has always been a central concept of ethics.[26] It is traditionally defined as the value that something has in itself.[27] The question of what has intrinsic value and what it is exactly has been discussed since

[17] Cf. Attfield (2012). P. 72ff.

[18] Cf. Schlosser (2015) P. 1.

[19] Cf. University of Texas (2022). P. 1f.

[20] Cf. University of Texas (2022). P. 1f.

[21] Cf. Velasquez et al. (2014). P. 1ff.

[22] Cf. Velasquez et al. (2014). P. 1ff.

[23] Cf. Velasquez et al. (2014) P. 1ff.

[24] Cf. Velasquez et al. (2014). P. 1ff.

[25] Cf. Velasquez et al. (2014). P. 1ff.

[26] Cf. Zimmerman and Bradley (2002). P. 1ff.

[27] Cf. Zimmerman and Bradley (2002). P. 1ff.

Plato (428-347 before the commo era).[28] It may be argued that the welfare of moral entities is intrinsically valuable, creating the necessity to protect the welfare of moral entities.[29] According to this, reasons for actions to ensure the welfare of a moral entity are created by the presence of value.[30] Yet it must be stated that value may have different strengths or degrees and therefore a relationship between a high degree of value and strong reasons can be observed. [31] Obligation describes the existence of overwriting reasons for an action, with the only exception being a conflict between two obligations.[32] This simply means that an obligations is a moral requirement for acting in a way that is morally right.[33] Even if two obligations are in contradiction to each other there will be strong evidence to comply with one of them.[34] However, it would be wrong to conclude that obligations protect the well-being of every holder of moral standing, as obligations will apply more easily when a high degree of value exists.[35] If the interests of different holders of moral standing contradict each other, then the value of the welfare of the weaker moral entity is often outweighed by the value of the welfare of the stronger moral entity.[36] An example: a governmental organization is using AI to monitor the constituents of its country in order to enhance the national security and public order. Due to this, the citizens of the country live safely which is maximizing its welfare of the whole population. In the wake of this surveillance, repression is exercised which leads to a decrease in welfare of an individual. Here we can see a conflict between the task of the governmental agency to keep its population safe and an individuals need for privacy. Another example can be animals that eat plants, as the welfare of the plants is in conflict with the animals need to eat.[37] Keeping all this in mind, it is important to state that not everything that we undertake is an

[28] Cf. Zimmerman and Bradley (2002). P. 1ff.

[29] Cf. Coakley (2017a). P. 17ff.

[30] Cf. Coakley (2017a). P. 17ff.

[31] Cf. Coakley (2017a). P. 17ff.

[32] Cf. Attfield (2012). P. 72ff.

[33] Cf. Attfield (2012). P. 72ff.

[34] Cf. Attfield (2012). P. 72ff.

[35] Cf. Attfield (2012). P. 72ff.

[36] Cf. Coakley (2017a). P. 17ff.

[37] Cf. Coakley (2017a). P. 17ff.

obligation or a duty, other reasons to motivate the execution of actions do exist.[38] This leads to a distinct difference between rightness and obligation, while there is a connection between obligations and the morally right action.[39] This stems from the definition of an obligation as outlined above and may be boiled down to the thought that an act that is obligatory is also right.[40] This must be regarded as the status of obligation in which actions are morally obligatory all things considered.[41] Meaning that the balance of reasons creates an obligation for an action.[42] This simply implies that when all things are considered in the status of equilibrium of reasons favor a specific course of action, then this action is morally right.[43]

Furthermore, obligations can also be defined as duties, for example keeping promises and refraining from useless violence.[44] It is paramount to remember that obligations can clash, thus making it impossible to comply with every obligation that may be present. [45] If this happens, the obligations cannot rely on the all-things-considered principle in the balance of reasons to decide the morally appropriate course of action.[46] Instead, they must rely on an other-things-being equal-approach (ceteris paribus).[47] This is generally the case when important moral questions arise as the equilibrium of reasons for the alternatives is in balance.[48] One of these important moral questions could be AI surveillance systems, as their possible upside must be weighed against their ethical challenges. Basically types of actions that are obligations in the ceteris paribus sense are that way because of the particular circumstances of the actions, meaning that normative theory aims at describing what set of circumstances makes an action obligatory and overrides other obligations if they

[38] Cf. Attfield (2012). P. 72ff.

[39] Cf. Attfield (2012). P. 72ff.

[40] Cf. Attfield (2012). P. 72ff.

[41] Cf. Attfield (2012). P. 72ff.

[42] Cf. Attfield (2012). P. 72ff.

[43] Cf. Attfield (2012). P. 72ff.

[44] Cf. Attfield (2012). P. 72ff.

[45] Cf. Attfield (2012). P. 72ff.

[46] Cf. Attfield (2012). P. 72ff.

[47] Cf. Attfield (2012). P. 72ff.

[48] Cf. Attfield (2012). P. 72ff.

are conflicting each other.[49] Oftentimes we associate the morally right alternative as the only viable alternative, but this is wrong as more than one action can be morally right even if it does not seem this way at first glance.[50] In AI surveillance these alternatives could be different recommendations by the AI surveillance system or the introduction of AI surveillance systems might be one of the morally right alternatives. We are able to accomplish the desired outcome in two or more different ways that bear equal moral weight and therefore we must not assume that there is one grand solution for every moral dilemma we currently face that normative ethical theory can present to us.[51] Specifically a holistic normative theory should deliver a broad overview of alternatives that are both morally right and wrong and help us differentiate them from each other.[52]

Certain actions may be morally desirable but not obligatory.[53] These actions are defined as actions that are laudable but cannot be morally required or expected of a moral entity and are called supererogatory acts.[54] They typically go beyond obligations and duties.[55]

2.2 Consequentialism

This section will explain consequentialism, as it is an important cornerstone of normative ethics, afterwards it will discuss utilitarianism which is an important consequentialist theory.[56] Most of the aspects of consequentialism that will be discussed are central pillars of utilitarianism. However, it is important to explain consequentialism first as it is a basis of utilitarian thinking. Consequentialism, as the name suggests, determines the morality of an action by the quality of its consequences.[57] It is a deontic theory, meaning that it tries to guide and judge our choices and actions.[58]

[49] Cf. Attfield (2012). P. 72ff.

[50] Cf. Attfield (2012) P. 75ff.

[51] Cf. Attfield (2012). P. 75ff.

[52] Cf. Attfield (2012). P. 75ff.

[53] Cf. Attfield (2012). P. 75ff.

[54] Cf. Attfield (2012). P. 75ff.

[55] Cf. Attfield (2012). P. 75ff.

[56] Cf. Pirie-Griffiths (2016). P. 1ff.

[57] Cf. Coakley (2017a). P. 17ff.

[58] Cf. Alexander and Moore (2007). P. 1ff.

Each variant has its own direct and indirect theories, leading to a different assessment of consequences for the same action.[59] Consequentialism typically deals with actions and not moral agents, while it is clear that an action naturally needs to be connected to a moral entity, or more generally, a moral agent.[60] Therefore it is crucial to incorporate the moral agent when assessing the morality of an action.[61] It is a general consensus between consequentialists that the morally best action is agent neutral, meaning that the desirable state of affairs is a state that all agents have a motivation to achieve.[62] This is true without regard to the exercise of moral agency by a moral agent.[63]

Naturally, our actions have impacts that can make a difference in a wide array of scenarios. [64] As briefly outlined above, consequentialism tries to assess these impacts based on their moral quality in order to find the best alternative for a certain scenario, especially when obligations or duties are in contrast to each other.[65] In order to assess the moral quality of an alternative all impacts of an decision should be known.[66] However, this is implausible as the impacts of our actions may develop over a long period of time, leaving the later questions completely unpredictable.[67] Therefore consequentialism should be regarded as a concept in which all foreseeable consequences are relevant to the moral assessment of an action.[68] Given the fact that consequences and motivations are not as easily comprehendible as they seem, double effects are possible.[69] A good example for a double effect can be found in the domain of AI. One may come to the conclusion that it is ethically and morally wrong to use AI surveillance technology to repress a population or to monitor an individual without any justifying means. However, if the AI is used to monitor individuals that are suspected

[59] Cf. Coakley (2017a). P. 17ff.

[60] Cf. Coakley (2017a). P. 17ff.

[61] Cf. Coakley (2017a). P. 17ff.

[62] Cf. Alexander and Moore (2007). P. 1ff.

[63] Cf. Alexander and Moore (2007). P. 1ff.

[64] Cf. Attfield (2012). P. 77ff.

[65] Cf. Attfield (2012). P. 77ff.

[66] Cf. Attfield (2012). P. 77ff.

[67] Cf. Attfield (2012). P. 77ff.

[68] Cf. Attfield (2012). P. 77ff.

[69] Cf. Attfield (2012). P. 77ff.

of planning a crime and that crime is then subsequentially foiled because of the use of AI surveillance systems and many innocent people are saved this way, the use of AI surveillance could be ethically allowable in this case. One especially interesting thought from the example before: we can clearly see two different uses for AI surveillance systems, that may be judged differently from an ethical point of view, whereas the underlying intention is the same. Intentions play an important role in consequentialism because they are paramount in the appraisal of actions.[70] Despite this, intentions do not always define the identity of an action, moral entities do not always reflect correctly on why they are acting in a certain way.[71]

Until now we have solely talked about actions and their assessment of an ethical perspective. It is vital to point out that not deciding to take action is also an action in itself.[72] To account for this possibility, consequentialism contains the principle of acts and omission.[73] It states that the predicted consequences of not acting in a certain way (omission) are not morally relevant, even if the consequences of acting are morally relevant.[74]

As the understanding for consequentialism has been laid, it is important to understand how consequentialism evaluates moral agents.[75] Many different ways to identify a moral agent exist in consequentialism, this thesis will adopt a basic approach to identify moral agents. Henceforth, morally good agents are moral agents that undertake actions that must be assessed as morally right and morally bad agents tend to act morally wrong in the view of consequentialism.

2.2.1 The utilitarian school of thought

Utilitarianism is a very popular ethical theory.[76] The most popular utilitarian theorists were Jeremy Bentham (1748-1832) and John Stuart Mill (1806-1873).[77] At its core it therefore also follows the

[70] Cf. Attfield (2012). P. 79ff.

[71] Cf. Attfield (2012). P. 79ff.

[72] Cf. Attfield (2012). P. 79 ff.

[73] Cf. Attfield (2012). P. 79ff.

[74] Cf. Attfield (2012). P. 79ff.

[75] Cf. Coakley (2017b). P. 17ff.

[76] Cf. Sidgwick (1874). P. 17ff.

[77] Cf. Nathanson (2018). P. 1ff.

principle of judging the moral quality of an action by its effects.[78] The scope however is different, utilitarianism argues that the raison d'être of morality is to increase the quality of live for everyone affected by a decision by increasing the amount of happiness or pleasure.[79] Consequently, it can be inferred that the reduction of pain and suffering also increase the well being of the affected individuals. Yet this hedonistic view must be expanded. Individuals can have desires that do not enhance their welfare directly.[80] For example, the fulfillment of any desire that an individual might have constitutes a benefit to the individual.[81] Simply put, utilitarianism strives at maximizing a metric that is called utility, for everyone involved.[82] Other goods that should secondarily be promoted in utilitarianism are fairness, justice and equality.[83]

The proponents of utilitarianism have been remarkably orientated to deliver practical changes that promote happiness or at least reduce suffering.[84] Bentham and Mill have advocated animal rights during a time at which laws for animal protection were not signed into codified law in any country.[85] In addition, utilitarians have led campaigns to strengthen the rights of women, including the admittance to universities.[86] Mill encouraged the freedom of expression and thought, urging governments to not interfere with the privacy of the population as long as they did not harm other individuals.[87] In political philosophy, utilitarians support democratic governments.[88] They commonly state that democracy is combining the interests of the government with the general interests of its constituents.[89] Utilitarians argue that the greatest individual liberties for individuals lead to

[78] Cf. Nathanson (2018). P. 2ff.

[79] Cf. Nathanson (2018). P. 2ff.

[80] Cf. Hooker (2003) P. 1ff.

[81] Cf. Hooker (2003) P. 1ff.

[82] Cf. Tännsjö (2022). P. 18ff.

[83] Cf. Hooker (2003) P. 1ff.

[84] Cf. Lazari-Radek and Singer (2017). P. 1ff

[85] Cf. Lazari-Radek and Singer (2017). P. 1ff.

[86] Cf. Lazari-Radek and Singer (2017). P. 1ff.

[87] Cf. Lazari-Radek and Singer (2017). P. 1ff.

[88] Cf. Duignan and West (2020).

[89] Cf. Duignan and West (2020).

maximized welfare, as individuals are typically the most fitting judges of their own utility and welfare.[90] But, what exactly is utility?

In the history of ethics and philosophy, utility has always played a vital role.[91] Etymologically, utility describes the usefulness of an object or a situation.[92] According to Bentham, nature has placed humankind under the rule of pain and pleasure.[93] He further states that pain and pleasure govern us in establishing a standard of what is right and what is wrong.[94] One can conclude from these assumptions that pain and pleasure and therefore utility is omnipresent. Bentham´s principle of utility captures this omnipresence. Further sharpening the term of utility, Bentham defines utility as a property that applies to anything describing an object or individuals' potential to produce benefit, advantage, pleasure, happiness or good.[95] In ethics, it has far evolved beyond the point of being an attribute that is used in a description or in an analysis.[96] Utility has evolved to be an underlying concept of moral and political philosophy and ethics.[97] However, this change in the perception of utility creates an opportunity to assess the morality of an action in a more precise and flexible way, paving the way to improving our understanding of morality.[98] It can be a pillar in various reforms of how we assess morality in order to increase human happiness.[99] Instead of being forced to obey inflexible rules and precepts, utility allows the development of realistic and practical guidelines to assess actions and omissions.[100] In this aspect, utility is seen as a liberating doctrine by utilitarian

[90] Cf. Duignan and West (2020).

[91] Cf. Crimmins (2017). P. 555f.

[92] Cf. Crimmins (2017). P. 555f.

[93] Cf. Bentham (1781). P. 1ff.

[94] Cf. Bentham (1781). P. 1ff.

[95] Cf. Bentham (1781). P. 1ff.

[96] Cf. Crimmins (2017). P. 555ff.

[97] Cf. Crimmins (2017). P. 555ff.

[98] Cf. Crimmins (2017). P. 555ff.

[99] Cf. Pohlman (1984). P. 11ff.

[100] Cf. Crimmins (2017). P. 555ff.

thinkers, allowing to better society as a whole.[101] Jeremy Bentham noted that utility plays a critical role in assessing the morality of actions.[102]

This metric also applies to decreasing the amount of pain and unhappiness.[103] That means that we can distinguish consequentialism and utilitarianism by their axiology, meaning their theory of how to create “good”.[104] Utilitarianism follows a monistic axiology, stating that utility is the only thing that is good for its own sake.[105] Typically utilitarianism rejects moral codes that arise from tradition or are orders given by sole leaders and supernatural entities.[106] Therefore, utilitarianism aims at creating the greatest good for the majority of individuals involved. Utilitarianism can be applied to actions, policies, laws and character traits, at the end of the consideration utilitarian thinkers choose the action that maximizes utility.[107] In line with John Stuart Mill, who argues that actions are right to the degree in which they produce utility, or what he calls the greatest happiness principle.[108]

First, the specific terminology of the utilitarian way of thinking will be defined. A typical utilitarian approach is to use moral terms in a slightly technical sense.[109] Actions must be right or wrong, if an action is not right then utilitarians ultimately regard it as wrong. [110] Obligations have been defined earlier and the same understanding of obligations applies to utilitarianism. Given the understanding of the terminology, we can formulate the utilitarian criterion of rightness for actions. **An action is morally correct if and only if, in the respective situation, no alternative that results in a higher degree of utility, exists**.[111]

[101] Cf. Crimmins (2017). P. 555ff.

[102] Cf. Bentham (1781). P. 1ff.

[103] Cf. Nathanson (2018). P. 2f

[104] Cf. Crimmins (2017). P. 555ff.

[105] Cf. Crimmins (2017). P. 555ff.

[106] Cf. Nathanson (2018). P. 1f.

[107] Cf. Nathanson (2018). P. 2ff.

[108] Cf. Mill (1863). P. 12ff.

[109] Cf. Tännsjö (2022). P. 17ff.

[110] Cf. Tännsjö (2022). P. 17ff.

[111] Cf. Tännsjö (2022). P. 17ff.

At first glance, utilitarianism seems to be a straight-forward theory but that is deceiving.[112] Before attempting to find out, which action maximizes utility, a few concepts must be defined.[113] First of all, the definition of what is regarded as morally acceptable and morally unacceptable should be established.[114] Furthermore, the stakeholders that are affected by the action must be identified.[115] Lastly, foreseeable and unforeseeable consequences of the action or the omission should be considered.[116]

To determine what is morally acceptable and what is morally unacceptable, e.g., what is “good” and what is “bad”, classical utilitarianism generally applies a hedonistic view.[117] Hedonism states that the only things worth pursuing in life are pleasure and happiness.[118][119] They are defined as intrinsic goods because they further produce happiness by themselves.[120] In contrast to this, important aspects of life, for example food, beverages and personal freedom, are instrumental in achieving pleasure and happiness.[121] On the negative side, everything that reduces happiness and pleasure and/or creates pain is morally unacceptable or undesirable.[122] Nevertheless, there is a vivid discussion among utilitarian thinkers how to define what is morally acceptable and what is morally unacceptable as many of them reject the ideas of hedonism, but this discussion will not be a subject of this thesis.[123]

In order to identify the stakeholders that are affected by the action, it is paramount to distinguish between individual interests, group-interests and the interest of everyone affected by the action or

[112] Cf. Nathanson (2018). P. 2ff.

[113] Cf. Nathanson (2018). P. 2ff.

[114] Cf. Nathanson (2018). P. 2ff.

[115] Cf. Nathanson (2018). P. 2ff.

[116] Cf. Nathanson (2018). P. 2ff.

[117] Cf. Nathanson (2018). P. 2ff.

[118] Cf. Nathanson (2018). P. 2ff.

[119] Cf. Tännsjö (2022). P. 17ff.

[120] Cf. Nathanson (2018). P. 2ff.

[121] Cf. Nathanson (2018). P. 2ff.

[122] Cf. Nathanson (2018). P. 2ff.

[123] Cf. Nathanson (2018). P. 2ff.

omission.[124] Individual interests are occurring when an individual, which is a moral entity, only considers how to maximize its own utility.[125] In this case the utilitarian way of thinking only applies to the decision about which action maximizes the utility for a single moral entity, it is aimed at judging how the possible actions affect a single person´s interest and does not take the interests of other people into account. [126] Yet, actions and their outcomes often affect groups of moral entities.[127] The metric for the well-being of a group is the aggregated total of the interests of all its members.[128] It is important to note that all interests are weighted equally.[129] Additionally, the scope must be widened to every moral entity who is affected by the action or omission.[130] Utilitarian theory focuses on the calculation of the utility of laws, actions and policies from an impartial point of view.[131]

The prediction of foreseeable and unforeseeable consequences of an action or an omission in utilitarian theory faces the same problem as the similar process in consequentialism, as the consequences are hardly determinable in space and time.[132] Usually, the actual consequences go beyond what was foreseen at the time of the decision.[133] Therefore, coming to the right conclusion may proof difficult. Due to this, there is a disagreement whether the foreseeable or the unforeseeable consequences should serve as a metric to determine the moral quality of actions.[134]

At this point, special attention should be dedicated to the application of the maximin principle by John Rawls in his famous book *A Theory of Justice* (1971).[135] Rawls positions his theory of justice around individuals that are hindered by what he calls “the veil of ignorance” in a hypothetical

[124] Cf. Nathanson (2018). P. 2ff.

[125] Cf. Nathanson (2018). P. 2ff.

[126] Cf. Nathanson (2018). P. 2ff

[127] Cf. Nathanson (2018). P. 2ff.

[128] Cf. Nathanson (2018). P. 2ff.

[129] Cf. Nathanson (2018). P. 2ff.

[130] Cf. Nathanson (2018). P. 2ff.

[131] Cf. Nathanson (2018). P. 2ff.

[132] Cf. Nathanson (2018). P. 2ff.

[133] Cf. Nathanson (2018). P. 2ff.

[134] Cf. Nathanson (2018). P. 2ff

[135] Cf. Rawls (1971) P. 1ff.

situation. This hypothetical situation is the "original position", according to Rawls.[136] However, these individuals have to choose among various alternatives of actions that are mutually exclusive.[137] Rawls states that the individuals will use the maximin principle to find the action with the highest moral quality, based on utility.[138] The maximin principle, etymologically, a fusion of maximum and minimum, directs our attention to the best possible outcome in the worst conditions.[139] Meaning that the minimum will be maximized. Rawls compares the maximin rule with the maximization principle of utilitarianism to argue that the worst outcome of a maximization of utility may very well be a life barely worth living for most affected individuals.[140] His application of the maximin rule does still ensure good living conditions for the affected individuals, as the minimum is maximal.[141]

2.2.2 Differences between act utilitarianism and rule utilitarianism

Act and rule utilitarianism are two important variants of utilitarianism.[142] In that, they agree that the overall aim should be the creation of the best outcomes for everybody.[143] This divide in utilitarian thinking exists since the 1950s, when this terminology was first used by the philosopher, Richard Brandt.[144] Other terms that may be used to describe it, such as direct utilitarianism for act utilitarianism und subsequently indirect utilitarianism for rule utilitarianism.[145]

[136] Cf. Olatunji (2008). P. 65ff.

[137] Cf. Olatunji (2008). P. 65ff.

[138] Cf. Olatunji (2008). P. 65ff.

[139] Cf. Olatunji (2008). P. 65ff.

[140] Cf. Olatunji (2008). P. 65ff.

[141] Cf. Rawls (1971) P. 1ff.

[142] Cf. Nathanson (2018). P. 2ff.

[143] Cf. Nathanson (2018). P. 2ff.

[144] Cf. Nathanson (2018). P. 2ff.

[145] Cf. Nathanson (2018). P. 2ff.

Act utilitarianism is, in comparison with rule utilitarianism, quite straightforward. Act utilitarians simply argue that, everything we do must follow the principle of maximizing utility with our action.[146] The correct action in any given situation is always the action that maximizes utility compared to other possible actions.[147] Thus, act utilitarianism combines consequentialism and welfarism, meaning that a moral agent is always required to commit to the action that maximizes utility.[148] In other words, the moral permissibility or impermissibility of an action is determined only by the value of its consequences, if no alternative action generates more utility.[149] It is vital to add that act utilitarianism is an agent-neutral theory.[150] The identification of a moral agent makes no difference to the principle of maximizing utility.[151]

Rule utilitarians believe in the importance of moral rules to maximize utility.[152] Justice, fairness and equality are maximized via a set of justified moral rules.[153] This means that rules should be selected on the basis of their consolidated net benefits and actions should always be based on the selected set of rules.[154] In other words, the rules are justified by their consequences.[155] These rules must incorporate every individual that is affected by the action. In order to achieve this, they typically follow a two-step approach.[156] The first part of this approach states that a specific action is morally justified if it is in accordance with a justified moral rule.[157] But what exactly is a justified moral rule? This is where the second part of the rule utilitarians approach comes into play, defining that a moral rule is justified if its inclusion into our moral code maximizes utility in comparison to

[146] Cf. Nathanson (2018). P. 2ff.

[147] Cf. Nathanson (2018). P. 2ff.

[148] Cf. Hooker et al. (2022). P. 40

[149] Cf. Crimmins (2017). P. 555ff.

[150] Cf. McNaughton (2007). P. 1ff.

[151] Cf. McNaughton (2007). P. 1ff.

[152] Cf. Nathanson (2018). P. 2ff

[153] Cf. Hooker (2003) P. 1ff.

[154] Cf. Crimmins (2017). P. 555ff.

[155] Cf. Hooker (2003) P. 2ff.

[156] Cf. Nathanson (2018). P. 2ff.

[157] Cf. Nathanson (2018). P. 2ff.

other rules or the absence of rules.[158] This means, that a rule that is part of our moral code is internalized.[159] Thus, the assessment of morality of individual actions or omissions should be conducted by referencing that action to our codex of moral rules.[160] This means that rule utilitarianism defines three main components. The first component describes what acts are ethically right or wrong while the second component defines the procedure that should be used (set of utility maximizing rules based on internalized moral rules).[161] The remaining third component forms the conditions under which sanctions for ethically wrong actions are appropriate.[162] Additionally, the codex of moral rules should be continuously improved by adopting new moral rules that maximize utility for society.[163] However, the definition of rule utilitarianism as a system of rules to guide everyday moral decisions and ethically difficult decisions is possible as well.[164]

Rule utilitarianism can be expressed via full rule utilitarianism.[165] Full rule utilitarianism contains criteria for all three components.[166] That means that full rule utilitarianism defines an action as morally wrong if it is forbidden by rules that maximize utility when obeyed. Additionally, full rule utilitarianism claims that moral agents should base their decision-making on a set of rules that maximizes utility, just as in normal rule utilitarianism.[167] Lastly, full rule utilitarianism establishes that the conditions under which sanctions for unethical behavior are justified, stem from the set of rules.[168] The decision procedure in full rule utilitarianism is based on the set of rules that maximize utility for the society, stating that the utility may already be maximized if, the majority of a society adheres to the set of rules.[169]

[158] Cf. Nathanson (2018). P. 2ff.

[159] Cf. Crimmins (2017). P. 555ff.

[160] Cf. Nathanson (2018). P. 2ff.

[161] Cf. Hooker (2003) P. 2ff.

[162] Cf. Hooker (2003) P. 2ff.

[163] Cf. Nathanson (2018). P. 2ff.

[164] Cf. Crimmins (2017). P. 555ff.

[165] Cf. Hooker (2003) P. 2ff.

[166] Cf. Hooker (2003) P. 2ff.

[167] Cf. Hooker (2003) P. 2ff.

[168] Cf. Hooker (2003) P. 2ff.

[169] Cf. Hooker (2003) P. 3ff.

Full rule utilitarianism can be formulated based on the utility that actually results from the set of rules either evidence based or rationally inferred, boiled down to the comparison of actual vs. expected net utility benefit.[170] The actual vs. expected good approach formulates full rule utilitarianism based on the actual or the expected utility the set of rules creates. The utility of a set of rules is calculated via identifying the value of each possible outcome of an action that is happen in adherence to the rules or disobeying them.[171] The value or disvalue is then multiplied with the probability of the occurrence of the respective outcome.[172] From here on out, rule utilitarianism in this paper is defined as full rule utilitarianism.

Both forms of utilitarianism could help answer the two main questions of this thesis. It may be possible to create insights about the ethical permission of AI in mass surveillance tools by leveraging a utilitarian perspective. Before this decision is made, deontological theories will have the stage.

2.3 Deontology

As mentioned in the beginning of Section 2.1 consequentialism is a cornerstone of normative ethics. Yet, deontology could just as well as consequentialism serve as the theoretical groundwork of this thesis. This subchapter will take a closer look at this ethical concept, similar to what was done before with consequentialism. Therefore, tangible deontological theories will be discussed before assessing their usability for this thesis in the next section.

The most famous deontological thinkers are Immanuel Kant (1724-1804) and William David Ross (1877-1971).[173] Deontology as a word derives from the Greek words for duty and study.[174] Similar to consequentialism, deontology aims at unveiling which choices are morally required, forbidden or permitted.[175] To achieve this, deontology defines a number of distinct duties.[176] These duties

[170] Cf. Hooker (2003) P. 3ff.

[171] Cf. Hooker (2003) P. 3ff.

[172] Cf. Hooker (2003) P. 3ff.

[173] Cf. Crimmins (2017). P. 490f.

[174] Cf. McNaughton (2007). P. 1ff.

[175] Cf. Alexander and Moore (2007). P. 1ff.

[176] Cf. McNaughton (2007). P. 1ff.

define that certain actions are intrinsically right while other duties are intrinsically wrong.[177] This implies that deontological theories assess the morality of an action by different criteria as the state of affairs these actions result in.[178] Typically, deontological theories incorporate two classes of duties, that have their origin in the social and personal relationships that a moral agent has.[179] Leading to the fact that deontological theories argue that some choices, no matter how positive their impact would be for the good of society, are morally forbidden.[180] It is vital to state, that no formal definition of deontology exists but it is generally regarded as the counterpart of consequentialism.[181] It is a normal process that people develop more than one theory to solve complex moral dilemmas. The shortcomings of consequentialism will be reflected later, but for now it is important to note that some ethical theorists claim that consequentialism, and subsequently utilitarianism, cannot be applied to all areas of ethics. [182] In contrast to utilitarianism, which is assessing the morality of actions based on the value (utility) of the consequences, a deontological theory focusses on judging the morality of an action built on its adherence to certain rules. The greatest contrast to consequentialism is, as previously outlined, is that some actions cannot be assessed by the moral quality of their outcomes.[183] In this sense of deontology, what makes an action morally right is its conformity with a social norm.[184] These norms must simply be obeyed by each moral agent, the right should have a priority over the good, thus an act that is not in accordance with right may not be undertaken even if it produces good.[185][186] Regarding moral agents, deontology is an agent-depending moral theory, as the specification of the duties is interconnect with the environment of the agent.[187] However, some versions of deontology, especially the one that are based on rights and

[177] Cf. McNaughton (2007). P. 1ff.

[178] Cf. Alexander and Moore (2007). P. 1ff.

[179] Cf. McNaughton (2007). P. 1ff.

[180] Cf. Johnson (2017). P. 33ff.

[181] Cf. Crimmins (2017). P. 114ff.

[182] Cf. Attfield (2012). P. 85ff.

[183] Cf. Alexander and Moore (2007). P. 1ff.

[184] Cf. Johnson (2017). P. 33ff.

[185] Cf. Alexander and Moore (2007). P. 1ff.

[186] Cf. Johnson (2017). P. 33ff.

[187] Cf. McNaughton (2007). P. 1ff.

not duties, may be considered as being agent-neutral.[188] The focus of this thesis will be on the agent-neutral deontological theories as they are easier to differentiate from consequentialism.[189]

One promising area of application for deontology is that of justice and rights.[190] Universal rights such as the right to live to be physically unharmed may be too important to be judged based on consequences or on the common good. [191] Consequentialism could make the application of these rights fluctuate from case to case. [192] Thus, several voices argue that the recognition of these universal rights is not adequately ensured by consequentialist theories such as utilitarianism.[193]

Upholding rights includes respecting moral rules that prohibit certain types of treatment. [194] The moral entities to which these rules apply are the bearer of the respective right that rests on the underlying moral rule. [195] The value of this relationship between rights and moral rules becomes clear as soon as the question is raised, whether the this contributes to the general good of society.[196] As the absence of this relationship between rights and moral rules would mean that no society would be secure from types of arbitrary mistreatment.[197]

A very popular deontological theory is the deontological pluralism. William David Ross (1877-1971) was a Scottish philosopher, who developed a new way of deontological ethics that rivaled the views of Immanuel Kant and utilitarianism.[198] Over the course of his life, Ross rejected Kantian deontological ethics and utilitarianism, stating that both of them are mere over-simplifications of the moral life, as they fail to take a variety of moral attitudes into account.[199][200] With the fact that

[188] Cf. Alexander and Moore (2007). P. 1ff.

[189] Cf. Alexander and Moore (2007). P. 1ff.

[190] Cf. Attfield (2012). P. 85ff.

[191] Cf. Attfield (2012). P. 85ff.

[192] Cf. Attfield (2012). P. 85ff.

[193] Cf. Attfield (2012). P. 85ff.

[194] Cf. Attfield (2012). P. 85ff.

[195] Cf. Attfield (2012). P. 85ff.

[196] Cf. Attfield (2012). P. 85ff.

[197] Cf. Attfield (2012). P. 85ff.

[198] Cf. Skelton (2010). P. 1ff.

[199] Cf. Skelton (2010). P. 1ff.

[200] Cf. Ross (2007). P. 1ff.

lying is always morally wrong in Kantian deontology, he argues that Kant wrongly boils down important aspects of decisions into a right or wrong mentality which does not do justice to complex moral decisions.[201] Kant´s high level of abstraction neglects relevant factors in assessing an action or omission for their morality.[202] Furthermore, Ross criticizes Kantian deontological ethics for establishing moral worth as the sole motivator and metric for decisions.[203]

Ross also blames utilitarianism as oversimplifying and misrepresenting important moral decisions.[204] He criticizes that utilitarianism is based on the maximization of utility, which means basing the theory onto a single basic value.[205] According to Ross, this misrepresents the understanding of moral deliberation.[206] He therefore opposes the utilitarian maximization principle, while not arguing that utilitarianism is fundamentally wrong.[207] From Ross´s perspective utilitarianism is counter-intuitive, acting in contrast to common sense ethics, and not holistic, as it does not incorporate all moral duties and complications.[208][209]

Based on this critique of utilitarianism and Kantian deontological ethics, William David Ross developed a new form of deontological ethics, the deontological pluralism, that incorporate prima facie (at first sight) duties, as Ross called them. Deontological pluralism is an anti-consequentialist theory, as it is based on duties and not on outcomes.[210] In this sense pluralist describe the fact that a number of different fundamental rules are established in the deontological pluralism.[211] These fundamental moral rules are the prima facie duties that are pillar of deontological pluralism that also sets the theory apart from other theories.[212] It is important to mention that the prima facie

[201] Cf. Skelton (2010). P. 1ff.

[202] Cf. Skelton (2010). P. 1ff.

[203] Cf. Skelton (2010). P. 1ff.

[204] Cf. Skelton (2010). P. 1ff.

[205] Cf. Skelton (2010). P. 1ff.

[206] Cf. Skelton (2010). P. 1ff.

[207] Cf. Simpson (2019). P. 1ff.

[208] Cf. Simpson (2019). P. 1ff.

[209] Cf. Ross (2007). P. 1ff.

[210] Cf. Simpson (2019). P. 9ff.

[211] Cf. Simpson (2019). P. 9ff.

[212] Cf. Simpson (2019). P. 9ff.

duties may come into conflict with each other.[213] The prima facie duties are the major innovation that deontological pluralism brings to the table, substituting absolute or exceptionless rules.[214][215] The prima facie duties according to William David Ross are:

1. Fidelity: The moral agent should strive to fulfill promises and be honest
2. Reparation: A moral agent should strive to make amends for caused damage or the wronging of another
3. Gratitude: Moral agents should return services to those from whom they have accepted beneficiary services in the past
4. Non-maleficence: Moral agents should refrain from harming others in any way
5. Beneficence: Moral agents should be kind to each other, promoting each other's well-being
6. Self-improvement: Moral agents should strive to improve their health, wisdom, security and well-being
7. Justice: Benefits and burdens should be equally distributed between the moral agents

The actual significance of the prima facie duties may be difficult to grasp at a first glance, Ross himself admitted that the term prima facie is unfortunate, while still being the most precise term to describe what prima facie duties are.[216] Prima facie duties are in fact not be confused with obligations. [217] Instead, every single duty relies on a separate and discrete moral ground and specifies an argument in favor or against an action.[218] The prima facie duties are, therefore never absolute.[219] That means that the considerations that stem from the prima facie duties must we weighted and balanced against each other to determine the morally correct course of action.[220] Furthermore, there is no hierarchical structure among the prima facie duties, since context and backstory have a huge

[213] Cf. Simpson (2019). P. 9ff.

[214] Cf. Simpson (2019). P. 9ff.

[215] Cf. Skelton (2010). P. 1ff.

[216] Cf. Skelton (2010). P. 1ff.

[217] Cf. Skelton (2010). P. 1ff.

[218] Cf. Skelton (2010). P. 1ff.

[219] Cf. Simpson (2019). P. 9ff.

[220] Cf. Skelton (2010). P. 1ff.

impact in the execution of the duties to identify the action with the highest moral quality.[221] Yet, some of the duties tend to overwrite other. Exemplary, the duty to be non-maleficent seems to trump the other duties.[222] Apart from AI, most people would agree that stealing food from a rich person to donate it to starving children is ethically permissible, even if it violates the principle of non-maleficence.[223] An AI surveillance system, however well and ethically permissible, if possible, is set up, may break the duty of non-maleficence, depending on how it collects data.

The non-hierarchical structure of Ross´s theory entails that a moral agent can have multiple moral obligations in contrast to a single imperative.[224] It is apparent that these obligations can come into contrast with each other which is a core takeaway from deontological pluralism.[225] In such cases, Ross argues, there will always be a duty that has a certain urgency and should therefore be prioritized.[226]

The deontological pluralism may serve as the theoretical basis for the aim of this thesis. The prima facie duties could be used to develop a framework for the ethical and responsible use of AI-based mass surveillance tools. However, some drawbacks exist that will be a subject in the next section.

2.4 A critical comparison of consequentialist and deontological theories

As the theoretical groundwork gas now been laid, it is time to conduct a critical discussion of the ethical frameworks based on their advantages and disadvantages. The selection of the ethical framework will be at the end of this section. Additionally, all the discussed ethical concept will be evaluated. Due to this discussion, utilitarianism, more to the point: rule utilitarianism, will be chosen as the framework. Therefore, the result of this section will be a commitment to utilitarianism and an explanation of this commitment, which that can help us answer the question, whether AI-based mass surveillance tools can be used in an ethically correct way and how this application may look.

[221] Cf. Simpson (2019). P. 9ff.

[222] Cf. Simpson (2019). P. 9ff.

[223] Cf. Simpson (2019). P. 9ff.

[224] Cf. Simpson (2019). P. 9ff.

[225] Cf. Simpson (2019). P. 9ff.

[226] Cf. Simpson (2019). P. 9ff.

Consequentialism will be discussed first. Over the decades it has been subject to criticism for two main reasons that have been articulated by several voices.[227] The two reasons, one of which has already been briefly mentioned earlier, may seem paradox.[228] Several voices have criticized that consequentialism is too demanding while not being demanding enough.[229] In order to understand these two major criticisms, it is crtitical to assess each one of them separately.[230] The first one, regarding extreme demandingness, is usually concerned with the fact that consequentialism does not know any realm for moral permission, supererogation and moral indifferences, meaning that everything is either required or forbidden.[231] Additionally, consequentialism does not grant partiality toward a moral entities family, friends or other preferences.[232] In the eyes of its critics, this reduces it to a self-effacing moral theory.[233] The second reason for criticism is that consequentialism may allow too much.[234] This has already been mentioned earlier, meaning that vital rights may not be protected by consequentialist thinking. In certain cases, this could mean that innocents are harmed in order to create a greater good for the society.[235] In the right environment, the focus on consequences may justify a wide array of acts, even if they harm others.[236] It does not matter how harmful an act may be as long as the numbers of beneficiaries is higher that the number of harmed moral entities.[237] If an AI system is effectively repressing a group of people in a respective country, consequentialism would still assess it as ethically permissible, if the people that are benefitting from the surveillance are more than individuals than the repressed group of people.

After having regarded the drawbacks of consequentialism, it is important to take a look at the advantages of it as well. Most deontological theories do not seem to draw the line between what is

[227] Cf. Alexander and Moore (2007). P. 1ff.

[228] Cf. Alexander and Moore (2007). P. 1ff.

[229] Cf. Alexander and Moore (2007). P. 1ff.

[230] Cf. Alexander and Moore (2007). P. 1ff.

[231] Cf. Alexander and Moore (2007). P. 1ff.

[232] Cf. Alexander and Moore (2007). P. 1ff.

[233] Cf. Alexander and Moore (2007). P. 1ff.

[234] Cf. Alexander and Moore (2007). P. 1ff.

[235] Cf. Alexander and Moore (2007). P. 1ff.

[236] Cf. Alexander and Moore (2007). P. 1ff.

[237] Cf. Alexander and Moore (2007). P. 1ff.

morally wrong or right. [238] Therefore, consequentialism offers a viable explanation for moral institutions, even those that deontological theories have trouble defining, because it states a clear line of what is morally permissible and what is not morally permissible.[239] Evidently, this can be illustrated with the example of two conflicting promises.[240] In consequentialism it is easier to determines which promise to break, it will always be the one which has, upon completion, less impact on the common good.[241] Contrary to this, deontological theories may have a problem in defining between which promise to keep, respectively break.[242]

Utilitarian thinkers often claim that their moral concept has an edge over competing non-consequentialist theories because it bases its assumptions on the consequences of conduct while incorporating utility.[243][244] To break this down, utilitarianism and the outcomes of its application to moral decisions are based on empirically determinable facts, offering to settle moral dilemmas on objective grounds.[245] So in a way, utilitarians claim that there is a big contrast between their theory and other moral theories.[246] However, this argued contrast may not hold up to critical reasoning, because critics have voiced concern that the sound assessment of morality may not be established on the consequences of an action as they are not easily foreseeable.[247]Additionally, utilitarianism is often criticized for a number of reasons.[248] First, it is often stated that the moral implications of utilitarianism are opaque, which means they are incredibly difficult to determine.[249] This opacity

[238] Cf. Sinnott-Armstrong (2003). P. 2ff.

[239] Cf. Sinnott-Armstrong (2003). P. 2ff.

[240] Cf. Sinnott-Armstrong (2003). P. 2ff.

[241] Cf. Sinnott-Armstrong (2003). P. 2ff.

[242] Cf. Sinnott-Armstrong (2003). P. 2ff.

[243] Cf. Hooker et al. (2022). P. 105ff.

[244] Cf. Lyons (2022). P. 1ff.

[245] Cf. Hooker et al. (2022). P. 105ff.

[246] Cf. Hooker et al. (2022). P. 105ff.

[247] Cf. Hooker et al. (2022). P. 105ff.

[248] Cf. Hooker et al. (2022). P. 105ff.

[249] Cf. Lyons (2022). P. 1ff.

is created by the circumstance that the utilitarian criterion of assessing the morality of an action is difficult to determine.[250]

Utilitarianism, as a type of consequentialism, is based on an consequence orientated framework, raising the question, which consequences have value based on their empirical value.[251] These aspects of critic deal with the value judgement of utilitarianism, that is needed to determine the moral assessment of actions in utilitarianism.[252] Until this criterion of moral assessment is identified, the utilitarian thinking cannot be applied to solve moral problems.[253] While utilitarianism seems to be straight forward and the maximization principle seems to be impervious to critic, the application of utilitarian theory has been subject to skeptical considerations.[254] While applying any utilitarian theory, no more than one action is morally permissible due to the maximization principle.[255]

Moreover, the potentially relevant consequences of the actions that were not assessed as maximizing utility will remain unknown, rendering the utilitarian principle as an unreliable procedure to solve morally ambiguous situations, even possibly rendering utilitarianism impossible to apply because we can hardly judge the long term effects of the moral actions and their alternatives.[256][257] Usually, it is always imaginable for an individual to act ethically right, as an action that cannot be performed must not be done.[258] Building on this, all actions hat are morally right in the view of utilitarianism are performable.[259] Nevertheless, since the evaluation of the moral actions and their alternatives does not always yield a clear result due to the reasons discussed above, utilitarianism may be impossible to apply.[260]

[250] Cf. Hooker et al. (2022). P. 105ff.

[251] Cf. Hooker et al. (2022). P. 105ff.

[252] Cf. Hooker et al. (2022). P. 105ff.

[253] Cf. Lyons (2022). P. 1ff.

[254] Cf. Lyons (2022). P. 1ff.

[255] Cf. Hooker et al. (2022). P. 105ff.

[256] Cf. Tännsjö (2022). P. 24ff.

[257] Cf. Hooker et al. (2022) P. 105ff.

[258] Cf. Tännsjö (2022) P. 24ff.

[259] Cf. Tännsjö (2022) P. 24ff.

[260] Cf. Tännsjö (2022) P. 24ff.

The neutrality of utilitarianism is also often stated a potential drawback, as it devalues the importance of preferences and personal relationships.[261] Based on the circumstances of the situation, utilitarianism may force a moral agent to disregard the individuals who are close to him, for example family and close friends.[262] If we always use the utilitarian maximization principle, do we become callously calculating individuals?[263]

Indubitably, utilitarianism might consider everyone equally, but that does not mean that everyone is treated in the same way.[264] This could be strong evidence that utilitarianism does not sufficiently address the question of equality. Who will profit from the maximized utility?[265] The distribution of welfare could benefit those, who are well off in their economic standing.[266]

As a consequentialist theory, utilitarianism inherits some perceived flaws from consequentialism.[267] Critics typically voice that utilitarianism is too demanding while also being too permissive.[268] A critic that was mentioned and discussed before briefly. However, the demandingness of utilitarianism differs from the demandingness of classical consequentialism. The critique that utilitarianism is overly demanding, materializes in a quick example.[269] As long as there is suffering in the universe, utility has not been maximized, and we are supposed to alleviate the suffering, where it is within our power to do so.[270] In an extreme way, this would require an individual to donate a massive amount of his wealth, rob a bank and donate this money as well.[271] This example may

[261] Cf. Levin (2019) P. 2ff.

[262] Cf. Levin (2019) P.2ff.

[263] Cf. Tännsjö (2022) P. 24ff.

[264] Cf. Levin (2019) P. 2ff.

[265] Cf. Tännsjö (2022) P. 24ff.

[266] Cf. Tännsjö (2022) P. 24ff.

[267] Cf. Tännsjö (2022) P. 24ff.

[268] Cf. Tännsjö (2022) P. 24ff.

[269] Cf. Tännsjö (2022) P. 24ff.

[270] Cf. Tännsjö (2022) P. 24ff.

[271] Cf. Tännsjö (2022) P. 24ff.

seem absurd but it is quite popular with the critics of utilitarianism, such as the American philosopher Peter Unger.[272]

Having discussed the potential drawbacks of utilitarianism, it is now time to shine a light on the advantages of utilitarianism. The most apparent strength of the theory is that, at first glance, utilitarianism will render it comparatively easy to come to a morally justified decision, as it provides a clear path for assessing the morality of an action.[273] Especially, in contrast to deontological theories, that do not always provide an apparent way to act morally, this strength becomes clear.[274] The impartiality of utilitarianism, that was mentioned before as a possible drawback, is on the other hand, another asset of the theory.[275] Thus, being able to assess the interests of everyone in an equal way when judging an action for its moral quality.

Regarding the aforementioned drawbacks of the theory, utilitarians may counter that all the voiced critiques are focused on the short-term impact of utilitarianism.[276] But what would a utilitarian answer to the critics of the theory? No utilitarian would agree with the criticism that the theory is impossible to apply.[277] A typical counter-argument by utilitarians is that the scope of this criticism is off. [278]

Utilitarianism aims at guiding the whole of a society to produce the best consequences that are possible. So, it is not aimed at ensuring every individual is always acting in the morally right way.[279] The negative impact of the impartiality of utilitarianism was prominently radically countered by the Cambridge philosopher Henry Sidgwick (1838-1900). Sidgwick argued that if the impartiality of utilitarianism does impact individuals in a negative way, then utilitarianism should be disregarded for the respective decision altogether.[280] Even if the maximation principle is correct

[272] Cf. Tännsjö (2022) P. 24ff.

[273] Cf. Levin (2019) P. 2ff.

[274] Cf. Levin (2019) P. 2ff.

[275] Cf. Levin (2019) P. 2ff.

[276] Cf. Levin (2019) P. 2ff.

[277] Cf. Tännsjö (2022) P. 24ff.

[278] Cf. Tännsjö (2022) P. 24ff.

[279] Cf. Tännsjö (2022) P. 24ff.

[280] Cf. Tännsjö (2022) P. 24ff.

but if the consequences of involving the maximization principle reduces utility, then it must be overlooked altogether.[281]

Equality in utilitarianism is according to utilitarians ensured by establishing a system that ensures that the focus of the maximization principle is on the individuals that are worse off than other individuals.[282] Naturally, a utilitarian would argue that not the utilitarian way of thinking is too demanding but the manmade heavy challenges of moral dilemmas such as poverty in certain parts of the world are in fact what is making the theory overly demanding.[283] Additionally, utilitarians argue that, while the obligations placed on individuals by utilitarianism are incredibly strict, they are ceteris paribus perfectly reasonable.[284]

As mentioned earlier, utilitarianism is also often criticized for allowing too much. Nevertheless, utilitarians typically counter this criticism by stating that murder should generally be outlawed and crimes against humanity must always be prosecuted and punished.[285] Even if a murder maximizes welfare for certain reasons, utilitarians claim that it still cannot be condoned and must still be punished in some way.[286] Otherwise no one in a society would be safe from arbitrary acts of violence in the name of the maximization principle.[287]

Act utilitarianism may be regarded as a highly demanding theory.[288] Unsurprisingly, similar to consequentialism, it does not incorporate the personal preferences of an agent.[289] It does not matter if someone enjoys a specific hobby or engages in a certain personal project, if these actions are not maximizing utility they must be substituted with actions that do.[290] The same is true for any action

[281] Cf. Tännsjö (2022) P. 24ff.

[282] Cf. Tännsjö (2022) P. 24ff.

[283] Cf. Tännsjö (2022) P. 24ff.

[284] Cf. Tännsjö (2022) P. 24ff.

[285] Cf. Tännsjö (2022) P. 24ff.

[286] Cf. Tännsjö (2022) P. 24ff.

[287] Cf. Tännsjö (2022) P. 24ff.

[288] Cf. Crimmins (2017) P. 1f.

[289] Cf. Crimmins (2017) P. 1f.

[290] Cf. Crimmins (2017) P: 1f.

a moral agent can undertake.[291] Additionally, supererogatory actions do not exist in act utilitarianism due to its maximizing nature.[292] **This maximizing nature is also responsible to open the decision making process of an agent to the coercion of others, making what an agent is supposed to do highly sensitive to the actions and intentions of other moral agents, no matter how evil these intentions are.**[293] Furthermore, act utilitarianism is epistemic, leading to the fact that an agent must perform difficult forecasts on the long-term effects of his or actions.[294] Additionally, the agents must estimate how much time must be allocated to the decision making process of every decision that will occur down the road.[295] Adding to this, an action that seems morally required, may be prohibited by act utilitarianism due to some remote long term effects, meaning that a moral agent will always have a lot of trouble in deciding whether an action is morally right or wrong.[296] This may lead to wrong answers to moral questions, permitting a wide array of actions that are, in fact, morally wrong. For example, if an AI surveillance system that has a repressive purpose that, apart from the exercised repression, increased the quality of life in other important aspects, act utilitarianism would see the repression as permissible. However, this rigidity is also one of the main drivers of establishing trust into these rules. The aforementioned points clarify that act utilitarianism is a very demanding moral concept.[297]

Despite its flaws, act utilitarianism shows some unique advantages. Act utilitarianism is the purest form of utilitarianism, demanding moral agents to act in a way that maximizes utility.[298] If we project this maximization principle onto a whole society, act utilitarianism maximizes utility.[299] On the flipside of rejecting rule based societies, act utilitarianism broadens the scope of an action by looking at its context while assessing it.[300] This can increase the welfare of a society because

[291] Cf. Crimmins (2017) P. 1f.

[292] Cf. Crimmins (2017) P. 1f.

[293] Cf. Crimmins (2017) P. 1f.

[294] Cf. Crimmins (2017) P. 1f.

[295] Cf. Crimmins (2017) P. 1f.

[296] Cf. Crimmins (2017) P. 1f.

[297] Cf. Hooker et al. (2022) P. 40ff.

[298] Cf. Nathanson (2018) P. 4f.

[299] Cf. Nathanson (2018) P. 4f.

[300] Cf. Nathanson (2018) P. 4f.

rules that do more harm than good can be ignored when making a decision.[301] Another advantage of act utilitarianism is that it can give us objective answers in the often perceived exclusively subjective ream of morality.[302] It provides us with the tools to assess whether certain moral believes are false.[303] Additionally, act utilitarians often argue that their theory is not completely understood by its critics.[304] They typically voice concern that act utilitarianism is misinterpreted and does not support wrong answers.[305] Normally, they claim that wrong answers are not maximizing utility and are therefore not supported by act utilitarianism.[306] Moreover, act utilitarians argue that the answers that critics label as wrong are in fact not wrong but hint at incorrect underlying values in common sense morality.[307]

Rule utilitarianism is often accused of irrationally supporting rule-based systems, even if they do not maximize utility.[308] Moreover, critics argue that while act utilitarianism and rule utilitarianism seem different, rule utilitarianism collapses into act utilitarianism upon closer review.[309] This criticism may seem counterintuitive at first glance, therefore it makes sense to investigate it more closely.[310] In order to understand it, it is paramount to understand what differentiates rule utilitarianism and for example popular deontological theories such as Kant´s categorical imperative[311]. Immanuel Kant claims that lying is always wrong, no matter the circumstances.[312][313] In a utilitarian

[301] Cf. Nathanson (2018) P. 4f.

[302] Cf. Nathanson (2018) P. 4f.

[303] Cf. Nathanson (2018) P. 4f.

[304] Cf. Nathanson (2018) P. 4f.

[305] Cf. Nathanson (2018) P. 4f.

[306] Cf. Nathanson (2018) P. 4f.

[307] Cf. Nathanson (2018) P. 4f.

[308] Cf. Nathanson (2018) P. 4f.

[309] Cf. Nathanson (2018) P. 4f.

[310] Cf. Nathanson (2018) P. 4f.

[311] The categorical imperative has not been discussed in depth. However, it is mentioned in this chapter as can be utilized to clearly illustrate the difference between utilitarianism and deontology. Additionally, it is used to discuss some commonly voiced critics to deontology.

[312] Cf. Nathanson (2018) P. 4f.

[313] Cf. Kant (2017) P. 107ff.

doctrine, Kant´s view would be regarded as overly rigid.[314] A utilitarian would argue that a lie may be permitted if it increases utility.[315] To illustrate further, a rule utilitarian would create a moral code that incorporates a list of rules dealing with the question if a lie is morally permissible or not.[316] These rules would have the character that aims at maximizing utility but also brings about a collapse to act utilitarianism with its strict focus on acting in a way that maximizes utility.[317] Additionally, critics of rule utilitarianism claim that the utility based rules that form the moral code in rule utilitarianism do not have a reasonable degree of flexibility incorporated.[318] Especially, the aspect that the degree of flexibility must be reasonable in this case is of importance.[319] If the rules become too flexible, a collapse into act utilitarianism could occur.[320] On the other hand, if the rules are to rigid, the complexities of life would not be taken into account and people could face a high degree of difficulty trying to understand the rules.[321] This is often called the “rule-worship” objection, also stating that rule utilitarians will always stick to the rules, even when the rules do not maximize utility.[322] An often negatively reviewed aspect of rule utilitarianism is that it may not be applied to find the right answers to complex moral problems.[323] Especially the areas of justice and rights and the subsequent applications of AI surveillance systems in this area may be rough waters for the concept of rule utilitarianism, because it solely focuses on developing a moral code that maximizes utility. Yet, regarding this assumption to be true, and if it bears truth, in any given case it may be wrong, meaning that circumstances can create a situation in which exercising repression via AI surveillance systems may in fact maximize utility in rule utilitarianism and would therefore be the mandatory action in rule utilitarianism while still seeming to be the morally wrong choice.[324]

[314] Cf. Nathanson (2018) P: 4f.

[315] Cf. Nathanson (2018) P. 4f.

[316] Cf. Nathanson (2018) P. 4f.

[317] Cf. Nathanson (2018) P. 4f.

[318] Cf. Nathanson (2018) P. 4f.

[319] Cf. Nathanson (2018) P. 4f.

[320] Cf. Nathanson (2018) P. 4f.

[321] Cf. Nathanson (2018) P: 4f.

[322] Cf. Hooker (2003) P. 3ff.

[323] Cf. Nathanson (2018) P. 4f.

[324] Cf. Nathanson (2018) P. 4f.

Nevertheless, rule utilitarians have addressed these skepticisms and it must be stated that rule utilitarianism also offers advantages.[325] They typically claim that act utilitarianism is an extremist form of utilitarianism from the perspective of every day morality.[326] The often negatively connotated obedience to rules may be countered by stating that maximizing utilitarianism comes with a focus on rules rather than acts.[327] This means that from the perspective of some rule utilitarians, an optimal code would not collapse into act utilitarianism under optimal circumstances, in which every moral agent follows complies with the rules.[328] Furthermore, rule utilitarians often counter the perceived collapse into act utilitarianism, that is brought forward by the critics, by incorporating the moral agent.[329] Rule utilitarians claim that this critique does not account for the highly developed moral agent the modern human is.[330] According to them, it is simply not logical, that this highly developed moral agent that is able to conceive and implement a wide variety of moral theories should conform to another moral code, that is not necessarily equivalent to act utilitarianism.[331] Furthermore, rule utilitarians usually counter the collapse-argument by arguing that the moral codex does not maximize utility by requiring that the rules are obeyed in any case, but by the general acceptance of the moral codex.[332] The rule utilitarians defend their position by stating that a higher level of general welfare is achieved by moral agents agreeing to cooperate based on a moral code.[333] The defense to the "rule-worship" criticism is that rule utilitarianism can actually endorse the breaking of rules, if it means that a negative outcome is prevented.[334] This mechanism is based on

[325] Cf. Hooker et al. (2022) P. 40ff.

[326] Cf. Hooker et al. (2022) P. 40ff.

[327] Cf. Hooker et al. (2022) P. 40ff.

[328] Cf. Hooker et al. (2022) P: 40ff.

[329] Cf. Hooker et al. (2022) P. 40ff.

[330] Cf. Hooker et al. (2022) P. 40ff.

[331] Cf. Hooker et al. (2022) P. 40ff.

[332] Cf. Hooker (2003) P. 3ff.

[333] Cf. Hooker et al. (2022) P. 40ff.

[334] Cf. Hooker (2003) P. 3ff.

exceptions to the rules of the moral codex in rule utilitarianism.[335] This will also ensure a higher degree of flexibility.[336]

Having extensively discussed the pros and cons of the consequentialist theories, the deontological theories will be the focus now. Deontological ethics in general are often criticized for requiring its own non-consequentialist framework of rationality.[337] This model of rationality must be a viable alternative to the consequentialist frameworks.[338] This creates the irrationality, as critics of deontological ethics claim, that deontology incorporates duties that may not be the morally best choice.[339] According to the skeptics, deontology will always remain paradox until it develops this underlying model of rationality.[340] Additionally, deontology is often negatively reviewed for not consisting of formulated texts, which paves the way to the question of authority in deontological ethics.[341] Even if a holistic general text would exist, deferring one´s judgement to the written judgment of the supposed holistic text is, at first glance, paradoxical.[342] Such a general text would have a religious character. The mentioned deference must be justified by deontologists, which is a difficult question and a fierce discussion among deontologists.[343] Taking a closer look at this discussion would not benefit this thesis, therefore it will not be regarded further. Another point, the critics of deontological ethics usually bring forth deals with that, if certain circumstances exist, deontological ethics demand that an individual's categorical obligations require the individual to act in a way that creates a morally worse state of affairs.[344] Additionally, it is vital for deontologists to mediate the conflicts that originate from contrasting duties and rights.[345] Kant and Ross have famously tried to solve this problem with their respective work but a definitive solution to this problem that silences

[335] Cf. Hooker (2003) P. 3ff.

[336] Cf. Hooker (2003) P. 3ff.

[337] Cf. Alexander and Moore (2007) P. 4f.

[338] Cf. Alexander and Moore (2007) P. 4f.

[339] Cf. Alexander and Moore (2007) P. 4f.

[340] Cf. Alexander and Moore (2007) P. 4f.

[341] Cf. Alexander and Moore (2007). P. 4f

[342] Cf. Alexander and Moore (2007) P. 4f.

[343] Cf. Alexander and Moore (2007) P. 4f.

[344] Cf. Alexander and Moore (2007) P. 4f.

[345] Cf. Alexander and Moore (2007) P. 4f.

all the critics is yet to be found.[346] Moreover, a paradox, typically regarded as the paradox of relative stringency, is typically viewed negatively by the skeptics of deontological thinking.[347] The paradox of relative stringency arises from the fact that all deontological duties are categorical., while still asserting that certain duties are more stringent than others.[348] Stringency, which may be defined as the degree of wrongness of the duty, originates from two considerations.[349] The premier consideration is that duties of different stringency may be balanced against each other if there is a contrast between them.[350] That means that a duty, that would have the moral weight to solve this conflict if duties can be stringent.[351] Secondly, when punishment is dealt to a moral agent, this severity of this punishment is typically based on the stringency in violation of the deontological theory.[352] Therefore, not all violations are punished equally.[353]

Lastly deontological ethics may pave the way for disastrous outcomes, which are not solely limited to thought experiments.[354] A popular example is that a moral agent is faced with the choice of torturing another moral agent in order to prevent a terrorist attack that could have millions of innocent victims.[355] The strict compliance to deontological ethics would force the moral agent to not torture and to therefore let the terrorist attack occur.[356] This point of critic is highly controversial, as a lot of individuals would judge the not-torturing as the right choice, while others see the possibility of a terror attack as enough to justify the torture of the suspect.

[346] Cf. Alexander and Moore (2007) P. 4f.

[347] Cf. Alexander and Moore (2007) P. 4f.

[348] Cf. Alexander and Moore (2007) P. 4f.

[349] Cf. Alexander and Moore (2007) P. 4f.

[350] Cf. Alexander and Moore (2007) P. 4f.

[351] Cf. Alexander and Moore (2007) P. 4f.

[352] Cf. Alexander and Moore (2007) P. 4f.

[353] Cf. Alexander and Moore (2007) P. 4f.

[354] Cf. Alexander and Moore (2007) P. 4f.

[355] Cf. Alexander and Moore (2007) P. 4f.

[356] Cf. Alexander and Moore (2007) P. 4f.

In contrast to every consequentialist theory, especially consequentialism, deontology is not impartial.[357] It therefore incorporates the preferences of a moral agent, meaning that deontology establishes a limit on the demandingness of its duties.[358] Due to this, deontology avoids one key criticism of consequentialist theories of being overly demanding, being more in accordance with our everyday lives.[359] Moreover, deontology leaves room for supererogatory actions.[360] This sets it apart from utilitarianism, which only knows actions that are either morally required or forbidden.[361] Deontological thinkers typically use a variety of arguments to counter the critical assumption that the application of deontological ethics can lead to moral disasters.[362] This thesis will only mention the two most popular counter arguments. The first one is that harms should not be aggregated, effectively denying the existence of moral disasters.[363] The second argument is to define a threshold in deontological ethics that defends deontological norms, up to the point of where the consequences of this defense become too dire.[364] However, this would in reality mean that consequentialism is used to solve moral disasters. Typically, threshold deontology is aimed at refuting claims that deontological ethics are fanatic, which a strict compliance to Kantian ethics may be regarded as.[365]

People who oppose deontological pluralism, typically criticize the theory for being overly complicated.[366] Additionally, utilitarians often see deontological pluralism as not systematic enough.[367] But even other deontologists such as John Rawls claimed that without taking into account how the

[357] Cf. Alexander and Moore (2007) P. 4f.

[358] Cf. Alexander and Moore (2007) P. 4f.

[359] Cf. Alexander and Moore (2007) P. 4f.

[360] Cf. Alexander and Moore (2007) P. 4f.

[361] Cf. Alexander and Moore (2007) P. 4f.

[362] Cf. Alexander and Moore (2007) P. 4f.

[363] Cf. Alexander and Moore (2007) P. 4f.

[364] Cf. Alexander and Moore (2007) P. 4f.

[365] Cf. Alexander and Moore (2007) P. 4f.

[366] Cf. Skelton (2010) P. 2ff.

[367] Cf. Skelton (2010) P. 1ff.

plurality of normative principles have to be compared to each other regarding priority, using an ethical criterium, it is sometimes claimed that a rational discussion is hardly possible.[368][369]

The introduction of prima facie duties must be seen as a major advance in settling the dispute between utilitarians and non-utilitarians, blazing the trail for a more modern version of deontology.[370] The innovation brought by the work of W. D. Ross helps avoiding and solving cases in which absolute deontology is unable to come to a different conclusion than the one that is, ceteris paribus, wrong.[371] Furthermore, Ross does not share the concern that Rawls brought forward.[372] He counters Rawls objections by stating that Rawls theory of justices endorses absolutism and could therefore produce counterintuitive results.[373][374]

It is now time, having discussed the disadvantages and advantages of the presented ethical theories. to select the guiding principle of this thesis. As mentioned before, utilitarianism, more precisely rule-utilitarianism, will be applied to answer the central questions of the thesis. In being a consequentialist theory, utilitarianism may be regarded as overly permissive and not demanding enough, yet it creates a clear metric to decide what is morally permissible and what is morally prohibited, as examined before. However, critics often voice that the theory relies on the hardly predictable outcome of actions, rendering it opaque and making it value judgement vulnerable to further criticism. Its impartiality is a strength and a weakness at the same time, as it does not take personal preferences into account but enables the moral agent to execute impartial decisions that benefit a greater number of individuals. Utilitarians commonly counter these critics by inferring that their scope is off and by stating that they only deal with the short-term consequences of utilitarian decisions.

Act utilitarianism is the purest form of utilitarianism but also extremely demanding. Additionally, the decision-making process of a moral agent is open to manipulation by other individuals. Despite

[368] Cf. Skelton (2010) P. 1ff.

[369] Cf. Rawls (1971) P. 186ff.

[370] Cf. Skelton (2010) P. 1ff.

[371] Cf. Skelton (2010) P. 1ff.

[372] Cf. Skelton (2010) P. 1ff.

[373] Cf. Rawls (1971) P. 186ff.

[374] Cf. Skelton (2010) P. 1ff.

the seemingly straight-forward approach of act utilitarianism, it is using a challenging decision-making process. Yet, act utilitarianism excels at maximizing utility. Rule utilitarianism is often criticized for be overly rigid and rule orientated, a critic that is typically countered by comparing it with deontological ethics that can be even more rigid. In addition, rule utilitarianism is often blamed for collapsing into act utilitarianism, yet its defenders claim that an optimal set of rules would not collapse into act utilitarianism. Furthermore, rule utilitarians claim that the value of the set of rules does not entirely rest in the obedience of the individuals to in the general acceptance of the code, which also prevents the collapse into act utilitarianism. The set of rules also achieves higher levels of welfare. Nevertheless, the rules can be broken to prevent a negative outcome.

When comparing rule utilitarianism to deontological ethics, it becomes clear that rule utilitarianism is not as strict as it seems. Furthermore, deontology needs its own framework and model of rationality that can be difficult to develop. Adding to this, no general formulated texts exist and conflicts between duties and rights can be unsolvable. The deontological pluralism by William David Ross has been paving the way for the modern versions of deontology. Hitherto, its critics claim that it is overly complicated and not systematic enough. Conflicts between the prima facie duties can be extremely difficult to solve.

Furthermore, current attempts at regulating AI such as the European Union´s AI regulation framework proposal may be motivated by rule utilitarianism as the seek to set a code of rules that regulates the use of AI in certain areas.[375] The framework will be discussed in-depth in Section 3.5. Nevertheless, other motives are combined in the framework as well.[376]

To conclude the critical comparison, **utilitarianism and especially rule utilitarianism will be the guiding principle of this thesis.** Due to its aim at maximizing utility via a set of rules it is the adequate framework to address the two main questions, this thesis is trying to answer. **Rule utilitarianism can be utilized well to develop a framework that examines if AI surveillance systems can be ethically permissible and a set of rules that guide AI engineers in designing morally acceptable AI surveillance solutions because it is based on a set of internalized moral rules.** The collapse into act utilitarianism can be prevented by designing these frameworks in an

[375] Cf. European Parliament (2022) P. 1ff.

[376] Cf. European Parliament (2022) P. 1ff.

optimal way that ensures the general acceptance of the frameworks that pose as set of rules and guidelines. The high capability of the modern human as the moral entity or agent leads to an optimal set of rules. Also, the general acceptance of the framework can already maximize utility in comparison to it not existing. Furthermore, rule utilitarianism counters the rule worship objection by actually endorsing the breaking of some rules if necessary. A fitting example was mentioned with the EU´s AI regulatory framework proposal. However, rule utilitarianism has some pitfalls, such as missing flexibility that results in rule-worship, that must be avoided while developing a framework that ensures ethically permissible AI surveillance solutions.

3 The AI framework of the thesis

> *Intelligence is central to what it means to be human. Everything that civilisation has to offer is a product of human intelligence.*[377]

After the previous chapter has dealt with the ethical framework of this thesis in-depth, this chapter is aimed at establishing the required understanding of AI, before attempting to assess if AI mass surveillance tools are usable in an ethical way and how this application may shape up. To achieve this the problems of defining AI will be discussed and a high-level definition of AI will be illustrated. Following this, a general understanding of what AI is will be inferred from this. Next, the vital connection between data and AI will be examined. Different types of data and how they link to AI will be discussed. The difference between automation and autonomy will be illustrated, supervised learning and unsupervised learning will be described. Additionally, machine learning and deep learning will be discussed. The technical basics of AI are going to be closely examined as well with a focus on backpropagation and gradient descent. Afterwards, the myriad of possible uses of an AI will be discussed. The moral challenges that the dawn of AI brings about will be studied after the possible uses have been illustrated. To conclude this section.

Computer scientists and engineers have trouble to agree on a common and holistic definition of AI.[378] This is partly owed to the complexity of AI and partly to the myriad of possible definitions of intelligence.[379] Furthermore, AI is intangible and can be quite abstract, which is a further hurdle to reaching a commonly agreed upon definition. In addition, another problem with defining AI lies in the grandiosity, that is usually connotated with AI, which creates unreasonable expectations and implies a high level of capability and autonomy.[380] Even the terminology is actively discussed by

[377] Cf. Hawking (2018) P. 180.

[378] Cf. Cummings (2017) P. 2.

[379] Cf. Cummings (2017) P. 2.

[380] Cf. Morgan et al. (2020) P. 8f.

computer scientists and engineers, with many voices arguing that the term computational intelligence is more precise.[381] However, as artificial intelligence is the most common term for the field, it will be used in this thesis.

For this thesis intelligence will be defined according to Robert J. Sternberg who defines intelligence as goal directed behavior that includes high degree of adaptability.[382] A general high-level definition of AI is that AI is intelligence exhibited by machines in contrast to natural intelligence which is presented by animals including humans.[383] A general understanding of AI then states that AI describes the ability of a computer system to successfully complete tasks that would normally require human intelligence.[384] AI gives technical systems the ability to perceive their environment and take the circumstances of the environment into account.[385] Additionally, the term does not account the development of technologies and their rising capabilities.[386] The seemingly straightforward definition mentioned before, actually captures this inconsistence by arguing that once a technology becomes common enough, the task that it performs does not require human intelligence anymore.[387] Therefore, the program that solves this task ceases to be an AI at this point.[388]

A very important distinction regarding AI is that between automation and autonomy. Autonomy is created, by an act of delegation of a specific task. Individuals either perform it on their own or delegate it to another individual or, in the case of AI, to another entity.[389] Focus on the case of delegation from an individual to an AI, that means the delegation of a specific task to an entity. This delegation assigns the entity the task without restraining it in regard to how the task needs to be done. The result of this is autonomy.[390] If the autonomy is restrained by a set of rules, depending

[381] Cf. Cummings (2017) P. 2.

[382] Cf. Sternberg (2000) P. 1ff.

[383] Cf. Poole et al. (1998) P. 1ff.

[384] Cf. Morgan et al. (2020) P. 8f.

[385] Cf. European Parliament (2020) P. 1ff.

[386] Cf. Morgan et al. (2020) P. 8f.

[387] Cf. Morgan et al. (2020) P. 8f.

[388] Cf. Morgan et al. (2020) P. 8f.

[389] Cf. Morgan et al. (2020) P. 23f.

[390] Cf. Morgan et al. (2020) P. 23f.

on the strictness of the rules, the result can be defined as automation.[391][392] For example, the entity could be bound by if/then/else programming statements that form a structure of rules.[393] This could lead to the outcome that an autonomous system that has been provided with the same input as an automated system, the autonomous system will come to different conclusions compared to the automated system.[394] Deviations from the predicted behavior are not defects, as in automated systems, but rather decisions made by the entity based on the available data of the situation.[395] This means that we can easily predict the outcome of an automized system in a specific situation.[396] Autonomy however, is closely connected with unpredictability as we cannot envisage how the entity will behave.[397][398] The conclusion from this fact is that autonomy is heavily associated with self-responsibility, negating the need for monitoring.[399] This evolution to fully autonomous systems means that the responsibility for an respective action shifts from the operators/creators of an entity to the entity itself.[400][401] Thus, for the creators of the technical system, the dependability of the system is vital.[402] The aspect of dependability is typically based on understanding the entity and its processes in detail, that describes the ability to trace, how and why the entity came to a specific decision.[403]

[391] Cf. Cummings (2017) P. 3ff.

[392] Cf. Morgan et al. (2020) P. 23f.

[393] Cf. Cummings (2017) P. 3ff.

[394] Cf. Cummings (2017) P. 3ff.

[395] Cf. Cummings (2017) P. 3ff.

[396] Cf. Adler (2019) P. 1ff.

[397] Cf. Cummings (2017) P. 3ff.

[398] Cf. Morgan et al. (2020) P. 9.

[399] Cf. Cummings (2017) P. 3ff.

[400] Cf. Cummings (2017) P. 3ff.

[401] Cf. Morgan et al. (2020) P. 9f.

[402] Cf. Cummings (2017) P. 3ff.

[403] Cf. Cummings (2017) P. 3ff.

3.1 Data as the heart of AI

The availability of data is a key to any AI and its training. Each step of developing an AI is closely related to the availability of data.[404] **In short, AI cannot exist without data**.[405] Large amounts of data are usually grouped into data sets for testing, evaluation and training which lead to deployment.[406] An AI begins with its conception and the identification of the problem/task it is supposed to solve.[407] The availability and the quality of the data determines the quality of the AI, which leads to the argument that data is the key driver behind AI development.[408]

In order to get the data, it must be sourced, which can be costly and difficult.[409] Furthermore, even in today´s data driven economy, data sourcing is often negatively connotated.[410] Issues of transparency, data protection and privacy often arise.[411] After the sourcing, data must be organized and refined, to assesses whether the sourced data is, regarding quantity and quality, sufficient for use in an AI application.[412] It is important to note that data quantity does not equal data quality, further methods of refinement or "data cleaning" may be needed before the data can be used in an AI.[413] Depending on the amount of data, this can be a time-consuming task. If the data covers, movement patterns for example, it is crucial that the movement patterns completely cover the respective area or all the individuals that are supposed to be monitored.

The organization and state of data are highly variable, generally data can be sourced in two states, structured and unstructured.[414] Structured data can be added to data models with the function of

[404] Cf. Sveinsdottir (2020) P. 9ff.

[405] Cf. Boucher (2020) P. 3ff.

[406] Cf. Boucher (2020) P. 3ff.

[407] Cf. Sveinsdottir (2020) P. 9ff.

[408] Cf. Boucher (2020) P. 3ff.

[409] Cf. Sveinsdottir (2020) P. 9ff.

[410] Cf. Sveinsdottir (2020) P. 9ff.

[411] Cf. Sveinsdottir (2020) P. 9ff.

[412] Cf. Sveinsdottir (2020) P. 9ff.

[413] Cf. Sveinsdottir (2020) P. 14ff.

[414] Cf. Sveinsdottir (2020) P. 14ff.

standardizing relations between data elements.[415] It can be found in databanks and is defined by the fact that the relationships between the elements originate from their position compared to other elements.[416] It is typically organized as quantitative data.[417] Additionally, structured data is often accompanied by a description of its structure and purpose.[418]

Unstructured data describes a state of data that is commonly referred to as “big data” and does not follow any organization according to a data model.[419] Unstructured data is commonly organized as qualitative data.[420] The unstructured data is unprocessed and may be the result of machine-led processes such as closed-circuit-television surveillance.[421] Closed-circuit-television surveillance footage for example has its own internal structure, e.g. timestamps, but does not have a defined relationship between its data elements.[422] In order to analyze unstructured datasets, the amount of preprocessing is to be done is higher than compared to structured datasets.[423] Unstructured data is much more common than structured data.[424]

Furthermore, data can be divided into different data types, according to certain criteria[425] These data types have a huge effect on how the potential AI might function, once it is fully developed. The first data type, that will be covered, is provided data. Provided data is data that, as the name suggests, is provided by individuals, who are aware of this provision and consent to it.[426] It is often highly personal and has a high degree of identifiability, meaning that access to it is likely to be restricted.[427] Oftentimes, provided data is gathered for a specific purpose and is therefore structured

[415] Cf. Sveinsdottir (2020) P. 14ff.

[416] Cf. Sveinsdottir (2020) P. 14ff.

[417] Cf. IBM (2021b) P. 1ff.

[418] Cf. Sveinsdottir (2020) P. 14ff.

[419] Cf. IBM (2021b) P. 1ff.

[420] Cf. IBM (2021b) P. 1ff.

[421] Cf. Sveinsdottir (2020) P. 14ff.

[422] Cf. Sveinsdottir (2020) P. 15ff.

[423] Cf. Sveinsdottir (2020) P. 15ff.

[424] Cf. IBM (2021b) P. 1ff.

[425] Cf. Sveinsdottir (2020) P. 15ff.

[426] Cf. Sveinsdottir (2020) P. 15ff.

[427] Cf. Sveinsdottir (2020) P. 15ff.

data.[428] Moving on from provided data, observed data is data that is gathered by observation.[429] Individuals are not necessarily aware of the data collection and may not be able to give their consent.[430] This observed data is produced and processed by AI surveillance applications. Thus, ethical problems may arise that will be the main subject of Section 4.4. The behaviors of individuals can be surmised in data sets which may be analyzed by an artificial neural network. Depending on the source, observed data can be structed or unstructured, issued with data quality may arise.[431] Observed data is a common output of surveillance applications such as closed-circuit-television cameras and facial recognitions software.[432] **As such, AI based mass surveillance tools deal, at least to a certain degree, with observed data.** This implies, that AI surveillance systems must have the capability to analyze structured and unstructured data.

Another type of data is derived data.[433] Derived data is gathered by processing and/or transformation of other data that has been made available by different sources.[434] Consequently, this means that data can be augmented to serve a use beyond its original intention.[435] In this sense, derived data can be used to gain insights from AI surveillance tools. Adding to derived data, inferred data is sourced by applying mathematical and/or statistical models to the available data to generate insights that can be used for predictive purposes.[436] An AI that is applied in surveillance operations therefore receives observed data which it can transform into derived data to generate insights about individuals, geographical movement patterns of individuals would be a possible example for these insights. The derived data can then be transformed into inferred data by the network to predict the behavior of individuals.

[428] Cf. Sveinsdottir (2020) P. 15ff.

[429] Cf. Sveinsdottir (2020) P. 15ff.

[430] Cf. Sveinsdottir (2020) P. 15ff.

[431] Cf. Sveinsdottir (2020) P. 15ff.

[432] Cf. Sveinsdottir (2020) P. 15ff.

[433] Cf. Sveinsdottir (2020) P. 15ff.

[434] Cf. Sveinsdottir (2020) P. 15ff.

[435] Cf. Sveinsdottir (2020) P. 15ff.

[436] Cf. Sveinsdottir (2020) P. 15ff.

AI training is typically conducted with training data, which is organized in a test set and a training set.[437] The training of an algorithmic model is ultimately realized via the ability of this network to generalize of the new data.[438] The training set serves the purpose of refining and training the AI in the application of different techniques to solve the defined problem or to reach its aim.[439] Quality and quantity of the available data should be at the highest possible level at this point, as any mistakes in the data drastically hinder the learning process of the network.[440] Any biases in the data should be rooted out as well.[441] The test set data is then used to evaluate the network and the learning process.[442]

The evaluation of the learning process is vital, as a lot of mistakes can be made in the relationship between the AI model and the data. In general, when evaluating an algorithmic model, the error should be minimized, thus meaning that the error of the generalization of the model should not bear any significance compared to the error when generalizing the model to unseen data. Overfitting is a common concept in this relationship that can effectively defeat the purpose of any AI.[443] Overfitting describes a situation in which a model fits perfectly against its training data.[444] Once this happens, the algorithm cannot operate truthfully against unseen datasets.[445] In opposition to this, a model can also be underfitted, if it has not been trained sufficiently or if the input variables do not prove significant enough to find a significant relationship between the input and output variables.[446] Therefore, underfitting also generalizes poorly against unseen data.[447]

[437] Cf. Sveinsdottir (2020) P. 15ff.

[438] Cf. IBM (2021a) P. 1ff.

[439] Cf. Sveinsdottir (2020) P. 15ff.

[440] Cf. Sveinsdottir (2020) P. 15ff.

[441] Cf. Sveinsdottir (2020) P. 15ff.

[442] Cf. Sveinsdottir (2020) P. 15ff.

[443] Cf. IBM (2021a) P. 1ff.

[444] Cf. IBM (2021a) P. 1ff.

[445] Cf. IBM (2021a) P. 1ff.

[446] Cf. IBM (2021a) P. 1ff.

[447] Cf. IBM (2021a) P. 1ff.

3.2 Artificial intelligence, machine learning and deep learning

This section will examine the vital distinction between AI, machine learning and deep learning. In the beginning of AI development, there was symbolic AI.[448] Symbolic AI is a collective description for approaches that hardcode the required knowledge and experience into an AI.[449] This was typical for the early stage of AI development up until the end of the 1990s.[450] Symbolic AI is still widely in use today, as it is relatively simple while still producing desired outcomes.[451] However, symbolic AI is not the state of the art. In the early 2000s, symbolic AI was succeeded by machine learning and data-driven AI.[452] Machine learning describes the general process how a computer can learn from data.[453] The outcome of a machine-learning process is the execution of a specific task by an algorithm that was not explicitly programmed to do so.[454] The computer system recognizes patterns in the data and performs predictions on these recognized patterns, with statistics being the driving force behind machine learning.[455] As mentioned before, AI is a complex system that has the ability to mimic human cognitive capabilities. In order to achieve this ability, AI has many subsets that each deal with an element of its intelligence.[456][457][458] Machine learning is a subset of AI, thus enabling a computer system to improve itself based on experience.[459] Other subsets of AI include Planning, general intelligence, social intelligence, perception, knowledge representation, logical

[448] Cf. Boucher (2020) P. 2.

[449] Cf. Boucher (2020) P. 2.

[450] Cf. Boucher (2020) P. 2.

[451] Cf. Boucher (2020) P. 2.

[452] Cf. Boucher (2020) P. 2.

[453] Cf. Microsoft (2021) P. 1.

[454] Cf. Wolfewicz (2021) P. 1ff.

[455] Cf. Wolfewicz (2021) P. 1ff.

[456] Cf. Poole et al. (1998) P. 1ff.

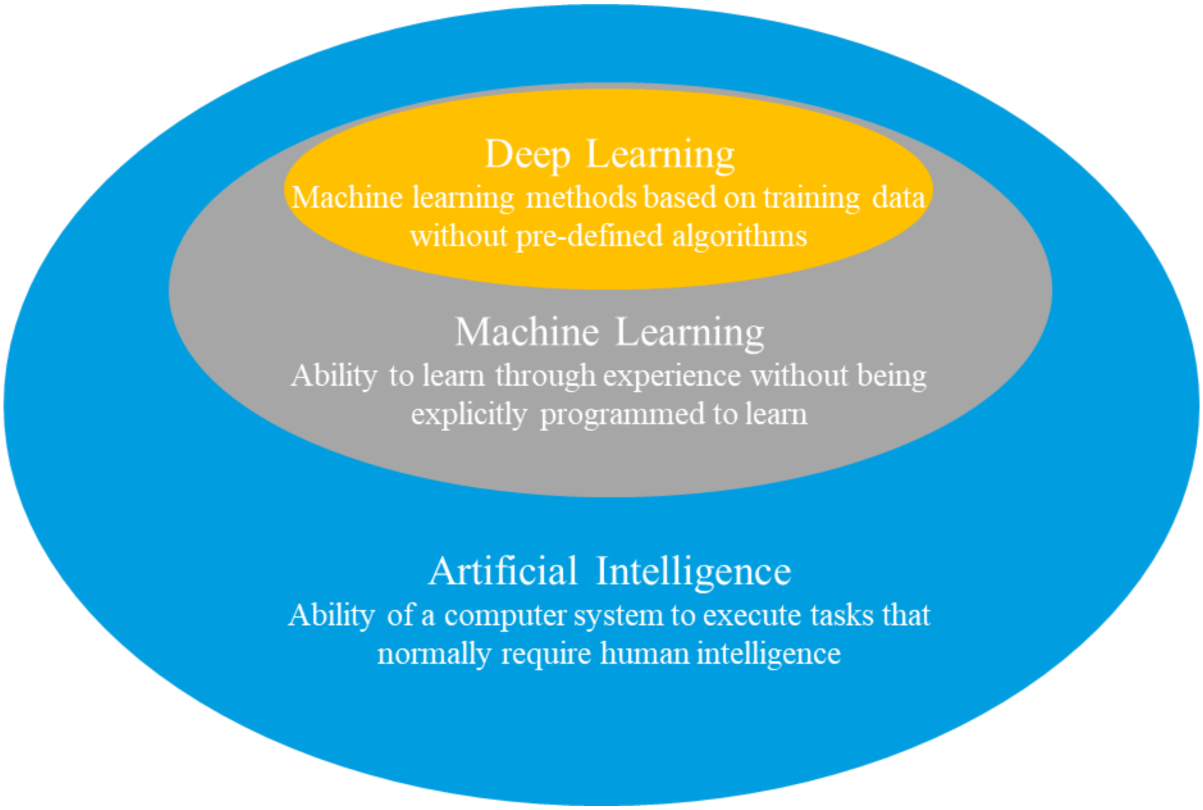
[457] Cf. Nilsson (1998) P. 1ff.

[458] Cf. Russell and Norvig (2003) P. 1ff.

[459] Cf. Wolfewicz (2021) P. 1ff.

problem-solving and robotics.[460][461][462][463] Deep learning is a subsection of machine learning which enables a machine to learn based on training data without any predefined algorithms.[464] Figure 1 shows the classification of AI processes.

Figure 1 Classification of AI, machine learning and deep learning


Deep Learning
Machine learning methods based on training data without pre-defined algorithms
Machine Learning
Ability to learn through experience without being explicitly programmed to learn
Artificial Intelligence
Ability of a computer system to execute tasks that normally require human intelligence

Source: Own illustration according to Morgan et al. (2020), P.10

Machine learning can be clustered into supervised learning and unsupervised learning, reinforcement learning and hybrid learning.[465] Artificial neural networks can be utilized for these purposes.[466] Artificial neural networks primarily consist of simple interconnected processing elements,

[460] Cf. Russell and Norvig (2003) P.1ff.

[461] Cf. Luger and Stubblefield (1993) P. 1ff.

[462] Cf. Nilsson (1998) P. 1ff.

[463] Cf. Poole et al. (1998) P. 1ff.

[464] Cf. Morgan et al. (2020) P. 9f.

[465] Cf. Delua (2021) P. 1ff.

[466] Cf. IBM (2020) P. 1ff.

units or nodes. They are loosely based on the structure of the human brain.[467][468] The neurons and their electro-chemical connections of the human brain are mirrored by classification algorithms called perceptron, that mimic their function.[469] The processing ability of the network stems from the strength of connections between the units or nodes, typically regarded as weight.[470] These weights are obtained and strengthened from a process of adaptation or learning derived from a dataset.[471] Artificial neural networks can be structured feedforward or recurrent.[472] Feedforward neural networks show the characteristic that the units or nodes are not structured in a cycle, there is only one direction in which the data can move through the nodes.[473] Therefore, there is no memory within the artificial neural network.[474][475] In contrast to this, recurrent neural networks show connections between nodes or units that shape up in a graph along a temporal sequence which enables recurrent neural networks to exhibit dynamic behavior, because the data goes through a loop.[476][477][478]

Similar to the human brain, the input needs to have a certain strength in order to activate the artificial neural network. This can be achieved via a binary classifier, which is utilized as a classification algorithm.[479] A binary classifier is a function that answers the question, whether an input, typically expressed via a vector of numbers, can be sorted into a specific class.[480] Originally, the perceptron algorithm was developed in 1958 by the American psychologist Frank Rosenblatt.[481] Albeit, that it

[467] Cf. Boucher (2020) P. 14ff.

[468] Cf. Boucher (2020) P. 14ff.

[469] Cf. Weiß (2021) P. 228.

[470] Cf. Gurney (2014) P. 13ff.

[471] Cf. Gurney (2014) P. 13ff.

[472] Cf. Eliasy and Przychodzen (2020) P. 5ff.

[473] Cf. Eliasy and Przychodzen (2020) P. 5ff.

[474] Cf. Eliasy and Przychodzen (2020) P. 5ff.

[475] Cf. Zell (1996) P. 73.

[476] Cf. Eliasy and Przychodzen (2020) P. 5ff.

[477] Cf. Zell (1996) P. 73.

[478] Cf. Tealab (2018) P. 334ff.

[479] Cf. Gurney (2014) P. 13ff.

[480] Cf. Freund and Schapire (1999) P. 2.

[481] Cf. Freund and Schapire (1999) P. 2.

is a simple algorithm, it or its evolutions are still widely utilized in deep learning networks.[482] Similar to a biological neuron, a perceptron is activated, causing electrical activity, by a stimuli from another perceptron, meaning that a stimuli is transferred from perceptron to perceptron.[483][484] Yet, it is not set in stone a stimuli is transferred by a perceptron, as a perceptron is a decision rule, that is defined with a mathematical expression.[485] This mathematical expression can tell us, if the perceptron is activated or not.[486]

In the artificial neural network, inputs are translated into signals that pass through a network of perceptrons to generate useful outputs.[487] The stimuli first arrive in the input layer and are processed in the hidden layers of the network.[488] To stick with the aforementioned examples of self-driving cars, traffic signs would need be incorporated into the neural network, so that the car can act correctly to the signs.[489] A STOP sign for example, would be incorporated into the neural network via feature engineering.[490] The hidden layers of the neural network would detect the features of the STOP sign and generate the correct output.[491] The artificial neural network is, however, not aware of itself, it is a simplified model of the brain that is able to tell us what it thinks a STOP sign is.[492] This shows that the structure of the artificial neural network is vital, as the number of hidden layers is proportional to the degree of abstractness that the conceptualizations of the artificial neural network can achieve.[493]

[482] Cf. Freund and Schapire (1999) P. 2.

[483] Cf. Freund and Schapire (1999) P. 2.

[484] Cf. Weiß (2021) P. 228.

[485] Cf. Weiß (2021) P. 229.

[486] Cf. Weiß (2021) P. 229.

[487] Cf. Boucher (2020) P. 2ff.

[488] Cf. Boucher (2020) P. 2ff.

[489] Cf. Wolfewicz (2021) P. 1ff.

[490] Cf. Wolfewicz (2021) P. 1ff.

[491] Cf. Wolfewicz (2021) P. 1ff.

[492] Cf. Boucher (2020) P. 2ff.

[493] Cf. Boucher (2020) P. 2ff.

Supervised learning is a method of machine learning that utilizes labeled datasets.[494] These datasets practically supervise the algorithms of the machine learning application.[495] In supervised learning, algorithms require input labels, which may be used to track the performance of a specific algorithm.[496] Structured data is often used in supervised learning or in artificial neural networks with limited capabilities.[497] Bear in mind that supervised learning specifically requires structured data. It excels at inferring or finding relationships between data elements.[498] Supervised learning can be split into two categories, classification and regression.[499] Classification aims at training algorithms to accurately sort data into specific categories, such as classifying individuals on surveillance footage by their optical properties.[500] Furthermore, regression is a type of supervised learning that examines the relationship between dependent and independent variables via an algorithm for example.[501] Regression models are best fitting for predicting numerical values based on existing data.

In contrast to supervised learning, unsupervised learning applies algorithms to analyze and sort unlabeled data sets with the goal of discovering hidden patterns and relationships in the data without requiring human assistance and/or human intervention.[502] It is a technique to gain meaningful insight from unstructured data, which is often referred to as “big data”.[503] Unsupervised learning applications are commonly applied for three main tasks: clustering, association and dimensionality

[494] Cf. Delua (2021) P. 1ff.

[495] Cf. Delua (2021) P. 1ff.

[496] Cf. Weiß (2021) P. 10ff.

[497] Cf. Sveinsdottir (2020) P. 14ff.

[498] Cf. Sveinsdottir (2020) P. 14ff.

[499] Cf. Delua (2021) P. 1ff.

[500] Cf. Delua (2021) P. 1ff.

[501] Cf. Delua (2021) P. 1ff.

[502] Cf. Delua (2021) P. 1ff.

[503] Cf. Sveinsdottir (2020) P. 14ff.

reduction.[504] Clustering groups unlabeled data focused on their parallels or disparities.[505] Association uses different rules to identify relationships between variables in a specific dataset.[506] Dimensionality reduction is applied when the number of features (dimensions) in a dataset is too high, reducing the data input into a manageable size at the same time as preserving data quality.[507]

A feedforward artificial neural network can be trained under supervision by analyzing the difference between the input and the output.[508] This difference is regarded as an error that must be minimized, as the minimization of the error infers that the artificial neural network has improved.[509] In feedforward neural networks this can be achieved via an algorithm that is called back propagation.[510] The operation of back propagation aims at correcting the error by propagating backwards through the artificial neural network.[511] Back propagation tries to determine the gradient of the error function while incorporating the weights of the artificial neural network.[512] The gradient of the function describes its direction of its steepest descent.[513] Afterwards, the perceptrons are adapted, that means that the weights are optimized to reduce the error, with the correction process starting at the output layer.[514][515] This process can be only examined using mathematics.[516] Another approach to find the error could be, theoretically, to generate a set of artificial neural networks with every possible combination of perceptrons, and to test these combinations against each other with

[504] Cf. Delua (2021) P. 1ff.

[505] Cf. Delua (2021) P. 1ff.

[506] Cf. Delua (2021) P. 1ff.

[507] Cf. Delua (2021) P. 1ff.

[508] Cf. Delua (2021) P. 1ff.

[509] Cf. Boucher (2020) P. 6ff.

[510] Cf. Boucher (2020) P. 5ff.

[511] Cf. Boucher (2020) P. 5ff.

[512] Cf. Goodfellow et al. (2016) P. 1ff.

[513] Cf. Weiß (2021) P. 272.

[514] Cf. Weiß (2021) P. 268ff.

[515] Cf. Boucher (2020) P. 6ff.

[516] Cf. Boucher (2020) P. 6ff.

a preset of labeled data.[517] Practically however, the high number of possible combinations renders this approach unfeasible.[518]

A more selective and smarter approach is the gradient descent.[519] This method calculates the errors via an algorithm of numerical mathematics. The goal is to start with arbitrary weights and change them to minimize the error between input and output.[520] To achieve this, the direction of movement is the direction of the steepest descent of the error function.[521] The learning rate governs the intensity of the movement.[522] It is a configurable hyperparameter, typically having a small positive value between 0 and 1.[523] The learning rate dictates how quickly the model can adapt to a specific problem.[524] The challenge in configuring the learning rate correctly, lies in its importance for the training of the model.[525] A learning rate that is too high can cause the model to come to suboptimal solutions, while a learning rate that is too low can cause the learning process to be stuck.[526] In order to make the results of this more tangible to human operators, they can be plotted into a chart, which is called error landscape because the altitude of the error representing its gradient.[527] This error landscape aims to find the global optimum, in which the error is minimal, yet this is the optimal outcome that is not guaranteed in any way.[528] Figure 2 shows an exemplary error landscape.

[517] Cf. Boucher (2020) P. 6ff.

[518] Cf. Boucher (2020) P. 6ff.

[519] Cf. Boucher (2020) P. 6ff.

[520] Cf. Weiß (2021) P. 273.

[521] Cf. Weiß (2021) P. 273.

[522] Cf. Weiß (2021) P. 273.

[523] Cf. Brownlee (2019) P. 1ff.

[524] Cf. Brownlee (2019) P. 1ff.

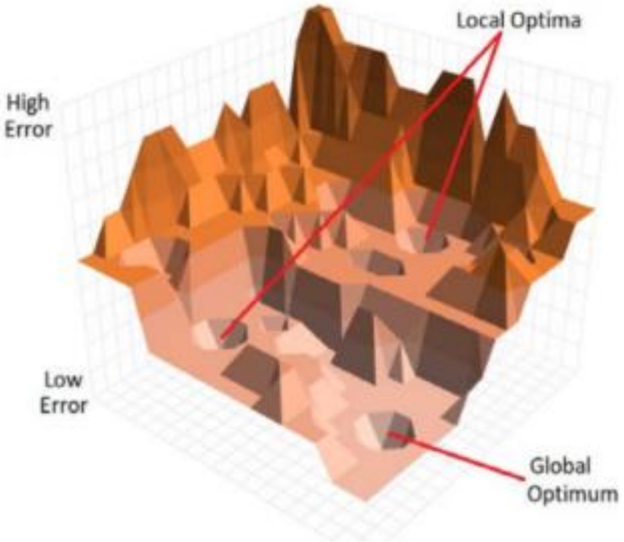
[525] Cf. Brownlee (2019) P. 1ff.

[526] Cf. Brownlee (2019) P. 1ff.

[527] Cf. Boucher (2020) P. 6ff.

[528] Cf. Boucher (2020) P. 6ff.

Figure 2 Exemplary Error Landscape



Source: Boucher (2020), P.6

A good example to understand how an AI learns from the gradient descent is that of a hiker that is trying to find his way back into the valley.[529] However, it is foggy so that the hiker cannot see more than one meter into every direction, so he has to scan his surroundings to find the steepest descent.[530] Once found, he moves into that direction and arrives one plateau closer to the valley where he repeats the process.[531] In the same way an artificial neural network can be generated from a random point in the error landscape, which error is already calculated.[532] Additionally, the error of possible adjustments to the artificial network is also known, this means that the adjustment that offers the best improvement is known as well.[533] This adjustment is implemented and the process is repeated at the next perceptron, enabling the artificial neural network to gradually improve itself.[534] Please note that other mechanics of supervised learning exist, such as decision trees which

[529] Cf. Boucher (2020) P. 6ff.

[530] Cf. Boucher (2020) P. 6ff.

[531] Cf. Boucher (2020) P. 6ff.

[532] Cf. Boucher (2020) P. 6ff.

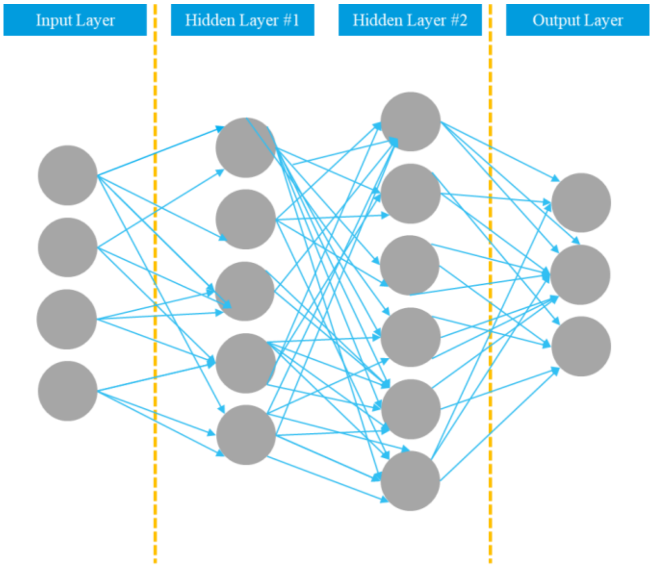
[533] Cf. Boucher (2020) P. 6ff.

[534] Cf. Boucher (2020) P. 6ff.

can be culminated in random forests, linear systems, regressions, classifications and binary tagging.[535][536]

Deep learning processes are mathematically complex algorithms, that may be regarded as the evolution of machine learning.[537] Similar to the relationship of AI and machine learning, deep learning is a subset of machine learning.[538] They analyze date similar to how a human analyzes data, via layered structures of algorithms that form the aforementioned neural network.[539] Figure 3 illustrates an example of how a deep learning neural network may look.

Figure 3 Example structure of a deep learning neural network



Source: Own illustration according to Wolfewicz (2021), P.3

The input and output layer of the neural network are the observable processes of the whole network.[540] Hidden layers may be characterized as the engine room of the network, meaning that the calculations of the algorithms mainly happen in these layers.[541] The higher the number of hidden

[535] Cf. Weiß (2021) P. 154 ff.

[536] Cf. Zhang et al. (2022) P. 24ff.

[537] Cf. Wolfewicz (2021) P. 1ff.

[538] Cf. IBM (2020a) P. 1ff.

[539] Cf. Wolfewicz (2021) P. 1ff.

[540] Cf. Wolfewicz (2021) P. 1ff.

[541] Cf. Wolfewicz (2021) P. 1ff.

layers, the deeper the network.[542] Typically, a neural network may be considered a deep neural network if it possesses at least two hidden layers.[543] To conclude, the main difference between machine learning and deep learning is the need for human intervention in the leaning process and in the labeling of the data beforehand and the data requirements.[544] Deep learning networks require a much higher amount of data, while being able to operate with less human intervention than a machine learning network.[545]

Furthermore, machine learning is not limited to supervised and unsupervised learning. Other types of machine learning are reinforcement learning and hybrid learning. Reinforced learning is usually conducted in a dynamic environment in which the algorithm, that is typically called agent, follows a predefined aim.[546][547] On the way of reaching this aim, the network constantly adapts itself to the dynamic environment via a reward and punishment approach.[548] The last approach is hybrid learning, combining supervised and unsupervised learning.[549] This approach may be used to create a hybrid deep neural network, which mimics the function of the human brain more precisely than other networks.[550]

Another significant influence on AI is probability, chance impacts AI.[551] As such, probability is also closely connected to machine learning and deep learning.[552] As AI often aims to rationally predict future data and thereby future events, these models are based on assumptions and this is where uncertainty comes into play as any model will face uncertainty when predicting future outcomes.[553] This uncertainty can materialize in many forms, for example blurry surveillance footage

[542] Cf. Wolfewicz (2021) P. 1ff.

[543] Cf. Wolfewicz (2021) P. 1ff.

[544] Cf. Wolfewicz (2021) P. 1ff.

[545] Cf. Wolfewicz (2021) P. 1ff.

[546] Cf. Weiß (2021) P. 11.

[547] Cf. Fumo (2017) P. 2ff.

[548] Cf. Weiß (2021) P. 11.

[549] Cf. Mathew et al. (2021) P. 4ff.

[550] Cf. Mathew et al. (2021) P. 4ff.

[551] Cf. Ghahramani (2015) P. 1ff.

[552] Cf. Ghahramani (2015) P. 1ff.

[553] Cf. Ghahramani (2015) P. 1ff.

or faulty parameters of a regression model can drastically increase the amount of uncertainty a model has to deal with.[554] Additionally, uncertainty about the used model itself also exists.[555] Probability distributions can be used to represent every uncertain unobserved quantity in a model and how they interact with the data.[556] Next, the basic rules of probability theory are applied to conclude the unseen quantities from the existing data.[557]

In the future, AI may be paired with robotics.[558] This may lead to robots that can conduct difficult and dangerous tasks.[559] Albeit, that robotics is not a field of AI, there are still a lot of synergies that may be reaped by a combination of the technologies.[560] note that this is also the path to lethal autonomous weapon systems.[561][562] Moreover, AI paired with quantum computing could reap the simultaneity of quantum computing to elevate processing power to new heights.[563]

3.3 Some possible uses of AI

Apart from using AI in mass surveillance applications, countless other uses of the technology are feasible. Humans have always strived to improve the quality of their lives across all its facets, AI will not be an exemption from this.[564] AI will experience an evolution, becoming more and more fluid while adding human qualities to its algorithms. In the future, quantum computing can be the paired with AI to solve the most pressing problems and reveal the greatest mysteries humanity encounters such as climate change, diseases, war, poverty, famine, deep-space exploration and maybe even the origins of our universe. Therefore, it is vital to understand the enormous potential

[554] Cf. Ghahramani (2015) P. 1ff.

[555] Cf. Ghahramani (2015) P. 1ff.

[556] Cf. Ghahramani (2015) P. 1ff.

[557] Cf. Ghahramani (2015) P. 1ff.

[558] Cf. Boucher (2020) P. 14ff.

[559] Cf. Boucher (2020) P. 14ff.

[560] Cf. Boucher (2020) P. 14ff.

[561] Cf. Hawking (2018) P. 183ff.

[562] Cf. Boucher (2020) P. 14ff.

[563] Cf. Boucher (2020) P. 14ff.

[564] Cf. Adams (2017) P. 1ff.

of AI. While the aforementioned examples may suffer from grandiosity and exaggeration, AI can have a myriad of possible uses, which will be briefly outlined in this section.

When the term AI is used, the imagination often refers to embodied AIs, with self driving cars being a prominent example of this.[565] Additionally, drones and robots are possible examples.[566] In contrast to embodied AIs, AI can also be a software, such as virtual assistants, pattern recognition software and search engines.[567] Popular examples of today are the personal assistants sold by Apple and Amazon. Albeit, those assistants are just the first step for the development and are therefore not truly AIs but advanced machine-learning algorithms that rely on deep learning.[568] The number of possible uses of AI is incredibly high and has the potential to have a huge impact on our everyday life.[569] Yet, AI is not a futuristic dream, AI based applications are already in use, the extent of this use will be analyzed at a later stage. Today, AI tools, or applications that contain at least some aspects of AI, are for example commonly used in advertising, web search applications, personal assistants, translators, smart homes, infrastructure, cars and cybersecurity.[570] Self-driving cars typically use the aforementioned deep learning networks to detect obstacles, traffic signs and other road users.[571] AI also played a role in the global effort to fight the spread of the Covid-19 pandemic by assisting in generating and analyzing thermal pictures in airports and other public places.[572]

Furthermore AI may help in the diagnosis of an infection by examining computerized tomography lung scans and by tracking statistical data regarding the spread of the disease.[573] Healthcare is one of the biggest potential fields of operation for AI.[574] Due to its excellence in dealing with huge quantities of data, AI may be able to discover new patterns that lead to breakthroughs in modern

[565] Cf. European Parliament (2020) P. 1ff.

[566] Cf. European Parliament (2020) P. 1ff.

[567] Cf. European Parliament (2020) P. 1ff.

[568] Cf. Adams (2017) P. 1ff.

[569] Cf. European Parliament (2020) P. 1ff.

[570] Cf. European Parliament (2020) P. 1ff.

[571] Cf. Wolfewicz (2021) P. 1ff.

[572] Cf. European Parliament (2020) P. 1ff.

[573] Cf. European Parliament (2020) P. 1ff.

[574] Cf. European Parliament (2020) P. 1ff.

medicine.[575] Moreover, AI can increase the quality of emergency response, by for example quickly identifying the danger of cardiac arrest of a patient.[576] Bear in mind that, the use of AI for medical purposes is also a topic that must be assessed for its ethical implications. In manufacturing, AI has the potential of increasing productivity and efficiency while lowering health hazards for workers.[577] In the agricultural industry, AI can help in creating a sustainable nutrition system by for example minimizing the need for fertilizers and pesticides.[578] AI-powered robots could identify and remove weeds and insects, lowering the need for potentially harmful chemicals.[579][580] Additionally, AI may be utilized for military purposes, autonomous systems may change the face of future wars.[581] Typical military applications of AI include image recognition, various analyses and autonomous weapons.[582] AI has a huge potential in economics as well. Major levers, that AI can utilize, are automation and data analytics.[583] Repetitive activities may vanish, and further insights may be generated.[584] Furthermore, AI applications in business can predict customer and market behaviour and developments, generating additional insights for a specific company.[585] On top of this, AI can tailor advertisements and other marketing communications to certain individuals, improving their success rate.[586] AI will penetrate every aspect of economics and of our society.[587] To conclude, AI can significantly augment our intelligence and enable us to solve the biggest challenges in almost all aspects of life.[588]

[575] Cf. European Parliament (2020) P. 1ff.

[576] Cf. European Parliament (2020) P. 1ff.

[577] Cf. European Parliament (2020) P. 1ff.

[578] Cf. European Parliament (2020) P. 1ff.

[579] Cf. European Parliament (2020) P. 1ff.

[580] Cf. Adams (2017) P. 1ff.

[581] Cf. Morgan et al. (2020) P. 8ff.

[582] Cf. Morgan et al. (2020) P. 9f.

[583] Cf. Kleinings (2022) P. 1ff.

[584] Cf. Kleinings (2022) P. 1ff.

[585] Cf. Kleinings (2022) P. 1ff.

[586] Cf. Kleinings (2022) P. 1ff.

[587] Cf. Hawking (2018) P. 183ff.

[588] Cf. Hawking (2018) P. 183ff.

3.4 Some ethical challenges of AI

The aforementioned uses of AI create their own ethical problems and dilemmas. Some of these ethical problems are general to AI, some a specific to certain applications. Ethical challenges that are specific to AI surveillance systems will be discussed in Section 4.4. The major ethical challenges of AI in general will be the subject of this section, examining the possible pitfalls of this technology and what it could mean for human life.

The first major ethical issue of AI, that will be discussed, is unemployment, often described as the disappearance of jobs.[589] The history of work has always been closely associated with automation, as the hierarchy of work is connected with the complexity of the occupied role.[590][591] The physical work of the pre-industrial age has slowly but steadily been substituted by the cognitive labour of today.[592] AI could jumpstart a similar evolution, that leads to a high unemployment rate.[593] The incredible potential of AI for automation can substitute repetitive jobs.[594][595] AI-based surveillance for example, could substitute the jobs of the private security industry. If no one must monitor the footage of closed-circuit-television cameras, as this may be done by an AI, these jobs may vanish. This raises the question, how individuals are going to spend their time. Most of us are still selling their time in order to generate income.[596] If this becomes impossible due to AI, huge ethical problems arise.[597] Similar to jobs in the private security sector, truck drivers around the globe could be rendered useless by the emergence of autonomous vehicles.[598] These to examples are systematic for the whole development of the future of work. Every sector of the economy and society will be penetrated by AI as examined in Section 3.3. What do we do with masses of people that create no viable use for our society? Furthermore, what can the people do, that lose their jobs due to AI?

[589] Cf. Stahl (2021) P. 39ff.

[590] Cf. Harari (2016) P. 413ff.

[591] Cf. Bossmann (2016)P. 1ff.

[592] Cf. Harari (2016) P. 301ff.

[593] Cf. Bossmann (2016) P. 1ff.

[594] Cf. Hawking (2018) P. 188ff.

[595] Cf. Pazzanese (2020) P. 1ff.

[596] Cf. Bossmann (2016) P. 1ff.

[597] Cf. Bossmann (2016) P. 1ff.

[598] Cf. Bossmann (2016) P. 1ff.

Furthermore, what would it mean if a part of a society loses its usefulness for the society? If we keep in mind what was discussed in Section 2.2.1, we must ask ourselves what utilitarianism would judge as ethically permissible in this case. This development creates a multitude of challenges that penetrates into the foundations of our society.

Another ethical problem created, or at least increased, by AI is inequality.[599] The majority of companies still pay their workers by the hour.[600] The use of AI may enable companies to drastically reduce their reliance on human workforce, leading to the fact that the individuals with ownership of AIs will earn a majority of the money in our economy.[601][602] Even today, a widening wealth gap can be observed, a development that will only be advanced by the dawn of AI.[603]

Additionally, it remains to be seen, how AI may affect our behaviour and interaction between each other.[604] AI based applications are becoming more and more efficient in mimicking human behaviour and modelling human conversations and relationships.[605][606] In 2015, a bot already cracked the Turing challenge, convincing more than half of the humans who interacted with it, that it was human.[607] Interaction with artificial bots will surely increase in the future, as they are able to pour virtually unlimited resources into building relationships and creating trust.[608] AIs are already able to activate the reward center of the human brain, in the future it may be able to direct human behaviour via subtle nudges to induce desired behaviour.[609] This possible application paired with the insights AI-based mass surveillance can deliver about individuals can lead to the abuse of powers by governmental agencies, the appearance of biases in AI surveillance predictions, the manipulation of individuals based on the insights the AI generates, the increasing risk of data breaches, th

[599] Cf. Stahl (2021) P. 39ff.

[600] Cf. Bossmann (2016) P. 1ff.

[601] Cf. Harari (2016) P. 431ff.

[602] Cf. Larbey et al. (2020) P. 9ff.

[603] Cf. Harari (2016) P. 413ff.

[604] Cf. Stahl (2021) P. 39ff.

[605] Cf. Bossmann (2016) P. 1ff.

[606] Cf. UNESCO (2019) P. 10ff.

[607] Cf. Bossmann (2016) P. 1ff.

[608] Cf. Bossmann (2016) P. 1ff.

[609] Cf. Bossmann (2016) P. 1ff.

looming singularity and the neutralization of the anonymity of the internet. Further details of these ethical challenges of AI surveillance technology will be discussed in Section 4.4.

Furthermore, how do protect ourselves from the mistakes an AI can still make? As an artificial neural network learns, for example but not exclusively via a combination of backpropagation and the gradient descent, it is bound to make mistakes.[610] AI, with all its advantages, can be fooled in ways, humans could never be misled by manipulating the inputs.[611][612] Looking at the aforementioned example of STOP signs, AI can fail to correctly identify the STOP sign, if is sightly covered by stickers or graffiti.[613] Furthermore, facial recognition systems can currently be fooled by colorful glasses.[614] Moreover, individuals could manipulate it to reach their own goals and desires.[615] To add to this, AIs are created by humans and therefore may show the same racial or non-racial biases that humans do.[616] This inheritance of biases will also be a topic in this section. In the field of surveillance, where the output of the AI may have huge implications for the life of an individual, these problems turn into a huge ethical challenge. In addition, AI as a product of human engineering, is not free from errors that are not dictated by the outputs or the manipulation of the inputs but by faulty AI engineering.[617][618] The scope of the AI solution dictates its size, a bigger scope of the AI solution will lead to the dramatic increase of negative effects that stem from ostensibly small error rates.[619] These negative effects can lead to faulty outputs of AI.[620] A variety of reasons can be responsible for those faulty outputs, moreover AI systems rely on probabilistic elements, which add uncertainty to the outputs of these elements.[621] Also, the an overall insufficient data quality

[610] Cf. Stahl (2021) P. 39ff.

[611] Cf. Bossmann (2016) P. 1ff.

[612] Cf. Goodfellow et al. (2014) P. 1ff.

[613] Cf. Eykholt et al. (2017) P. 1ff.

[614] Cf. Sharif et al. (2016) P. 1528ff.

[615] Cf. Bossmann (2016) P. 1ff.

[616] Cf. Bossmann (2016) P. 1ff.

[617] Cf. Human Rights Council (2021).

[618] Cf. Stahl (2021) P. 39ff.

[619] Cf. Human Rights Council (2021).

[620] Cf. Human Rights Council (2021).

[621] Cf. Human Rights Council (2021).

can be responsible for faulty outputs of AI.[622] Besides, unrealistic expectations and unclear definitions of the goals that an AI is supposed to accomplish, can lead to the utilization of AI systems that are not fit to reach these predefined goals.[623]

AI also raises a challenge to security and peace around the world.[624] This is not limited to lethal autonomous weapons systems but also the damage AI can cause in information warfare and cyber warfare.[625] How do we ensure that humans are still responsible for the actions of an AI? And what do we do, if AI comes to conclusions that are detrimental to human interests? Bear in mind that this does not mean that an AI acts malicious, moreover, it means that the machine solved its task in a way that was not predicted or intended by its creators or that its goals are not in line with ours.[626] For surveillance purposes, this could mean that an AI tasked with lowering the crime rate in a respective area could come to the conclusion that the whole area must be on lock down, recommending to kill everyone who breaks that lockdown.[627] The AI would have completed its task of lowering the crime rate but not in the way its creators would have intended.

The problem of AI security goes hand in hand with the issue of singularity.[628] Human intelligence and ingenuity has brought us to the top of the food chain, enabling us to construct tools that counter our disadvantage in physical prowess.[629] Adding to this, humans can adapt and learn via cognitive processes which leads to bigger and better tools, that allow us to be the dominant species on earth.[630] According to Moore´s Law, computers amplify their speed and memory capacity by a factor of two every eighteen months.[631] It therefore should not be surprising that AI will very likely outsmart us at some point in the upcoming century.[632] This raises the issue of singularity, that

[622] Cf. Human Rights Council (2021).

[623] Cf. Human Rights Council (2021).

[624] Cf. Stahl (2021) P. 39ff.

[625] Cf. Bossmann (2016) P. 1ff.

[626] Cf. Hawking (2018) P. 183ff.

[627] Cf. Bossmann (2016) P. 1ff.

[628] Cf. Stahl (2021) P. 39ff.

[629] Cf. Bossmann (2016) P. 1ff.

[630] Cf. Bossmann (2016) P. 1ff.

[631] Cf. Hawking (2018) P. 183ff.

[632] Cf. Hawking (2018) P. 183f.

describes the advent of an AI that is more capable than its human creators, an AI that outsmarts us.[633] The potential of an AI to learn and improve itself could lead to an exponential rise in artificial intelligence ultimately resulting in machines that exceed our cognitive abilities by a huge margin.[634] An AI tasked with mass surveillance that becomes independent, and has goals that are not in line with the goals of its creators, from its human creators is a terrifying thought.

Furthermore, AI may start a new arms race.[635] Autonomous weapon systems can choose and eliminate their own targets, creating huge ethical challenges as no human is responsible for the actions of the AI.[636] The consequences of such an arms race may be disastrous.[637] Especially the combination of lethal autonomous weapons systems and surveillance AIs in authoritarian states could be extremely problematic.

Another ethical challenge AI inherits from its human creators is bias.[638][639] Biases are a well-documented aspect of human thinking patterns, presenting thought patterns that can skew a decision towards a certain group of people that show distinctive traits.[640][641] These biases can heavily affect the outcomes of human decisions, and lead to harmful results.[642] Due to the continuing advancements of AI in the last decades, the question in how far human biases find their way into AI systems has gained considerable weight.[643] These biases can be against a specific group of people which leads to a racially biased AI system [644] In the context of surveillance technology a bias could lead to racially motivated recommendations of the AI surveillance system which leads to the wrongful overrepresentation of specific ethnic groups in the findings of the system. Yet, AI has the ability

[633] Cf. Stahl (2021) P. 39ff.

[634] Cf. Hawking (2018) P. 183ff.

[635] Cf. Hawking (2018) P. 183ff.

[636] Cf. ICRC (2021) P. 1ff.

[637] Cf. ICRC (2021) P. 1ff.

[638] Cf. Manyika et al. (2019) P. 1ff.

[639] Cf. Human Rights Council (2021) P. 3ff.

[640] Cf. Mehrabi et al. (2019) P. 1ff.

[641] Cf. Manyika et al. (2019) P. 1ff.

[642] Cf. Manyika et al. (2019) P. 1ff.

[643] Cf. Manyika et al. (2019) P. 1ff.

[644] Cf. Mehrabi et al. (2019) P. 1ff.

to counter human biases but it can just as well amplify these biases by deploying them in scale in areas where its application is sensitive.[645] Bias can find its way into algorithms and then into AI systems by a number of ways.[646][647] The datasets used to train the AI can contain biased human data of reflect historical inequalities, this can also happen due to faulty data sourcing.[648] Common biases from data are the measurement bias, the omitted variable bias, the representation bias, the aggregation bias, the sampling bias, the longitudinal data fallacy and the linking bias.[649][650] The measurement bias arises from how we choose to measure particular features, owing its name to a faulty weighting of certain structures in a dataset.[651] While the omitted variable bias is born when one or more variables that are vital for a data model are left out.[652] Another bias is the representation bias, that comes from misconceptions in the way the data about a population is sourced.[653] The last typical bias, the aggregation bias, arises when wrong conclusions about individuals are drawn from observing the entire populace.[654] Besides data as the source of bias, AI system can also exhibit biases due to design choices, even if the training set contains unbiased data.[655] The result of biased AI systems is that existing biases in the data are enlarged and perpetuated, as the outcomes of the algorithms that form the AI further fuel the existing bias.[656] This feedback loop between data, the algorithm and user interaction leads to a myriad of biases.[657] User behavior is in turn also strongly modulated by algorithms, meaning that biases can perpetuate from the AI to the user.[658] Another

[645] Cf. Manyika et al. (2019) P. 1ff.

[646] Cf. Green (2020) P. 1ff.

[647] Cf. Manyika et al. (2019) P. 1ff.

[648] Cf. Manyika et al. (2019) P. 1ff.

[649] Cf. Mehrabi et al. (2019) P. 4ff.

[650] Cf. Dobelli (2011) P. 98ff.

[651] Cf. Mehrabi et al. (2019) P. 4ff.

[652] Cf. Mehrabi et al. (2019) P. 4ff.

[653] Cf. Mehrabi et al. (2019) P. 4ff.

[654] Cf. Mehrabi et al. (2019) P. 4ff.

[655] Cf. Mehrabi et al. (2019) P. 4ff.

[656] Cf. Mehrabi et al. (2019) P. 4ff.

[657] Cf. Mehrabi et al. (2019) P. 4ff.

[658] Cf. Mehrabi et al. (2019) P. 4ff.

important factor of AI bias is that training sets for AI are generated by humans.[659] Any known or unknown bias, the human tasked with creating the training set may has will be inherited by the AI.[660]

Neuroscientists are still trying to figure out how humans consciously experience their world.[661] However, in the last decades it became clear that humans share basic reward and aversion loops with even simple animals.[662] These reward and aversion loops are also used in reinforced learning of artificial neural networks, improved performance is rewarded in a certain way.[663] At its current state, these systems are quite superficial but as they are becoming more and more lifelike, the question of how the ethically correct treatment of AI materializes becomes vital.[664] This is extremely important for the approach of reinforced learning. Does an AI feel pain if it gets a negative input and how do we deal with the deletion of AIs that are no longer the state of the art?[665]

Last but not least, another crucial challenge of AI is its environmental impact. AI can help combat climate change, but on the way to this, AI must be transformed into environmentally friendly AI.[666] GPT-3, which is a language processing AI, is estimated to have used the rough equivalent of energy during its training process that would be required to drive a car over the distance from the earth to the moon and back.[667] Another example, the AI AlphaGo, which managed it to defeat a champion of the game Go, required roughly 50,000 times the power a human brain needs for the same task.[668] Yet, AI does not just prove adversely to the climate by its energy consumption, commonly rare raw materials are needed for the hardware.[669] The industrial exploitation of these resources leads to pollution. In addition, the industrial exploitation of these resources may lead to the mistreatment

[659] Cf. Mehrabi et al. (2019) P. 4ff.

[660] Cf. Mehrabi et al. (2019) P. 4ff.

[661] Cf. Bossmann (2016) P. 1ff.

[662] Cf. Bossmann (2016) P. 1ff.

[663] Cf. Bossmann (2016) P. 1ff.

[664] Cf. Bossmann (2016) P. 1ff.

[665] Cf. Bossmann (2016) P. 1ff.

[666] Cf. Khatry et al. (2021) P. 2ff.

[667] Cf. Khatry et al. (2021) P. 2ff.

[668] Cf. Larbey et al. (2020) P. 28f.

[669] Cf. Ligozat et al. (2021) P. 3ff.

of people in the process. Furthermore, the manufacturing of the system, transport of materials and components and the dismantling of the AI all further add to the environmental pollution caused by AI.[670]

3.5 The current pace of AI development and its regulation

AI development and how the regulation of AI is evolving are the focus of this section. In order to understand how fast the research and development of AI is generating new results a viable metric must be established. In this section the number of scientific publications, published between 2000 and 2020, dealing with AI and AI related issues will serve as this metric. The quantitative data of this section, covering the publications and the citations of these publications, is sourced from Zhang et al. (2021) in which it is gathered from Microsoft Academic Graph and Elsevier/Scopus databank.[671] Additionally, the technical capability of today´s AI applications will be examined.

The research and development of AI began in the early 1950s, when the technology occupied the minds of theoretical mathematicians and the pioneers of computer science.[672] From this paper based theoretical age of AI development, AI has evolved into a huge research discipline which already yields commercial applications.[673] The economically most important nations of this globe, China, Russia, the United States and the EU, are investing vast amounts of money and dedication into AI development and into the question of how to address the risks that are associated with AI and into the issue of how to regulate AI.[674][675][676] For example, the EU has issued a proposal of a regulatory framework on AI, which aims to offer developers, deployers and users of AI clear requirements and restrictions for specific AI applications.[677] Parallel to this, the proposals ambition is to reduce the administrative and financial burdens for business, especially small- to mid-level businesses.[678]

[670] Cf. Ligozat et al. (2021) P. 3ff.

[671] Cf. Zhang et al. (2021) P. 15ff.

[672] Cf. Zhang et al. (2021) P. 15ff.

[673] Cf. Zhang et al. (2021) P. 15ff.

[674] Cf. Zhang et al. (2021) P. 15ff.

[675] Cf. Stanton et al. (2019) P. 6.

[676] Cf. European Parliament (2022) P. 1ff.

[677] Cf. European Parliament (2022) P. 1ff.

[678] Cf. European Parliament (2022) P. 2ff.

Ultimately, the proposal is supposed to ensure that that individuals can trust AI.[679] The framework follows a risk-based approach, sorting the specific uses of AI into risk categories, which are unacceptable risk, high risk, limited risk and minimal to no risk.[680] In connection to the different risk categories, the EU is planning to publish clear set of rules for every category.[681] These rules will address the risk that is created by specific applications of AI and set clear boundaries for AI systems, especially in the higher risk categories.[682] Moreover, the framework will propose an assessment of conformity before a specific AI system is deployed, while also offering clear governance of AI within the EU.[683] Additionally, the framework aims at defining clear requirements for the high risk AI systems while also providing a list if these systems and makes oversight and regulation easier.[684] Furthermore, it will provide a governance structure at the European level and the national level of the member states.[685] Systems that are sorted into the category of unacceptable risk are considered a threat to the safety, way of live and rights of people.[686] According to the regulatory framework, these systems must be banned without any exception.[687] An example for these system is an AI surveillance system that operates without any restriction, and is therefore fully autonomous. Moving on, high risk systems are defined as critical infrastructure systems, educational systems, employment related systems, systems that are in use in essential and private services, applications in law enforcement and administrative processes.[688] These high risk systems are subjected to regulatory obligations before they can be deployed, such as e.g. risk assessment and mitigation approaches, high standard of quality of the used data, ensured traceability of how the AI came to a decision.[689] These rules however, may be subject to exception in some cases, which are strictly

[679] Cf. European Parliament (2022) P. 2ff.

[680] Cf. European Parliament (2022) P. 2ff.

[681] Cf. European Parliament (2022) P. 2ff.

[682] Cf. European Parliament (2022) P. 2ff.

[683] Cf. European Parliament (2022) P. 2ff.

[684] Cf. European Parliament (2022) P. 2ff.

[685] Cf. European Parliament (2022) P. 2ff.

[686] Cf. European Parliament (2022) P. 2ff.

[687] Cf. European Parliament (2022) P. 2ff.

[688] Cf. European Parliament (2022) P. 2ff.

[689] Cf. European Parliament (2022) P. 1ff.

defined.[690] For AI-based mass surveillance systems, such an exception may be to look for a missing child or to prosecute an individual responsible for serious criminal offences.[691] Limited risk and minimal or no risk systems are, for example chat bots, AI-enabled video games and spam filters.[692] The minimal or no risk systems are freely usable under the framework, most f AI applications that are currently in use in the European union fall into this risk-category.[693]

Regarding high-risk systems, the framework follows a four-step approach to decide whether the AI can be used or not. Step one is the development of the AI systems, while step two is the conformity assessment.[694] Next up, in step three, the AI must be registered in a database of the European union.[695] Step four intends the signing of a declaration of conformity after which the system can be placed on the market.[696]

In addition, this risk-based regulatory framework is just a small part of a large package of regulatory guidelines, the EU is going to publish in the next years.[697] Another framework that will be a part of this thesis are the guidelines on trustworthy AI that were developed by the European Commission, which also serves as a basis for the aforementioned risk-based framework. The number of scientific publications, such as peer-reviewed journal articles, patents, books and conference papers, that have seen the light of day in recent years has risen at a staggering rate.[698]

The number of peer-reviewed scientific papers has increased by a factor of 12 in the timeframe of 2000-2019.[699] Figure 4 shows this growth. The number of journal publications in 2020 about AI related topics is 5.4 times more than it was in 2000.[700] Figure 5 illustrates these developments.

[690] Cf. European Parliament (2022) P. 1ff.

[691] Cf. European Parliament (2022) P. 1ff.

[692] Cf. European Parliament (2022) P. 1ff.

[693] Cf. European Parliament (2022) P. 1ff.

[694] Cf. European Parliament (2022) P. 1ff.

[695] Cf. European Parliament (2022) P. 1ff.

[696] Cf. European Parliament (2022) P. 1ff.

[697] Cf. European Parliament (2022) P. 1ff.

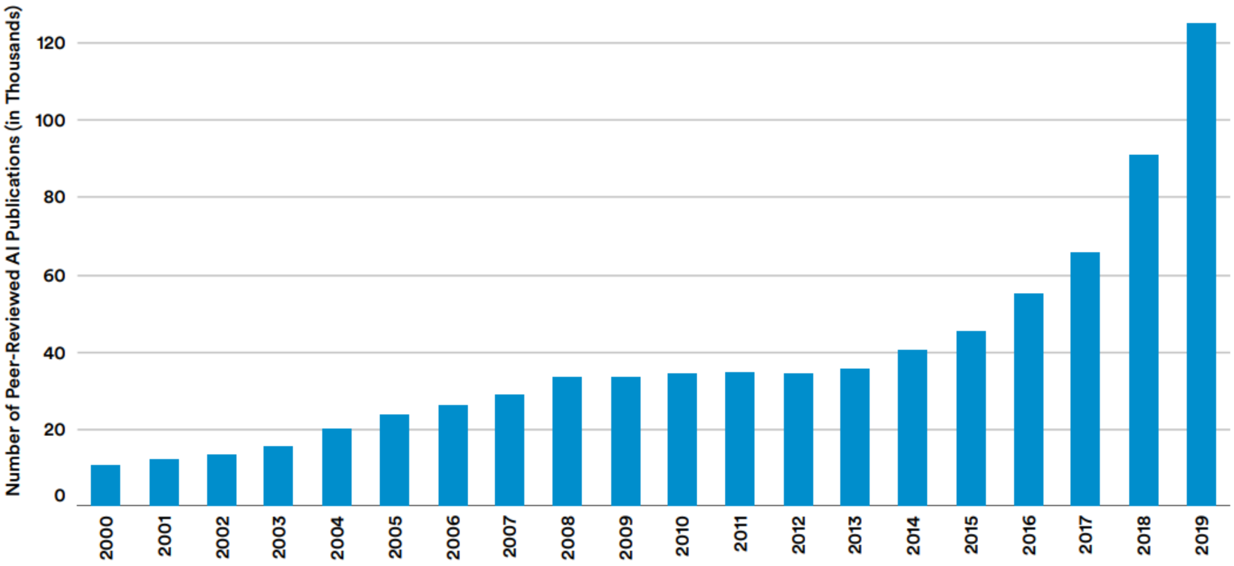
[698] Cf. Zhang et al. (2021) P. 15ff.

[699] Cf. Zhang et al. (2021) P. 15ff.

[700] Cf. Zhang et al. (2021) P. 15ff.

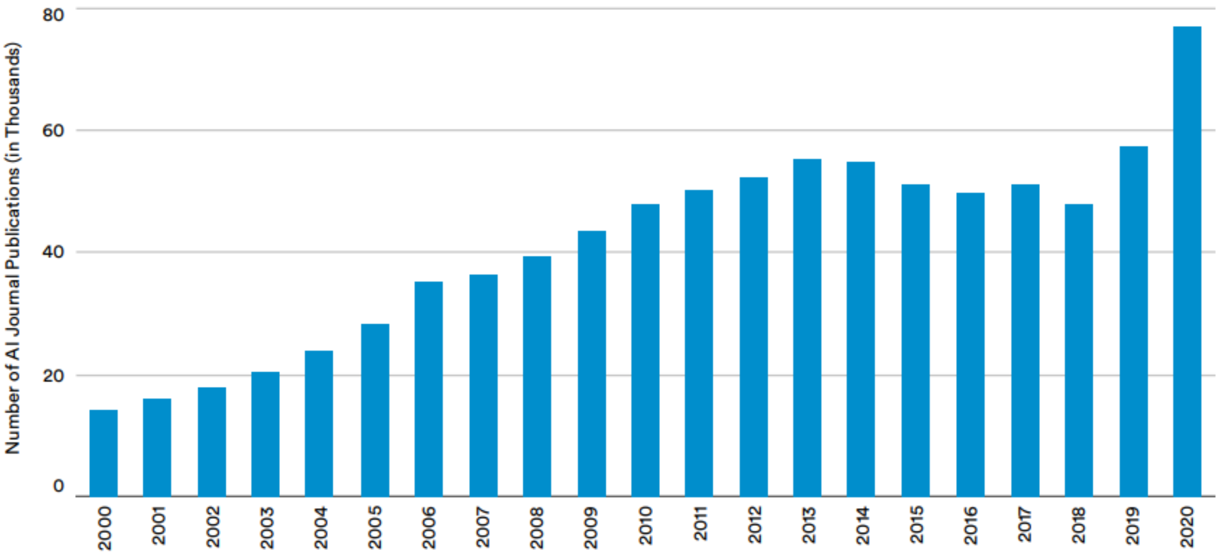
Adding to this development, the attendance at conferences about AI and AI related topics has risen as well.[701] Especially the switch to virtual conferences has ignited this growth.[702]

Figure 4 Number of peer reviewed publications about AI (2000-2020)



Source : Zhang et al. (2020), P.18

Figure 5 Number of journal publications about AI (2000-2020)



Source : Zhang et al. (2020), P.25

[701] Cf. Zhang et al. (2021) P. 15ff.

[702] Cf. Zhang et al. (2021) P. 36ff.

Similar to the rise of AI in the academic discussion, the capability of AI systems has increased dramatically as well, compared to its humble beginnings.[703] These technological advances have allowed AI systems, or at least systems that incorporate artificial neural networks, to experience widespread deployment and use.[704] Today, AI based applications can be found in computer vision, translators and theorem proving applications.[705] State of the art AI systems are able to compose audio, images and text at a high standard, leaving humans with the difficult task of differentiating outputs of an AI with products of human creation.[706] The range of further developments based on these capabilities is enormous, paving the way for AI applications that can successfully execute tasks that are traditionally associated with human intelligence.[707] Computer vision has seen a steady maturation since its conception in the 1960s.[708] Commonly computer vision includes object recognition, body language estimation and semantic segmentation.[709] This development spurred the research and development of autonomous cars, medical image analysis and surveillance applications, leading to a continuous industrialization of computer vision.[710] An enabler of computer vision is machine learning, more accurately deep learning, which is also evolving quickly.[711] A majority of the companies invested into AI are dedicating vast amounts of money in the chase for technological breakthroughs.[712]

[703] Cf. Zhang et al. (2021) P. 36ff.

[704] Cf. Zhang et al. (2021) P. 44ff.

[705] Cf. Zhang et al. (2021) P. 44ff.

[706] Cf. Zhang et al. (2021) P. 44ff.

[707] Cf. Zhang et al. (2021) P. 44ff.

[708] Cf. Zhang et al. (2021) P. 44ff.

[709] Cf. Zhang et al. (2021) P. 44ff.

[710] Cf. Zhang et al. (2021) P. 44ff.

[711] Cf. Zhang et al. (2021) P. 44ff.

[712] Cf. Zhang et al. (2021) P. 44ff.

4 The advent of AI in surveillance technology

AI is bound to transform surveillance. Traditionally, surveillance describes the activity of tracking the actions and movements of an individual.[713] This chapter will have the use of AI in mass surveillance tools as its focus. After having examined the ethical basics and the basics of AI, this chapter will shine a light one the use of AI in surveillance technology, discussing different technologies as well as enabling technologies. Next, the global proliferation of AI surveillance technology will be studied. Afterwards, the conflict between basic human rights and AI surveillance technology will be discussed, focusing on aspects that could allow governmental organizations to use AI surveillance technology in accordance with basic human rights.

4.1 The use of AI in surveillance technology and its opportunities

Innovations in technology are not just the origin of a new age of communication and connectedness, they also create new opportunities for surveillance by governmental organizations, thus generating new possibilities for governmental intervention into the lives of its constituents.[714] This intervention may serve national security and law enforcement interests, enabling a government to prevent and prosecute serious crimes and threats to its national security with an unparalleled effectiveness.[715] This section will investigate the possible uses of AI in surveillance technology and its upside, discussing opportunities that can be beneficial to individuals and to society as a whole. Nations, more precisely their governments, apply AI surveillance technology to accomplish a wide range of goals.[716] Commonly used are smart city platforms, facial and behavioral recognition systems, smart policing and communication surveillance.[717]

The first common AI surveillance systems that will be discussed are smart city platforms. Smart city platforms incorporate an enormous number of sensors in a specific city, which transmit real-time data that can be used to coordinate the delivery of services, increase the efficiency of city management and to enhance public safety, creating opportunities to better a society as a whole.

[713] Cf. Nouri (2020) P. 1ff.

[714] Cf. La Rue (2013) P. 4f.

[715] Cf. La Rue (2013) P. 4f.

[716] Cf. Feldstein (2019) P. 16ff.

[717] Cf. Feldstein (2019) P. 16ff.

However, this sensory network also offers surveillance opportunities. The sensor network of smart cities heavily relies on closed-circuit television cameras with facial and behavioral recognition features, police body cameras and other sensors connected to command centers that analyze the data and gain insights from it.[718] These insights can be utilized to prevent or respond to crimes and emergencies and to ensure that the citizens are safe.[719] Additionally, these insights can be used to combat traffic congestion and enhance the speed of administrative processes.[720] Therefore, smart cities are technology intensive urban centers that depend on the real time data from a myriad of sensors to function.[721] According to the Chinese company Huawei, smart cities include video surveillance, video communication, integrated command and control systems, big data applications and a secure public safety cloud.[722] For the individual, the advantages of the smart city, apart from enhanced security, can be the availability of real-time traffic information and other information related services such as the current business of public administration offices. The major advantage for the society would be the enhanced security. Governments can use the real-time data to predict and counter traffic jams, crimes and to amend the operations schedule of their public offices based on the demand.

The aforementioned facial recognition systems are not just a part of smart cities but a key technology on their own. These systems read the attributes of a face of an individual using facial recognition software. At first the image of a specific face is scanned into the system from a photo or a video, which allows the system to recognise the face in crowds or even from the silhouettes of a person.[723] Main attributes that the facial recognition system is utilizing to create an individual signature are the distance between the eyes and the distance from the forehead to the chin.[724] This signature is a mathematical expression that is then compared to other signatures in the database.[725] It is one of the most widespread use-cases of AI and is already attracting a sizeable commercial

[718] Cf. Feldstein (2019) P. 16ff.

[719] Cf. Feldstein (2019) P. 16ff.

[720] Cf. Feldstein (2019) P. 17ff.

[721] Cf. Feldstein (2019) P. 17ff.

[722] Cf. Feldstein (2019) P. 17ff.

[723] Cf. Moraes et al. (2021) P. 159ff.

[724] Cf. Moraes et al. (2021) P. 159ff.

[725] Cf. Moraes et al. (2021) P. 159ff.

market with interest from governments and militaries around the world.[726] Biometric technology can be linked to cameras and databases to make a connection between live footage and the information it sources from various databases.[727] Yet, not all facial recognition systems rely on individual identification via database matching, another method of matching for these systems is the use of demographic trends.[728] Facial recognition systems are highly intrusive, collecting data about facial features without consent from the individuals, to whom the data belongs.[729] These recognition algorithms experience constant improvement.[730]

An element of facial and behavioral recognition is human pose estimation, which is an omni use AI capability.[731] Human pose estimation studies the position of human body parts, analyzing and predicting behaviors of individuals from surveillance footage.[732] Another important element of facial and behavioral recognition is semantic segmentation.[733] Semantic segmentation describes the process of classifying every pixel of an image with a corresponding label, to isolate individuals or objects that appear in a picture or video footage.[734] In contrast to image classification, semantic segmentation aims at isolating certain aspects of a picture or video footage, that may be defined in advance, in order to enable a thorough examination.[735] Currently, a good example for semantic segmentation is a large scale dataset named Cityscapes. Cityscapes includes scenes from diverse urban environments recorded during different seasons (spring, summer and fall).[736] 25000 pictures from 50 different cities are the heart of the dataset, that is used to train artificial neural networks to understand urban environments.[737] Closely linked to facial and behavioral recognition is object

[726] Cf. Zhang et al. (2021) P. 61.

[727] Cf. Feldstein (2019) P. 16ff.

[728] Cf. Feldstein (2019) P. 16ff.

[729] Cf. Feldstein (2019) P. 16ff.

[730] Cf. Zhang et al. (2021) P. 54ff.

[731] Cf. Zhang et al. (2021) P. 54ff.

[732] Cf. Zhang et al. (2021) P. 54ff.

[733] Cf. Zhang et al. (2021) P. 54ff.

[734] Cf. Zhang et al. (2021) P. 54ff.

[735] Cf. Zhang et al. (2021) P. 54ff.

[736] Cf. Zhang et al. (2021) P. 54ff.

[737] Cf. Zhang et al. (2021) P. 54ff.

detection, which could for example detect concealed weapons and, in combination with body pose estimation and semantic segmentation, may predict crimes. An important sidenote is that AI is also becoming more and more capable in detecting and analyzing language, thus language recognition could be a viable addition to facial and behavior recognition systems.[738] Facial and behavioral recognition systems are already in use today.[739] Today, facial and behavioral recognition systems are already in use in Malaysia, where facial recognition body cameras have been issued to the police force.[740]

Smart policing is spurred by the idea of feeding immense amounts of data, including geographic location, criminal records, biometric data and social media feeds of an individual into an algorithm.[741] This historical data is then combined with real-time data to infer predictions about the type, time, location, the perpetrators(s) and the potential victims of a crime.[742] Artificial neural networks monitor individuals and raise an alarm if certain criteria are met that may be indicators for a crime, effectively surveying the individuals of a population around the clock.[743] This algorithm may then be utilized to prevent and respond to crime, and even predict future crimes, which is an advantage for the whole society.[744] Smart policing aims at enhancing the performance, efficiency and fairness of a police force while also optimizing costs.[745] With the continuing proliferation of facial recognition systems and ever rising amounts of data that can be openly sourced, police forces can have access to an amount of data hitherto undreamt of.[746]

Communication surveillance can be used as a tool to monitor private conversations in public spaces, in private spaces, in the internet and in telecommunication.[747] Technological advancements

[738] Cf. Zhang et al. (2021) P. 56ff.

[739] Cf. Feldstein (2019) P. 18ff.

[740] Cf. Feldstein (2019) P. 18ff.

[741] Cf. Feldstein (2019) P. 18ff.

[742] Cf. Yang (2015) P. 2.

[743] Cf. Feldstein (2019) P. 18ff.

[744] Cf. Feldstein (2019) P. 18ff.

[745] Cf. U.S. Department of Justice (2013) P. 1.

[746] Cf. Feldstein (2019) P. 18ff.

[747] Cf. La Rue (2013) P. 10ff.

have increased its effectiveness while simultaneously lowering the costs of communication surveillance. [748] Especially AI, with is capability to process huge amounts of data is a driver in this area. Communication surveillance can be differentiated into targeted and untargeted communications surveillance.[749] Targeted communication surveillance techniques are focused on an individual´s private communications.[750] Governmental organization may monitor telecommunication in real-time or record phone calls for later analysis.[751] The location of an individual can be tracked and the received and sent text messages can be intercepted.[752] Furthermore, online activity can be monitored and analyzed as well.[753] AI surveillance technology can blaze new trails in the sector of targeted communication surveillance, as an AI can be trained to monitor individuals in all aspects of media activity. Additionally, an AI may be able to infiltrate computer systems, in order to turn on microphones or cameras of computers and/or mobile phones. AI allows governmental organizations to automate tracking and monitoring functions, thus eliminating potential principal-agent loyalty issues.[754] Such a loyalty conflict could occur, if the individuals that operate in the name of the government try to seize power for themselves.[755] This application of AI can also be used to predict future dangers to the public order or to the national security, creating the opportunity to keep the population in a specific country safe. The added value for the individual would be the enhanced safety, as threats can be identified early.

In addition to targeted communication surveillance, AI systems may be utilized to deploy mass communication surveillance.[756] AI allows governmental organizations to broadly intercept and monitor communication within in a population, filtering out communication that are indicators for possible threats to the national security and/or the public order.[757] The capability of AI in speech

[748] Cf. La Rue (2013) P. 10ff.

[749] Cf. La Rue (2013) P. 10ff.

[750] Cf. La Rue (2013) P. 10ff.

[751] Cf. La Rue (2013) P. 10ff.

[752] Cf. La Rue (2013) P. 10ff.

[753] Cf. La Rue (2013) P. 10ff.

[754] Cf. Feldstein (2019) P. 13ff.

[755] Cf. La Rue (2013) P. 10ff.

[756] Cf. La Rue (2013) P. 10ff.

[757] Cf. La Rue (2013) P. 10ff.

and text recognition can be a key to mass communication surveillance.[758] Especially, social networking sites may contain a myriad of information an governmental agency can utilize for its purposes.[759] AI applications enable governments to cast a wider surveillance network than ever before.[760] AI systems do not experience fatigue and can execute surveillance operations around the clock every day.[761] As AI systems commonly rely on huge datasets, that almost always include personal data, the widespread collection of data is incentivized.[762] A growing number of business is implementing online services with the aim of collecting as much data as possible.[763] Social media companies are collecting and monetizing personal data in an ever rising scope.[764] This data collection takes place in intimate, private and public spaces, with the data being traded vividly between different organizations and individuals.[765] The result are datasets that contain an amount of information about individuals unparalleled in history.[766] Again, the individual can profit from enhanced public order and national security.

Aside from these major areas of AI surveillance technology, it is also significant to take a look at enabling technologies of AI surveillance.[767] These enabling technologies provide critical capabilities that are essential for AI surveillance applications, such as smart cities, facial and behavioral recognition systems and smart policing, to serve their purpose.[768] Advanced facial and behavior recognition software could not exist in a useful way without cloud computing and 5G-Networks.[769] The other enabling technologies are automated border control systems, the internet of things and

[758] Cf. La Rue (2013) P. 10ff.

[759] Cf. La Rue (2013) P. 13.

[760] Cf. Feldstein (2019) P. 13ff.

[761] Cf. Feldstein (2019) P. 13ff.

[762] Cf. Human Rights Council (2021) P. 3ff.

[763] Cf. Human Rights Council (2021) P. 3ff.

[764] Cf. Human Rights Council (2021) P. 3ff.

[765] Cf. Human Rights Council (2021) P. 3ff.

[766] Cf. Human Rights Council (2021) P. 3ff.

[767] Cf. Feldstein (2019) P. 16ff.

[768] Cf. Feldstein (2019) P. 21ff.

[769] Cf. Feldstein (2019) P. 21ff.

other AI technology such as digital government and research centers.[770] These enabling technologies will be discussed now.

Cloud computing is on the rise.[771] In its essence it is an advancement of classical webhosting and can enable computers to access decentral processing plants and storage solutions.[772] This access is typically happening through a network, most commonly the internet.[773] Cloud computing penetrates all areas of technology, from GPS navigation, to social media and email communication.[774] It is a technology that is primarily focused on enabling network access whenever necessary to a common pool of processing and storage resources.[775] Today, more and more countries are switching to cloud computing for their governmental data storage requirements, that are not classified.[776] Cloud computing offers unique advantages such as improved collaboration, lower maintenance costs and a good accessibility.[777]

Another enabling technology are automated border control systems, these are mainly applied in international airports and border checkpoints.[778] These systems use multi-model biometric matching, utilizing facial recognition systems in combination with electronic passports and other biometric documents.[779] The data that is generated this way is then cross-referenced with databases to identify individuals that may pose a security risk.[780] Additionally, governments generate data to train their facial and behavioral recognition systems this way and they are always informed about the whereabouts of certain individuals.[781] The EU has concluded field testing an application that is

[770] Cf. Feldstein (2019) P. 21ff.

[771] Cf. Feldstein (2019) P. 21ff.

[772] Cf. Labes (2012) P. 2ff.

[773] Cf. Feldstein (2019) P. 21ff.

[774] Cf. Feldstein (2019) P. 21ff.

[775] Cf. Feldstein (2019) P. 21ff.

[776] Cf. Feldstein (2019) P. 21ff.

[777] Cf. Jaiswal (2022a) P. 1.

[778] Cf. Dumbrava (2021) P. 1ff.

[779] Cf. Dumbrava (2021) P. 4ff.

[780] Cf. Dumbrava (2021) P. 1ff.

[781] Cf. Feldstein (2019) P. 21ff.

called iBorderCtrl in Greece, Latvia and Hungary by the end of 2019.[782] This system screens migrants when they cross the border, asking them about their background, with the answers fed into an AI-based lie detection application.[783] iBorderCtrl is based on reading facial expressions and to conclude moral states from them, rendering a leading decision based on its conclusion.[784] iBorderCtrl is able to enable a quicker and more detailed border control for third country citizens passing into the EU, reducing the cost and time for every individuals that travels into the EU.[785]

The reality, that an increasing number of devices is linked to the internet, creates new possibilities for the internet of things which make it an enabling technology of AI surveillance.[786] The internet of things offers advantages such as an efficient resource allocation and utilization, the minimization of human effort and efficiency gains.[787] This allows the data of a myriad of devices to be shared in a cloud for analytic purposes.[788] The internet of things, as soon as the problem of interoperability of devices is overcome, can transform every device in a network into omnipresent surveillance instruments.[789] A TV that has a built in camera and a microphone that are programmable by its vendor, can be used to listen to every telephone communication that happens adjacent to it, no matter how strong the encryption of the communication is.[790] In the beginning of 2019 for example, Amazon´s smart speaker system Echo, has sparked controversy, as it was discovered that, Amazon employees had listened to conversations collected by the smart speakers.[791] This observed data was used by Amazon to gain insights without the consent of the individuals the data belonged to.[792] Facebook and Google have operated in a similar way via their applications on mobile devices.[793]

[782] Cf. Dumbrava (2021) P. 17.

[783] Cf. Feldstein (2019) P. 21ff.

[784] Cf. Feldstein (2019) P. 21ff.

[785] Cf. European Commission (2019) P. 1ff.

[786] Cf. Feldstein (2019) P. 21ff.

[787] Cf. Jaiswal (2022b) P. 1.

[788] Cf. Feldstein (2019) P. 21ff.

[789] Cf. Feldstein (2019) P. 21ff.

[790] Cf. Feldstein (2019) P. 21ff.

[791] Cf. Feldstein (2019) P. 21ff.

[792] Cf. Feldstein (2019) P. 21ff.

[793] Cf. Feldstein (2019) P. 21ff.

Autonomous cars may also be reconfigured into mobile surveillance technology, using its built-in cameras to spot number plates and faces while driving around.[794]

4.2 The global proliferation of AI surveillance systems

The proliferation of AI surveillance continues globally, according to the AI Global Surveillance Index by Steven Feldstein (2019) which examines this proliferation in 176 countries.[795] Out of these 176 countries, at least 75 are actively using one of the mentioned AI surveillance technologies.[796] The countries examined in the AI Global Surveillance Index are not limited to authoritarian states, liberal democracies are also among the users of AI surveillance technology.[797] Figure 7 illustrates this proliferation using a map of the world.

This thesis will follow a structure based on basic political systems, authoritarianism and democracy to discuss the proliferation of AI surveillance technology. Authoritarian governments are rejecting political plurality while relying on a strong central power, commonly a party or the military, to preserve the political status quo.[798] Authoritarian administrations can be autocratic or oligarchic, which describes if one individual or organization holds all the power or if a small group of people or organizations wields the power.[799] Authoritarianism reduces the separation of powers, democratic voting and the rule of law. Limited levels of political rights for the individual are a key attribute of authoritarianism as well. The proliferation of AI surveillance technology in authoritarian, and semi-authoritarian, states will be discussed first. Afterwards, the role of liberal democracies in the proliferation of AI surveillance tech will be examined.

[794] Cf. Feldstein (2019) P. 23ff.

[795] Cf. Feldstein (2019) P. 23ff.

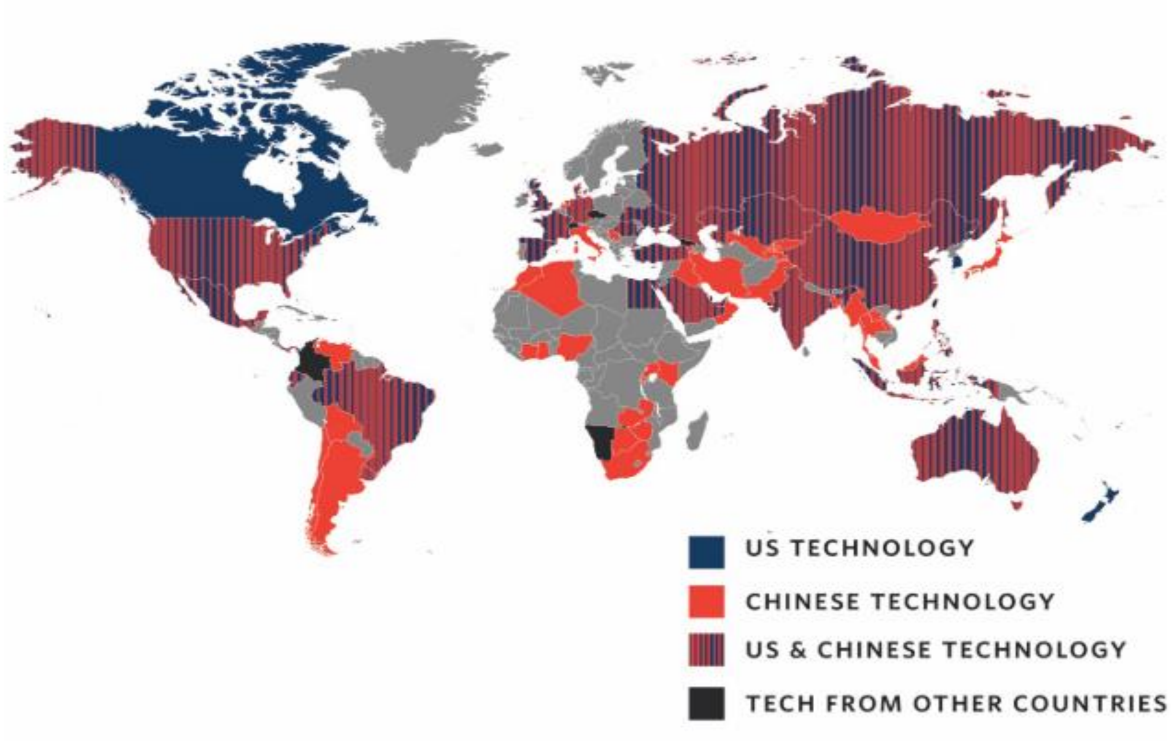
[796] Cf. Feldstein (2019) P. 23ff.

[797] Cf. Feldstein (2019) P. 23ff.

[798] Cf. Cerutti (2017) P. 17.

[799] Cf. Frantz and Ezrow (2011) P. 17.

Figure 6 Proliferation and origin of AI surveillance technology



Source: Feldstein. (2019), P.3

Mostly, authoritarian states are located around the Persian Gulf as well as southeast and central Asia, north to central Africa and the northern parts of Latin America.[800] South and Central Asia as well as the Americas are heavily gunning up with AI surveillance technology, while sub-Saharan Africa is slowly developing AI surveillance capabilities.[801][802] This may not be surprising, given the disparities in technological developments between the regions. Nevertheless, Chinese Companies are actively expanding to Africa, not just with surveillance technology but also with broadband internet access.[803] The Chinese company Huawei, has been on the frontier of the development of smart cities, currently marketing smart cities as safe cities.[804] The company connects its smart city technology explicitly to regional security challenges.[805] For the Middle East Huawei states that

[800] Cf. Armstrong (2019) P. 1.

[801] Cf. Feldstein (2019) P. 8ff.

[802] Cf. Fontes and Perrone (2021) P. 3.

[803] Cf. Feldstein (2019) P. 8ff.

[804] Cf. Feldstein (2019) P. 8ff.

[805] Cf. Feldstein (2019) P. 8ff.

smart cities can prevent religious extremism and in Latin America, Huawei states that its systems can combat organized crime.[806] A real world example of smart cities can be seen in Saudi Arabia´s Makkah region, where a crowd control system is in use that is supposed to increase the safety and the security of the pilgrims in the region.[807][808] Data is sourced via surveillance cameras and wristbands that contain personal information of the individuals, such as GPS and medical data.[809]

A country that is often associated with AI surveillance is the People´s Republic of China. Therefore, it should not be surprising that China is a key provider of AI surveillance technology.[810] Additionally, China operates more than 626 million facial recognition cameras.[811][812] The Chinese technology can be found in at least 63 countries, with Huawei alone supplying AI surveillance capability to at least 50 countries.[813] Evidently, a considerable overlap between the Chinese Belt and Road Initiative and the efforts of Chinese Companies to market their AI technology exists.[814] Chinese companies, including the aforementioned Huawei, are the primary supplies on the market.[815] On the market, China is constantly pushing to consolidate its leadership position, making China a driver of the proliferation of AI surveillance applications.[816] Due to this, China is often accused of employing its authorities to work directly with the companies that conduct research and development in the area of AI surveillance applications, with the purpose of exporting “authoritarian tech” to similar governments and to liberal democracies to spread its influence while also gaining the ability to monitor the populations of other countries.[817] Zimbabwe and Venezuela, both

[806] Cf. Feldstein (2019) P. 8ff.

[807] Cf. Feldstein (2019) P. 8ff.

[808] Cf. Nouri (2020) P. 1ff.

[809] Cf. Feldstein (2019) P. 8ff.

[810] Cf. Feldstein (2019) P. 8ff.

[811] Cf. Fontes and Perrone (2021) P. 3.

[812] Cf. Nouri (2020) P. 1ff.

[813] Cf. Feldstein (2019) P. 8ff.

[814] Cf. Feldstein (2019) P. 8ff.

[815] Cf. Feldstein (2019) P. 8ff.

[816] Cf. Sahin (2020) P. 3.

[817] Cf. Sahin (2020) P. 1ff.

identified as violators of human rights, are key importers of Chinese AI surveillance technology.[818] The gathered data is then sent back to China.[819] Yet, China is also exporting its surveillance technology to liberal democracies and to companies based in liberal democracies. Authoritarian states rarely supply their needs for AI surveillance technology from a single source.[820] Yet, there is a reason why Chinese companies are especially scrutinized. Huawei is the market leader by a huge margin, with its technology linked to a lot of countries worldwide.[821] Additionally, several start-ups that are specialized on AI surveillance technology are based in China.[822] The company is actively expanding in markets with a huge number of, at least partly and in some aspects, authoritarian states such as sub-Saharan Africa.[823] Additionally, the company is offering ongoing technological and logistical support.[824] Currently, Huawei is engaged in 75 smart city projects around the world, experiencing an unprecedented growth in its AI surveillance related business line.[825] From 2017 to 2018 Huawei has increased its global reach from 40 to more than 90 countries in which its smart city technologies have been introduced.[826] Huawei´s marketing model operates with the direct pitching of its technologies to national security agencies, working in close cooperation with Chinese banks that increase the attractivity of these pitches with subsidized loans.[827] This increases the dependency of a foreign country on the People´s Republic of China, mandating contracting with Chinese companies.[828] Additionally, there is a significant doubt, that Huawei is as independent from the Chinese government as it claims.[829]

[818] Cf. Sahin (2020) P. 7.

[819] Cf. Sahin (2020) P. 7.

[820] Cf. Feldstein (2019) P. 8ff.

[821] Cf. Sahin (2020) P. 7.

[822] Cf. Sahin (2020) P. 7.

[823] Cf. Feldstein (2019) P. 8ff.

[824] Cf. Feldstein (2019) P. 8ff.

[825] Cf. Feldstein (2019) P. 8ff.

[826] Cf. Feldstein (2019) P. 8ff.

[827] Cf. Feldstein (2019) P. 8ff.

[828] Cf. Feldstein (2019) P. 13ff.

[829] Cf. Feldstein (2019) P. 13ff.

The use of smart policing is steadily rising in authoritarian states, such as Laos, Qatar and Zimbabwe.[830] and Moreover, China has been on the forefront of smart policing, most likely using an integrated joint operations platform.[831][832] The integrated joint operations platform collects data from closed-circuit-television systems, facial recognition systems and from invasive devices that eavesdrop on traffic within private wireless networks.[833] Data regarding license plates and identification cards, that are scanned in police controls or checkpoints, bank statements and health of an individual are also facilitated into the integrated joint operations platform.[834] Additionally, Chinese authorities also add genetic data of the population of certain regions.[835] The integrated joint operations platform´s algorithm then scans through this data in search of suspectedly threatening patterns, which leads to an individual being brought in for questioning.[836]

Liberal democracies are not only among the major exporters of AI surveillance technology, they are key users of these technologies as well.[837][838] Interestingly, the liberal democracies of Europe are also deploying more and more AI surveillance capability with automated border controls posing as the most popular use of the technology.[839] They are deploying AI surveillance systems to control their borders, apprehend criminals and monitor the behavior of their citizens.[840] The negative connotation of this use, however, does not necessarily mean that these countries are utilizing AI surveillance technology to repress their population.[841] The EU border control systems that were illustrated in Section 4.1 are typically focused on migrants that want to enter the EU.[842] Yet, this is a

[830] Cf. Feldstein (2019) P. 25ff.

[831] Cf. Sahin (2020) P. 6.

[832] Cf. Human Rights Watch (2018) P. 1ff.

[833] Cf. Sahin (2020) P. 6.

[834] Cf. Feldstein (2019) P. 13ff.

[835] Cf. Feldstein (2019) P. 13ff.

[836] Cf. Human Rights Watch (2018) P. 1ff.

[837] Cf. Feldstein (2019) P. 10ff.

[838] Cf. Sahin (2020) P. 8.

[839] Cf. Feldstein (2019) P. 10ff.

[840] Cf. Feldstein (2019) P. 10ff.

[841] Cf. Feldstein (2019) P. 10ff.

[842] Cf. Dumbrava (2021) P. I.

blurry line, as advanced democracies are facing difficulties in balancing their security interests with the protection of civil rights and liberties.[843] For example, an increasing number of cities in the US have begun to utilize AI surveillance technology.[844][845] From airborne drones with facial and behavioral recognition capabilities in Baltimore, Maryland to advanced multisensory towers on the US-Mexico border.[846] France is using Chinese technology supplied by ZTE in the sea adjacent city of Marseille to increase the security of the city by lowering the crime rate.[847] This shall be realized by a network of intelligent closed-circuit television cameras with facial and behavioral recognition ability, that is interconnected via an operations center.[848] Additionally, Huawei provided the French city of Valenciennes with a smart city demonstration model that can detect unusual movements and crowd formations.[849] Countries such as Germany, France, the United Kingdom and the US, are also exporting AI surveillance technology to unsavory governments.[850] As mentioned before, Saudi Arabia is using technology supplied by Huawei to build smart cities, but without the cloud serves from Google and BAE Systems mass surveillance systems, this project would not work.[851]

Smart policing has received considerable attention in liberal democracies after the United States have jumpstarted its development in 2009.[852] The first common technology was PredPol, launched in 2012.[853] The predictive policing systems work on massive data aggregation and analyses, trying to estimate where future crimes may take place and who the perpetrator may be.[854] Based on this, one should not be surprised, that the PredPol predictive analytics program is already widely used

[843] Cf. Feldstein (2019) P. 10ff.

[844] Cf. Fontes and Perrone (2021) P. 3.

[845] Cf. Sahin (2020) P. 9.

[846] Cf. Feldstein (2019) P. 20ff.

[847] Cf. Feldstein (2019) P. 20ff.

[848] Cf. Feldstein (2019) P. 20ff.

[849] Cf. Feldstein (2019) P. 20ff.

[850] Cf. Feldstein (2019) P. 20ff.

[851] Cf. Feldstein (2019) P. 20ff.

[852] Cf. Feldstein (2019) P. 20ff.

[853] Cf. Yang (2015) P. 4.

[854] Cf. Yang (2015) P. 2ff.

in the United States.[855] PredPol applies a forecasting algorithm that generates crime predictions in small geographical areas.[856] These predictions however, are limited to assumptions that crimes are more likely to occur during pinpointed houses in specific areas at specific times based on historical databases.[857]

Another example for already existing surveillance systems can be found in the West Bank territory in Israel.[858] Whenever Palestinians or inhabitants of the West Bank in general make a phone call, travel or post content on social media, they are likely monitored by Israeli governmental organizations.[859] This surveillance is achieved via microphones, cameras, drones and spy software.[860] A system that allows the Israeli security forces to identify and subsequently neutralize potential threats.

4.3 High-level juristic issues of AI surveillance

In order to get a first impression and develop an understanding of the ethical implications of AI, this section will examine what is currently regarded as lawful and unlawful use of AI surveillance technology in the literature. Additionally, a point of discussion will be the reasons any government organization may have, that could legitimate surveillance in general.

AI surveillance law is a patchwork of different policies around the globe. China for example has established new data protection laws which are designed to protect the personal data of individuals.[861] Interestingly however, these laws do not mention governmental agencies.[862] In the US the states and even the cities are not united when it comes to laws governing AI surveillance.[863] In

[855] Cf. Yang (2015) P. 4.

[856] Cf. Yang (2015) P. 4.

[857] Cf. Yang (2015) P. 1ff.

[858] Cf. Harari (2018) P. 1ff.

[859] Cf. Harari (2018) P. 1ff.

[860] Cf. Harari (2018) P. 1ff.

[861] Cf. Junck et al. (2021) P. 1ff.

[862] Cf. Fontes and Perrone (2021) P. 3.

[863] Cf. Sahin (2020) P. 8.

Oakland, San Francisco and San Diego, facial and behavioral recognition systems are strictly forbidden, while Detroit allows the restrained use of the technology. The EU has already published a regulatory framework proposal for AI as discussed in Section 3.5. Additionally, the EU has issued a framework to assess the ethical permissibility of AI applications which will be examined in Section 5. Yet, despite these publications, the EU still needs to find its balance in the governance of AI surveillance technology.[864] Due to this legal heterogenous legal background, this section will examine the high-level juristic issues of AI surveillance systems based on the universal declaration of human rights because it has been established as the standard for basic human rights as it is widely internationally recognized as the basis of upholding basic human rights, in an apolitical document that bridges religions, political ideologies and cultures.[865][866]

Unlawful acts are often commonly referred to as crimes, yet no simple and generally applicable definition exists. These acts are punishable by a government or any other legitimated authority. Possible approaches to a general definition of crime all show the following aspects. They distinct that an action becomes unlawful once it is declared as such by the relevant and applicable law. Additionally, these approaches stress that an unlawful act is harmful to the whole society.[867][868]

Basic human rights are held by every human being, no matter the race, sex, ethnicity, nationality, language religion or other personal features.[869] A basic human right for example is the right to freedom of opinion and expression which is codified in the universal declaration of human Rights and the International Covenant on Civil and Political Rights.[870] Every human being has the right to hold opinions without interference and to gather information through any kind of media.[871]

[864] Cf. Sahin (2020) P. 8.

[865] Cf. United Nations (2021) P. 1ff.

[866] Cf. Akkad (2012) P. 1ff.

[867] Cf. Cane and Conaghan (2009) P. 263.

[868] Cf. Scott and Marshall (2009) P. 1ff.

[869] Cf. United Nations (2022) P. 1ff.

[870] Cf. La Rue (2013) P. 6ff.

[871] Cf. La Rue (2013) P. 6ff.

Privacy is indisputably recognized as a fundamental human right, which is also recognized in the universal declaration of human rights.[872] Privacy may be defined as the notion that every individual should have an area of autonomous development, liberty and interaction without outside interference.[873] This right also includes decision making authority who holds information regarding the individual and how that information is utilized.[874] To exercise the right to privacy in communications, individuals must be able to ensure that their communications remain private, secure and possibly anonymous.[875] Therefore, privacy of communication assumes that individuals are able to exchange information in a space that cannot be accessed by other member of the society such as their government.[876] A necessary key of this is that the individuals are able to verify that their communication is read by no one but the intended recipients.[877] The right to privacy is an representation of human dignity and serves the purpose of protecting human autonomy and the personal identity of an individual.[878] Governments are required to refrain themselves from any violation of the right to privacy, additionally, they are obliged to protect and promote the right to privacy within their jurisdiction.[879] This infers a governmental duty to use adequate legislative measures to protect individuals from intrusion in their privacy.[880] Businesses have an obligation to respect all internationally codified human rights, meaning that they have to refrain from either infringing on human rights and/or adverse impacts on human rights created by their actions.[881]

[872] Cf. La Rue (2013) P. 6ff.

[873] Cf. La Rue (2013) P. 6ff.

[874] Cf. La Rue (2013) P. 6ff.

[875] Cf. La Rue (2013) P. 6ff.

[876] Cf. La Rue (2013) P. 6ff.

[877] Cf. La Rue (2013) P. 6ff.

[878] Cf. Human Rights Council (2021) P. 1ff.

[879] Cf. Human Rights Council (2021) P. 3ff.

[880] Cf. Human Rights Council (2021) P. 3ff.

[881] Cf. Human Rights Council (2021) P. 3ff.

Yet the right to privacy can be restricted, if necessary.[882] However, these restrictions must follow a set of predefined rules, that guarantee that these restrictions are neither arbitrary nor unlawful.[883][884] Any restriction must be justified by codified law.[885] In addition, the essence of the human right to privacy is never a subject to any restriction.[886] Any indiscretion that comes with implementing the restriction must be avoided.[887] A restriction is necessary if it is vital to the legitimate aim that is the reason of existence for the restriction.[888] Furthermore, restrictive measures must be guided by and conform to the principle of proportionality, determining that the instrument that must be preferred is the least intrusive one.[889]

Having established these definitions, it is central to state that surveillance by governmental organizations can happen in accordance with the right to freedom of opinion and expression and the right of privacy. Governments can have legitimate reasons for surveillance operations that do not have the aim of enforcing political repression and limit individual freedoms.[890] Yet, the advent of the internet and its subsequent evolution has given governments new ways to monitor individuals.[891] The international human rights law contains three principles to assess the lawfulness of any given surveillance operation.[892] These three principles are derived from the rules that allow the restriction of privacy of an individual that are the justification by codified law, the protection of the essence of the human right to privacy and the principle of proportionality. Therefore, they are quite similar to them.

[882] Cf. La Rue (2013) P. 6ff.

[883] Cf. Human Rights Council (2021) P. 3ff.

[884] Cf. La Rue (2013) P. 6ff.

[885] Cf. La Rue (2013) P. 6ff.

[886] Cf. La Rue (2013) P. 6ff.

[887] Cf. La Rue (2013) P. 6ff.

[888] Cf. La Rue (2013) P. 6ff.

[889] Cf. La Rue (2013) P. 6ff.

[890] Cf. Feldstein (2019) P. 11ff.

[891] Cf. Feldstein (2019) P. 11ff.

[892] Cf. Feldstein (2019) P. 11ff.

The first principle requires that the domestic law of a country allows the surveillance operation.[893] These legal requirements must be clear, precise, publicly accessible, comprehensive and non-discriminatory.[894] Second, stemming from the last rule that could allow the restriction of privacy of an individual, the surveillance operation must be in proportionality with the situation and international legal standards.[895] Third, the surveillance operation must be justified by legitimate aims.[896] Yet, what are legitimate aims? Governmental organizations typically justify surveillance operations with national security and public order concerns.[897] However, these claims may be too broad, and could restrict an individual's right to freedom of opinion and expression.[898] Legitimate surveillance operations are requiring governments to establish a robust and independent system to oversee the surveillance operations in order to guarantee that a specific surveillance operation is necessary.[899]

These three principles lead to the fact that the legal standards to conduct surveillance operations are difficult for any government to meet.[900] This also applies to liberal democracies with a strong rule of law and robust oversight institutions.[901] Many inferences and predictions of AI can effect the right to privacy, including the autonomy of individuals and their right to have an identity of their own.[902] Moreover, they raise the questions concerning other basic human and personal rights such as the right to freedom of thought and opinion and the right to freedom of expression.[903]

The in Section 3.4 mentioned possibility of faulty outputs by an AI can lead to a multitude of human rights violations.[904] These faulty outputs could lead to the false marking of an individual as

[893] Cf. Feldstein (2019) P. 11ff.

[894] Cf. Kaye (2019) P. 10ff.

[895] Cf. Feldstein (2019) P. 11ff.

[896] Cf. Feldstein (2019) P. 11ff.

[897] Cf. Feldstein (2019) P. 11ff.

[898] Cf. Feldstein (2019) P. 11ff.

[899] Cf. Feldstein (2019) P. 11ff.

[900] Cf. Feldstein (2019) P. 11ff.

[901] Cf. Feldstein (2019) P. 11ff.

[902] Cf. Human Rights Council (2021) P. 3ff.

[903] Cf. Human Rights Council (2021) P. 3ff.

[904] Cf. Human Rights Council (2021) P. 3ff.

a threat to national security or public order.[905] This individual would then be punished for an act he or she did not commit.

The data sets themselves could infringe on the right to privacy as they can contain a multitude of historic personal data, such as criminal records and statistics of police interventions in certain neighborhoods.[906] The storage and subsequent use of this data could be still be in accordance with the right to privacy, as restrictions may apply. It all depends on the three principles outlined above, if the codified law allows the storage, the proportionality is given and the data storage is justified by legitimate aims. The legitimate aims are commonly the upholding of the public order or questions of the national security. The problem here should therefore be clear, certain governmental actions could enable governmental agencies to storage personal data while not infringing on the privacy of their constituents.

In the same context, AI can be biased by the way it has been trained.[907] Biases can find their way into the AI system via the training data and/or via the design of the AI as explored in Section 3.4. A special area of concern for biases are behavioural recognition systems, as they create a biometric profile of every individual they monitor.[908] These biometric profiles contain key attributes of the individual, revealing information that enable a governmental agency to differentiate individuals or ethnic groups that may be flagged more often by an AI surveillance system. The interventions that the state or governmental organization undertakes could therefore be based on a biased decision which could negatively affect the justification of this intervention. Yet, these interventions, which may include warrants, arrests and prosecution possibly leading to convictions, could just as well happen in accordance to the human right to privacy. The respective government just needs to ensure that it has acted in respect to the codified laws of the specific state, that the surveillance operation meets the principle of proportionality and has a legitimate aim. Nevertheless, the interventions by a governmental agency that are based on the insights of AI surveillance systems could must happen in accordance with the human right to a fair trial, the protection from arbitrary arrest and detention

[905] Cf. Human Rights Council (2021) P. 3ff.

[906] Cf. Human Rights Council (2021) P. 3ff.

[907] Cf. Human Rights Council (2021) P. 3ff.

[908] Cf. Human Rights Council (2021) P. 3ff.

as well as the right to live freely.[909] The opacity of AI decision making could enable governmental agencies to masquerade the true capabilities of their AI surveillance systems, especially in areas that suffer from a lack of transparency as their purpose dictates that, e.g. counter-terrorism forces.[910]

To conclude, AI surveillance systems may infringe upon basic human rights such as the right to privacy and the human right to freedom from arbitrary arrest and detention. Yet, these human rights have restrictions that governmental agencies can use to set up AI surveillance systems and AI surveillance networks that bypass these human rights in the correct conditions. **Countries with inadequate oversight and regulation of governmental agencies and/or authoritarian systems could certainly use these restrictions, albeit that they routinely dodge the obligations that stem from basic human laws.[911]** Thus, AI surveillance systems, ethical or not, could be used in accordance with, at least some, basic human rights.

4.4 Some Ethical challenges of AI surveillance technology

AI surveillance systems create a myriad of ethical challenges that will be highlighted in this part of the thesis. It will outline the most pressing ethical challenges, aiming at creating an understanding of why the use of AI surveillance systems, that might occur in accordance with basic human laws, and its continuing proliferation may be problematic. These ethical challenges will be further regarded with the perspective of rule utilitarianism in Section 5.

AI surveillance systems are closely correlated to an abuse of powers by governmental agencies. Liberal democracies such as the member states of the EU and the US are still trying to find a way in which AI surveillance technologies can benefit the safeguarding of their societies without exercising repression with these systems.[912] **Authoritarian states exploit AI surveillance software in**

[909] Cf. Human Rights Council (2021) P. 6ff.

[910] Cf. Human Rights Council (2021) P. 6ff.

[911] Cf. Feldstein (2019) P. 11ff.

[912] Cf. Bohai (2021) P. 1ff.

the name of national security and/or public order issues.[913] Countries such as China are intending to use AI for defense and social welfare purposes.[914] However, a malicious intent seems to be commonly connotated to these plans, given the current government of the People´s Republic. This malicious intent could materialize in the possibility of AI surveillance systems being used in ways that do not serve the public interest, such as silencing the political opposition.[915] Authoritarian states could abuse AI surveillance systems and turn them into all-seeing tools to control their populations giving the government absolute control about all aspects of their lives, violating important values such as fairness, integrity and transparency.[916] Such an AI would also allow governments to apply racial profiling explicitly monitoring specific ethnic groups of a population, further undermining fairness.[917] The more AI surveillance technology is utilized in this context, the more insights it can generate. It is possible to use biometric bracelets, as already in use in Saudi Arabia and mentioned in Section 4.2, that could monitor everything.[918] From speech to vital signs and brain activity, delivering incredible insights about humans that can be used to either manipulate individuals or to sanction them.[919] All forms of governments, not just autocracies and liberal democracies, could create huge, auto-populating data profiled for each and every one of their citizens.[920] Every single profile could contain millions of data points that are created by the individual's appearance in the surveilled space, allowing the granular scoring of every individual of the population in accordance with its loyalty to the government.[921] Individuals who favor the government could experience preferred treatment, further undermining fairness. Additionally, this can be another opportunity for a government to sanction the opposition to the government. Such actions

[913] Cf. Bohai (2021) P. 1ff.

[914] Cf. Bohai (2021) P. 1ff.

[915] Cf. Whitton (2001) P. 4ff.

[916] Cf. Bohai (2021) P. 1ff.

[917] Cf. Bohai (2021) P. 1ff.

[918] Cf. Harari (2018) P. 1ff.

[919] Cf. Harari (2018) P. 1ff.

[920] Cf. Andersen (2020) P. 1ff.

[921] Cf. Harari (2018) P. 1ff.

would also erode the legitimacy of any government as it is supposed to administer the laws and administrative power impartially.[922]

AI based facial recognition systems also suffer from biases and inaccuracies, reflected in false positives and false negatives.[923] False positives can happen when a signature (see Section 4.1) is incorrectly related to a signature in the database.[924] Furthermore, biases can spur these inaccuracies, pave the way for racial profiling and undermine trust into the AI surveillance systems.[925] This becomes clear when studying the results of the ActivityNet temporal action localization task, which is a video benchmark that tests how well an AI can comprehend, understand and label human activities.[926] The temporal action localization task demands AI to detect human activities in a 600 hour video sequence, focused on localization and recognition of the activity within the footage.[927] Figure 6 shows the activities that AIs had the most problems in correctly comprehending, understanding and labelling the action. Additionally, the figure shows, how AIs have improved regarding to the identification of the activities, from 2019 to 2020. Yet, it is clear that errors persist. This raises the question, if we can trust the recommendations of an AI powered facial and behavioral recognition surveillance system. For example, despite improvement, an AI surveillance system has trouble in differentiating, if someone is throwing darts or if that someone is throwing something else that may have a more malicious intend.

[922] Cf. Whitton (2001) P. 4ff.

[923] Cf. Moraes et al. (2021) P. 159ff.

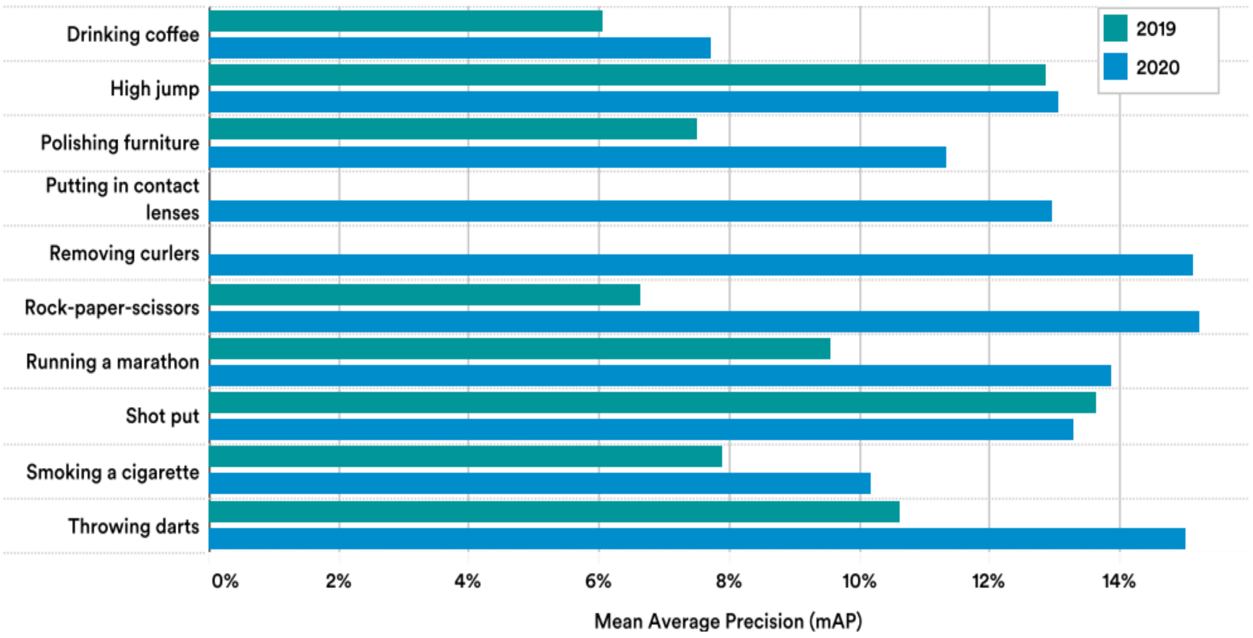
[924] Cf. Moraes et al. (2021) P. 159ff.

[925] Cf. Moraes et al. (2021) P. 159ff.

[926] Cf. Zhang et al. (2021) P. 1ff.

[927] Cf. Zhang et al. (2021) P. 1ff.

Figure 7 Hardest Activities for AI in ActivityNet (2019-2020)



Source : Zhang et al. (2020), P.59

The aggregated information about individuals or their desires that was collected by AI surveillance systems can be used to unknowingly to the individual erode his or her autonomy and rational decision-making process, thus enabling manipulation.[928] This manipulation can be conducted in a way in which the manipulator tries to benefit at the expense of the target.[929] The interaction of individuals with the digital world creates more and more data which allows more refined efforts of manipulation, individuals become more and more vulnerable to the nudges of the AI surveillance systems or its engineers/controllers.[930] While a nudge is defined as a change of environment that influences the behavior of an individual in a way that is beneficial to the individual, the line between nudge and manipulation may become blurry.[931] This is something that is not explicitly limited to governmental agencies, companies can use AI surveillance systems as well to manipulate individuals into buying their products and/or services, thus maximizing their sales in the future.[932]

[928] Cf. Müller (2020) P. 1ff.

[929] Cf. Noggle (2018) P. 7ff.

[930] Cf. Müller (2020) P. 1ff.

[931] Cf. Müller (2020) P. 1ff.

[932] Cf. Müller (2020) P. 1ff.

Political propaganda and misinformation can be tailored to individuals, threatening democracies or consolidating the power of authoritarian regimes.[933] Systematic political manipulation can seriously erode democratic institutions and may lead to regime changes that favor more repressive forms of government.[934] Furthermore, social media allows the direct addressing of individuals around the clock.

AI relies on datasets to function and these datasets must be stored somewhere. Even though AI does not exclusively rely on the processing of personal data, stored AI datasets can still contain data from which a behavioral pattern of an individual can be created.[935] This creates the imminent risk of data breaches.[936] These data breaches are not a moral challenge that stems from the use of AI, but from its continuing proliferation and today´s data driven economy.[937] Data breaches have occurred frequently in the past, exposing sensitive data regarding millions of individuals to malicious uses.[938] But these large datasets themselves are ethically challengeable because they allow for countless approaches of analysis and data sharing with third parties in exchange for monetary values, often leading to privacy infringements.[939] This violates the basic human right to privacy inasmuch as individuals are the sole sovereigns when it comes to deciding who is holding their data.[940] Such data sharing must be applicable for a restriction from the right to privacy. These restrictions must fulfill three criteria (justified by codified law, legitimacy, proportionality), as outlined in Section 4.3. Likewise, the sale of data to third parties can lead to identity theft, fraud, cyberterrorism, information warfare and extortion.[941] Data from various sources can be utilized to match the data with an individual, paving the way for the arbitrary and/or unlawful use of the data.[942] As mentioned before, AI does not necessarily need personal data to generate behavioral

[933] Cf. Müller (2020) P. 1ff.

[934] Cf. Noggle (2018) P. 7ff.

[935] Cf. Human Rights Council (2021) P. 3ff.

[936] Cf. Human Rights Council (2021) P. 3ff.

[937] Cf. Human Rights Council (2021) P. 3ff.

[938] Cf. Human Rights Council (2021) P. 3ff.

[939] Cf. Human Rights Council (2021) P. 3ff.

[940] Cf. European Commission (2021a) P. 1ff.

[941] Cf. Wanbil et al. (2016) P. 1ff.

[942] Cf. Human Rights Council (2021) P. 4ff.

insights about individuals.[943] If an AI gets access to the rights datasets, it can infer movement patterns und draw conclusions about individuals from those movement patterns.[944] These insights can be used to identify religious or political beliefs of individuals and predict the probability of future events.[945]

Additionally, the continuing proliferation of AI surveillance systems can normalize the use of AI surveillance technology and further stimulate the introduction of AI surveillance applications into our everyday lives.[946] Individuals tend to have a low interest in data protection and online privacy in social media.[947] Moreover, younger generations assess the growing reduction of online privacy as a part of the contemporary lives, having the opinion that the exposition of private data is necessary to be a part of the digital world.[948] Societies grow more and more indifferent to AI surveillance applications, reducing the oversight of governmental agencies.[949]

Concerning the issue of singularity, how would we handle an AI surveillance system that outsmarts us? A deep learning algorithm could come to conclusions that were not predicted by its creators, as it does not follow any predefined path while learning, as studied in Section 3.2. This adds to the general peace and security concerns, remember the example of an AI surveillance system that recommends a lock down in a specific area, but also endorsing the idea to kill every individual who breaks the lockdown, that was mentioned in Section 3.4. Apart from dystopian scenarios, such an AI system would have access to an enormous amount of data concerning individuals and institutions. Moreover, such an AI system could predict any attempts at shutting it off, possibly being a threat to humanity if its goals become in contradictive to our goals, a huge ethical challenge that was also discussed in Section 3.4 This may be a scenario for the future, yet it must be tackled and addressed. When designing an AI surveillance system that constantly improves itself without human intervention, shackles must be implemented at some point of the development process. How

[943] Cf. Human Rights Council (2021) P. 4ff.

[944] Cf. Human Rights Council (2021) P. 4ff.

[945] Cf. Human Rights Council (2021) P. 4ff.

[946] Cf. Barth and Jong (2017) P. 1038ff.

[947] Cf. Rainie (2018) P. 1ff.

[948] Cf. Rainie (2018) P. 1ff.

[949] Cf. Barth and Jong (2017) P. 1038ff.

these constraints effect the function of the AI must be kept in mind as well. Additionally, if a highly capable AI surveillance systems is autonomous in its decisions and we leave all the specific decisions, for example arrests, to it, we rely solely on the AI. This could erode human autonomy.[950] However, an AI that completely follows the wishes of its human creators may also be something we should be wary of.[951]

Furthermore, we must ask ourselves if AI based surveillance technology is proportional to its aims. As outlined earlier AI surveillance technology can fight and prevent crime but its application and deployment exposes everyone to the system, regardless if guilty or innocent.[952] This is placing every citizen in the situation of being observed, at least, in public spaces.[953]

Lastly, AI surveillance technology has the power to completely neutralize the anonymity of the internet, that enables individuals to access information without fear of repercussions and to communicate safely.[954] In reality, it is not as easy as sometimes depicted in the general discussion to find out who is responsible for a specific communication or which computer belongs to an individual.[955] AI can eradicate this anonymity via data analysis.

[950] Cf. Boucher (2020) P. 30ff.

[951] Cf. Harari (2018) P. 1ff.

[952] Cf. Fontes and Perrone (2021) P. 8ff.

[953] Cf. Fontes and Perrone (2021) P. 8ff.

[954] Cf. La Rue (2013) P. 13.

[955] Cf. La Rue (2013) P. 13.

5 Rule utilitarianism and AI surveillance technology

This chapter will combine the ethical groundwork and the description of AI surveillance technology to determine whether AI surveillance technology can be used in an ethical way and how this ethical way of AI surveillance may materialize. The selected ethical framework of rule utilitarianism will be applied to investigate, whether AI surveillance can be used in an ethical way. After this, it will be studied how such an application, if it is ethically permissible, may look. However, AI surveillance is just a part of digital tools that could enable repression, information and communication technologies that may be used to intimidate and/or harass opposing individuals.[956] A holistic approach to AI surveillance tools requires to mention that AI is not a standalone system that may enable repression.[957] This must be kept in mind because it signifies that AI surveillance is just another tool that may be used with a malicious intent alongside and in connection with other tools.[958]

5.1 Framework proposal for ethical AI surveillance systems

The advantages of AI surveillance and the possible threats of the malicious use of the technology create a great amount of uncertainty when dealing with AI surveillance applications of any type. Companies and governmental agencies that try to reap the benefits of AI surveillance need a guideline in order to assess the ethical implications of their use of AI surveillance technology. However, the first question that needs to be adressed, before discussing the specifics of an AI surveillance system, is the question of the ethical permissibility of AI based surveillance applications. This question is also the first central question of this thesis. In order to attempt to answer it, the framework for trustworthy AI that was developed by the European Commission´s high-level expert group will be utilized. In this framework, the European Commission established a set of criteria. To meet these criteria, the European Commission defined seven key requirements that enable the assessment of the ethical permissibility of AI surveillance systems and that ensure that the three criteria are fulfilled. Therefore, this framework will allow individuals to reap the benefits of AI surveillance systems while also ensuring the ethical use of this immensely powerful technology,

[956] Cf. Feldstein (2019) P. 16ff.

[957] Cf. Feldstein (2019) P. 16ff.

[958] Cf. Feldstein (2019) P. 16ff.

hereby increasing the utility of a specific population. The framework may be able to prevent harm by creating an ethical AI surveillance system that is in accordance with basic human rights thereby adhering to justified moral rules that are internalized in our codex of moral rules.[959] From a rule utilitarian point of view, that means that the rules for AI that can be inferred from the framework are in accordance with our codex of moral rules, which means that the utility rises in comparison to a state of affairs in which no regulation is established. An AI surveillance system that may be regarded as ethically permissible and trustworthy must adhere to three criteria during its entire lifecycle.[960] This adherence is ensured via seven key requirements the ethical AI surveillance system must satisfy.

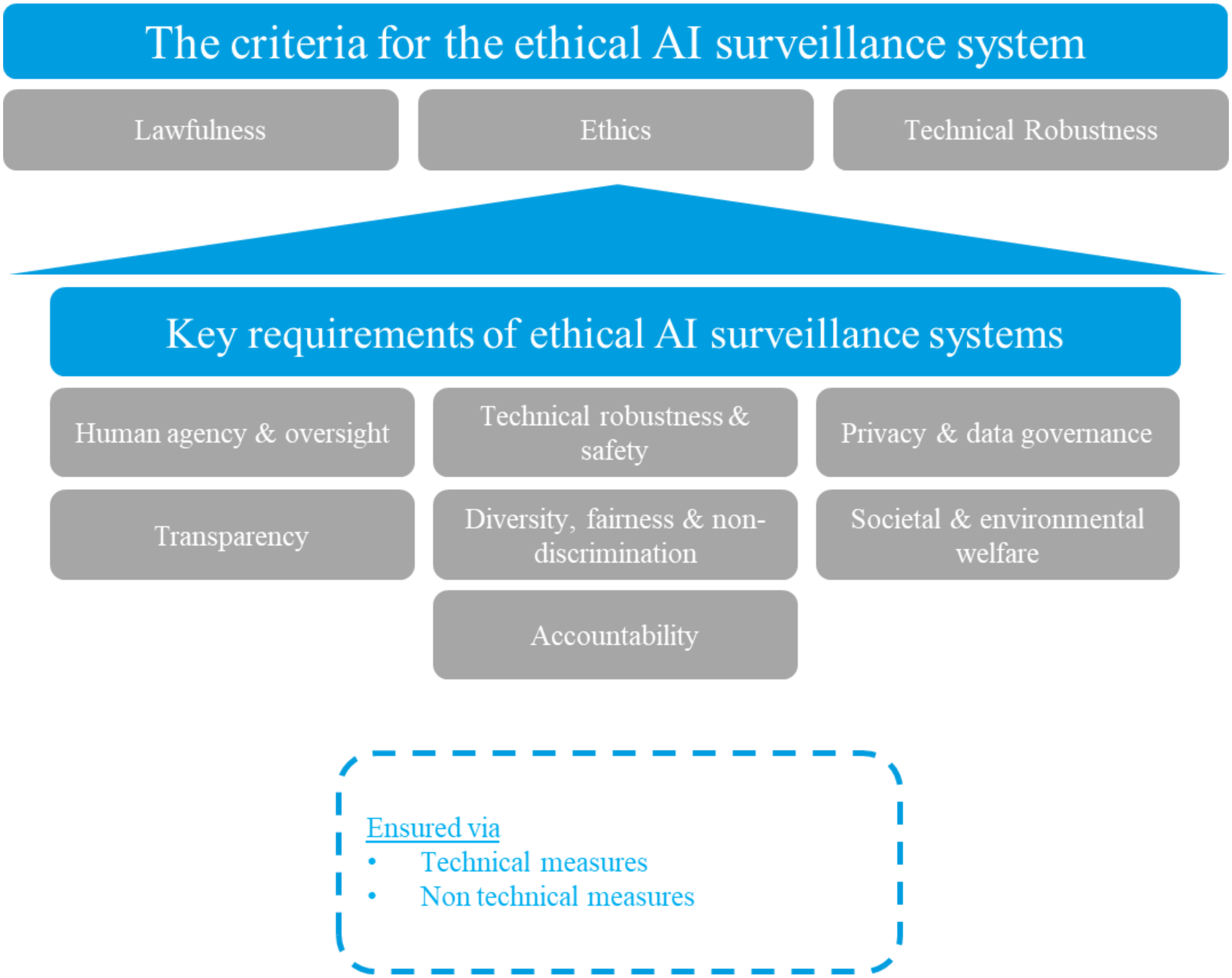
1. The system must function in compliance with codified law, both national and international.
2. The system itself must be ethical.
3. The system must be technically robust to not cause unintentional harm.

If an AI meets all three criteria, it may be regarded as ethically permissible, trustworthy and from the perspective of rule utilitarianism, utility increasing compared to a state in which no regulation exists. The utility is increasing because the set of rules allows for the utilization of the advantages of AI surveillance systems while probably providing a clear roadmap of how to steer clear of the pitfalls and ethical challenges while being in accordance with our moral codex of what is right and what is wrong. Figure 8 illustrates the framework for an ethical AI surveillance system.

[959] Cf. Smuha et al. (2019) P. 6ff.

[960] Cf. Sahin (2020) P. 11.

Figure 8 Framework for an ethical AI surveillance system


The criteria for the ethical AI surveillance system
Lawfulness
Ethics
Technical Robustness
Key requirements of ethical AI surveillance systems
Human agency & oversight
Technical robustness & safety
Privacy & data governance
Transparency
Diversity, fairness & non-discrimination
Societal & environmental welfare
Accountability
Ensured via
• Technical measures
• Non technical measures

Source: Own illustration according to Smuha et al. (2019) P. 8

The framework defines seven key requirements that are crucial for the realization of ethically permissible AI surveillance systems. These key requirements can be ensured via technical and non-technical measures such as an ethically permissible AI architecture and robust regulation.[961] The seven key requirements are according to the high-level-expert group of the European commission are:

[961] Cf. Smuha et al. (2019) P. 6ff.

- **Human agency and oversight:** Completely automated decision-making processes must not exist, moreover, human operators must always be able to critically assess the AI.[962] Humans must constantly have a capability to act. This means that human operators are necessarily in the loop or in command, allowing for intervention at every step of the decision-making process.[963]
- **Technical robustness and safety:** In this thesis, this is defined as a preventative approach to technical risks with the goal of designing a predictable AI, predictable in the sense that the AI surveillance system behaves as intended by its creators.[964] Additionally, technical robustness and safety incorporates the resilience of an AI surveillance system against attacks from the outside.
- **Privacy and data governance:** The definition of privacy was laid out om Section 4.3. To recap: privacy requires that every individual should have an area of independent development, liberty and interaction without outside intrusion.[965] Data governance incorporates the enforcement of standards and procedures of how the data is sourced, used and distributed and saved. It describes the management of the availability, usefulness, integrity and safety of obtainable data.[966]
- **Transparency:** Transparency describes an insight into the code, logic, model goals, variables and input-output relationship of the AI system.[967]
- **Diversity, fairness and non-discrimination:** Diversity may be defined as the occurrence of variances within a group of individuals.[968] Many different definitions of fairness exist. In this thesis fairness will be defined as treating similar individuals in a similar way.[969] The absence of discrimination is understood as the just and fair treatment of all individuals.

[962] Cf. Smuha et al. (2019) P. 13ff.

[963] Cf. Boucher (2020) P. 39ff.

[964] Cf. Smuha et al. (2019) P. 14ff.

[965] Cf. La Rue (2013) P. 6ff.

[966] Cf. Stedman and Vaughan (2022) P. 1ff.

[967] Cf. Koene et al. (2019) P. 4.

[968] Cf. Tan (2019) P. 1ff.

[969] Cf. Saxena et al (2019) P. 99ff .

- **Societal and environmental welfare:** This requirement states that the AI surveillance system should also incorporate the broader society, other living beings and the environment into account in order to benefit all individuals, including future generations.[970][971]
- **Accountability:** In the context of this thesis accountability describes that a set of mechanisms, characteristics and best practices that form a holistic structure of governance. This structure of governance ensures that humans will always be accountable for the decisions of the ethical AI surveillance system, guaranteeing that humans are responsible for the results of the ethical AI surveillance system.[972]

These seven key requirements must be addressed and continuously certified through the complete life cycle of the AI surveillance system via technical and non-technical measures.[973] They must be followed by an AI surveillance system to make sure that it meets the three main criteria and may therefore be regarded as ethically permissible and trustworthy. Yet, these requirements have to be examined for their potential of increasing utility.

Human agency and oversight is a requirement that demands AI surveillance systems act in accordance with basic human rights, fostering these fundamental rights and thereby also obeying to internalized moral rules.[974] Additionally, it promotes human agency enabling individuals to critically challenge the assumptions of the AI system.[975] That constitutes that the outputs of the ethical AI surveillance system cannot be used effectively unless they have been verified by a human.[976] This includes that the AI must not manipulate or coerce individuals in any way, which would reduce their agency.[977] The oversight aspect of this refrains the AI surveillance system from undermining

[970] Cf. Smuha et al. (2019) P. 19ff.

[971] Cf. Ligozat et al. (2021) P. 4ff.

[972] Cf. Koene et al. (2019) P. 4.

[973] Cf. Larbey et al. (2020) P. 73ff.

[974] Cf. Larbey et al. (2020) P. 73ff.

[975] Cf. Smuha et al. (2019) P. 13ff.

[976] Cf. European Commission (2020) P. 21.

[977] Cf. Smuha et al. (2019) P. 13ff.

human autonomy which could cause adverse effects.[978][979] An AI surveillance system should not be used, if this use is not increasing the utility of every involved individual. The utility may not be increased if the use of the AI surveillance system results in unnecessary harm to individuals or if the AI systems infringes upon basic human rights as these rights are based on the moral codex and therefore internalized. The human interventions must always be able to override decisions made by the AI surveillance system.[980] For the ethical AI surveillance system, this could result in the establishment of a control and command center in which human operators control every aspect of the operation of the AI surveillance system. An AI surveillance oversight institution could further enhance human agency and divide the oversight between the different institutions.

Ensuring the trust into the outputs of an AI surveillance system, the system must meet the requirement of technical robustness and safety.[981] This is connected to the principle of prevention of harm.[982] It should not come to unpredicted conclusions such as the lockdown example mentioned in Section 3.4. An essential component of this requirement is the resilience to attacks from third parties and the security of the AI surveillance systems.[983][984] These attacks from third parties can be aimed at the databases of the AI system, the operating model and/or the underlying infrastructure of the system.[985][986] A third party must not be allowed to enter the database and extract the personal data of individuals. This is especially important for AI surveillance systems such as smart policing applications, as they rely on a huge amount of valuable historical personal data to infer predictions about crimes. If someone changes this data, the outputs of the AI could reduce the utility for everyone involved. Police forces could make false arrests based on manipulated data. It is important for all the other systems of AI surveillance applications as well. Additionally, a change in behavior

[978] Cf. European Commission (2020) P. 15.

[979] Cf. Smuha et al. (2019) P. 13ff.

[980] Cf. Larbey et al. (2020) P. 73ff.

[981] Cf. European Commission (2020) P. 20.

[982] Cf. Smuha et al. (2019) P. 14ff.

[983] Cf. European Commission (2020) P. 21.

[984] Cf. Boucher (2020) P. 39ff.

[985] Cf. Smuha et al. (2019) P. 14ff.

[986] Cf. Boucher (2020) P. 37ff.

may be forced this way, thus alternating the decisions made by the AI surveillance system.[987] The corporeal elements of the AI should be protected for example. Copies of essential elements of the operating model might be stored in networks that are not connected to the internet, thereby enabling the operators of the system to use these copies as a failsafe. Nevertheless, AI surveillance systems must be designed for their specific purpose, a dual-use of the system in other applications should not be possible.[988] Dual-use could result in the utilization of the AI system in a way that reduces utility. Last of all, the AI surveillance system must exhibit reliability and reproducibility.[989] Reliability describes that the system works the right way with different inputs in different situations, thus giving its the creators the chance to prevent unintended harm.[990] For surveillance operations this is particularly vital as they deal with a myriad of different inputs and different situations, e.g., facial recognition systems have to scan a huge number of faces on different times a day. Reproducibility deals with the outputs of the AI system. These outputs must always be the same, if the input parameters remain unchanged.[991]

A further key requirement that is connected to the principle of prevention of harm is the requirement for privacy and data governance.[992] The personal data that fuels an AI surveillance system as well as the insights the systems generate about individuals must be protected throughout the entire lifecycle of the AI system.[993] The General Data Protection Regulation must be taken into account when dealing with personal data.[994] The access to the data should only happen if needed and only qualified personnel should be allowed to conduct this access.[995] Additionally, the quality and integrity of the data must be guarded as these aspects are paramount to the performance of the AI

[987] Cf. Smuha et al. (2019) P. 14ff.

[988] Cf. Smuha et al. (2019) P. 14ff.

[989] Cf. Smuha et al. (2019) P. 14ff.

[990] Cf. Smuha et al. (2019) P. 14ff.

[991] Cf. Smuha et al. (2019) P. 14ff.

[992] Cf. Smuha et al. (2019) P. 14ff.

[993] Cf. European Commission (2020) P. 19.

[994] Cf. Boucher (2020) P. 24ff.

[995] Cf. Smuha et al. (2019) P. 14ff.

surveillance system.[996] For example, a reporting standard could be established that permits the valuation of the data quality and integrity in a regular interval.

The next key requirement is transparency which shows a strong connection with the principle of explicability.[997] Transparency incorporates the traceability, explainability and the communication of the AI with individuals.[998] Regarding traceability, the structure of the data sets, beginning with the sourcing of the data, and how this data is used by the AI surveillance system to come to its output must be documented.[999][1000] This documentations should be published in regular intervals, creating a reporting standard that enhance the transparency of the AI surveillance system. This also shows synergy effects with auditability and oversight. Explainability is concerned with communicating the technical aspects of the AI and the decisions of its creators, e.g., to monitor a specific part of a city, to everyone who is affected by this surveillance in any way. [1001][1002][1003] These explanations should be tailored to the different groups that are targeted by the monitoring. The last aspect of transparency is communication.[1004] Upon communicating with individuals, the AI should not mask itself as human as humans have a right to know if they are communicating with an artificial system.[1005][1006] The abilities, limitations and shackles of the AI surveillance system must be communicated as well.[1007] The whole approach to communication should be human centered.[1008] A reporting standard that must be published and that contains high-level technical details of the ethical AI surveillance system are possible here as well.

[996] Cf. Smuha et al. (2019) P. 14ff.

[997] Cf. Smuha et al. (2019) P. 14ff.

[998] Cf. Smuha et al. (2019) P. 14ff.

[999] Cf. Boucher (2020) P. 39ff.

[1000] Cf. Larbey et al. (2020) P. 47ff.

[1001] Cf. Boucher (2020) P. 39ff.

[1002] Cf. Smuha et al. (2019) P. 14ff.

[1003] Cf. Larbey et al. (2020) P. 33ff.

[1004] Cf. Smuha et al. (2019) P. 14ff.

[1005] Cf. Smuha et al. (2019) P. 14ff.

[1006] Cf. Boucher (2020) P. 39ff.

[1007] Cf. Smuha et al. (2019) P. 14ff.

[1008] Cf. European Commission (2020) P. 3.

With the aim of creating an AI that possibly increases the utility of everyone involved, the requirement of diversity, fairness and non-discrimination must be met as well to ensure the equal treatment of all the individuals affected. This equal treatment makes sure that the possible maximization of utility incorporates every individual. Equal rights and equal treatment are paramount to this requirement, linking it closely to the principle of fairness.[1009] Unfair bias must be avoided, regardless of how they find their way into the AI surveillance system.[1010] The ethical AI surveillance system must be trained with real life data, while closely monitoring the system for any biases. Once biases are found they must be rooted out and the way in which they entered the system must be traced to prevent the biases from entering again. A robust testing and validation regime should be established in the early stages of training. Unintended discrimination and prejudices would reduce utility as these discriminations would not lead to the greatest good. Additionally, bias reduces equality. justice and fairness.

Societal and environmental welfare, as a requirement, is connected to the principle of fairness and the principle of prevention of harm.[1011] The ethical AI surveillance system should be trained to make choices that have a positive impact on the environment.[1012] The life cycle and the supply chain of the AI system must be set up as ecologically as possible.[1013][1014] Including its power supply. Societal welfare describes that the holistic use of AI surveillance systems should be strictly monitored in regard to its impact on the social behavior of individuals.[1015] During the early development of the AI, it must be ensured that no individuals are exploited while gathering the rare resources that are required. This could be managed by another oversight institution that also regulates the extraction of resources that are required to build an AI. Furthermore, it necessitates that institutions, society and democracy must be understood by the ethical AI surveillance system.[1016] This includes

[1009] Cf. Smuha et al. (2019) P. 14ff.

[1010] Cf. Smuha et al. (2019) P. 14ff.

[1011] Cf. Smuha et al. (2019) P. 19ff.

[1012] Cf. European Commission (2020) P. 5.

[1013] Cf. Ligozat et al. (2021) P. 4ff.

[1014] Cf. Smuha et al. (2019) P. 19ff.

[1015] Cf. Smuha et al. (2019) P. 19ff.

[1016] Cf. Smuha et al. (2019) P. 19ff.

democratic and electoral processes.[1017] Similar to the human decision-making process, democracy must not be eroded by the AI surveillance system, as democracy is in itself utility-maximizing, as mentioned in Section 2.2.1. This raises the probability that the framework increases utility for everyone involved.

The last requirement is accountability, it is closely linked to the principle of fairness.[1018] It demands that mechanisms are created and implemented that ensure the responsibility and accountability of the AI surveillance system.[1019][1020] An example for this could be that a human operator must always be the one to make the final decision. Auditability is also an aspect in this context, that allows the assessment of algorithms, data and strategy processes of the AI.[1021][1022] Trade-offs must be approached cautiously, making sure that trade-offs that do not increase utility are not implemented. Consequently, trade-offs must always be chosen in the way that increases utility by adhering to internalized moral rules that are expressed via this framework.

The framework makes it possible that the three criteria for trustworthy and ethically permissible AI surveillance systems are met. These requirements aim at increasing the utility, justice, fairness and equality of every individual involved, as they are selected according to internalized moral rules such as basic human rights and the values democracies uphold. Basic human rights and the values of democracies may be defined as justified and internalized moral rules, as they increase utility in comparison with other moral rules or the absence of moral rules, as mentioned in Section 2.2.1 **This infers that an AI surveillance system can be ethically justified, if it is designed in accordance to the presented framework, which answers the first question of this thesis.** Subsequently this system must be designed in adherence to the seven key requirements.

[1017] Cf. Smuha et al. (2019) P. 19ff.

[1018] Cf. Smuha et al. (2019) P. 19ff.

[1019] Cf. Smuha et al. (2019) P. 19ff.

[1020] Cf. Larbey et al. (2020) P. 47ff.

[1021] Cf. Smuha et al. (2019) P. 19ff.

[1022] Cf. European Commission (2020) P. 25.

5.2 Technical and non-technical approaches to the ethical AI surveillance system

After having discussed the framework proposal for ethical AI surveillance systems, it will now be examined how to ensure the adherence to the key requirements of the framework. In general, technical and non-technical approaches exist. The technical approaches begin with the architecture of an ethically permissible AI surveillance system.[1023] The system architecture must reflect the seven requirements of ethically permissible AI surveillance systems, as they are in adherence with internalized moral rules. A way to realize this could be a set of whitelist rules that the system must obey at any time.[1024] Parallel to the whitelist rules, a set of blacklist rules should be implemented as well, representing rules the system must never break.[1025] These lists could be oversighted by an ethics board that continuously revises and improves them. A separate process should continuously monitor the systems obedience to these rules. It may also prove beneficial to introduce a third set of rules, greylist rules, to enhance the flexibility of the system. These rules may be broken by the system if a human operator authorizes the system to do so. The human operator could be a judge who issues a warrant, based on the grounds of sufficient evidence, for limited telecommunications surveillance of a specific individual because this individual may be a threat to national security or public order, as this individual is planning a terrorist attack. This case would also pose a restriction to the basic human right to privacy as the codified law would allow such an action, the principle of proportionality would be met and legitimate aims would appropriate the AI surveillance action. AI engineers and designers must ensure that the system is constrained in its ability to draw conclusions, they would have to implement these restrictions early in the development process. These constraints should not decrease the performance of the AI surveillance system, moreover, they should be designed in a way that the system cannot draw conclusions that transgress the seven requirements. Thus, ensuring that the AI surveillance system obeys to ethics and rule of law by design. This requires that the system comprehends the abstract principles of ethics and law and

[1023] Cf. Smuha et al. (2019) P. 20ff.

[1024] Cf. Smuha et al. (2019) P. 20ff.

[1025] Cf. Smuha et al. (2019) P. 20ff.

recognizes their connection to its activity and purpose.[1026] Additionally, the AI surveillance system must have a fail-safe mechanism and a mechanism to force a shut-down at any time.[1027][1028]

In order to implement transparency, the ethical AI surveillance system must incorporate a structure that allows individuals to comprehend its decision-making processes (transparency, traceability). However, this may be a fine line as artificial neural networks, especially deep leaning networks, can be incredibly complex, as discussed in Section 3.2, and any structure that aims at explaining these networks is consequently complex as well. Furthermore, the structure may hinder the performance of the ethical AI surveillance system which would in turn mean that the utility may not be increased by the system. Further research must be conducted in this field to find a fitting solution. A current and promising field of research is the field of Explainable AI.[1029] Nevertheless, any ethical AI surveillance system must be tested with realistic training data, ensuring that it becomes more predictable once it is practically implemented.[1030][1031] The testing and validation must be started as early as possible by a wide group of individuals.[1032] During this testing it should be deliberately attempted to break into the system to uncover vulnerabilities.[1033]

Non-technical approaches are an important part of the set-up of the ethical AI surveillance system. These non-technical approaches are designed to secure and maintain the ethical permissibility of the ethical AI surveillance system; thus, it is preferable to conduct these approaches on a continuous basis.[1034] Regulation is the first non-technical approach, as it already exists today.[1035] Regulation needs to be enforced by neutral agents.[1036] For example, if the executive branch of a government deploys AI surveillance, it must be regulated by the judicative and the legislative. Otherwise,

[1026] Cf. Smuha et al. (2019) P. 20ff.

[1027] Cf. Smuha et al. (2019) P. 20ff.

[1028] Cf. Larbey et al. (2020) P. 70ff.

[1029] Cf. Smuha et al. (2019) P. 21ff.

[1030] Cf. Smuha et al. (2019) P. 21ff.

[1031] Cf. Larbey et al. (2020) P. 36ff.

[1032] Cf. Smuha et al. (2019) P. 21ff.

[1033] Cf. Smuha et al. (2019) P. 21ff.

[1034] Cf. Smuha et al. (2019) P. 21ff.

[1035] Cf. European Commission (2021b) P. 5ff.

[1036] Cf. European Commission (2021b) P. 5ff.

it may be abused which would prove detrimental to the maximization of utility because it would definitely not maximize justice, fairness or equality. This oversight could also be provided by an impartial organization that is created with the intent of certifying the ethical use of AI surveillance technology, thus possibly enhancing accountability.[1037] Accountability can also be increased by tailoring the framework examined in this thesis to the needs of an organization or institution. The use of AI surveillance technology by companies must also be regulated, laws and guidelines play a paramount role in this context.[1038][1039] Guidelines and laws must be monitored via effective key performance indicators.[1040] The organization that deploys the ethical AI surveillance system should define the intentions and goals of this deployment and infer standards to ensure effective human control of the AI.[1041][1042] Effective standards should be based on the framework and on internalized moral rules to maximize utility. Further tools of regulation can be accreditation systems, professional codes of ethics and basic human rights which were discussed in Section 4.3.[1043] As mentioned, these standards must always adhere to internalized moral rules.

To summarize, the ethical AI surveillance system must adhere to the three criteria that were developed by the European Commission (lawfulness, ethical in itself and technical robustness) of ethical AI deployment **by following the seven requirements**, which were also developed by the European Commission[1044], human agency & oversight, technical robustness & safety, privacy & data governance, transparency, diversity, fairness and non-discrimination, societal & environmental welfare and accountability, that build the foundation of these criteria. **These criteria and subsequently the requirements are based on internalized moral rules thereby possibly increasing utility, equality, justice and fairness compared to a state in which no rules are in effect.**

[1037] Cf. Smuha et al. (2019) P. 22ff.

[1038] Cf. European Commission (2021b) P. 1ff.

[1039] Cf. Larbey et al. (2020) P. 51ff.

[1040] Cf. Smuha et al. (2019) P. 22ff.

[1041] Cf. Smuha et al. (2019) P. 22ff.

[1042] Cf. Larbey et al. (2020) P. 51ff.

[1043] Cf. Smuha et al. (2019) P. 22ff.

[1044] Cf. Smuha et al. (2019) P. 4ff.

Specifically, the ethical AI system must have a robust architecture of its underlying artificial neural network, incorporating white-black- and greylist rules. The learning nature of the artificial neural network of the ethical AI surveillance system must be taken into account by introducing restrictions that enhance the predictability of the system. The ethical AI surveillance system must understand the abstract constructs of ethics and law and make the link from these abstract constructs to the actions it engages in. Failsafe, shut-down and explanation structures are mandatory as well. Validation and testing must start the dawn of the ethical AI system, using realistic data to train it to enhance the predictability of the system. Key performance indicators must be defined and monitored to control the performance of the ethical AI surveillance system. All of these technical approaches guarantee that the ethical AI surveillance system obeys internalized moral rules which is the optimal condition in rule utilitarianism. Non-technical approaches of the ethical AI surveillance system deal with regulation which must be based on a system of checks and balances ensuring impartial oversight, tailored ethical frameworks, standards that increase utility, fairness equality and justice and methods of certification.

Yet, it will be extremely difficult to bring the ethical AI surveillance system into accordance with environmental regulations, especially regarding the raw resources that are required and the power supply of the ethical AI surveillance system. Moreover, authoritarian states may choose not to limited themselves in their application of AI surveillance technology, which still makes the governmental abuse of powers, as mentioned in Section 4.4, possible. Also, data breaches remain a security concern.

The proposed framework can be combined with the risk-based AI regulatory framework proposal of the EU, which was explained in Section 3.5, to further increase the values rule utilitarianism upholds. While the framework for the ethical AI surveillance system can be applied to design and engineer the ethical AI surveillance system, the regulatory framework proposal can be used to assign risk-categories to the AI surveillance systems, also delivering clear guidelines for the deployment of the system based on its risk category.

6 Conclusion

This thesis studied the ethical implications of AI-based mass surveillance tools. It examined two central questions. The first one was whether AI surveillance systems and tools can be used in and ethically permissible way. The second central question asked, was how such an ethically permissible system may materialize. To form holistic arguments to answer these questions, the basics of possible ethical frameworks was discussed. Theories of normative ethics were discussed and the scope was focused on utilitarianism. Utilitarianism strives to find the ethically right action via the maximization of a metric defined as utility, with utility being the amount of happiness or reduction of pain an action creates for everyone affected. Parallel to utility, the theory also tries to maximize justice, fairness and equality. Act and rule utilitarianism were discussed as well. They represent two approaches to practically deploy utilitarianism. Rule utilitarianism aims at maximizing utility via a set of rules that is chosen by their net consolidated benefits. The general acceptance of this set of rules maximizes utility, fairness, equality and justice. Deontological theories were examined as well. In this context, deontological pluralism, according to William David Ross was studied.

After these foundations, the advantages and disadvantages of the theories were discussed regarding their application in this thesis. In the end, rule utilitarianism was chosen as it is adaptable into a framework of ethical AI surveillance technology, due to its incorporation of internalized moral rules which dictate the guidelines for ethical AI surveillance systems.

Afterwards, AI was the focus of discussion. The problems of defining AI were studied and a definition of intelligence was sketched out. From this definition of intelligence, a high-level definition of AI was inferred, defining AI as intelligence exhibited by machines. From this, this thesis´s general understanding of AI was deducted, seeing AI as the capability of a computer to successfully complete assignments that would commonly demand human intelligence. Next, data and its connection to AI was illustrated, identifying data as the heart of AI. Following this, AI machine learning and deep learning were examined in depth, differentiating them from each other while shining a light on artificial neural networks and how they function. Supervised- and unsupervised learning were defined, while also briefly discussing reinforcement learning and hybrid learning. Backpropagation and the gradient descent were explained. Probability, as an important influence on AI, was reviewed as well.

The possible uses of AI in general and its advantages were examined next. Afterwards, some ethical challenges AI creates were investigated as well. AI can be the solution to almost every challenge in every field humanity faces today. The current pace of AI development and its regulation were the next center of the discussion. In this aspect, the ever-rising number of published scientific papers on AI was established as a metric to understand the pace of development. A huge increase was detected, showing the huge attention AI research enjoys in science. A possible regulation approach is the AI regulatory framework, that offers a risk-based approach to sort AI into different categories that have to fulfill different criteria in order to be certified for use.

Based on this AI surveillance technology was examined. **AI shows a lot of potential in surveillance technology, possibly ushering in a new age of public safety and national security.** Furthermore, it can help tackle a myriad of challenges the cities and communities of tomorrow may face. Authoritarian regimes are key players in the proliferation of AI surveillance technology. Chinese companies such as Huawei are leaders of the global market for AI surveillance applications, also enhancing the Chinese sphere of influence. Liberal democracies are also heavily exporting and utilizing smart city and behavioral/facial recognition systems today. Moreover, companies that are based in liberal democracies supply Chinese companies with enabling technology.

The connection between basic human rights, such as the right to privacy was discussed. It must be stated that AI surveillance actions could happen in accordance with this basic human right if the principle for restrictions of the right are met. Yet, these laws are subject to restrictions, if certain requirements are met. These restrictions can be a source of repression in the future as governmental agencies can easily create circumstances in which these demands for the restriction are met.

These developments are the harbinger of a vast number of ethical challenges that can lead to the use of AI surveillance technology in way that allows governmental agencies to repress the population of a respective country. Further ethical challenges are privacy infringements, the manipulation of individuals based on the insights of the AI surveillance system, bias, eroding of human decision authority and singularity related aspects.

In order to asses the ethical permissibility of AI surveillance systems, a framework that is based on rule utilitarianism was adapted from the regulatory framework proposal of the high-level-expert group of the European Commission. It consists of three central criteria which can possibly lead to the establishment of the ethical AI surveillance system, which are:

1. The system must function in obedience with codified law, both national and international.
2. The system itself must be ethical.
3. The system must be technically robust to not cause unintended harm.

These criteria are met via a set of seven key requirements. These key requirements ensure the ethical use of AI surveillance technology. They can be enforced via technical and non technical measures. **To answer the first central question, an AI surveillance system that adheres to the key requirements and therefore to the three central criteria may be ethically permissible**. To answer, the second central question, the tangible conception of the ethical AI system must be ensured by an AI architecture that is in accordance with basic laws and fundamental concepts of ethics. Additionally, the ethical AI surveillance system must adhere to the seven key requirements of the framework.

AI surveillance technology has unrivaled potential to improve society as a whole. However, the pitfalls are steep, authoritarian governments can utilize AI surveillance systems to exercise repression and data breaches will always remain a security concern. The racing proliferation of AI surveillance technology mandates us to find a way to ethically deploy this technology. The presented framework of this thesis should be regarded as a start for further considerations of the regulation of AI surveillance technology. However, the efforts of regulation must catch up with the development and proliferation of this technology in order to reap its astounding benefits.

References

Adams, R.l. 2017. 10 Powerful Examples Of Artificial Intelligence In Use Today. *Forbes*, 10 January. https://www.forbes.com/sites/robertadams/2017/01/10/10-powerful-examples-of-artificial-intelligence-in-use-today/?sh=29b35012420d. Accessed: 21 April 2022.

Adler, Rasmus. 2019. Autonomous or merely highly automated – what is actually the difference? *Fraunhofer IESE*, 13 December. https://www.iese.fraunhofer.de/blog/autonomous-or-merely-highly-automated-what-is-actually-the-difference/. Accessed: 15 April 2022.

Akkad, Dania. 2012. Human Rights: The Universal Declaration vs The Cairo Declaration. https://blogs.lse.ac.uk/mec/2012/12/10/1569/. Accessed: 2 June 2022.

Alexander, Larry, and Michael Moore. 2007. Deontological Ethics. https://plato.stanford.edu/entries/ethics-deontological/. Accessed: 31 March 2022.

Andersen, Ross. 2020. China's Artificial Intelligence Surveillance State Goes Global. *The Atlantic*, 29 July. https://www.theatlantic.com/magazine/archive/2020/09/china-ai-surveillance/614197/. Accessed: 11 May 2022.

Andre, Claire, and Manuel Velasquez. 1991. Who Counts? https://www.scu.edu/mcae/publications/iie/v4n1/counts.html. Accessed: 26 March 2022.

Armstrong, Martin. 2019. The State of Democracy. *Statista*, 19 July. https://www.statista.com/chart/18737/democracy-index-world-map/. Accessed: 30 May 2022.

Attfield, Robin. 2012. Normative Ethics. In *Ethics. An overview*, Ed. Robin AttfieldLondon, New York: Continuum.

Barth, Susanne, and Menno D.T. de Jong. 2017. The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review. *Telematics and Informatics* 34 (7): 1038–1058. doi: 10.1016/j.tele.2017.04.013.

Bentham, Jeremy. 1781. *An Introduction to the Principles of Morals and Legislation*.

Bohai, Luhai. 2021. Ethical Concerns of Combating Crimes with AI Surveillance and Facial Recognition Technology. *Towards Data Science*, 3 June. https://towardsdatascience.com/ethical-concerns-of-combating-crimes-with-artificial-intelligence-surveillance-and-facial-a5eb7a09abb1. Accessed: 11 May 2022.

Bossmann, Julia. 2016. Top 9 ethical issues in artificial intelligence. https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/. Accessed: 7 March 2022.

Boucher, Philip. 2020. *Artificial intelligence. How does it work, why does it matter,and what can we do about it?* Brussels: European Parliament.

Brownlee, Jason. 2019. Understand the Impact of Learning Rate on Neural Network Performance. *Machine Learning Mastery,* 24 January. https://machinelearningmastery.com/understand-the-dynamics-of-learning-rate-on-deep-learning-neural-networks/. Accessed: 27 May 2022.

Burton, Robert A. 2015. How I Learned to Stop Worrying and Love A.I. https://perma.cc/EU3V-QEV4. Accessed: 5 May 2022.

Cane, Peter, and Joanne Conaghan. 2009. *The new Oxford companion to law*. Oxford: Oxford Univ. Press.

Cerutti, Furio. 2017. *Conceptualizing politics. An introduction to political philosophy*. Abingdon, Oxon, New York, NY: Routledge.

Coakley, Mathew. 2017a. Consequentialism and the moral agent question. In *Motivation Ethics*, Ed. Mathew CoakleyBloomsbury Publishing Plc.

Coakley, Mathew (Ed.). 2017b. *Motivation Ethics*. Bloomsbury Publishing Plc.

Crimmins, James E. (Ed.). 2017. *Bloomsbury encyclopedia of utilitarianism*. Place of publication not identified: Bloomsbury Academic.

Cummings, M.L. 2017. *Artificial Intelligence and the Future of Warfare*. London.

Delua, Julianna. 2021. Supervised vs. Unsupervised Learning: What’s the Difference?, 12 March. https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning. Accessed: 24 May 2022.

Dobelli, Rolf. 2011. *Die Kunst des klaren Denkens. 52 Denkfehler, die Sie besser anderen überlassen*. München: Hanser.

Duignan, Brian, and Henry R. West. 2020. Utilitarianism - Effects of utilitarianism in other fields. https://www.britannica.com/topic/utilitarianism-philosophy/Effects-of-utilitarianism-in-other-fields. Accessed: 15 May 2022.

Dumbrava, Costica. 2021. *Artificial intelligence at EU borders. Overview of applications and key issues*. Brussels.

Eliasy, Ashkan, and Justyna Przychodzen. 2020. The role of AI in capital structure to enhance corporate funding strategies. *Array* 6: 100017. doi: 10.1016/j.array.2020.100017.

European Commission. 2019. Intelligent Portable Border Control System | iBorderCtrl Project | Fact Sheet | H2020 | CORDIS | European Commission. https://cordis.europa.eu/project/id/700626. Accessed: 11 May 2022.

European Commission. 2020. WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1603192201335&uri=CELEX:52020DC0065. Accessed: 15 April 2022.

European Commission. 2021a. *Ethics and data protection*. Brussels.

European Commission. 2021b. *REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS*.

European Parliament. 2020. What is artificial intelligence and how is it used? https://www.europarl.europa.eu/news/en/headlines/society/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used. Accessed: 15 April 2022.

European Parliament. 2022. Regulatory framework on AI. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai. Accessed: 7 March 2022.

Eykholt, Kevin, Ivan Evtimov, Earlence Fernandes, Bo Li, Amir Rahmati, Chaowei Xiao, Atul Prakash, Tadayoshi Kohno, and Dawn Song. 2017. *Robust Physical-World Attacks on Deep Learning Models*.

Feldstein, Steven. 2019. *The Global Expansion of AI Surveillance*. Washington D.C.

Fontes, Catarina, and Christian Perrone. 2021. *Ethics of surveillance: harnessing the use of live facial recognition technologies in public spaces for law enforcement*.

Frantz, Erica, and Natasha M. Ezrow. 2011. *Dictators and dictatorships. Understanding authoritarian regimes and their leaders*. New York, NY, London: Continuum.

Freund, Yoav, and Robert E. Schapire. 1999. Large Margin Classification Using the Perceptron Algorithm. *Machine Learning* 37 (3): 277–296. doi: 10.1023/A:1007662407062.

Fumo, David. 2017. Types of Machine Learning Algorithms You Should Know. *Towards Data Science,* 15 June. https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861. Accessed: 24 May 2022.

Ghahramani, Zoubin. 2015. Probabilistic machine learning and artificial intelligence. *Nature* 521 (7553): 452–459. doi: 10.1038/nature14541.

Goodfellow, Ian, Aaron Courville, and Yoshua Bengio. 2016. *Deep learning*. Cambridge, Massachusetts: The MIT Press.

Goodfellow, Ian J., Jonathon Shlens, and Christian Szegedy. 2014. *Explaining and Harnessing Adversarial Examples*.

Green, Brian Patrick. 2020. Artificial Intelligence and Ethics: Sixteen Challenges and Opportunities. https://www.scu.edu/ethics/all-about-ethics/artificial-intelligence-and-ethics-sixteen-challenges-and-opportunities/. Accessed: 25 May 2022.

Gurney, Kevin. 2014. *An Introduction to Neural Networks*. Hoboken: Taylor and Francis.

Harari, Yuval Noah. 2018. Yuval Noah Harari on Why Technology Favors Tyranny. *The Atlantic,* 30 August. https://www.theatlantic.com/magazine/archive/2018/10/yuval-noah-harari-technology-tyranny/568330/?gclid=EAIaIQobChMIo63o7OXw6gIVkYr-ICh36mAW_EAAYASAAEgLEs_D_BwE. Accessed: 11 May 2022.

Harari, Yuval Noaḥ. 2016. *Homo Deus. A brief history of tomorrow*. London: Harvill Secker.

Hawking, Stephen. 2018. *Brief answers to the big questions*. New York: Bantam Books.

Hooker, Brad. 2003. Rule Consequentialism. https://plato.stanford.edu/entries/consequentialism-rule/#TwoWayArgForRulCon. Accessed: 12 April 2022.

Hooker, Brad, Elinor Mason, and Dale E. Miller (Eds.). 2022. *Morality, Rules, and Consequences. A Critical Reader*. Edinburgh: Edinburgh University Press.

Human Rights Council. 2021. *Annual report of the United Nations High Commissioner for Human Rights and reports of the Office of the High Commissioner and the Secretary-General. Promotion* *and protection of all human rights, civil, political, economic, social and cultural rights, including the right to development*.

Human Rights Watch. 2018. China: Big Data Fuels Crackdown in Minority Region. https://www.hrw.org/news/2018/02/26/china-big-data-fuels-crackdown-minority-region. Accessed: 28 April 2022.

IBM. 2020. Deep Learning, 1 May. https://www.ibm.com/cloud/learn/deep-learning?mhsrc=ibm-search_a&mhq=deep%20learning. Accessed: 24 May 2022.

IBM. 2021a. Overfitting. https://www.ibm.com/cloud/learn/overfitting. Accessed: 27 May 2022.

IBM. 2021b. Structured vs. Unstructured Data: What's the Difference? https://www.ibm.com/cloud/blog/structured-vs-unstructured-data. Accessed: 3 June 2022.

ICRC. 2021. *ICRC position on autonomous weapon systems*.

Jaiswal, Sonoo. 2022a. Advantages of Cloud Computing. https://www.javatpoint.com/advantages-and-disadvantages-of-cloud-computing. Accessed: 30 May 2022.

Jaiswal, Sonoo. 2022b. IoT Advantages and Disadvantages. https://www.javatpoint.com/iot-advantage-and-disadvantage. Accessed: 30 May 2022.

Johnson, Matthew. 2017. *Getting Multiculturalism Right: Deontology and the Concern for Neutrality*.

Junck, Ryan D., Bradley A. Klein, Akira Kumaki, Ken D. Kumayama, Steve Kwok, Stuart D. Levi, James S. Talbot, Eve-Christie Vermynck, and Siyu Zhan. 2021. China's New Data Security and Personal Information Protection Laws. https://www.skadden.com/Insights/Publications/2021/11/Chinas-New-Data-Security-and-Personal-Information-Protection-Laws. Accessed: 2 June 2022.

Kant, Immanuel. 2017. *Immanuel Kant Werke I. Vorkritische Schriften bis 1768*. Darmstadt: WBG - Wissenschaftliche Buchgesellschaft.

Kaye, David. 2019. *Promotion and protection of the right to freedom of opinion and expression. Note by the Secretary-General*. New York.

Khatry, Sarah, Edward Kwartler, Kay Firth-Butterfield, and Mark Caine. 2021. What would it take to make AI ‘greener’? https://www.weforum.org/agenda/2021/09/make-ai-greener-climate-solution-cop26-technology/. Accessed: 25 May 2022.

Kleinings, Hanna. 2022. 8 Applications of Artificial Intelligence in Business. https://levity.ai/blog/8-uses-ai-business. Accessed: 22 April 2022.

Koene, Ansgar, Christopher Wade Clifton, Yohko Hatada, Helena Webb, Menisha Patel, Caio Machado, Jack LaViolette, Rashida Richardson, and Dillon Reisman. 2019. *A governance framework for algorithmic accountability and transparency. Study*. Brussels: European Parliament.

La Rue, Frank. 2013. *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, Frank La Rue*. New York.

Labes, Stine. 2012. *Grundlagen des Cloud Computing. Konzept und Bewertung von Cloud Computing*. Berlin: Univ.-Verl. der TU Berlin.

Larbey, Ruth, Eleanor Bird, Jasmin Fox-Skelly, Nicola Jenner, Emma Weitkamp, and Alan Winfield. 2020. *The ethics of artificial intelligence. Issues and initiatives : study Panel for the Future of Science and Technology*. Brussels: European Union.

Lazari-Radek, Katarzyna de, and Peter Singer. 2017. *Utilitarianism. A very short introduction*. Oxford: Oxford University Press.

Levin, Noah. 2019. Utilitarianism- Pros and Cons. https://human.libretexts.org/Bookshelves/Philosophy/Introduction_to_Ethics_(Levin_et_al.)/04%3A_Happiness/4.03%3A_Utilitarianism-_Pros_and_Cons_(B.M._Wooldridge). Accessed: 10 April 2022.

Ligozat, Anne-Laure, Julien Lefèvre, Aurélie Bugeau, and Jacques Combaz. 2021. *Unraveling the Hidden Environmental Impacts of AI Solutions for Environment*.

Luger, George F., and William A. Stubblefield. 1993. *Artificial intelligence. Structures and strategies for complex problem solving*. Redwood City, Calif.: Benjamin/Cummings Publ.

Lyons, David. 2022. The Moral Opacity of Utilitarianism. In *Morality, Rules, and Consequences. A Critical Reader*, Eds. Brad Hooker, Elinor Mason, and Dale E. MillerEdinburgh: Edinburgh University Press.

Manyika, James, Jake Silberg, and Brittany Presten. 2019. What Do We Do About the Biases in AI? https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai. Accessed: 4 May 2022.

Mathew, Amitha, P. Amudha, and S. Sivakumari. 2021. Deep Learning Techniques: An Overview. In *Advanced Machine Learning Technologies and Applications. Proceedings of AMLTA 2020,* Eds. Aboul Ella Hassanien, Roheet Bhatnagar, and Ashraf Darwish, 599–608Singapore: Springer Singapore; Imprint Springer.

McNaughton, David. 2007. Deontological ethics. In *Routledge encyclopedia of philosophy online. All site license & consortia/,* Ed. Preston King[Place of publication not identified]: Routledge.

Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2019. *A Survey on Bias and Fairness in Machine Learning*.

Microsoft. 2021. Artificial Intelligence vs. Machine Learning. https://azure.microsoft.com/en-us/overview/artificial-intelligence-ai-vs-machine-learning/#introduction. Accessed: 21 April 2022.

Mill, John Stuart. 1863. *Utilitarianism*. Kitchener, Ontario, Canada: Batoche Books Limited.

Moraes, Thiago Guimarães, Eduarda Costa Almeida, and José Renato Laranjeira de Pereira. 2021. Smile, you are being identified! Risks and measures for the use of facial recognition in (semi-)public spaces. *AI and Ethics* 1 (2): 159–172. doi: 10.1007/s43681-020-00014-3.

Morgan, Forrest E., Benjamin Boudreaux, Andrew J. Lohn, Mark Ashby, Christian Curriden, Kelly Klima, and Derek Grossman. 2020. *Military applications of artificial intelligence. Ethical concerns in an uncertain world*. Santa Monica, Calif.: RAND Corporation.

Müller, Vincent C. 2020. *Ethics of Artificial Intelligence and Robotics*.

Nathanson, Stephen. 2018. Act and Rule Utilitarianism. https://iep.utm.edu/util-a-r/. Accessed: 11 February 2022.

Nilsson, Nils J. 1998. *Artificial Intelligence. A new synthesis*. San Francisco, Calif: Morgan Kaufmann Publishers.

Noggle, Robert. 2018. *The Ethics of Manipulation*.

Nouri, Steve. 2020. How AI Is Making An Impact On The Surveillance World. *Forbes,* 4 December. https://www.forbes.com/sites/forbestechcouncil/2020/12/04/how-ai-is-making-an-impact-on-the-surveillance-world/. Accessed: 28 May 2022.

Olatunji, Oyeshile. 2008. A Critique of the Maximin Principle in Rawls' Theory of Justice. *Humanity & Social Sciences Journal* (3): 65–69. https://idosi.org/hssj/hssj3(1)08/8.pdf. Accessed: 21 April 2022.

Pazzanese, Christina. 2020. Ethical concerns mount as AI takes bigger decision-making role. *Harvard Gazette,* 26 October. https://news.harvard.edu/gazette/story/2020/10/ethical-concerns-mount-as-ai-takes-bigger-decision-making-role/. Accessed: 25 May 2022.

Pirie-Griffiths, Olivia. 2016. Ethics Explainer: Consequentialism. *The Ethics Centre,* 15 February. https://ethics.org.au/ethics-explainer-consequentialism/. Accessed: 28 March 2022.

Pohlman, H. L. 1984. *Justice Oliver Wendell Holmes & Utilitarian Jurisprudence*. s.l.: Harvard University Press.

Poole, David L., Alan K. Mackworth, and Randy Goebel. 1998. *Computational intelligence. A logical approach*. New York, Oxford: Oxford Univ. Press.

Rainie, Lee. 2018. Americans' complicated feelings about social media in an era of privacy concerns. *Pew Research Center,* 27 March. https://www.pewresearch.org/fact-tank/2018/03/27/americans-complicated-feelings-about-social-media-in-an-era-of-privacy-concerns/. Accessed: 30 May 2022.

Rawls, John. 1971. *A Theory of Justice. Original Edition*. Cambridge, MA: Harvard University Press.

Ross, William David. 2007. *The right and the good*. Oxford: Clarendon Press.

Russell, Stuart J., and Peter Norvig. 2003. *Artificial intelligence. A modern approach ; [the intelligent agent book*. Upper Saddle River, NJ: Prentice Hall.

Sahin, Kaan. 2020. The West, China, and AI surveillance. *Atlantic Council,* 18 December. https://www.atlanticcouncil.org/blogs/geotech-cues/the-west-china-and-ai-surveillance/. Accessed: 8 February 2022.

Saxena, Nripsuta Ani, Karen Huang, Evan DeFilippis, Goran Radanovic, David C. Parkes, and Yang Liu. How Do Fairness Definitions Fare? In *Conitzer, Hadfield et al. (Hg.) 2019 – Proceedings of the 2019 AAAI/ACM,* 99–106.

Schlosser, Markus. 2015. Agency. https://plato.stanford.edu/entries/agency/. Accessed: 3 June 2022.

Scott, John, and Gordon Marshall. 2009. *A dictionary of sociology*. Oxford: Oxford University Press.

Sharif, Mahmood, Sruti Bhagavatula, Lujo Bauer, and Michael K. Reiter. 2016. Accessorize to a Crime. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security,* Ed. Edgar Weippl, 1528–1540New York, NY: ACM.

Sidgwick, Henry. 1874. *The Methods of Ethics*.

Simpson, David. 2019. William David Ross (1877-1971). https://iep.utm.edu/ross-wd/. Accessed: 7 April 2022.

Sinnott-Armstrong, Walter. 2003. Consequentialism. https://plato.stanford.edu/entries/consequentialism/#ArgCon. Accessed: 5 April 2022.

Skelton, Anthony. 2010. William David Ross. https://plato.stanford.edu/entries/william-david-ross/. Accessed: 7 April 2022.

Smuha, Nathalie, European Commission, and Independent High-Level Expert Group on Artificial Intelligence. 2019. *ETHICS GUIDELINES FOR TRUSTWORTHY AI*. Brussels.

Stahl, Bernd Carsten. 2021. Ethical Issues of AI. *Artificial Intelligence for a Better Future:* 35–53. doi: 10.1007/978-3-030-69978-9_4.

Stanton, Charlotte, Vivien Lung, Nancy Zhang, Minori Ito, Steve Weber, and Katherine Chalet. 2019. *What the Machine Learning Value Chain Means for Geopolitics*.

Stedman, Craig, and Jack Vaughan. 2022. What Is Data Governance and Why Does It Matter? https://www.techtarget.com/searchdatamanagement/definition/data-governance. Accessed: 3 June 2022.

Sternberg, Robert J. 2000. *Handbook of intelligence*. Cambridge: Cambridge University Press.

Sveinsdottir, Thordis. 2020. *The Role of Data in AI*.

Tan, Tina Q. 2019. Principles of Inclusion, Diversity, Access, and Equity. *The Journal of Infectious Diseases* 220 (220 Suppl 2): S30-S32. doi: 10.1093/infdis/jiz198.

Tännsjö, Torbjörn. 2022. *Understanding Ethics*. Edinburgh: Edinburgh University Press.

Tealab, Ahmed. 2018. Time series forecasting using artificial neural networks methodologies: A systematic review. *Future Computing and Informatics Journal* 3 (2): 334–340. doi: 10.1016/j.fcij.2018.10.003.

U.S. Department of Justice. 2013. *Smart Policing Initiative. Data. Analysis. Solutions.*

UNESCO. 2019. *PRELIMINARY STUDY ON THE ETHICS OF ARTIFICIAL INTELLIGENCE*. Paris.

United Nations. 2021. Universal Declaration of Human Rights | United Nations. https://www.un.org/en/about-us/universal-declaration-of-human-rights. Accessed: 20 May 2022.

United Nations. 2022. Human Rights | United Nations. https://www.un.org/en/global-issues/human-rights. Accessed: 10 May 2022.

University of Texas. 2021. Utilitarianism - Ethics Unwrapped. https://ethicsunwrapped.utexas.edu/glossary/utilitarianism. Accessed: 11 February 2022.

University of Texas. 2022. Moral Agent - Ethics Unwrapped. https://ethicsunwrapped.utexas.edu/glossary/moral-agent. Accessed: 31 March 2022.

Velasquez, Manuel, Claire Andre, Thomas Shanks, and Michael, J., Meyer. 2014. What is a right? https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/rights/. Accessed: 29 March 2022.

Wanbil, Lee W., Wolfgang Zankl, and Henry Chang. 2016. An Ethical Approach to Data Privacy Protection. *ISACA Journal* 6: 1–9. https://www.isaca.org/-/media/files/isacadp/project/isaca/articles/journal/2016/volume-6/an-ethical-approach-to-data-privacy-protection_joa_eng_1216.pdf. Accessed: 30 May 2022.

Weiß, Christian. 2021. *Data Science*.

Whitton, Howard. 2001. IMPLEMENTING EFFECTIVE ETHICS STANDARDS IN GOVERNMENT AND THE CIVIL SERVICE. https://www.oecd.org/mena/governance/35521740.pdf. Accessed: 30 May 2022.

Wolfewicz, Arne. 2021. Deep learning vs. machine learning – What’s the difference? https://levity.ai/blog/difference-machine-learning-deep-learning. Accessed: 21 April 2022.

Yang, Fei. 2015. Predictive Policing. In *Oxford research encyclopedia of criminology and criminal justice*, Ed. Henry N. PontellNew York, NY: Oxford University Press.

Zell, Andreas. 1996. *Simulation neuronaler Netze*. Zugl.: Stuttgart, Univ., Habil.-Schr., 1994. Bonn: Addison-Wesley.

Zhang, Daniel, Saurabh Mishra, Brynjolfsson, John Etechemedy, and Deep Ganguli. 2021. *The AI Index 2021 Annual Report*. Stanford.

Zhang, Aston, Zachary C. Lipton, Mu Li, and Alexander J. Smola. 2022. *Dive into Deep Learning*.

Zimmerman, Michael J., and Ben Bradley. 2002. *Intrinsic vs. Extrinsic Value*.

Table of Figures

Affidavit

I, Marvin Wiesenthal, hereby declare that I have written this thesis independently, without outside help and using only the sources and aids indicated. All text passages that I have taken verbatim or in spirit from published or unpublished sources are identified as such. The work has not been submitted in the same or similar form to any other examination authority.

Mülheim an der Ruhr, 07.06.2022

M. Wiesenthal

Marvin Wiesenthal