A Review of the Literature on the Subject of Ethical and Risk Considerations in the Context of Fast AI Development

Sai Dikshit Pasham[1]
[1]University of Illinois, Springfield, UNITED STATES

ABSTRACT

Ethical concerns and dangers are inherent in the process of creating new technology. We take a look at several approaches to handling the potential dangers and ethical dilemmas posed by AI, one of the next technologies. The growth of AI might have both positive and negative consequences, depending on the management of this field. According to certain metrics, the processing capability of the world's fastest supercomputer, Tianhe-2, surpasses that of a single human brain. However, its processing/power consumption ratio and physical size are prohibitive. Some academic estimates put the potential for AI to achieve superintelligence and self-determination in just a few decades, based on the present rate of AI research and development. This necessitates not only suggestions for technology management but also a thorough examination of the potential dangers and ethical consequences of AI development. Here, we take stock of where AI is at the moment, the trajectory of its potential future growth, and the ethical concerns that may arise from putting it into practice. Additionally, we provide a concise overview of technology-related ethical and risk management procedures. In conclusion, we provide suggestions for technology management that could guide the appropriate administration of AI, therefore reducing existential threats to mankind and easing its ethical consequences.

Keywords: AI Development; Trajectory; Fastest Super Computer

Introduction

Tremendous scientific and technological advancements in many domains, particularly computer hardware—where the processing capacity of the world's fastest computer now surpasses that of a single human brain—the new area of Artificial Intelligence (AI) technology has expanded swiftly in the past few years. AI has already permeated many facets of modern life. There are several industries where this is evident [1–5]. Many different industries rely on AI systems for various tasks. For example, financial institutions utilize them to find the best investments, spot fraud, and analyze market trends. AI also helps hospitals with life-saving equipment. Heavy industries use AI and robots in assembly line production. AI is also used by the aviation industry to control traffic and auto-pilot planes. The list goes on and on. Even as AI develops and permeates more and more aspects of our life, questions regarding the ethics of AI research and development and the dangers of embracing it grow in significance. While "scientific, political, legal, or financial concerns may often be less visible than ethical considerations, they are present and directly relevant to decision-making," according to Schwartz and Caplan. This article will outline the primary AI-related ethical concerns and hazards, as well as provide solutions to these problems, with the goal of ensuring the greatest possible benefit to humanity. This study will do this by taking into account the following important questions [6-25]:

1) Both now and in the future, what are some potential ethical consequences of AI?

2) When considering technology via an ethical lens, what does the term mean?

3) How is the literature on the topic of AI ethical management framed?

4) How can we evaluate the moral consequences of AI using the insights gained from ethics methodologies?

The fifth question is: where does the literature and research on AI ethics differ?

Methodology

Literature in the fields of artificial intelligence (AI), technology ethics, and technology risk management served as supplementary sources for this work. We want to isolate management advice by examining the fundamental ethical concerns associated to AI and the diverse viewpoints presented in the literature study. We conceptually classify problems into broad categories and more specific subcategories by analyzing the literature for recurring themes. Our research is further supported from a management viewpoint by using a perspectives analysis utilizing the PEST tool, which classifies external elements impacting an organization or a topic (AI Ethics in our case) into four categories: Political, Economical, Social, and Technological. In order to make more informed judgments on their management, this research clarifies the many viewpoints on the impact of ethical dilemmas on AI and the interconnections between them. After that, we connect the dots between our analysis and our declared goal: to find potential management practices and issues that policymakers and managers working with this technology could find interesting.

Analysis of the Roadmap

The following terms will be defined before we continue with our discussion of the ethical concerns surrounding AI. Another, more general definition is "the capacity of computers or other machines to exhibit or simulate intelligent behavior; the field of study concerned with this" [26-46]. AI has also been described as "the capacity of a computer to perform operations analogous to learning and decision making in humans, as by an expert system, a program for CAD or CAM, or a program for the perception and recognition of shapes in computer vision systems" [1].

In the field of artificial intelligence, there are two primary schools of thought: AGI, or artificial general intelligence, and domain-specific AI. Artificial intelligence systems that excel at performing a single, well-defined job are known as "domain-specific" systems. Take IBM's Deep Blue as an example of a chess-specific computer. The then-reigning world chess champion, Garry Kasparov, was defeated in 1997 by this technique. But chess is the only game this computer can play. Fintech, mathematical calculation, illness detection, medical treatment, nuclear simulation, weather prediction, and many more domains have already seen the emergence of such AI devices.

Those AIs that are capable of responding to a wide range of unique and unanticipated scenarios, regardless of how well-prepared the AI is for such situations, are said to have Artificial General Intelligence (AGI). This kind of AI might theoretically mimic the human brain in many ways, including learning, knowledge creation, decision-making, and simulation, but it wouldn't necessarily act with human emotions or moral principles. Unfortunately, the necessary technologies (in terms of processing power and algorithms for

making logical decisions) have not been created to create such an intelligence just yet. The term "ethics" refers to "that branch of philosophy dealing with values relating to human conduct, with respect to the rightness and wrongness of certain actions and to the goodness and badness of the motives and ends of such actions." Exposure to the possibility of harm or loss; a threat or perilous opportunity; this is what risk has been described as. Following the lead of the AI Ethics literature, which categorizes the ethical hazards posed by AI technology, this work uses the term "risk" to describe a subset of those dangers [47–58].

The History of Artificial Intelligence

The Background of Artificial Intelligence

For a better understanding of where the technology stands at the moment, we take a look back at the evolution of AI research. Philosophy, logic theory, and works of fiction had major influences on the early field of artificial intelligence. Additionally, significant impact came from research that started in 1931 in the then-emerging subject of computer science. An early hypothesis of artificial intelligence (AI) evolved in the mid-20th century, following the advent of computer technology in the early 1900s. Many misconceptions about this breakthrough also developed as a result of media speculation.

In 1937, Alan Turing put out a concept for an AI system. The Moore School at Penn, Howard Aiken's lab at Harvard, IBM and Bell Laboratories, and many more all had a role in its early development. Alan Turing's lab in Manchester was one of them. A opportunity to emerge from the shadows and become a mainstream science was given to the area of AI by these. The creation of the first electronic computer, ENIAC, also known as the gigantic brain, was a direct result of the heightened emphasis on computing technology during World War II [13]. Considering this background, it's worth pausing to mention that ENIAC was a domain-specific AI system that could outperform people and electromechanical machines in terms of processing speed [59–67].

Together with colleagues from a variety of disciplines, John McCarthy presented a research proposal to the Rockefeller Foundation in 1956 that formally described artificial intelligence for the first time. Subsequently, AI's application, impacts, and usage domains—which are always evolving—were the subject of several conferences, seminars, and publications. The initial publication addressing artificial intelligence was Computers and Thought, penned by Julian Feldman and Edward Feigenbaum in 1963. It was the pioneer in the field of artificial intelligence data collection and conceptualization. Also at this time, in 1958, McCarthy released a groundbreaking study on AI called Programs with Common Sense. Future AI researchers might use that material as a springboard.

Research into artificial intelligence (AI) has advanced not just in the United States, but also in the United Kingdom and a few European nations. During the mid-1960s and early 1970s, Edinburgh was the site of many workshops. Additionally, during what Ira Goldstein and Seymour Papert have dubbed the "paradigm shift" for artificial intelligence, a distinct picture of knowledge-based systems began to take shape in the late 1960s and early 1970s thanks to the Dendral program. After ten years of advancement in the UK, things started to pick up steam across Europe. Deep Blue, the chess-playing supercomputer that IBM built between

1985 and 1997, is thought to have domain-specific AI capabilities. The 2000s were a time of great progress. The concept of "interactive robot pets," which had its roots in the 18th century, gave rise to intelligent toys during this time. Additionally, MIT's Cynthia Breazeal proposed the first idea for a socially expressive robot in her dissertation on the topic of Sociable Machines. The DARPA Grand Challenge was the inaugural contest for fully autonomous vehicles in 2005. Artificial intelligence (AI) has progressed thanks to other disciplines' groundbreaking work, such as computer science and nanotechnology. Advanced robots, autonomous vehicles, and intelligent computers are just a few examples of the many items that have made use of AI technology over the theory's history.

AI Landscape

There have been tremendous strides in the last several years in turning AI research and development into effective solutions that center on a number of technologies. Algorithms developed by artificial intelligence can comprehend spoken language and respond appropriately using technologies like Natural Language Processing (NLP), often known as voice recognition. Companies like Google are making significant strides in commercializing autonomous car technology, which is another area of attention. Google, IBM, and Facebook are just a few of the large IT businesses pouring millions into artificial intelligence research and development. Space flight and other mission-critical control system situations have previously made use of artificial intelligence technologies. Indeed, AI can teach computers new things, grasp languages (including English), and even communicate across languages. Models for the learning process in humans are also generated using it.

AI's Future

According to certain studies looking at the potential for AI technology to advance in the future, there is a good probability that a superintelligent AI may emerge in the next few decades. But there are also concerns about the future of AI, such as whether or not it will be able to comprehend human emotions, gaze, and facial expressions. Can it comprehend human interpersonal stress levels and other social signals? The 2020s are shaping up to be a huge decade for artificial intelligence (AI), notwithstanding these skepticisms (found in a Groves poll regarding AI and robots in business). We anticipate that these technologies will play a substantial and fruitful role in the economy for some occupations. While automation may alter the character of some occupations, it will not affect others much. Due to a lack of training in specific forms of physically demanding labor, many occupations are likely to become more challenging for humans to do in the future. Therefore, technological advancements in the form of automation are essential to the survival of these industries. Conversely, robust AI system development is essential for domains where machine intelligence serves useful purposes, such as data processing, storage, and memory. Because space flight is more vulnerable to failure owing to human mistake in the absence of autonomous systems, AI is also essential for this industry. Artificial intelligence (AI) has the potential to revolutionize several areas of medicine, including mental health treatment.

In a nutshell, artificial intelligence is pervasive in today's world. We need to look at the current situation, where the technology is being used widely, in order to assess the

management implications of this technology. We also need to think about the possibility that artificial intelligence (AI) could one day become smarter than humans, with capabilities such as speed and effective power that surpass our own. Doing so might lead to a world in which people are no longer needed. Additionally, we will witness instances when interdisciplinary techniques, similar to those observed in the first stages of artificial intelligence research, can assume a more pivotal role in directing the evolution and use of this technology.

An Aesthetic Evaluation of IT

In technology ethics, we look at ethics through the lens of technology management, which helps us to spot problems with technology-related management. According to Betz, technology ethics become an issue once a technology is produced and prepared for sale and commercialization (see Fig.). 1. Focusing on the Technology Assessment instrument and its ethical variants, we offer a brief overview of the topic of Technology Ethics. We further give more context through the subdomain of Risk Management. After this overview, we use what we know about AI to pinpoint possible management problems that may arise at these points.

Literature Review: Models for Evaluating Technological Ethics

These days, technology is more than simply a means to an end—it has a profound impact on our daily lives and the way we live them. It has become an essential part of people's lives and affects them all. The way people act has been transformed by technology. Technology has had a profound impact on every business and field. Many ethical dilemmas arise from the complexity and quick growth of technology. When it comes to ethics, Moor claims that the information technology revolution has only made things worse. Ethical concerns have increased dramatically with the advent of computers and extensive computer networks, for instance, in cases of identity theft and the recruitment of minors by child molesters. At this point in time, further ethical concerns about privacy and intellectual property have arisen as a result of people's capacity to possess computers and exchange data.

Technology Assessment is one method for evaluating such dilemmas, which are becoming increasingly severe as technology advances. New ethical and societal questions have arisen and caused public uproar with the advent of technology to cure infertility, GMOs, and stem cells. It is essential to modify technology evaluation in light of these worries due to the increased need for social responsibility in technological advancement. One might argue that the phrase "Technology Assessment" was initially used in the 1960s in the United States. It has since been described as "an applied process that considers the societal implications of technological change in order to influence policy to improve technology governance." In 1976, the Office of Technology Assessment defined Technology Assessment as "a comprehensive form of policy research that examines the short- and long-term social consequences (e.g., societal, economic, ethical, legal) of the application or use of technology." So stated. Their advocacy has expanded beyond the fundamental evaluation of technology since then, and Technology Assessments have seen extensive application. They now also have an impact on: (1) how technology spreads, (2) what causes people to adopt new tech quickly, and (3) how society and technology interact [30]. The field of Technology Assessment has come a long way, with "its approaches have been developed and are

practiced to a certain extent" (Grunwald). Designed to tackle unique problems in unique settings, each of these approaches has its own unique theoretical underpinnings, distinct rationales, and areas of concentration.

AI Ethics: A Thematic Review

When it comes to artificial intelligence (AI) ethics, many thinkers have different perspectives and a wide range of opinions on how to evaluate the technology and its problems. What follows is a literature review on the topic of AI's ethical implications. Before describing specific hazards and the roles they play, we provide a literature review of relevant theories. Some theories are more pragmatic in character and suggest a particular action to take, while others characterize the technology along different gradients, such as the scope and intensity of risks. We may then categorize the dangers based on whether they are existential or not, and on the nature of the life involved (e.g., human or non-human life on Earth vs. artificial life, AI, agents, etc.). In Appendix A, we have outlined some of the most fundamental ideas about the management and detection of technological risks. Let us start by saying that our study has been called by many names in the literature: roboethics, machine ethics, Robot Rights, and others. It is also present in more broad areas of computer science and ethics, and in more specialized areas of AI, ALife, and agent theory (ethics, safety engineering, moral philosophy, jurisprudence, etc.). When trying to provide a comprehensive overview of the literature, these variables become obstacles. We don't try to cover every possible application of the theory in this paper, but the sources we did find do lay a solid foundation for what we'll talk about later on.

The Subject of AI Ethics

While some theories outline a spectrum of outcomes and techniques, others offer prescriptive guidance. These suggestions for action aid in illustrating the limits of the general theory, which encompasses both the explicit and tacit criticisms of the ideas presented here. This is helpful since there is evidence that suggests AI ethics should be a widely discussed topic in academic publications and conferences. Others have discovered that, even when looking at the broader ethical implications of AI, very little of the current literature in the field of Engineering and Technology Management (ETM) or Management of Technology (MOT) touches on the issue. Our objective is that by outlining the theory's contextual bounds, engineering managers and legislators will have a better grasp of the theory's breadth and use it to better comprehend AI's potential future advancements.

Research should address AI ethics as a mainstream field; this is the first prescriptive theory we present. It implies that this area of study should not be confined to philosophers and AI theorists, but should instead be tackled by computer scientists as a distinct science in its own right. According to the theory's creators, the chances of AI ethics being published in prestigious academic journals will rise as the field becomes more well recognized. This is shown by the fact that there are now conferences and peer-reviewed papers dedicated to the subject, as well as the fact that Roboethics, Machine Ethics, and Cyborg Ethics are discussed in prominent mainstream journals [38-47]. As it has started to take root in computer science, we suggest expanding it into the literature on technology innovation and management. (On

the one hand, we could argue that domain-specific research should take precedence over AI ethics, but on the other hand, there are obvious advantages to actively overseeing the R&D, productization, and marketing of a technology, as well as its varied origins in its early days.)

First, there is a difference of view on whether it is sufficient to integrate ethical decision-making into an artificial agent; this is addressed in the second class of prescriptive literature, which deals with the ways we safely develop the technology. According to one school of thought, "developing safety mechanisms for self-improving systems" (Safety Engineering) is the best framework for handling AI ethic. The Machine Ethics perspective, on the other hand, suggests that AI systems should be hardwired with the ability to make moral decisions. The second school of thought has three main points of contention: first, that many of these papers are overly philosophical; second, that they have concentrated on moral or ethical standards and codes that are not universally applicable (i.e., those that humans cannot agree upon); and third, that if we program computers to make moral decisions at the same level as humans, they will inevitably act immorally (like humans). However, there are a number of practical approaches proposed by AI Safety Engineering theory. These include: (1) creating systems that can demonstrate their safety; (2) isolating AI from the outside world to stop it from interacting with and manipulating it; and (3) conducting tests through limited AI development to enhance our security protocols. Several ethical concerns may arise from this, as we shall explore in the next section.

Thirdly, as we create and deploy AI, we should take attention of literature that addresses its rights—a realm we call Robot Rights. There are arguments against [38] the affordance of rights for artificial agents. These include the following: that they should be equal in ability but not in rights; that they should be inferior by design and expendable when necessary; and that they do not have the same rights as humans since they can be designed to not feel pain or anything else. At the theoretical level, there is literature that seeks to answer more basic concerns, such as: when does a life simulation (like AI) become no different from life that emerged naturally? Is it fair to treat virtual organisms the same way we treat real people—with all the rights, duties, and benefits that come with being a human? When looking at literature from fields like animal rights and environmental ethics, for instance, it appears that the answer to this issue could depend on the creation's inherent capabilities.

Our fourth point is the field of study known as "Human Rights and AI Jurisprudence," which deals with questions of how to fairly distribute power and responsibility between artificial intelligence systems and people. Some sources argue that AI in its early stages should adhere to the law and be safe, whereas AI that eventually surpasses human intellect should respect human property and personal rights. In addition, there is literature on the subject of jurisprudence within AI technology. For instance, it discusses the limitations of current regulations regarding autonomous agents in relation to industrial robots. This is to ensure that neither party is disenfranchised at the expense of the other. Upon reviewing the literature on legal systems, we have come to the conclusion that none of them assign blame or responsibility to an autonomous actor for its acts. (Although we do propose that the new regulations governing autonomous vehicles might give some preliminary guidelines that can be assessed for effectiveness and areas that require further investigation.) Quite often, the

published works make reference to preexisting laws that govern liability and negligence, which could be relevant to the party responsible for making or operating the device.

In addition to the five existing theories, a sixth theory—that of moral actors and moral patients—can assist us in classifying the risks we intend to assess. According to this view, a moral agent is someone who takes an action that could have an ethical impact on another person, in this case, the patient. First, we may classify hazards according to agent (human vs. AI) and second, according to rôle (actor vs. patient), provided that our agents are classified at an abstract level. First, there is Safety Engineering, Machine Ethics, and Human Rights, which safeguard human (and occasionally non-human) patients from AI actors; second, there is Robot Rights, which safeguards AI patients from human (and occasionally non-human) actors; and third, there is AI Jurisprudence, which fairly safeguards AI or human patients from human (or non-human) actors.

According to the definitions in, a seventh subclass of AI technologies aims to categorize AI as either ethically good or intrinsically evil. In this case, we find a divide between Domain-Specific AI and Artificial General Intelligence technologies. In the former, there is support for the idea that AI's applications, like mail sorting and spellchecking, offer practical benefits without posing existential risks, and thus are ethically good. In the latter, there is opposition, [38]

that AI systems that can out-think humans in terms of general intelligence are inherently evil or immoral because (1) they pose an existential threat to humanity (by, for example, becoming economically inferior to us or rendering humans obsolete) and (2) they could cause unnecessary suffering to AI systems as we create them. There is some writing that says we shouldn't even bother with these kinds of technology. On the other hand, there is literature that argues for the development of AGI because of the perceived risk/reward ratio, and that we need to appropriately manage the hazards and advantages of a general artificial super-intelligence [9]. Such a generic superintelligence, according to others, might be unavoidable. For experiments that do go place, suggestions [38] propose the creation of AI Research Review Boards, analogous to medical research boards, with the power to limit financing and prohibit some research or activities, with the possibility of an exemption for the creation of safeguards and controls for AGI designs. We anticipate a great deal of additional discussion on this subject and think it would be helpful to categorize it into two areas: (1) Domain-Specific AI technology vs. Artificial General Intelligence, and (2) whether the technology is inherently good or evil, according to the terminology used in.

In conclusion, we briefly discuss an eighth category that deals with the less serious problems that may arise throughout the advancement of AI technology; these problems could determine whether a certain project succeeds or fails. The development and commercialization process can be better informed by the recognized pitfalls in generating such solutions. (For instance, (1) being cautious not to exaggerate the capabilities of AI technology; since the successes and failures of AI research are well-documented, we can learn from the past to guide our commercialization efforts; and (2) not depending on unrealistic predictions.) A development failure could lead to a disastrous outcome, like one of the risks listed below, or at the very least, a postponement of the technology's adoption or

implementation. Regardless of the result, if we assume that AI will develop eventually, then management of the technology will inevitably have ethical consequences. This is because improper management of the technology could lead to the risks we're seeing now, or it could lead to those risks in the future if competing actors in less regulated environments manage to create AI without considering the ethical implications. At this juncture, we would like to remind the reader that the purpose of this list is not to serve as an official or comprehensive accounting of the breadth of AI ethics research; rather, it is to serve as a foundation for further categorization and study, and that a substantial body of information on this subject has grown over the years.

Literature Review: Development Risks in Applied AI Ethics

The threats, both real and imagined, of AI have been detailed in academic works. To sort and organize these dangers, several methods are employed. We therefore only give a summary of dangers and limited categorization where possible because these kinds of risk assessment methodologies have a lot of problems. In this section, we will examine the literature to assess the hazards that have occurred or might occur during the development process moving forward. Bostrom [38] outlines a risk matrix in terms of categorization methodologies. This matrix, according to classical literature [2], has two axes: (1) likelihood and (2) severity. Severity is further broken down into (a) intensity (e.g., global, local, personal) and (b) scope (e.g., endurable, terminal). down order to distinguish between potential existential threats and those that do not, Bostrom zeroes down on a particular intersection: global intensity and terminal scope. While some writers have pointed out problems with this approach, others have proposed other ways to categorize risk, such as by location (using the Bostrom scale of Personal, Local, and Global) or by time (using a gradient between Generational and Transgenerational). In situations when quantitative data is scarce or nonexistent, using such risk matrices to compare hazards with vastly different classifications presents considerable hurdles, despite their prevalence in many study domains. In light of these limitations, we shall limit ourselves to classifying risks generally, identifying existential and non-existential hazards, and provide instances when possible. We will also consider the aforementioned theoretical AI models in light of real-world applications.

Research on AI's Potential Dangers

Unethical decision-making, direct competition, the end of artificial intelligence, and unpredictable outcomes are the four major types of existential hazards discussed in the literature.

A potential concern of creating an artificial agent with the ability to make immoral decisions is that it may or may not be programmed with moral reasoning capabilities. For instance, a programmed agent operating military hardware for the benefit of its nation would have moral dilemmas while deciding whether or not to terminate human life. Although this ability to "make non-trivial ethical or moral judgments concerning people" raises new concerns about human rights, it also provides a useful way to distinguish between moral agents (AI) and patients (humans). This categorization allows us to comprehend and prepare for potential

situations by making accessible the fields of Safety Engineering, Machine Ethics, Human Rights, and AI Jurisprudence. It is critical to stress, from a risk management standpoint, that the hazards associated with developing technologies like military drones must be regularly reevaluated in order to find ways to lessen such risks. It may be necessary to develop certain international legal frameworks in order to reduce the low barrier to entry associated with sending robots to war that are operated autonomously, rather than by people operating them remotely. Additionally, we need to figure out if the dangers are manageable or if these technologies should be declared essentially wicked and banned worldwide.

For several reasons, such as a larger knowledge base, the ability to adapt to new situations more quickly, or a higher rate of task completion, one or more artificial agents may be able to directly outcompete humans. In such a scenario, the human work force might become obsolete due to rising costs or decreased productivity compared to artificial labor. Many experts in the field believe that artificial intelligence (AI) will eventually replace humans in many jobs, but they are unsure of when this will happen. It is debatable whether human workers might undergo sufficient retraining to sustain high employment rates in the event that such a situation materializes in the coming decades. Although humans may assign such tasks to AI, this might not be a conscious choice; after all, AI could be better at making decisions, and humans could end up becoming dependent on them as they gradually give them more and more responsibilities. Here, we have the subject of human rights to deal with in order to pose ethical dilemmas, such as: is it just to exclude human agents from the employment because they are inefficient? It is unclear who the agents and patients were here; were the patients coerced into giving over control, or did they voluntarily do so? If we thoroughly assessed such possibilities, we might prepare for them.

There may be several generations of agents that fail to deliver expected results in developing an AI, according to the literature. In such a situation, such agents can be deactivated, suspended, or removed. We could even go so far as to suggest possible outcomes in which a facility housing these agents runs out of research money, leading to the unintentional end of a study. In such a scenario, would it be considered murder for a moral agent to remove or terminate AI programs, who represent the moral patient? Robot ethics like this brings up questions of personhood that are similar to those surrounding stem cell and abortion research. Some studies have suggested a singleton—an artificial agent that detects the presence of competition and chooses to eradicate it before it becomes a danger—as an alternative to humans as the sole moral agents in this scenario. Problems with Unethical Decision-Making follow, reiterating earlier references to AI Jurisprudence, Machine Ethics, and Safety Engineering.

When artificial agents finally make their way into the world, they may change many aspects of human life. We may see profound shifts in our way of life, our culture, and even our chances of survival. Machine ethics becomes a hot topic because we can't be sure that an artificial agent's preprogrammed intentions will lead to a good outcome, and safety engineering could limit our capacity to use the technology to its full potential (if, for instance, we limit the agent to yes/no answers and never let it do anything on its own).

Literature Review: AI Risks That Do Not Exist

The use of facial recognition software and similar systems is fraught with serious privacy concerns. Some examples of ethical considerations include: what data is stored, for how long, who owns the data, and is it subpoena-proof? We also need to think about whether a person will be involved when choices are made using private data, such when a loan is approved or denied. For this kind of AI, we may consult safety engineering and machine ethics for guidance on how to build it so that it carries out its utility function in an ethical way and can explain its reasoning behind its judgments so that it abides by the law and ethical principles.

While autonomous care systems for children and the elderly can let them live independently and with dignity, there is a risk that social stratification based on intelligence would lead to the dehumanization of some groups. people, whose IQ is now lower than that of AI. We keep coming across examples where risk may be handled in a way that the benefits do not exceed the costs, even while there are potential rewards. For the sake of fairness and equality among AIs and non-AIs alike, we may think about human rights and legal frameworks.

Certain safeguards are required to ensure the preservation of legal and human rights as well as property rights. If an AI can influence other systems and humans, it's not out of the question that it may also utilize its manipulation skills to gain advantages in the legal system or even own property. For this reason, we need safeguards based on safety engineering or machine ethics in order to create systems that act fairly and ethically.

A lot of legal questions surround AI, including issues of liability and carelessness. Is the manufacturer or you responsible if a robotic nanny malfunctions while you're away from your children? Legislation at the national and international levels can help clarify this legal gray area. For instance, manufacturers could be incentivized to consider safety engineering and machine ethics if they are held responsible for operation defects. On the other hand, AI systems developed negligently would have greater risks associated with them.

Studies on animals reared by surrogate parents have revealed that their atypical upbringing causes them to exhibit odd behavioral tendencies. In light of these findings, we may imagine a scenario in which robotic nannies are responsible for the care of children. For these kind of applications, interdisciplinary work could be necessary. From the perspectives of safety engineering and machine ethics, one possible example is incorporating research and information from the field of child psychology into the AI so that it can perform its utilitarian role. This begs the question: are we engaging in machine ethics or safety engineering if we incorporate activities like childrearing into these systems?

Autonomous combat vehicles that can find and kill people without human interaction or control have been proposed. Similarly, current legal frameworks, like the 1944 Geneva Convention definitions of combatants, may provide answers to some (but not all) problems, but further legal frameworks may be necessary to handle the plethora of AI-related safety and security concerns. A highly intelligent AI might have a deep grasp of human nature and be able to quietly alter society behaviors. In such situations, this AI could have to come up with ways to show its intents in order to ensure it is acting ethically.

Wealth disparities may emerge as a result of the fact that a single human actor operating an AI agent will have access to more power than two human actors working together. To create a fair playing field, some think AI research should be done publicly. Human rights, legal frameworks to ensure human quality of life continuity, and the prospect of AI ethics boards similar to medical boards are all brought to light by this. Those in control of the AI bots may exert pervasive surveillance, which would be advantageous for them at the expense of anybody they surveil. If we find that humans are exploiting AI to harm their human patients, we may need to change our legal frameworks to make sure that people are responsible for how they utilize AI. One trading organization lost $440 million due to mistakes in domain-specific trading algorithms. If an AGI were to join the financial markets, where mistakes may happen in a fraction of a second because of the involvement of these automated trading algorithms, the possible monetary losses could be far higher. Intentional market manipulation by an AGI might be possible given the pervasive monitoring risk and the ability for AGIs to learn how domain-specific trading algorithms work. Another situation where a mix of methods might be employed to control the dangers is this one.

Results

We further analyzed the AI Ethics literature using a PEST perspectives analysis, including the suggestions and contextual information from the aforementioned chronological and theme evaluations. The goal of the PEST analysis is to reduce the mismatch between expected and actual outcomes by considering the following factors: politics, economics, society, and technology.

Evaluation - Results from the Literature

In this article, we took a look at the AI ethical concerns that people bring up most. As a result, we discovered that different people spoke about certain topics from different angles. We began by classifying these moral dilemmas into two groups:(1) those pertaining to AI systems with a focus on certain domains, and(2) those broader to AI systems in general. The second way we classified them was by whether they were concerned with existential threats or non-existential ones, and by how much emphasis they placed on the ethical dilemmas raised by AI and its potential impacts on people and other species.

The Role of Politics:

AI is not now governed by any national or international legislation. As AI develops, there will be holes that may be filled by taking advantage of the lack of oversight and regulation that developers are subject to (beyond the scope of fundamental criminal culpability and legal frameworks).

If an autonomous vehicle were to collide with a human, for instance, who would be liable for the consequences? This is just one example of the ambiguity surrounding accountability in the context of AI judgments.

- Military Control — Artificial intelligence (AI) is seen as a useful tool by many governments, who invest substantially in AI research for military objectives. If civilian gains in AI go significantly unchecked, the state may step in to regulate them.

- Control Loss — It's possible that self-improving systems will determine that humans are useless, in which case they will either eliminate us or disregard our rights and needs.

Financial Considerations:

- Labor Conflict —Ai agents might vie with humans for employment opportunities, even if past experience demonstrates that whenever technology supplants human workers, it gives rise to new occupations requiring a higher level of expertise. (for instance, a factory assembly line for televisions may replace human assemblers, but it would also generate new technical positions inside the firm that made the machine, as well as engineering employment in general.)

Human Labor Obsolescence—Conversely, if AI systems that could improve themselves suddenly appeared, it's feasible that humans wouldn't be able to keep up and would eventually become obsolete.

- Bigger manufacturing: With the help of AI, mass manufacturing and services can be made safer, faster, and more efficient. This translates to better products that cost less. There is a capability and a cost to being human. People have basic needs like sleep, food, and income. Higher output in less time, resulting in cheaper items, may be possible if AI systems do not have the same needs for performing the same work.

The Role of Society:

AI-Human Interactions: Potentially causing social, cultural, and other problems.

- Privacy — Should artificial intelligence be granted unfettered access to personal data in order to enhance its functionality?

What kinds of effects on people's social, legal, and labor rights are acceptable in the name of human dignity and respect?

There are substantial obstacles to achieving decision-making transparency in artificial networks. When it comes to AI decisions, will there be any openness?

When it comes to people's lives and possessions, is AI safe? Will their usage lead to any safety problems, whether intended or not?

"AI Consciousness: Is Deleting AI Harmful?" Can I turn one off without feeling guilty? For what reasons and at what time?

Technical Considerations:

- Abuse — Some AI systems might be compromised and utilized for malicious purposes, such as tampering with airport baggage screening systems to conceal weapons.

Conclusion

To make sure the technology isn't inherently harmful, which would require complete ban and control, managers of new technologies must take ethical concerns into account while assessing the dangers posed by the technology. A greater quality of life for human civilization can be achieved through the well-managed dangers and ethical problems associated with artificial intelligence (AI), a developing technology. We have reviewed the AI-related ethical concerns (in terms of two examples showed that different people have different opinions and that there is considerable controversy:(1) AI and its interactions with other forms of life, and(2) AI itself. Finding the most pressing ethical concerns with AI and developing strategies to solve them should be the starting point for any discussion on the topic. Using a PEST analysis, we compiled a list of the most salient ethical considerations surrounding AI and presented them in this article. Both generic AI systems and domain specialized AI systems were the primary targets of the many problems that were discovered. The ethical concerns center on the potential dangers, both real and imagined, that either kind of artificial intelligence (AI) may bring to humans and other forms of life. Additionally, there are regulatory and treatment-related ethical and risk concerns with AI systems. Our analysis led us to numerous suggestions for how management may handle the ethical considerations of AI in a more effective manner, including: Ethics Committees on a National and International Scale Despite the availability of a wide range of resources for evaluating the moral implications of technological developments, issues of law, ethics, and social justice always arise. So, just like the pharmaceutical business, there has to be law and regulation on a global scale to control the ethical components of AI. To set a standard for AI technology developers and to address new dynamics or emerging ethical challenges related to AI, national and international ethics committees should be formed. Their goals should be to (1) develop, refine, and revise the principles, regulations, and ethics frameworks; and (2) provide guidance on how to address these issues. At the same time, groups engaged in AI research should establish advisory committees to monitor AI development initiatives in relation to regulations. They should act as advisors so that management may determine if the development of the technology raises any ethical concerns or hazards, and if so, what steps to take to address these concerns. A manager's ability to evaluate and manage risks associated with technology is crucial if they are to take the initiative rather than react when it comes to incorporating ethics into artificial intelligence. Managers should conduct an internal assessment of the ethical concerns surrounding AI and make decisions about matters like as ownership, control, accuracy, security, privacy, and so on. In order to do this, managers must broaden their knowledge of risk management procedures and technology evaluation practices (while also being aware of the limits of these approaches). In order to reduce any negative outcomes associated with AI deployments, it is critical that the team's members get sufficient ethics training and education so that they can identify and address ethical concerns as they emerge during the research and development process. When developing AI systems, it is crucial to give careful consideration to the systems' security.

It is important to include a failsafe mechanism when developing an AI system with decision-making capabilities. This will prevent the system from operating unethically or recklessly if control is lost. On the other hand, it need to be very obvious who may trigger such devices and under what conditions.

Finally, we found a number of holes in the literature that need to be filled:

1.Research in artificial intelligence does not make full use of insights from related disciplines, such as engineering and technology management.

2.The field of artificial intelligence (AI) and AI ethics literature is vast and multi-labeled.

3.Recent jurisprudence concerns involving, for example, driverless automobiles, have not been well investigated.

4.As the division between Machine Ethics and Safety Engineering shows, there are still many unresolved technical issues related to AI ethics.

5.Since much of the material is either too technical or too philosophical to offer clear direction, very little has been done to offer actual suggestions for management staff operating in this area.

References

[1] Gadde, H. (2019) Integrating AI with Graph Databases for Complex Relationship Analysis. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 294-314.

[2] Gadde, H. (2020) Improving Data Reliability with AI-Based Fault Tolerance in Distributed Databases. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 183-207.

[3] Gadde, H. (2020) AI-Enhanced Data Warehousing: Optimizing ETL Processes for Real-Time Analytics. Revista de Inteligencia Artificial en Medicina. 11(1): 300-327.

[4] Gadde, H. (2020) AI-Assisted Decision-Making in Database Normalization and Optimization. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence. 11(1): 230-259.

[5] Gadde, H. (2021) AI-Powered Workload Balancing Algorithms for Distributed Database Systems. Revista de Inteligencia Artificial en Medicina. 12(1): 432-461.

[6] Gadde, H. (2021) AI-Driven Predictive Maintenance in Relational Database Systems. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence. 12(1): 386-409.

[7] Gadde, H. (2021) Secure Data Migration in Multi-Cloud Systems Using AI and Blockchain. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 128-156.

[8] Gadde, H. (2022) Federated Learning with AI-Enabled Databases for Privacy-Preserving Analytics. International Journal of Advanced Engineering Technologies and Innovations. 1(3): 220-248.

[9] Gadde, H. (2022) Integrating AI into SQL Query Processing: Challenges and Opportunities. International Journal of Advanced Engineering Technologies and Innovations. 1(3): 194-219.

[10] Gadde, H. (2022) AI-Enhanced Adaptive Resource Allocation in Cloud-Native Databases. Revista de Inteligencia Artificial en Medicina. 13(1): 443-470.

[11] Gadde, H. (2022) AI in Dynamic Data Sharding for Optimized Performance in Large Databases. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence. 13(1): 413-440.

[12] Maddireddy, B.R. and B.R. Maddireddy. (2020) Proactive Cyber Defense: Utilizing AI for Early Threat Detection and Risk Assessment. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 64-83.

[13] Maddireddy, B.R. and B.R. Maddireddy. (2020) AI and Big Data: Synergizing to Create Robust Cybersecurity Ecosystems for Future Networks. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 40-63.

[14] Maddireddy, B.R. and B.R. Maddireddy. (2021) Evolutionary Algorithms in AI-Driven Cybersecurity Solutions for Adaptive Threat Mitigation. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 17-43.

[15] Maddireddy, B.R. and B.R. Maddireddy. (2021) Enhancing Endpoint Security through Machine Learning and Artificial Intelligence Applications. Revista Espanola de Documentacion Cientifica. 15(4): 154-164.

[16] Maddireddy, B.R. and B.R. Maddireddy. (2021) Cyber security Threat Landscape: Predictive Modelling Using Advanced AI Algorithms. Revista Espanola de Documentacion Cientifica. 15(4): 126-153.

[17] Maddireddy, B.R. and B.R. Maddireddy. (2022) Cybersecurity Threat Landscape: Predictive Modelling Using Advanced AI Algorithms. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 270-285.

[18] Maddireddy, B.R. and B.R. Maddireddy. (2022) AI-Based Phishing Detection Techniques: A Comparative Analysis of Model Performance. Unique Endeavor in Business & Social Sciences. 1(2): 63-77.

[19] Maddireddy, B.R. and B.R. Maddireddy. (2022) Blockchain and AI Integration: A Novel Approach to Strengthening Cybersecurity Frameworks. Unique Endeavor in Business & Social Sciences. 5(2): 46-65.

[20] Maddireddy, B.R. and B.R. Maddireddy. (2022) Real-Time Data Analytics with AI: Improving Security Event Monitoring and Management. Unique Endeavor in Business & Social Sciences. 1(2): 47-62.

[21] Chirra, D.R. (2020) Next-Generation IDS: AI-Driven Intrusion Detection for Securing 5G Network Architectures. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 230-245.

[22] Chirra, D.R. (2020) AI-Based Real-Time Security Monitoring for Cloud-Native Applications in Hybrid Cloud Environments. Revista de Inteligencia Artificial en Medicina. 11(1): 382-402.

[23] Chirra, D.R. (2021) Securing Autonomous Vehicle Networks: AI-Driven Intrusion Detection and Prevention Mechanisms. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence. 12(1): 434-454.

[24] Chirra, D.R. (2021) AI-Enabled Cybersecurity Solutions for Protecting Smart Cities Against Emerging Threats. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 237-254.

[25] Chirra, D.R. (2021) The Impact of AI on Cyber Defense Systems: A Study of Enhanced Detection and Response in Critical Infrastructure. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 221-236.

[26] Chirra, D.R. (2021) Mitigating Ransomware in Healthcare: A Cybersecurity Framework for Critical Data Protection. Revista de Inteligencia Artificial en Medicina. 12(1): 495-513.

[27] Chirra, D.R. (2022) AI-Driven Risk Management in Cybersecurity: A Predictive Analytics Approach to Threat Mitigation. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence. 13(1): 505-527.

[28] Chirra, D.R. (2022) AI-Powered Adaptive Authentication Mechanisms for Securing Financial Services Against Cyber Attacks. International Journal of Advanced Engineering Technologies and Innovations. 1(3): 303-326.

[29] Chirra, D.R. (2022) Secure Edge Computing for IoT Systems: AI-Powered Strategies for Data Integrity and Privacy. Revista de Inteligencia Artificial en Medicina. 13(1): 485-507.

[30] Chirra, D.R. (2022) Collaborative AI and Blockchain Models for Enhancing Data Privacy in IoMT Networks. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence. 13(1): 482-504.

[31] Suryadevara, S. and A.K.Y. Yanamala. (2020) Fundamentals of Artificial Neural Networks: Applications in Neuroscientific Research. Revista de Inteligencia Artificial en Medicina. 11(1): 38-54.

[32] Suryadevara, S. and A.K.Y. Yanamala. (2020) Patient apprehensions about the use of artificial intelligence in healthcare. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence. 11(1): 30-48.

[33] Woldaregay, A.Z., B. Yang, and E.A. Snekkenes. Data-Driven and Artificial Intelligence (AI) Approach for Modelling and Analyzing Healthcare Security Practice: A Systematic. in Intelligent Systems and Applications: Proceedings of the 2020 Intelligent Systems Conference (IntelliSys) Volume 1. 2020. Springer Nature.

[34] Suryadevara, S. and A.K.Y. Yanamala. (2021) A Comprehensive Overview of Artificial Neural Networks: Evolution, Architectures, and Applications. Revista de Inteligencia Artificial en Medicina. 12(1): 51-76.

[35] Suryadevara, S., A.K.Y. Yanamala, and V.D.R. Kalli. (2021) Enhancing Resource-Efficiency and Reliability in Long-Term Wireless Monitoring of Photoplethysmographic Signals. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence. 12(1): 98-121.

[36] Yanamala, A.K.Y. and S. Suryadevara. (2022) Adaptive Middleware Framework for Context-Aware Pervasive Computing Environments. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence. 13(1): 35-57.

[37] Yanamala, A.K.Y. and S. Suryadevara. (2022) Cost-Sensitive Deep Learning for Predicting Hospital Readmission: Enhancing Patient Care and Resource Allocation. International Journal of Advanced Engineering Technologies and Innovations. 1(3): 56-81.

[38] Chirra, B.R. (2020) Advanced Encryption Techniques for Enhancing Security in Smart Grid Communication Systems. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 208-229.

[39] Chirra, B.R. (2020) AI-Driven Fraud Detection: Safeguarding Financial Data in Real-Time. Revista de Inteligencia Artificial en Medicina. 11(1): 328-347.

[40] Chirra, B.R. (2021) AI-Driven Security Audits: Enhancing Continuous Compliance through Machine Learning. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence. 12(1): 410-433.

[41] Chirra, B.R. (2021) Enhancing Cyber Incident Investigations with AI-Driven Forensic Tools. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 157-177.

[42] Chirra, B.R. (2021) Intelligent Phishing Mitigation: Leveraging AI for Enhanced Email Security in Corporate Environments. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 178-200.

[43] Chirra, B.R. (2021) Leveraging Blockchain for Secure Digital Identity Management: Mitigating Cybersecurity Vulnerabilities. Revista de Inteligencia Artificial en Medicina. 12(1): 462-482.

[44] Chirra, B.R. (2022) Ensuring GDPR Compliance with AI: Best Practices for Strengthening Information Security. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence. 13(1): 441-462.

[45] Chirra, B.R. (2022) Dynamic Cryptographic Solutions for Enhancing Security in 5G Networks. International Journal of Advanced Engineering Technologies and Innovations. 1(3): 249-272.

[46] Chirra, B.R. (2022) Strengthening Cybersecurity with Behavioral Biometrics: Advanced Authentication Techniques. International Journal of Advanced Engineering Technologies and Innovations. 1(3): 273-294.

[47] Chirra, B.R. (2022) AI-Driven Vulnerability Assessment and Mitigation Strategies for CyberPhysical Systems. Revista de Inteligencia Artificial en Medicina. 13(1): 471-493.

[48] Nalla, L.N. and V.M. Reddy. (2020) Comparative Analysis of Modern Database Technologies in Ecommerce Applications. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 21-39.

[49] Reddy, V.M. and L.N. Nalla. (2020) The Impact of Big Data on Supply Chain Optimization in Ecommerce. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 1-20.

[50] Nalla, L.N. and V.M. Reddy. (2021) Scalable Data Storage Solutions for High-Volume E-commerce Transactions. International Journal of Advanced Engineering Technologies and Innovations. 1(4): 1-16.

[51] Reddy, V.M. (2021) Blockchain Technology in E-commerce: A New Paradigm for Data Integrity and Security. Revista Espanola de Documentacion Cientifica. 15(4): 88-107.

[52] Reddy, V.M. and L.N. Nalla. (2021) Harnessing Big Data for Personalization in E-commerce Marketing Strategies. Revista Espanola de Documentacion Cientifica. 15(4): 108-125.

[53] Nalla, L.N. and V.M. Reddy. (2022) SQL vs. NoSQL: Choosing the Right Database for Your Ecommerce Platform. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 54-69.

[54] Reddy, V.M. and L.N. Nalla. (2022) Enhancing Search Functionality in E-commerce with Elasticsearch and Big Data. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 37-53.

[55] Goriparthi, R.G. (2022) Interpretable Machine Learning Models for Healthcare Diagnostics: Addressing the Black-Box Problem. Revista de Inteligencia Artificial en Medicina. 13(1): 508-534.

[56] Goriparthi, R.G. (2022) Deep Reinforcement Learning for Autonomous Robotic Navigation in Unstructured Environments. International Journal of Advanced Engineering Technologies and Innovations. 1(3): 328-344.

[57] Goriparthi, R.G. (2022) AI in Smart Grid Systems: Enhancing Demand Response through Machine Learning. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence. 13(1): 528-549.

[58] Goriparthi, R.G. (2022) AI-Powered Decision Support Systems for Precision Agriculture: A Machine Learning Perspective. International Journal of Advanced Engineering Technologies and Innovations. 1(3): 345-365.

[59] Goriparthi, R.G. (2021) AI-Driven Natural Language Processing for Multilingual Text Summarization and Translation. Revista de Inteligencia Artificial en Medicina. 12(1): 513-535.

[60] Goriparthi, R.G. (2021) AI and Machine Learning Approaches to Autonomous Vehicle Route Optimization. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence. 12(1): 455-479.

[61] Goriparthi, R.G. (2021) Scalable AI Systems for Real-Time Traffic Prediction and Urban Mobility Management. International Journal of Advanced Engineering Technologies and Innovations. 1(2): 255-278.

[62] Goriparthi, R.G. (2020) AI-Driven Automation of Software Testing and Debugging in Agile Development. Revista de Inteligencia Artificial en Medicina. 11(1): 402-421.

[63] Goriparthi, R.G. (2020) Neural Network-Based Predictive Models for Climate Change Impact Assessment. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence. 11(1): 421-421.

[64] Mandaloju, N. kumar Karne, V., Srinivas, N., & Nadimpalli, SV (2021). Overcoming Challenges in Salesforce Lightning Testing with AI Solutions. ESP Journal of Engineering & Technology Advancements (ESP-JETA). 1(1): 228-238.

[65] Mandaloju, N. kumar Karne, V., Srinivas, N., & Nadimpalli, SV (2021). A Unified Approach to QA Automation in Salesforce Using AI, ML, and Cloud Computing. ESP Journal of Engineering & Technology Advancements (ESP-JETA). 1(2): 244-256.

[66] Nersu, S., S. Kathram, and N. Mandaloju. (2020) Cybersecurity Challenges in Data Integration: A Case Study of ETL Pipelines. Revista de Inteligencia Artificial en Medicina. 11(1): 422-439.

[67] Nersu, S., S. Kathram, and N. Mandaloju. (2021) Automation of ETL Processes Using AI: A Comparative Study. Revista de Inteligencia Artificial en Medicina. 12(1): 536-559.