Journal of Database Management
Volume 31 • Issue 2 • April-June 2020

Artificial Intelligence (AI) Ethics:
Ethics of AI and Ethical AI

Keng Siau, Missouri University of Science and Technology, Rolla, USA
Weiyu Wang, Missouri University of Science and Technology, Rolla, USA

ABSTRACT

Artificial intelligence (AI)-based technology has achieved many great things, such as facial recognition, medical diagnosis, and self-driving cars. AI promises enormous benefits for economic growth, social development, as well as human well-being and safety improvement. However, the low-level of explainability, data biases, data security, data privacy, and ethical problems of AI-based technology pose significant risks for users, developers, humanity, and societies. As AI advances, one critical issue is how to address the ethical and moral challenges associated with AI. Even though the concept of "machine ethics" was proposed around 2006, AI ethics is still in the infancy stage. AI ethics is the field related to the study of ethical issues in AI. To address AI ethics, one needs to consider the ethics of AI and how to build ethical AI. Ethics of AI studies the ethical principles, rules, guidelines, policies, and regulations that are related to AI. Ethical AI is an AI that performs and behaves ethically. One must recognize and understand the potential ethical and moral issues that may be caused by AI to formulate the necessary ethical principles, rules, guidelines, policies, and regulations for AI (i.e., Ethics of AI). With the appropriate ethics of AI, one can then build AI that exhibits ethical behavior (i.e., Ethical AI). This paper will discuss AI ethics by looking at the ethics of AI and ethical AI. What are the perceived ethical and moral issues with AI? What are the general and common ethical principles, rules, guidelines, policies, and regulations that can resolve or at least attenuate these ethical and moral issues with AI? What are some of the necessary features and characteristics of an ethical AI? How to adhere to the ethics of AI to build ethical AI?

KEYWORDS

AI Ethics, Artificial Intelligence, Ethical AI, Ethics, Ethics of AI, Machine Ethics, Roboethics

1. INTRODUCTION

Some researchers and practitioners believe that artificial intelligence (AI) is still a long way from having consciousness and being comparable to humans, and consequently, there is no rush to consider ethical issues. But AI, combined with other smart technologies such as robotics, has already shown its potential in business, healthcare, transportation, and many other domains. Further, AI applications are already impacting humanity and society. Autonomous vehicles can replace a large number of jobs, and transform the transportation and associated industries. For example, short-haul flights and

DOI: 10.4018/JDM.2020040105

hospitality services along highways will be impacted if driverless cars enable passengers to sleep and work during the journey. AI-recruiters are known to exhibit human biases because the training data inherits the same biases we have as humans. The wealth gap created by the widening differences between return on capital and return on labor is posed to create social unrest and upheavals. The future of work and future of humanity will be affected by AI and plans need to be formulated and put in place. Building AI ethically and having ethical AI are urgent and critical. Unfortunately, building ethical AI is an enormously complex and challenging task.

1.1. What is Ethics?

Ethics is a complex, complicated, and convoluted concept. Ethics can be defined as the moral principles governing the behaviors or actions of an individual or a group of individuals (Nalini, 2019). In other words, ethics are a system of principles or rules or guidelines that help determine what is good or right. Broadly speaking, ethics can be defined as the discipline dealing with right versus wrong, and the moral obligations and duties of entities (e.g., humans, intelligent robots, etc.).

Ethics has been studied by many researchers from different disciplines. Most humans are familiar with virtue ethics since very young because it is a behavior guide instilled by parents and teachers to help children practice good conduct. Aristotle (Yu, 1998) believes when a person acts in accordance with virtue, this person will do well and be content. Virtue ethics is part of normative ethics, which studies what makes actions right or wrong. It can be viewed as overarching moral principles that help people resolve difficult moral decisions. As the interaction between humans, between humans and animals, between humans and machines, and even between machines is increasing, ethical theories have been applied to real-life situations, such as business ethics, animal ethics, military ethics, bioethics, and machine ethics. The study of ethics and ethical principles is constantly evolving and developing. Table 1 lists several ethics definitions given by researchers.

In the context of AI, the ethics of AI specifies the moral obligations and duties of an AI and its creators. Researchers have done much work studying human ethical issues. Many ethical frameworks can be used to direct human behaviors, such as actions and activities related to respect for individuals,

Table 1. Definition of ethics

Normative Ethics	Reference
Ethics is the capacity to think critically about moral values and direct our actions in terms of such values.	Churchill, 1999
Ethics is a set of concepts and principles that guide us in determining what behavior helps or harms sentient creatures.	Paul & Elder, 2006
Ethics is the norm for conduct that distinguishes between acceptable and unacceptable behavior. Ethics is the discipline that studies standards of conduct, such as philosophy, theology, law, psychology, or sociology. Ethics is a method, procedure, or perspective for deciding how to act and for analyzing complex problems and issues.	Resnik, 2011
Applied Ethics	
Computer ethics is the analysis of the nature and social impact of computer technology and the corresponding formulation and justification of policies for the ethical use of such technology.	Moor, 1985, p. 266
Machine ethics is concerned with giving machines ethical principles or a procedure for discovering a way to resolve the ethical dilemmas they might encounter, enabling them to function in an ethically responsible manner through their own ethical decision making.	Anderson and Anderson, 2011, p. 1

beneficence, justice, privacy, accuracy, ownership/property, accessibility, fairness, accountability, and transparency (Wang & Siau, 2018).

One of the best-known ethical frameworks is developed by Ken Blanchard and Norman Vincent Peale (Blanchard & Peale, 2011). The framework consists of three main questions: Is it legal? Is it fair? How does it make me feel? Another framework is the Markkula Center Framework, which identifies five approaches to dealing with ethical issues, including the utilitarianism approach, rights approach, fairness or justice approach, common good approach, and virtue approach (Markkula Center for Applied Ethics, 2015).

AI ethics, however, is a relatively new field and the subsequent parts of the paper will discuss the AI ethics – ethics of AI and ethical AI.

1.2. What is "Ethics of AI" and "Ethical AI"?

The ethics of AI is part of the ethics of advanced technology that focuses on robots and other artificially intelligent agents. It can be divided into roboethics (robot ethics) and machine ethics.

Roboethics is concerned with the moral behaviors of humans as they design, construct, use, and interact with AI agents, and the associated impacts of robots on humanity and society. In this paper, we consider it as ethics of AI, which deal with ethical issues related to AI, including ethical issues that may arise when designing and developing AI (e.g., human biases that exist in data, data privacy, and transparency), and ethical issues caused by AI (e.g., unemployment and wealth distribution). Further, as machines become more intelligent and may one day gain consciousness, we should consider robot rights -- the concept that people should have moral obligations towards intelligent machines. It is similar to human rights and animal rights. For instance, whether it is ethical to deploy intelligent military robots to dangerous battlefields or assign robots to dirty environments. The rights of liberty, freedom of expression, equality, and having thought and emotion belong to this category.

Machine ethics deals with the moral behaviors of Artificial Moral Agents (AMAs), which is the field of research addressing the design of artificial moral agents. As technology advances and robots become more intelligent, robots or artificially intelligent agents should behave morally and exhibit moral values. We consider the ethical behaviors of AI agents as ethical AI. Currently, the best known proposed rules for governing AI agents are the Three Laws of Robotics put forth by Issac Asimov in the 1950s (Asimov, 1950). First Law, a robot may not injure a human being or, through inaction, allow a human being to come to harm. Second Law, a robot must obey the orders given to it by human beings except when such orders would conflict with the First Law. Third Law, a robot must protect its existence as long as such protection does not conflict with the First or Second Law.

Table 2 depicts the two dimensions of AI ethics (i.e., ethics of AI and ethical AI) and how the two dimensions interact with AI, Human, and Society. The ethical interaction between AIs is new in this paper. This is especially important for AIs with consciousness. Not only should the AIs do no harm to humans and self-preserve, but it also should do no harm to other intelligent agents. Thus, the Three Laws of Robotics may need to be extended to take into account the interaction between intelligent AIs.

Understanding the ethics of AI will help to establish a framework for building ethical AI. Figure 1 shows the initial framework for building ethical AI.

Table 2. AI ethics

	AI	Human	Society
Ethics of AI	Principles of developing AI to interact with other AIs ethically	Principles of developing AI to interact with human ethically	Principles of developing AI to function ethically in society
Ethical AI	How AI should interact with other AIs ethically?	How AI should interact with humans ethically?	How AI should operate ethically in society?

Figure 1. Initial framework for building ethical AI

1.3. Why Should We Build Ethical AI?

Recently, Criminals used AI-based voice technology to impersonate a chief executive's voice and demand a fraudulent transfer of $243,000 (Stupp, 2019). This is not an isolated incident. PINDROP reported a 350% rise in voice fraud between 2013 and 2017 (Livni, 2019). AI voice impersonation being used for fraud is not the only concern. Deepfake, which is an approach to superimpose and synthesize existing images and videos onto source images or videos using Machine Learning (ML), is also becoming common. With deepfake, human faces could be superimposed on pornographic video content and political leaders can be portrayed in videos to incite violence and panic. Deepfake may also be used in the election cycle to influence and bias the American electorate (Libby, 2019). In 2017, researchers from the University of Washington created a synthetic Obama, using a neural network AI to model the shape of Obama's mouth (BBC News, 2017). Although there was no security threat from the University of Washington experiment, the demonstration illustrates what is possible with AI-altered videos. Fake news is another concern. For example, an AI fake text generator was deemed too dangerous to be released by its creators, Open AI, for fear of misuse. Undoubtedly, advanced AI agents could put individuals, companies, and societies at increased risk.

Human rights, such as privacy, freedom of association, freedom of speech, right to work, non-discrimination, and access to public services, should always be put in the first place. However, the growing use of AI in the criminal justice system may have a discrimination concern. The recidivism risk-scoring software used across the criminal justice system shows incidents of discrimination based on race, gender, and ethnicity. For instance, some defendants are falsely labeled as high risk because of their ethnicity.

The right to privacy, which is essential to human dignity, can also be affected by AI. As big data technology developed, the collection of data interferes with the rights to privacy and data security. For instance, ML models can accurately synthesize data and estimate personal characteristics, such as gender, age, marital status, and occupation, from cell phone location data. Another example is government surveillance. In the U.S., half of all adults are already in law enforcement facial recognition databases (Telcher, 2018), which threatens to end anonymity. Rights to freedom of expression, assembly, and association may accordingly be affected. Last but not least, the right to work and an adequate standard of living (Access Now, 2018) would be affected. Automation has resulted in job loss and job displacement in certain industries and the rapid advancement in AI would accelerate this trend.

2. REVIEW OF ETHICAL GUIDELINES, FRAMEWORKS, AND PRINCIPLES

Advanced AI will spark unprecedented business gains, but along the way, government and industry leaders will have to grapple with a smorgasbord of ethical dilemmas such as data privacy issues, machine learning bias, public safety concerns, as well as job replacement and unemployment rate problems. To guide their strategies in developing, adopting, and embracing AI technologies, organizations should consider establishing AI ethics frameworks/guidelines. Some institutions have started work on this issue and published some guidelines. Table 3 shows eight institutions that work on AI ethical frameworks or principles and their objectives. Table 4 shows the content of those ethical frameworks and principles.

Table 3. Institutions' works on AI ethics and their objectives

Resource	Objective
Future of Life Institute (2017)	This report emphasizes "do no harm". It requires the development of AI to benefit society, foster trust and cooperation, and avoid competitive racing.
International Association of Privacy Professionals (IAPP, 2018)	The proposed framework explores risks to privacy, fairness, transparency, equality, and many other issues that can be amplified by big data and artificial intelligence. They provide an overview of how organizations can operate data ethics and how to reflect ethical considerations in decision making.
Institute of Electrical and Electronics Engineers (IEEE, 2019)	The proposed design lays out practices for setting up AI governance structure, including pragmatic treatment of data management, affective computing, economics, legal affairs, and other areas. One key priority is to increase human well-being as a metric for AI progress. Besides, the IEEE principle requires everyone involved in the design and development of AI is educated to prioritize ethical considerations.
The Public Voice (2018)	The proposed guidelines aim to improve the design and use of AI, maximize the benefits of AI, protect human rights, and minimize risks and threats associated with AI. They claim that the guidelines should be incorporated into ethical standards, adopted in national law and international agreements, and built into the design of systems.
European Commission's High-Level Expert Group on AI (European Commission, 2019)	The guidelines are designed to guide the AI community in the development and use of "trustworthy AI" (i.e., AI that is lawful, ethical, and robust). The guidelines emphasize four principles: respect for human autonomy, prevention of harm, fairness, and explicability.
AI4People (Floridi et al., 2018)	This framework introduces the core opportunities and risks of AI for society; present a synthesis of five ethical principles that should undergird its development and adoption; and offer 20 concrete recommendations—to assess, to develop, to incentivize, and to support good AI—which in some cases may be undertaken directly by national or supranational policymakers.
United Nations Educational, Scientific, and Cultural Organization (UNESCO, 2017)	The proposed ethical principle aims to provide decision-makers with criteria that extend beyond purely economic considerations.
Australia's Ethics Framework (Dawson et al., 2019)	This ethics framework highlights the ethical issues that are emerging or likely to emerge in Australia from AI technologies and outlines the initial steps toward mitigating them. The goal of this document is to provide a pragmatic assessment of key issues to help foster ethical AI development in Australia.

In Table 5, we summarized the frequency of each factor in those frameworks shown in Table 4. We can see that different frameworks may include the same or similar factors, but also include different considerations. The study of ethical issues of AI is still a new area and more discussion is needed to finally establish the framework of building ethical AI. In the next section, we will discuss each ethical issue in detail.

3. REVIEW OF ETHICAL ISSUES IN AI

AI, at the present stage, is referred to as Narrow AI or Weak AI. Weak AI can do well in a narrow and specialized domain. The performance of narrow AI depends much on the training data and programming, which is closely related to big data and humans. The ethical issues of Narrow AI, thus, involve human factors.

"A different set of ethical issues arises when we contemplate the possibility that some future AI systems might be candidates for having the moral status" (Bostrom and Yudkowsky, 2014, p.5).

Table 4. Ethical frameworks and principles from eight institutions

Resource	Ethical Framework/Principle
Future of Life Institute (2017)	Safety, Failure Transparency, Judicial Transparency, Responsibility, Value Alignment, Human Values, Personal Privacy, Liberty and Privacy, Shared Benefit, Shared Prosperity, Human Control, Non-subversion, AI Arms Race
International Association of Privacy Professionals (IAPP, 2018)	Data ethics, Privacy, Bias, Accountability, Transparency, Human Rights
Institute of Electrical and Electronics Engineers (IEEE, 2019)	Human Rights, Well-being, Data Agency, Effectiveness, Transparency, Accountability, Awareness of Misuse, Competence
The Public Voice (2018)	Right to Transparency; Right to Human Determination; Identification Obligation; Fairness Obligation; Assessment and Accountability Obligation; Accuracy, Reliability, and Validity obligation; Data Quality Obligation; Public Safety Obligation; Cybersecurity Obligation; Prohibition on Secret Profiling; Prohibition on Unitary Scoring; Termination Obligation.
European Commission's High-Level Expert Group on AI (European Commission, 2019)	Human Agency and Oversight, Technical Robustness and Safety, Privacy and Data Governance, Transparency, Diversity, Societal and Environmental Well-being, Accountability
AI4People (Floridi et al., 2018)	Beneficence: promoting well-being, preserving dignity, sustaining the planet; Non-maleficence: privacy, security, monitoring AI advancement/capability; Autonomy: the power to decide; Justice: promoting prosperity, preserving solidarity; Explicability: enabling the other principles through intelligibility and accountability
United Nations Educational, Scientific, and Cultural Organization (UNESCO, 2017)	Human Dignity, Value of Autonomy, Value of Privacy, "Do no harm" Principle, Principle of Responsibility, Value of Beneficence, Value of Justice
Australia's Ethics Framework (Dawson et al., 2019)	Generates Net-benefits, Regulatory and Legal Compliance, Fairness, Contestability, Do No Harm, Privacy Protection, Transparency and Explainability, Accountability

They adopt the definition of moral status that "X has moral status = because X counts morally in its own right, it is permissible/impermissible to do things to it for its own sake." From this perspective, once AI has moral status, we should treat it not as a machine/system, but an object that has equal rights as humans. The technological singularity, when technological growth becomes uncontrollable and irreversible, is hypothesized to come as AI advances. If it happens, human civilization would be affected, and robot rights and consciousness should be considered. But these issues are beyond the consideration of this paper. The following discussion mainly focuses on ethical issues related to narrow AI.

Research on ethical issues of AI falls into three categories: features of AI that may give rise to ethical problems (Timmermans et al., 2010), human factors that cause ethical risks (Larson, 2017), and social impact of ethical AI issues.

3.1. Features of AI Give Rise to Ethical Issues

3.1.1. Transparency

Machine learning is a brilliant tool, but it is hard to explain the inner processing of machine learning, which is usually called the "black box". The "black box" makes the algorithms mysterious even to its creators. This limits people's ability to understand the technology, leads to significant information asymmetries among AI experts and users, and hinders human trust in the technology and AI agents. Trust is crucial in all kinds of relationships and a prerequisite reason for acceptance (Siau & Wang, 2018).

Table 5. Summary of factors of ethical frameworks

Factors\Institutions	FLI	IAPP	IEEE	TPV	EUCE	AI4P	UNESCO	AEF	Total
Responsibility/Accountability	1	1	1	1	1	1	1	1	8
Privacy	1	1		1	1	1	1	1	7
Transparency	1	1	1	1	1			1	6
Human Values/Do No Harm	1	1	1	1			1	1	6
Human Well-Being/Beneficence			1		1	1	1		4
Safety	1				1	1			3
Liberty/Autonomy	1					1	1		3
Human Control	1				1			1	3
Bias/Fairness		1		1				1	3
Shared Benefit	1							1	2
AI Arms Race	1		1						2
Justice						1	1		2
Prosperity	1								1
Effectiveness			1						1
Accuracy				1					1
Reliability				1					1
Diversity					1				1
Human Dignity							1		1
Regulatory And Legal Compliance								1	1

Further, because of the black box that humans are not able to interpret, AI may evolve without human monitoring and guidance. For example, in 2017, Facebook shut down an AI engine because they found that the AI had created its own unique language and humans could not understand the language (Bradley 2017). Whether humans can control AI agents is a big concern. Humans prefer AI agents to always do exactly what we want them to do. For instance, if a guest asks a self-driving taxi to drive to the airport as fast as possible, the taxi may not follow the traffic rules, but reach the airport at the fastest speed. This is not what the customer wants but what the customer asked for literally. However, considering this problem from another perspective, if we treat AI agents ethically, is it ethical that we control what actions they take and how they make decisions?

3.1.2. Data Security and Privacy

The development of AI agents relies heavily on the huge amount of data, including personal data and private data. Almost all of the application domains in which deep learning is successful, such as Apple Siri and Google Home, have access to mountains of data. With more data generated in societies and businesses, there is a higher chance to misuse these data. For instance, a health record always contains sensitive information, which if not adequately protected, a rogue institution could gain access to that information and harm the patients personally and financially. Thus, data must be managed properly to prevent misuse and malicious use (Timmermans et al., 2010). To keep data safe, each action to the data should be detailed and recorded. Both the data per se and the transaction record may cause privacy-related risks. It is, therefore, important to consider what should be recorded and who should take charge of the recording action, and who can have access to the data and records.

3.1.3. Autonomy, Intentionality, and Responsibility

Whether the robots are regarded as moral agents affect the interactions (Sullins, 2011). To be seen as real moral agents, robots have to meet three criteria: autonomy, intentionality, and responsibility (Sullins, 2011). Autonomy means that the machines are not under the direct control of any other agents. Intentionality means that machines "act in a way that is morally harmful or beneficial and the actions are seemingly deliberate and calculated." Responsibility means the machines fulfill some social role that carries with it some assumed responsibilities. In the very classic trolley case, the one who controls the trolley is the ethical producer (Allen et al., 2006). To continue to run on the current track and kill five workers or to turn to another track and kill a lone worker is a hard-ethical choice for humans. What choice should or would AI make? Who should be responsible for the AI's choice? The military robots that take charge of bomb disposal are ethical recipients. Is it ethical that humans decide the destiny of these robots? Human ethics and morality today may not be seen as perfect by future civilizations (Bostrom and Yudkowsky, 2014). One reason is that humans cannot solve all the recognized ethical problems. The other reason is that humans cannot recognize all the ethical problems.

3.2. Human Factors Give Rise to Ethical Issues

Human Bias, such as gender bias (Larson, 2017) and race bias (Koolen and Cranenburgh, 2017), may be inherited by AI. AI agents are only as good as the data human put into them.

AI agents are being trained by humans and using datasets made by humans, existing biases may be learned by AI agents and exhibited in real applications. Once biased data are used by the AI agent, the bias will become an ongoing problem. For instance, software used to predict future criminals showed bias against a certain race (Bossmann, 2016). The bias comes from the training data that contains human biases. Thus, figuring out how to program and train AI agents without human biases is critical.

3.2.1. Accountability

When an AI agent fails at a certain assigned task, who should be responsible. This may lead to what is referred to as "the problem of many hands" (Timmermans et al., 2010). When using an AI agent, an undesirable consequence may be caused by the programming codes, entered data, improper operation, or other factors. Who should be the responsible entity for the undesirable consequence -- the programmer, the data owner, or the end-users?

3.2.2. Ethical Standards

"The ultimate goal of machine ethics is to create a machine that itself follows an ideal ethical principle or set of principles" (Anderson and Anderson, 2007 p. 15). It is theoretically easy but practically hard to formulate ethical principles for AI agents. Without comprehensive and unbiased ethical standards, how can humans train a machine to be ethical? Further, how can we make certain that intelligent machines understand ethical standards in the same way that we do? (Wang and Siau, 2019a). For instance, if we program robots to always perform no harm, we should first make sure that the robots understand what harm is. This results in another problem -- what should be the ethical standards for harm? A global or universal level of ethics is needed. To put such ethics into machines, it is necessary to reduce the information asymmetries between AI programmers and creators of ethical standards. While attempting to formulate ethical standards for AI and intelligent machines, researchers and practitioners should try to better understand existing ethical principles so that they will be able to apply the ethical principles to research activities and help train developers to build ethical AIs (Wang and Siau, 2018).

3.2.3. Human Rights Laws

Without training in human rights laws, software engineers may write codes that violate and breach key human rights without even knowing it. It is crucial to teach human rights laws to software

engineers. Ensuring privacy-by-design is important and more cost-efficient than the alternatives. A better knowledge of human rights laws can help AI designers and engineers eliminate or at least alleviate the discrimination and invasion of privacy issues in AI.

3.3. Social Impact of Ethical Issues

3.3.1. Automation and Job Replacement

The debate on whether AI age and Industry 4.0 (Siau, Xi, and Zou, 2019; Wang and Siau, 2019b) will create more jobs or eliminate some jobs is still heated. Stories of factory workers being replaced by automated systems and robots abound. Some argue that AI will also create millions of new jobs with many of these jobs that are non-existence today. Nevertheless, the concern is still there about the future workforce disruptions in the age of AI, such as the cooperation between humans and AI agents. The labor market will be disrupted and transformed with AI. What is not entirely clear is the speed and scope of the change. The term "useless class" has been suggested by Harari (2016). Universal Basic Income (UBI) has been piloted in some countries and Freedom Dividend, which is a universal basic income for all American adults with no strings attached, is the campaign platform for a 2020 U.S. Presidential candidate Andrew Yang. The original intention of technology development is to assist humans and improve human lives. If automation and AI cause huge job replacement and unemployment, should we keep the rapid pace of technology development? Also, how can we protect human rights and human well-being while keeping up with the rapid evolutions and revolutions of technology?

3.3.2. Accessibility

Accessibility, as an ethical principle, refers to whether systems, products, and services are available, accessible, and suitable for all people, including the elderly, the handicapped, and the disabled. Considering the complexity of new technology and high-tech products, as well as the aging populations in some countries, the accessibility of new technology will directly affect human well-being. Technology development should benefit humans. But if only a portion of people benefit, is it ethical and fair? Consideration must be given to developing systems, products, and services that are accessible to all, and the benefits of advanced technology should be fairly distributed to all (Wang and Siau, 2019b).

3.3.3. Democracy and Civil Rights

Unethical AI results in the fragmentation of truth and eventual loss of trust, and loss of societal support for AI technology. The loss of informed and trusting communities dents the strengths of democracies. As democracies suffer and structural biases amplified, the free exercise of civil rights no longer remains uniformly available to all. AI ethics needs to take into consideration democracy and civil rights.

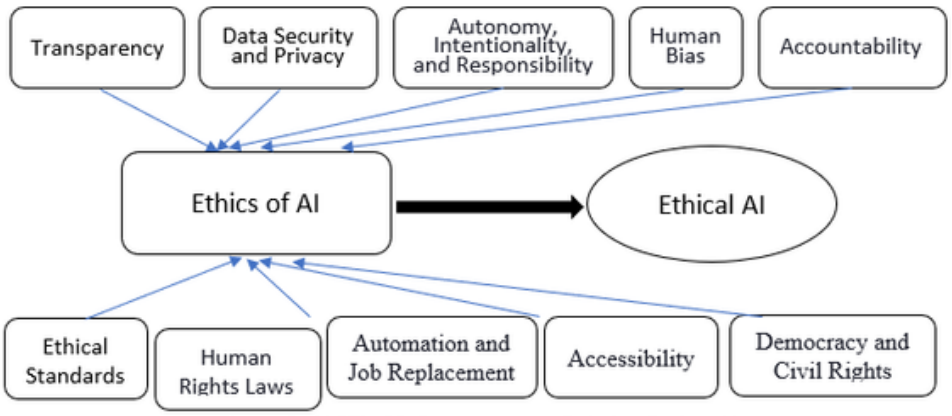
Figure 2 summarizes the above ethical issues in AI. Solving these issues properly will help to establish a framework for building ethical AI.

4. DISCUSSIONS

4.1. From Framework to Practice

Figure 2 establishes the framework for AI ethics listing the factors that need to be considered in defining the ethics of AI in order to build ethical AI. Even though defining the ethics of AI is multifaceted and convoluted, putting the ethics of AI into practice to build ethical AI is no easy feat too. What should ethical AI look like? In the simplest form, we may define that ethical AI should do no harm to humans. But, what is harm? What constitutes human rights? Many questions need to be answered before we can design and build ethical AI. ethical sensitivity training is required to make good ethical

Figure 2. AI Ethics: Framework of building ethical AI



decisions. In theory, AI should be able to recognize ethical issues. If AI is capable of making decisions, how can we design and develop an AI that is sensitive to ethical issues? Unfortunately, it is not easy to implement in practice and to realize. Long-term and sustained efforts are needed. Nonetheless, understanding and realizing the importance of developing ethical AI and starting to work on it step by step are positive steps forward.

Many institutions, such as Google, IBM, Accenture, Microsoft, and Atomium-EISMD, have started working on building ethical principles to guide the development of AI. In November 2018, the Monetary Authority of Singapore (MAS), together with Microsoft and Amazon Web Services, launched the FEAT principles (i.e., fairness, ethics, accountability, and transparency) for the use of AI. Academics, practitioners, and policymakers should work together to widen the engagement to establish ethical principles for AI design, development, and use.

With the frameworks and principles, protective guardrails to ensure ethical behaviors are needed. Good governance is necessary to enforce the implementation and adherence to those ethical principles, and a legal void is waiting to be filled by regulatory authorities (Hanna, 2019). Either based on case law or accomplished via legislative and regulatory obligations, these legal and regulatory instruments will be critical to the good governance of AI, which helps to implement and enforce ethics of AI to enable the development of ethical AI.

To protect the public, the U.S. has long enacted regulatory instruments, such as rules against discrimination, equal employment opportunity, HIPAA Title II, Commercial Facial Recognition Privacy Act, and Algorithmic Accountability Act. All these instruments would be useful in guiding the development of legal and regulatory policies and frameworks for AI ethics.

In addition to the legal and government rules, self-regulation plays an important role as well. Communication and information disclosure can help society as a whole to ensure the development and deployment of ethical AI. Fostering discussion forums and publishing ethical guidelines by companies, industries, and policymakers can help to educate and train the public in understanding the benefits of AI and dispelling myths and misconceptions about AI. Besides, having a better knowledge of legal frameworks on human rights, strengthening the sense of security, and understanding the ethical issues related to AI, can foster trust in AI and enable the development of ethical AI more efficiently and effectively.

4.2. Ways to Educate AI to Be Ethical

Moor (2006) indicates three potential ways to transfer AI to ethical agents: to train AI into "implicit ethical agents", "explicit ethical agents", and "full ethical agents". Implicit ethical agents mean

constraining the machine's actions to avoid unethical outcomes. Explicit ethical agents mean stating explicitly what action is allowed and what is forbidden. Full ethical agents mean machines, as humans, have consciousness, intentionality, and free will. An implicit ethical agent can restrict the development of AI. An explicit ethical agent is currently getting the most attention and is considered to be more practical (Anderson and Anderson, 2007). A full ethical agent is still an R&D initiative and one is not sure when a full ethical agent will be a reality.

When a full ethical agent is realized, how to treat an AI agent that has consciousness, moral sense, emotion, and feelings will be another critical consideration. For instance, is it ethical to "kill" (shut down) an AI agent if it replaces human jobs or even endangers human lives? Is it ethical to deploy robots into a dangerous environment? These questions are intertwined with human ethics and moral values.

4.3. Tradeoff Between AI Ethics and AI Advancement

President-elect of the European Commission made clear in her recently unveiled policy agenda that the cornerstone of the European AI plan will be to ensure that "AI made in Europe" is more ethical than AI made anywhere else in the world. The European Commission is not the only one that is concerned about AI ethics. Many countries are also working on AI ethics. U.S. agencies such as the Department of Defense and the Department of Transportation have launched their initiatives to ensure the ethical use of AI within their respective domains. In China, the government-backed Beijing Academy of Artificial Intelligence has developed the Beijing AI Principles that rival those of other countries, and the Chinese Association for Artificial Intelligence has also developed its own ethics guidelines. Many non-European countries, including the United States, have signed on to the Organization for Economic Co-operation and Development's (OECD) AI Principles focusing on "responsible stewardship of trustworthy AI."

However, the makers and researchers of AI at this time are likely to pay more attention to hard performance metrics, such as safety and reliability, or softer performance metrics, such as usability and customer satisfaction. More nebulous concepts like ethics are not yet the most urgent consideration – especially with the intense competition between companies and between nations.

Further, while some consumers may pay lip service to ethical design, their words do not match their actions. For example, among consumers who said they distrust the Internet, only 12% report using technological tools, such as virtual private network, to protect their data, according to a worldwide Ipsos survey (CIGI-Ipsos, 2019). Instead, the most important factors influencing consumers' purchasing decisions are still the price and quality. Right now, consumers care more about what AI can do rather than whether all AI's actions are ethical.

This situation may put companies and institutions which are developing AI in a tradeoff situation -- whether to focus on AI advancement to realize profit maximization, or to focus on AI ethics to ensure societal benefits from AI innovations.

5. CONCLUSION

Understanding and addressing ethical and moral issues related to AI is still in the infancy stage. AI ethics is not simply about "right or wrong", "good or bad", and "virtue and vice". It is not even a problem that can be solved by a small group of people. However, ethical and moral issues related to AI are critical and need to be discussed now. This research aims to call attention to the urgent need for various stakeholders to pay attention to the ethics and morality of AI agents. While attempting to formulate the ethics of AI to enable the development of ethical AI, we will also understand human ethics better, improve the existing ethical principles, and enhance our interactions with AI agents in this AI age. AI ethics should be the central consideration in developing AI agents and not an afterthought. The future of humanity may depend on the correct development of AI ethics!

REFERENCES

Allen, C., Wallach, W., & Smit, I. (2006). Why machine ethics? *IEEE Intelligent Systems*, *21*(4), 12–17. doi:10.1109/MIS.2006.83

Anderson, M., & Anderson, S. L. (2007). Machine ethics: Creating an ethical intelligent agent. *AI Magazine*, *28*(4), 15–26.

Anderson, M., & Anderson, S. L. (2011). *Machine ethics*. Cambridge University Press. doi:10.1017/CBO9780511978036

Asimov, I. (1950). Runaround. In I, Robot (The Isaac Asimov Collection Ed.). New York City: Doubleday.

Blanchard, K., & Peale, N. V. (2011). *The power of ethical management*. Random house.

Bossmann, J. (2016). Top 9 ethical issues in artificial intelligence. World Economic Forum. Retrieved from https://www.weforum.org/ethical-issues-in-AI

Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In The Cambridge handbook of artificial intelligence (pp. 316-334). Cambridge Press. doi:10.1017/CBO9781139046855.020

Bradley, T. (2017). Facebook AI Creates Its Own Language In Creepy Preview of Our Potential Future. Forbes. Retrieved from https://www.forbes.com/sites/tonybradley/2017/0 7/31/facebook-ai-creates-its-own-language-in-creepy-preview-of-our-potential-future/#45c65554292c

Churchill, L. R. (1999). Are We Professionals? A Critical Look at the Social Role of Bioethicists. *Daedalus*, 253–274. PMID:11645877

CIGI-Ipsos. (2019). 2019 CIGI-Ipsos Global Survey on Internet Security and Trust. Retrieved from http://www.cigionline.org/internet-survey-2019

Dawson, D., Schleiger, E., Horton, J., McLaughlin, J., Robinson, C., Quezada, G., . . . Hajkowicz, S. (2019). Artificial Intelligence: Australia's Ethics Framework. Data61 CSIRO, Australia.

European Commission. (2019). Ethics guidelines for trustworthy AI. Retrieved from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., & Vayena, E. et al. (2018). AI4People—an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707. doi:10.1007/s11023-018-9482-5 PMID:30930541

Future of Life Institute. (2017). Asilomar AI Principles. Retrieved from https://futureoflife.org/ai-principles/?cn-reloaded=1

Hanna, M. (2019). We don't need more guidelines or frameworks on ethical AI use. It's time for regulatory action. Brink the Edge of Risk. Retrieved from https://www.brinknews.com/we-dont-need-more-guidelines-or-frameworks-on-ethical-ai-use-its-time-for-regulatory-action/

Harari, Y. N. (2016). *Homo Deus: a brief history of tomorrow*. Random House.

IAPP. (2018). White Paper -- Building Ethics into Privacy Frameworks for Big Data and AI. Retrieved from https://iapp.org/resources/article/building-ethics-into-privacy-frameworks-for-big-data-and-ai/

IEEE. (2019). Ethically aligned Design. Retrieved from https://ethicsinaction.ieee.org/

Koolen, C., & van Cranenburgh, A. (2017). These are not the Stereotypes you are looking For: Bias and Fairness in Authorial Gender Attribution. Proceedings of the First ACL Workshop on Ethics in Natural Language Processing (pp. 12-22). Academic Press. doi:10.18653/v1/W17-1602

Larson, B.N. (2017). Gender as a variable in natural-language processing: Ethical considerations.

Libby, K. (2019). This Bill Hader Deepfake video is Amazing. It's also Terrifying for our Future. Popular Mechanics. Retrieved from https://www.popularmechanics.com/technology/security/a28691128/deepfake-technology/

Livni, E. (2019). A new kind of cybercrime uses AI and your voice against you. Quartz. Retrieved from https://qz.com/1699819/a-new-kind-of-cybercrime-uses-ai-and-your-voice-against-you/

Markkula Center for Applied Ethics. (2015). A framework for ethical decision making. Santa Clara University. Retrieved from https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/a-framework-for-ethical-decision-making/

Moor, J. H. (1985). What is computer ethics? *Metaphilosophy*, *16*(4), 266–275. doi:10.1111/j.1467-9973.1985.tb00173.x

Moor, J. H. (2006). The nature, importance, and difficulty of machine ethics. *IEEE Intelligent Systems*, *21*(4), 18–21. doi:10.1109/MIS.2006.80

Nalini, B. (2019). The Hitchhiker's Guide to AI Ethics. Medium. Retrieved from https://towardsdatascience.com/ethics-of-ai-a-comprehensive-primer-1bfd039124b0

BBC News. (2017). Fake Obama created using AI video tool [YouTube video]. BBC News. Retrieved from https://www.youtube.com/watch?v=AmUC4m6w1wo

Now, A. (2018). Human rights in the age of artificial intelligence. Accessnow.org. Retrieved from https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf

Paul, R., & Elder, L. (2006). *The Miniature Guide to Understanding the Foundations of Ethical Reasoning*. United States: Foundation for Critical Thinking Free Press. P. NP.

Resnik, D. B. (2011). What is ethics in research and why is it important. *National Institute of Environmental Health Sciences*, *1*(10), 49–70.

Siau, K., & Wang, W. (2018). Building Trust in Artificial Intelligence, Machine Learning, and Robotics. *Cutter Business Technology Journal.*, *31*(2), 47–53.

Siau, K., Xi, Y., & Zou, C. (2019). Industry 4.0: Challenges and Opportunities in Different Countries. *Cutter Business Technology Journal.*, *32*(6), 6–14.

Stupp, C. (2019). Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case. The Wall Street Journal. Retrieved from https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402

Sullins, J. P. (2011). When is a robot a moral agent. *Machine ethics*, 151-160.

Telcher, J. G. (2018). What do facial recognition technologies mean for our privacy? *The New York Times*. Retrieved from https://www.nytimes.com/2018/07/18/lens/what-do-facial-recognition-technologies-mean-for-our-privacy.html?nytapp=true&smid=nytcore-ios-share

The Public Voice. (2018). Universal Guidelines for Artificial Intelligence. Retrieved from https://thepublicvoice.org/ai-universal-guidelines/

Timmermans, J., Stahl, B. C., Ikonen, V., & Bozdag, E. (2010). The ethics of cloud computing: A conceptual review. *Proceedings of the IEEE Second International Conference Cloud Computing Technology and Science* (pp. 614-620). IEEE Press. doi:10.1109/CloudCom.2010.59

UNESCO. (2017). Report of World Commission on the Ethics of Scientific Knowledge and Technology on Robotics Ethics. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000253952

Wang, W., & Siau, K. (2018). *Ethical and moral issue with AI – a case study on healthcare robots. AMCIS 2019 proceedings*. Academic Press.

Wang, W., & Siau, K. (2019a). Artificial Intelligence, Machine Learning, Automation, Robotics, Future of Work and Future of Humanity: A Review and Research Agenda. *Journal of Database Management*, *30*(1), 61–79.

Wang, W., & Siau, K. (2019b). Industry 4.0: Ethical and moral Predicaments. *Cutter Business Technology Journal.*, *32*(6), 36–45.

Yu, J. (1998). Virtue: Confucius and Aristotle. *Philosophy East & West*, *48*(2), 323–347. doi:10.2307/1399830

Keng Siau is Chair of the Department of Business and Information Technology at the Missouri University of Science and Technology. Previously, he was the Edwin J. Faulkner Chair Professor and Full Professor of Management at the University of Nebraska-Lincoln (UNL), where he was Director of the UNL-IBM Global Innovation Hub. Dr. Siau also served as VP of Education for the Association for Information Systems. He has written more than 300 academic publications, and is consistently ranked as one of the top information systems researchers in the world based on the h-index and productivity rate. Dr. Siau's research has been funded by the U.S. National Science Foundation, IBM, and other IT organizations. He has received numerous teaching, research, service, and leadership awards, including from the International Federation for Information Processing Outstanding Service Award, the IBM Faculty Award, and the IBM Faculty Innovation Award. Dr. Siau received his Ph.D. in Business Administration from the University of British Columbia. He can be reached at siauk@mst.edu.

Weiyu Wang holds a Master of Science degree in Information Science and Technology and an MBA from the Missouri University of Science and Technology. Her research focuses on the impact of artificial intelligence (AI) on economy, society, and mental well-being. She is also interested in the governance, ethical, and trust issues related to AI. She can be reached at wwpmc@mst.edu.