Journal of Systemics, Cybernetics and Informatics (2023) 21(4), 42-50
https://doi.org/10.54808/JSCI.21.04.42

The Ethics of Artificial Intelligence in the Era of Generative AI

Vassilka D. KIROVA
vassilka.kirova@bell-labs-consulting.com
Nokia Bell Labs
New Providence, NJ, USA

Cyril S. KU
kuc@wpunj.edu
Department of Computer Science
William Paterson University
Wayne, NJ, USA

Joseph R. LARACY
joseph.laracy@shu.edu
Department of Systematic Theology
Department of Mathematics and Computer Science
Seton Hall University
South Orange, NJ, USA

Thomas J. MARLOWE
thomas.marlowe@shu.edu
Department of Mathematics and Computer Science
Seton Hall University
South Orange, NJ, USA

ABSTRACT

In the early 2020s, advances in transformer-based deep neural networks enabled the development and growth of a number of generative artificial intelligence (GenAI) systems notable for accepting natural language prompts as input. These include large language model chatbots such as ChatGPT, Bard, and others. GenAI has applications across a wide range of industries, including art, writing, software development, product design, healthcare, finance, gaming, and more. In this paper, we place these recent advances in a historical, cybernetic context. We analyze ethical issues that arise in the area of software engineering and cyber-physical systems. In addition, we explore AI-based challenges in healthcare and medicine, including a number involving GenAI. This research shows the importance of rigorous ethical analysis and resulting safeguards to address the emerging issues with AI.

Keywords: Artificial Intelligence, Generative AI, Large Language Models, Ethics, Healthcare, Cyber-physical Systems, Software Engineering, Cybernetics

1. INTRODUCTION

Generative AI can trace its origins in research back to the ELIZA chatbot and its relatives in the 1960's [1], which combined generative features with rule bases, but were primarily an entertaining research sidetrack. Other than these curiosities, there were no major developments until the early twenty-first century, with the connection to machine-learning models and algorithms based on neural nets and related technologies [2]. This changed with the release in 2022 of the first commercial GenAI systems, followed closely by the end of that year by ChatGPT, developed by the American research laboratory OpenAI. ChatGPT and its cousins have made "a big splash" across the whole of human endeavor and expression. Areas noticeably affected include, but are hardly limited to, education at all levels [3, 4], academic research [5], arts and the media [6], software engineering and computer-based systems [7], and health and medicine [8]. In each case there are benefits and opportunities, but also risks and ethical challenges. The rapid pace of developments in machine learning and data science, AI in general, and generative LLM models such as ChatGPT in particular, and the ever-wider range of their deployment, continue to receive attention, focusing on legal and ethical issues, and on standards, regulation, safeguards, reporting, and control (see for example [9, 10]).

This paper looks at the last two of the above areas as domains of special concern: software engineering both as the source of future generative AI developments and because of the ubiquity of software and cyber-physical systems in modern life, and health and medicine because the practice of medicine has always entailed ethical challenges. In addition, each of these areas already has well-understood structures for standards, safeguards, reporting, and control, facilitating discussion of the need for and nature of more stringent and modified implementation, and of additional safeguards.

The rest of this paper is structured as follows. Section 2 looks at the historical connections between AI and cybernetics and Section 3 offers an overview of the philosophical foundations of professional ethics. Section 4 then presents a capsule review of modern AI and ethical challenges. Sections 5 and 6 then consider two areas with emergent use of AI in applications, software engineering and cyber-physical systems, and then health and medicine. Section 7 follows with discussion, future directions, and conclusions.

2. AI AND CYBERNETICS

This revolution in computing is not simply of interest to current and future software developers, but also bears relevance for the interdisciplinary cybernetics community. Generative AI can be viewed as a fusion of *Markov* and *neural network* models with data analysis and training sets. A Markov model, named after the Russian mathematician Andrey Markov, is a stochastic model used to understand pseudo-random processes in which future states depend only on the current state, not on the events that occurred before it. A countably infinite sequence, in which the chain changes state at discrete time steps, produces a discrete-time Markov chain (DTMC). Generative AI uses such Markov chains to generate sequences of data, e.g., text, music, or images, by modeling the probability of transitioning from one state to another [11]. Neural networks are the backbone of deep learning algorithms and are inspired by the massive parallel computing capability of the human brain. A classical mathematical model for parallel computing is the petri net. The petri net, named after the German mathematician Carl Adam Petri [12], is a class of discrete event dynamical systems constructed by a directed bipartite graph with two types of elements: places and transitions [13].

Some of the earliest discussions around what is now referred to as AI took place at the Macy Cybernetics Conferences held between 1946 and 1953. There, visionaries such as William Ross Ashby, Heinz von Foerster, Warren McCulloch, John von Neumann, Claude Shannon, Norbert Wiener, Margaret Mead, and many others set out to establish the foundations for a general science of the workings of the human mind [14]. These early studies of interdisciplinarity led to modern systems theory, cybernetics, cognitive science, and artificial intelligence. In the first conference in New York City in 1946, participants discussed topics such as self-regulating and teleological mechanisms, simulated neural networks emulating the calculus of propositional logic, as well as anthropology and how computers might learn how to learn [15].

From a historical perspective, it is significant to note that the first Macy Conference was ten years prior to the Dartmouth Summer Research Project on Artificial Intelligence, a workshop widely considered to be the founding event of artificial intelligence as a field. At the same time, computer science was emerging as a stand-alone discipline [16]. Columbia University offered one of the first for-credit courses in computer science in 1946. The first degree program was created in the UK at the University of Cambridge in 1953. This was followed by the first computer science department in the United States at Purdue University in 1962 [17].

Computational research was also taking place in this period in the context of electrical engineering departments. This led to the first computer engineering degree program in the United States in 1971 at Case Western Reserve University [18]. Second-order cybernetics partially owes its origin to the Biological Computer Laboratory (BCL) at the University of Illinois Urbana-Champaign. It was founded in 1958 by electrical engineering professor Heinz von Foerster. BCL carried out research in the areas of self-organizing systems, bionics (e.g., artificial neurons and neural networks), and bio-inspired computing. Thus, since the mid-twentieth century, the fields of cybernetics, computing, and AI have always been highly intertwined [19].

3. PHILOSOPHICAL FOUNDATIONS OF PROFESSIONAL ETHICS

Research scientists, physicians, and engineers cannot remain on the level of emotivism (i.e., reducing moral judgments to mere expressions of feelings) or “bare minimum” legal compliance when faced with serious ethical issues, especially involving AI. Etymologically, the English word “ethics” (*ethica* in Latin) can be traced back to the ancient Greek noun, ἔθος (*ethos*), which denotes a “habit” or “custom.” There is also the derived Greek adjective, ἠθικός (*ethikos*), which means “moral” or “expressing character.” As an academic discipline, ethics, also known as moral philosophy, studies those acts of the will affirming or rejecting the order proposed by reason. It is a science, in the Aristotelian sense, that deals with the acts of the will in their order to each other and in their order to their end [20].

Ethics is a practical discipline that refers to human action with the purpose of being morally good. Whereas the natural sciences, and other theoretical sciences, are descriptive in character, concerned with empirical or phenomenological facts, ethics is prescriptive in character, concerned with normative values. As a normative value-based discipline, ethics is more concerned with what ought to be than what is.

Scientists, physicians, and engineers are traditionally encouraged to employ teleological, deontological, virtue, and other approaches to address ethical issues in their professions. Teleological ethics focuses on the τέλος (*telos*), i.e., end or goal of an act. Thus, teleological approaches to ethics judge an act to be right or wrong in relation to the outcome, i.e., I deliberately order my actions toward the realization of an outcome which I previously identified as good. For example, one might ask how the deployment of AI in a particular domain would ultimately affect human safety? Deontological approaches are rooted in δέον (*deon*), i.e., duty, and thus judge the ethics of an act based on the inherent rightness or wrongness, rather than of its consequences. For example, what is my responsibility to the client, patient, or public in developing or utilizing this particular AI system? Virtue ethics is based on the Latin term *virtus*, i.e., excellence of character and can be classified as a species of teleological ethics. It is established on the principle that one should act in a way that a virtuous person would act in this situation. For example, when confronting a particular challenge in healthcare privacy and data analytics, how would a person of the highest moral character act here? [21, 22, 23].

The classical Thomist approach to ethics is clearly applicable here as well. The fundamental three criteria are

- the **moral object** (i.e., the substance of the action, the end inherent to the chosen act),
- the **intention** of the acting subject (i.e., the reason or purpose, for which the act was chosen), and
- the **circumstances** or concrete conditions surrounding and touching upon the action (i.e., the totality of the reasonably anticipated consequences of the act).

The moral object is the substance or subject matter of a human act; it answers the question: “What have you done?” The intention answers the question, from the agent’s point of view, “Why is this act being done?” The circumstances of an action are defined as the particular conditions of each act that are outside the act, and yet in some way touch upon the act. To be a morally good act, all three elements must be good [24]. As Anderson et. al. point out, the principles of Catholic Social Teaching (CST), and some of its analogs from other traditions, can also be fruitfully applied to ethical issues that arise in engineering and technology [25].

Medical, engineering, and science curricula should address ethical issues in professional practice such as safety, liability, professional responsibility to clients or patients as well as employers, whistle-blowing, codes of ethics, and legal obligations. Such ethics courses should provide philosophical analysis based on various ethical theories and review numerous case studies. Given the issues that are arising around fairness, accountability, transparency, and privacy, a substantial number of classroom hours should be set aside for AI. For example, bias in AI, often resulting from biased training data, may lead to skewed outcomes and unjust discrimination. This can have very serious implications, particularly when AI is used in sensitive areas such as hiring, lending, criminal justice, or law enforcement. Thus, it is not surprising that a team of psychologists and cognitive scientists have called for the establishment of an independent scientific body to test and certify generative artificial intelligence, before the technology damages science and public trust [26].

4. AI AND ETHICS

4.1 Brief Overview of Modern AI

Modern AI has transitioned from theoretical and aspirational to practical and ubiquitous, touching almost every facet of our digital lives [27, 28, 29]. It encompasses a broad spectrum of algorithms, methodologies, and technologies. It enables machines to simulate human-like cognitive functions such as learning, reasoning, problem-solving, and language understanding. At the core of modern AI is machine learning (ML), which allows computers to learn from and make predictions or decisions based on data (usually large amounts – the so-called big data). Machine learning and a subset called deep learning improve these capabilities over time without being explicitly programmed. Deep learning uses neural networks modeled after the human brain (neurons) that can process complex patterns in big data. These technologies make generative AI possible, which are AI models designed to generate novel data, such as text, image, code, music, or other structured output, that is often indistinguishable from content created by humans [30, 31, 32]. AI today is not just about programming computers to execute tasks; it enables machines to learn and adapt. The proliferation of big data, increasing computational power, and algorithmic advancements have fueled this AI evolution, allowing machines to uncover or discover patterns and insights at scales and speeds beyond human capability.

4.2. Basics of Generative AI

Generative AI combines the above ingredients with a few more in a sophisticated machine-learning model. Interactive Large Language Models (LLMs) such as ChatGPT use (1) a natural language understanding module to determine the content of the prompt, (2) a training set comprised of information, perhaps including articles, stories, and conversations, (3) a model/architecture (e.g., a Markov model, generative adversarial networks (GAN), or Google's transformer architecture) to stochastically determine the next word in the response given previous words in the sentence or the line, together perhaps with (4) grammatical rules to ensure the response isn't jarring, and most importantly, (5) a machine-learning component that manages generation of the response. These are accompanied by a user interface to handle the required communication.

An interaction seeking a textual response proceeds as follows: The user generates a prompt (Step 1), which is passed to the LLM transformer engine (Step 2), The engine uses the learning from the model rules (Step 3.1) and training sets (Step 3.2) to generate a response (Step 4), which is then passed to the user (Step 5). See Figure 1.

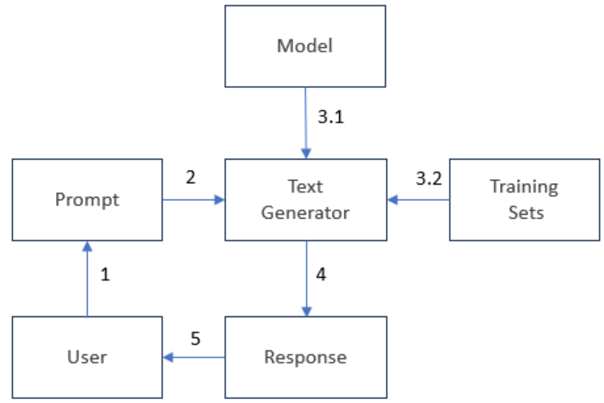

Figure 1. Sketch of an LLM interaction

There are of course variants. Generation of other artifacts, such as graphics, animation, videos, or music, proceeds similarly, each with its own rules for generation. Generation of spoken responses will typically entail the creation of the text, and then following rules for its enunciation. Generation of text in special forms, such as a sonnet or a model essay, will require either multiple examples in the training set, or a set of rules for the form, or both. Similarly, to use generative AI in a specialized context, such as finance, the training set will have to stress material on that topic. It should also be noted that one can use an LLM tool in an iterative or faceted fashion, by refining the prompt or by asking for a different aspect or stakeholder perspective, or for a different audience, possibly using prompt engineering.

4.3. Ethics of Applied AI

The ethics of AI concerns itself with the moral principles that govern the practical implementation of AI. As AI systems integrate into our daily lives and drive innovation in every industry, ethical questions naturally arise about the future of work, the potential displacement of jobs, and the new forms of digital divide. As mentioned in Section 3, ethics involves navigating the delicate balance between technological advancement and the preservation of human values such as privacy, fairness, transparency, and accountability. For example, there is a growing imperative to ensure that AI-driven decisions, especially in critical domains like healthcare, criminal justice, and employment, do not reflect or amplify societal biases, thus necessitating the development of fair and equitable AI. Privacy is another significant concern, as AI systems may contain sensitive personal information, prompting debates about consent and data protection. Navigating these challenges requires a collaborative approach, involving ethicists, scientists, technologists, policymakers, as well as the public, to develop AI that is not only intelligent and effective but also aligned with human values and ethical standards [33, 34, 35].

One example of the challenges is how humans can control the training of the generative AI models. For example, when is it appropriate to use the crowdsourcing, human-in-the-loop, or expert-in-the-loop approaches? These are different methods for integrating human judgment into AI systems [36]. The appropriateness of each depends on the specific requirements, context, and goals of the AI application. For instance, a hybrid model can be employed where the crowd sources initial data, human-in-the-loop is used for ongoing system improvements, and experts-in-the-loop are brought in for complex or high-risk decisions. The key is to balance the cost, speed, scalability, and quality requirements of the generative AI system with the appropriate level of human involvement.

Another issue raised or intensified by generative AI is the issue of "deepfakes," where images or videos are manipulated for self-aggrandizement, to create disinformation, or to bully or shame others, or even worse. While this was possible with some effort using tools such as Photoshop, generative AI makes it easy to create convincing deepfakes with minimal effort, resulting in serious emotional harm to the victims [37], or in some cases even suicide.

Various public and private organizations including the ACM and IEEE, the European Union, UNESCO, IBM, and Google, have already become active in this space, issuing statements of ethical principles, either general, or focused on software development and/or AI. Three of these are summarized in Table 1 below.

5. APPLICATION: SOFTWARE DEVELOPMENT AND CYBER-PHYSICAL SYSTEMS

The growing application of Artificial Intelligence and Machine Learning (AI/ML) tools, models and algorithms, and the recent rapid adoption of LLMs and Generative AI have multifaceted impact on Software Engineering (SE) and development of cyber-physical systems (CPS). Software engineering is the discipline of applying engineering principles and practices to the design, development, testing, maintenance, and evolution of software systems that run on various devices and platforms. Cyber-physical systems (CPS) are a subset of software systems that integrate computation, communication, and physical processes to achieve a desired functionality. Examples of CPS include industrial robots, medical devices, smart grids, smart homes, and autonomous vehicles.

UNESCO	IEEE	IBM
UNESCO ethical recommendations are based on specific core values such as human dignity and rights, promoting peace, and care for the environment. Based on these values, UNESCO specifies ten principles: 1. Proportionality and Do No Harm 2. Safety and Security 3. Right to Privacy and Data Protection 4. Multistakeholder and Adaptive Governance & Collaboration 5. Responsibility and Accountability 6. Transparency and Explainability 7. Human Oversight and Determination 8. Sustainability 9. Awareness and Literacy 10. Fairness and Non-discrimination [38]	The IEEE Standards Association (SA) has established a Global Initiative on the Ethics of Autonomous and Intelligent Systems. The IEEE approach is established on eight fundamental principles: 1. Human Rights, 2. Well-being, 3. Data Agency, 4. Effectiveness, 5. Transparency, 6. Accountability, 7. Awareness of Misuse, and 8. Competence [39]	IBM proposes three guiding values for AI: 1. The purpose of AI is to augment human intelligence, 2. Data and insights belong to their creator, and 3. Technology must be transparent and explainable. Leveraging insights from the 1979 Belmont Report, IBM defines three overarching principles for AI: 1. Respect for persons, 2. Beneficence, and 3. Justice, i.e., burdens and benefits may be distributed either by: a. Equal share, a. Individual need, a. Individual effort, a. Societal contribution, or a. Merit [40]

Table 1. Ethical Principles Statements from selected organizations

Software engineering is a complex and challenging process that requires creativity, analysis, logic, and technical skills. Software developers often face problems such as bugs, errors, inefficiencies, security challenges, and compatibility issues. Cyber-physical systems are often even more challenging to design, implement, and maintain, as they usually pose very strict quality requirements such as reliability, performance, and safety, integrate components from multiple domains, require diverse sets of skills, and involve a number of experts in the engineering process. Software systems, including CPS, have become ubiquitous in our society, they are embedded in our daily lives, and their quality, trustworthiness, and safety are critical properties to safeguard as we introduce new technologies.

AI/ML and generative AI technology offer numerous benefits to SE projects, such as automating tasks, enabling closed-loop operations, enhancing creativity, increasing productivity, and improving consistency and quality. These can be beneficial in most SE workflows, a topic we address in detail in an upcoming paper [41], from providing alternative designs to code generation and automated documentation. Combined with modern software practices such as agile development and other mature tools such as dependency tracking, they can lead to effective, speedy deliveries [42].

At the same time, Generative AI systems pose new risks and challenges. There are well-known problems of hallucination and the introduction of biases. There is also the lack of transparency, and the introduction of difficult-to-debug errors. These issues can impact the quality of many software engineering artifacts and require new added rigor in reviews, testing and verification and validation activities. For example, empirical studies have shown that they can generate insecure code [43]. There are two main contributing factors to the insecure code generation – first, using training datasets that don't represent the best practice in code security, often relying on published data, and second using evaluation metrics that are overly focused on functionality and not on non-functional properties as security and reliability. Further, there is also a risk of developing organizational and individual over-dependency on the ability of LLMs and Generative AI systems to cut through complex problems easily and thus slowly degrading the development and retention of selected cutting-edge knowledge and skills in the teams and organizations.

The use of LLMs and Generative AI in software engineering and development of CPSs raises additional ethical concerns that need to be addressed by all stakeholders, including software developers, AI experts and data scientists, CPS engineers, users, regulators, and society at large. Software professionals, data scientists and AI experts, who are involved in developing AI systems and integrating them into different complex applications, including CPSs, can have a strong impact and are expected to act as proactive agents that support meaningful applications of ethics in system design and development.

Many of the major ethical concerns are related to privacy and confidentiality, intellectual property rights, fairness, and trust. LLMs and Generative AI can potentially infringe on the privacy of individuals and organizations by accessing, processing, or generating sensitive or personal data without proper consent or knowledge. LLMs and Generative AI can also generate synthetic data or content that can impersonate or reveal the identity, behavior, or preferences of individuals or groups, such as "deepfakes," synthetic voices, or synthetic biometrics [44]. Therefore, it is essential to protect the privacy of data providers and data consumers, and to respect their rights and preferences regarding the collection, use, and disclosure of their data.

LLMs and Generative AI can potentially discriminate against or harm certain individuals or groups by producing or amplifying biased or unfair outputs. For this, it is crucial for software professionals and CPS engineers to prevent or mitigate any potential issues by applying ethical AI principles, by ensuring an unbiased and representative training set, and thorough SE practices, with reviews, testing, verification, and validation, including use of static and other analyses.

LLMs and Generative AI can potentially undermine or erode the trust of individuals and organizations in software and CPSs by producing or causing unreliable, inaccurate, or deceptive outputs. For example, LLMs and Generative AI can generate code, data, or content that can contain or introduce errors, bugs, or vulnerabilities that can affect the functionality, performance, or security of software, as discussed earlier, and very importantly, jeopardize the safety of cyber-physical systems. They could also generate code, data, or content that can mislead or manipulate, e.g., fake news or misinformation. Therefore, it is vital to ensure the trustworthiness and credibility of the outputs generated by LLMs and Generative AI, and to provide adequate information, explanation, and verification of the sources, methods, and limitations [45].

Protection of intellectual property is also a serious concern, in both incoming and outgoing directions. The former can occur when third-party proprietary code or data is included in the training set (or if the data set is enriched by searching online or adding examples from other completed projects); the latter may happen when enterprise confidential information is included in LLM-generated code or documentation. Adherence to specified properties and required constraints is another issue. Code generated by LLM models may not preserve these properties, including security properties such as access control. CPS systems introduce additional challenges. As important examples, safety always needs to be considered, and timeliness for real-time properties needs to be assured. Again, assuring such properties calls for rigorous review, testing, validation, and verification, including static analyses and platform testing.

To address these ethical considerations, some of the key strategies and best practices focus on the quality of data, models, and outputs, and on the establishment and application of globally accepted ethical principles and practices. Software engineers and CPS developers should be educated and trained for proper and ethical interaction with generative AI systems, components and artifacts. They should ensure the quality of the data used to train, test, or evaluate LLMs and Generative AI systems and applications. The data has to be relevant, accurate, consistent, and representative of the intended domain, task, and population. Further, the data has to be obtained, processed, and stored in a lawful, ethical, and transparent manner, respecting the privacy, consent, and ownership of the data providers. SEs and CPS developers should also ensure that the models used to implement LLMs, and Generative AI are validated, verified, and evaluated using appropriate methods, metrics, and benchmarks, and that the results are reported and documented in a clear, honest, and reproducible manner [45]. Further, the outputs have to be monitored, reviewed, and corrected by human experts, and the users have to be informed and educated about the source, nature, and limitations of the outputs.

Finally, SE and CPS developers should adhere to the established and emerging ethical principles and guidelines defined to govern the use of LLMs and Generative AI in SE and CPS development activities, for example, the ACM Code of Ethics and Professional Conduct, IEEE Ethically Aligned Design, the EU Ethics Guidelines for Trustworthy AI, the OECD Principles on AI, and Google's Responsible AI principles and operational practices [45].

Various public and private organizations are working actively in this space, and it is important that SE and CPS development community also engages with these organizations, the users, regulators, and society at large to ensure that the use of LLMs and Generative AI in SE and CPS development is aligned with the values, norms, and expectations of all stakeholders. It is also important that Software engineering students get introduced to these developments. A university course that addresses AI ethics should expose students to the work thus far by professional societies (e.g., ACM and IEEE), the private sector (e.g., Google and IBM), and global organizations (e.g., UNESCO.) Complementarily, ethics should be revisited within the software engineering course sequences, with a focus on ethical development, deployment, evolution, and impact on critical concerns such as security, safety, and privacy, and these concerns should be included in practical exercises and course projects [41].

In addition to generative AI's impacts on technology and software engineering practices and education, it also changes the management responsibilities of software engineering leaders. Gartner, for example, predicts that "50% of software engineering leader roles will explicitly require oversight of generative AI projects by 2025" [46]. Proper understanding of the benefits and risks of using generative AI will enable leaders to further increase the value of their teams by investing in training, up-skilling, and hiring individuals with required AI skill sets.

6. APPLICATION: HEALTH AND MEDICINE

Medicine is sometimes described as the most humanistic of the sciences and the most scientific of the humanities. For this reason, and others, the expanded use of AI in healthcare must be carefully considered by physicians, ethicists, and computer scientists. While medical students study cell biology, pharmacology, physiology, genetics, and other "hard sciences," medicine as a discipline is not reducible to a "lab science."

This is reflected in the various forms of the Hippocratic Oath taken by medical students. For example, students in American osteopathic medical schools take an oath based on the following core principles:

1. The body is a unit; a person is a unit of body, mind, and spirit.
2. The body is capable of self-regulation, self-healing, and health maintenance.
3. Structure and function are reciprocally interrelated.
4. Rational treatment is based on an understanding of these principles: body unity, self-regulation, and the interrelationship of structure and function [47].

These meta-principles obviously operate at a higher level than the principles governing simple passive diffusion in cells, i.e., when small molecules pass through the lipid bilayer of a cell membrane. Even the most advanced artificial neural network (pre-)trained using self-supervised learning and semi-supervised learning cannot take a Hippocratic oath and practice the aforementioned principles in a clinical setting.

AI-assisted diagnosis and care could be a powerful tool in the hands of an experienced physician trained to interact with it. The use of AI may improve access to medical information as well as assist in the interpretation of symptoms, tests, and imaging. For example, deep neural networks have been shown to diagnose skin cancer more accurately than a board-certified dermatologist using traditional methods [48]. AI could provide real time analytics and streamline certain tasks, thus reducing physician stress [49]. It has the potential to ease demands on caregivers and thus ease caregiver fatigue. AI can assist with the detection of drug interactions, help with the identification of high-risk patients, and streamline the coding of medical notes [8]. With the rising cost of healthcare, there is of course the desire that AI might lower the cost of care. However, healthcare management must resist pressure to drastically increase demands on staff or substantially decrease caregiver-patient interactions in response.

Healthcare AI faces many limitations. For example, human intelligence, e.g., a physician, can sense fear in a patient and respond with compassion and reassurance. An *artificial* intelligence cannot sense fear in a patient, at least in the same way. Questions arise such as: Might predictive AI be used in the future to "ration" healthcare? Who decides the "cut offs" for expensive medical interventions? How are the ethical biases of the software developers reflected in such systems? Is there sufficient informed consent when AI is involved [50]?

While much AI support for health and medicine relies on predictive rather than generative machine-learning-based models, or on older techniques such as rule-based systems, generative AI is also used [51]. LLMs such as ChatGPT rely significantly on publicly available data on the internet. It is well known that the internet is full of poor medical advice. How would the regurgitation of such information be handled by a novice clinician who has been instructed to trust the AI? As with software engineering and cyber-physical system applications, it would be far preferable to use a specialized and validated training set. Even then, one might prefer the use of a rule-based system, possibly with an LLM front-end to generate a readable explanation.

An experienced physician, especially in a primary care context, may have intimate and unique knowledge of a patient. He or she is therefore able to discern how some side effects from a particular medication or procedure may be intolerable for a patient. Could an AI-guided process offer this personal touch?

AI ethics issues specific to or exacerbated in health and medicine include inadvertent bias (based on race [52], sex, insurance, etc.), justice and fairness, privacy and confidentiality, autonomy (for both patients and medical staff) and informed consent, transparency and explainability, empathy and sympathy, safety and prevention of harm, security, and the need for public trust. Ethical, professional, and legal standards are also recursively entangled, as in the HIPAA standards for patient protection [53]. Thus, it is not surprising to see a call from World Health Organization (WHO) officials for "cautious optimism with safeguards" for AI deployment in the public health sector [55].

An editorial in the medical journal, *The Lancet*, succinctly expresses the need for "stringent regulation" and adequate training for physicians utilizing AI:

> AI could continue to bring benefits if integrated cautiously. It could change practice for the better as an aid—not a replacement—for doctors. But doctors cannot ignore AI. Medical educators must prepare health-care workers for a digitally augmented future. Policy makers must work with technology firms, health experts, and governments to ensure that equity remains a priority. Above all, the medical community must amplify the urgent call for stringent regulation [51].

For a structured literature review of this topic from 2021, see [55]. For a systematic scoping review on AI and health inequities in primary care from 2022, see [56].

The use of Generative AI will affect and likely exacerbate many of these challenges [8, 50], and require additional ethical safeguards, validation, and standards and regulation. Interactive LLMs are likely to be used in two principal modes: and as an advisor, resource, or annotator for medical personnel, or as a first-line medical advisor on the Web. Further, the latter may be on approved sites, through other "health-aware" or designated medical sites, or from ChatGPT or another generic LLM tool.

In the former, problems arise from both the training set and the responses. Incomplete data or bias may result in mistaken diagnosis or inappropriate treatment, compromising protection from harm as well as public trust. Privacy and confidentiality, even for individuals who are not seeking care, may be compromised in responses, or through lack of anonymity or of security for the training dataset. And the problem of hallucination and fabulation is always present, and may call for education and training so staff can recognize a possible problem and seek to verify details of the response.

In the latter use, there are additional risks. Responses will be seen by members of the public without deep medical background, leading to increased risk of harm, from misdiagnosis, improper or ineffective remedies, or omitting treatment altogether, as a result of bias or fabulation, compounded by possible misinformation or terminology overload in the latter two cases. Further, generic tools in particular may not be able to usefully expand or clarify responses.

7. CONCLUSIONS AND FUTURE WORK

From its beginnings, AI has challenged us to understand knowledge, cognition, creativity, and interaction, and to explore related legal and ethical concerns and boundaries. The dawn of the Generative AI era, characterized by the rapid spread of sophisticated systems capable of creating novel content, has further reshaped our understanding of creativity, authenticity, and intelligence. But not surprisingly, it has also called for closer and more robust consideration of legal issues, standards, and ethics.

As predictive and generative AI continue to penetrate every sector of our society, from the arts and journalism to science and education, these ethical considerations have become paramount. Issues of intellectual property, bias including misdirected advice, hallucination and the authenticity of generated content, the potential for misinformation, and the undermining of human labor are at the forefront of this debate. This paper has reviewed how serious issues arise in software engineering and cyber-physical systems as well as in medicine and healthcare.

It is clear that the world recognizes the problem of responsible and ethical AI. This paper is hardly the first to address the problem. Academics and philosophers continue to discuss these issues. Governments, professional organizations, and social action groups are studying the problem too, and proposing and in many cases implementing partial solutions via standards, regulations, laws, and domain-specific process and product safeguards, such as the testing, validation and verification discussed for software engineering. One important safeguard will be to require that appropriate and representative training sets be used whenever possible.

On the other hand, generative AI, combined with predictive models and other AI tools and approaches, has tremendous potential for positive applications. The substantial social and economic benefits are well-understood and likely, as with every other aspect of generative and predictive AI, to increase dramatically, leading to greater efficiency, better service, and higher quality and personalized results. The ethical position, even if it were feasible, thus cannot seek to abandon the use of generative AI and other AI-based approaches. From aiding researchers in generating innovative solutions to fostering unparalleled creativity, this technology is poised to be a catalyst for unprecedented achievement.

Rather, the results of their use must be subject to careful oversight, testing, validation, and verification, as appropriate to the domain of use, in part expressed through professional standards and government recommendations, policies, and regulations. The level of intervention can vary with the domain and problem, with the user, and with the intended use of the results. For example, some casual or creative use by adults for individual or small group consumption may need only the lightest hand. On the other hand, use in online medical consultation or in designing a safety-critical cyber-physical aerospace system may call for intense and repeated review, validation, and assessment.

One possible future development, suggested in Section 4, will be to couple generative AI tools with other tools and analyses, both generic and domain-specific, that can review its products and suggest changes, much in the way that recursive human interaction via clarifying prompts can result in sharper, more extensive, and more accurate responses (compare [42]). This in turn suggests that the study of prompt engineering [57] explores how to better use prompts to mitigate or eliminate some of the ethical risks. Another is to address specific problems, ranging from annoyances to major issues, perhaps by reintegrating something of a rule base, such as rules requiring that citations refer to literature that actually exists. Like all tools of immense power, its ethical use is contingent on humanity's collective wisdom.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support of and fruitful interactions with the Department of Mathematics and Computer Science, the Immaculate Conception School of Theology, and other faculty at Seton Hall University; Dr. Katherine Herbert of Montclair State University; Dr. Susu Nousala and the Collaborative Systemic Research Platform Institute (CSRP); and Dr. Nagib Callaos and the International Institute of Informatics and Systemics (IIIS). We would also like to thank Ms. Carol Marlowe for her indefatigable contributions to developing an extensive resource list, entries from which appear in the references, and Dr. Michael Kelly for his insights on issues in medicine and healthcare. The authors are grateful for the assistance of Mr. Paul Laracy with proofreading this manuscript.

REFERENCES

[1] "ELIZA," **Wikipedia** (Accessed 11/02/2023): https://en.wikipedia.org/wiki/ELIZA.
[2] M. White, "A Brief History of Generative AI," **Medium** (2023): https://matthewdwhite.medium.com/a-brief-history-of-generative-ai-cb1837e67106.
[3] D. Baidoo-Anu and L. Owusu Ansah, "Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning," **Journal of AI**, Vol. 7, No. 1 (2023), pp. 52-62: https://dergipark.org.tr/en/pub/jai/issue/77844/1337500.
[4] T. Wang, "Navigating Generative AI (ChatGPT) in Higher Education: Opportunities and Challenges," in C. Anutariya, D. Liu, Kinshuk, A. Tlili, J. Yang, M. Chang (eds.), **Smart Learning for A Sustainable Society**, ICSLE 2023, Lecture Notes in Educational Technology, Singapore: Springer: https://doi.org/10.1007/978-981-99-5961-7_28.
[5] Y.K. Dwivedi et al., "'So what if ChatGPT wrote it?' Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy," **International Journal of Information Management**, Vol. 71 (2023): https://doi.org/10.1016/j.ijinfomgt.2023.102642.
[6] J. Hutson and M. Harper-Nichols, "Generative AI and Algorithmic Art: Disrupting the Framing of Meaning and Rethinking the Subject- Object Dilemma," **Global Journal of Computer Science and Technology** (2023): https://digitalcommons.lindenwood.edu/faculty-research-papers/461.

[7] C. Ebert and P. Louridas, “Generative AI for Software Practitioners,” **IEEE Software**, Vol. 40, No. 4, pp. 30-38 (July/August 2023), doi: 10.1109/MS.2023.3265877.
[8] [6-4] P. Lee et al., “Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine,” **New England Journal of Medicine** (March 23, 2023), 388:1233-1239: https://www.nejm.org/doi/full/10.1056/NEJMsr2214184.
[9] The White House, “President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence,” (October 30, 2023): https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/.
[10] Federal Trade Commission, “Artificial Intelligence,” https://www.ftc.gov/industry/technology/artificial-intelligence.
[11] P.A. Gagniuc, **Markov Chains: From Theory to Implementation and Experimentation**, Hoboken, NJ: Wiley, 2017.
[12] C.A. Petri, **Kommunikation mit Automaten** (Ph.D. Thesis, University of Bonn, 1962).
[13] G. Petrişor, “Neural Networks and Petri Nets,” **Bulletin Mathématique de la Société des Sciences Mathématiques de Roumanie**, Nouvelle Série, Vol. 35 (83), No. 3/4 (1991), pp. 245-247.
[14] “History of Cybernetics – Chapter 2: The Coalescence of Cybernetics,” **American Society for Cybernetics**: https://asc-cybernetics.org/foundations/history2.htm.
[15] “The Macy Conferences,” **American Society for Cybernetics**: https://asc-cybernetics.org/foundations/history/MacySummary.htm.
[16] P. J. Denning et al., “Computing as a Discipline,” **Computer**, Vol. 22, No. 2, 1989, pp. 63–70, doi: 10.1109/2.19833.
[17] Purdue University, “Computer Science Pioneer Samuel D. Conte Dies at 85,” (July 1, 2002): https://www.cs.purdue.edu/about/conte.html.
[18] Case Western Reserve University, “Case School of Engineering History”: https://engineering.case.edu/about/historye.edu/about/history.
[19] A. Mueller, “A Brief History of the BCL,” **Austrian Journal of History**, Vol. 11, No. 1, 2000, pp. 9–30.
[20] T. Aquinas, **Commentary on Aristotle’s Nicomachean Ethics**, trans. C.I. Litzinger, Dumb Ox Publications, 1993.
[21] C.E. Harris et al., **Engineering Ethics: Concepts and Cases**, 6th ed., Cengage Learning: Boston, MA, 2018.
[22] K.D. O’Rourke, **Medical Ethics: Sources of Catholic Teachings**, 4th ed., Washington, DC: Georgetown University Press, 2011.
[23] P.C. Herbert and W. Rosen, **Doing Right: A Practical Guide to Ethics for Medical Trainees and Physicians**, 4th ed., Oxford, UK: Oxford University Press, 2019.
[24] **Catechism of the Catholic Church**, Part III, Chapter 1, Article 4.: https://www.vatican.va/archive/ENG0015/_INDEX.HTM.
[25] J. Anderson et al., “A Multi-Disciplinary Analysis of Catholic Social Teaching with Implications for Engineering and Technology, **Journal of Systemics, Cybernetics, and Informatics**, Vol. 18, No. 6, 2020, pp. 41–49: https://www.iiisci.org/journal/pdv/sci/pdfs/SA244NH20.pdf
[26] C.L. Bockting et al., “Living guidelines for generative AI—Why scientists must oversee its use,” **Nature**, October 19, 2023: https://www.nature.com/articles/d41586-023-03266-1.
[27] M. Wooldridge, **A Brief History of Artificial Intelligence: What It Is, Where We Are, and Where We Are Going**, New York: Macmillan Publishers, 2021.
[28] S. Lucci, S. M. Musa, and D. Kopec, **Artificial Intelligence in the 21st Century,** 3rd ed., Herndon, VA: Mercury Learning and Information, 2022.
[29] D.L. Poole and A.K. Mackworth, **Artificial Intelligence: Foundations of Computational Agents**, 3rd ed., Cambridge, UK: Cambridge University Press, 2023.
[30] J. D. Kelleher and B. Tierney, **Data Science**, Cambridge, MA: MIT Press, 2018.
[31] E. Alpaydin, **Machine Learning: The New AI**, Cambridge, MA: MIT Press, 2016.
[32] J. D. Kelleher, **Deep Learning**, Cambridge, MA: MIT Press, 2019.
[33] S. M. Liao (ed.), **Ethics of Artificial Intelligence,** Oxford Academic, 2020.
[34] C. Huang, Z. Zhang, B. Mao, and X. Yao, “An Overview of Artificial Intelligence Ethics,” **IEEE Transactions on Artificial Intelligence,** Vol. 4, Issue 4, pp. 799-819, 2022.
[35] J.A. Siqueira de Cerqueira, H.A. Tives, and E.D. Canedo, “Ethical Guidelines and Principles in the Context of Artificial Intelligence,” **Proceedings of the XVII Brazilian Symposium on Information Systems,** Article No. 36, 2021, pp. 1-8.
[36] E. Mosqueira-Rey, E. Hernandez-Pereira, D. Alonso-Rios, J. Bobes-Bascaran, and A. Fernandez-Leal, “Human-in-the-loop Machine Learning: A State of the Art,” **Artificial Intelligence Review**, Springer, Vol. 56, 2023, pp. 3005-3054: https://doi.org/10.1007/s10462-022-10246-w.
[37] S. Harris and A. Moshtaghian, “High schooler calls for AI regulations after manipulated pornographic images of her and others shared online,” **CNN** (November 4, 2023): https://www.cnn.com/2023/11/04/us/new-jersey-high-school-deepfake-porn/index.html.
[38] UNESCO, “Ethics of Artificial Intelligence”: https://www.unesco.org/en/artificial-intelligence/recommendation-ethics.
[39] IEEE, “The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems”: https://standards.ieee.org/industry-connections/ec/autonomous-systems/.
[40] IBM, “AI Ethics”: https://www.ibm.com/impact/ai-ethics.
[41] V.D. Kirova, C.S. Ku, J.R. Laracy, T.J. Marlowe, “Software Engineering Education Must Adapt and Evolve for an LLM (Large Language Model) Environment,” **ACM SIGCSE 2024 Technical Symposium on Computer Science Education**, Portland, OR, USA, March 2024 (to appear).
[42] C. Ebert and P. Lourida, “Generative AI for Software Practitioners,” **IEEE Software**, Vol. 40, No. 4, 2023, pp. 30-38.
[43] M.L. Siddiq and J.C.S. Santos, “Generate and Pray: Using SALLMS to Evaluate the Security of LLM Generated Code,” arXiv:2311.00889 [cs.SE], November 1, 2023.
[44] A. Spirling, “Why open-source generative AI models are an ethical way forward for science,” **Nature**, April 2023.
[45] Google, “Responsible AI practices”: https://ai.google/responsibility/responsible-ai-practices/, accessed November 5, 2023.
[46] “Gartner Says More Than 50% of Software Engineering Leader Roles Will Explicitly Require Oversight of Generative AI by 2025,” Gartner (August 30, 2023): https://www.gartner.com/en/newsroom/press-releases/2023-09-30-gartner-says-more-than-50-percent-of-software-

engineering-leader-roles-will-explicitly-require-oversight-of-generative-ai-by-2025.
[47] American Association of College of Osteopathic Medicine, “About Osteopathic Medicine”: https://www.aacom.org/become-a-doctor/about-osteopathic-medicine.
[48] M.J. Rigby, “Ethical Dimensions of Using Artificial Intelligence in Health Care,” **AMA Journal of Ethics**, Vol. 21, No. 2, 2019, E121-124: https://journalofethics.ama-assn.org/article/ethical-dimensions-using-artificial-intelligence-health-care/2019-02.
[49] Drexel University College of Computing and Informatics, “Pros & Cons of Artificial Intelligence in Medicine,” July 21, 2021: https://drexel.edu/cci/stories/artificial-intelligence-in-medicine-pros-and-cons/.
[50] S. Gerke et al., “Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare,” **Artificial Intelligence in Healthcare**, Cambridge, MA: Academic Press, 2020: 295–336.
[51] Editorial Board, “AI in Medicine: Creating a Safe and Equitable Future,” **The Lancet**, Vol. 402, No. 10401, August 12, 2023, p. 503: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(23)01668-9.
[52] J. Lagasse, “Healthcare chatbots may promote racist misinformation,” **Healthcare Finance**, October 2023: https://www.healthcarefinancenews.com/news/healthcare-chatbots-may-promote-racist-misinformation.
[53] J. Byron, “AI and HIPAA Privacy Concerns,” **American Institute of Healthcare Compliance**, August 3, 2023: https://aihc-assn.org/ai-and-hipaa-privacy-concerns/.
[54] S. Pujari et al., “Artificial intelligence for global health: cautious optimism with safeguards,” **Bulletin of the World Health Organization**, Vol. 101, No. 6, June 1, 2023, p. 364: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10225938/pdf/BLT.23.290215.pdf.
[55] S. Secinaro et al., “The role of artificial intelligence in healthcare: a structured literature review,” **BMC Medical Informatics and Decision Making**, 2021, Vol. 21, 125: https://doi.org/10.1186/s12911-021-01488-9.
[56] A. d’Elia et al., “Artificial intelligence and health inequities in primary care: a systematic scoping review and framework,” **Family Medicine and Community Health**, November 2022, Vol. 10, e001670: https://pubmed.ncbi.nlm.nih.gov/36450391/.
[57] The Prompt Artisan, “Lesson 9: Ethics and Responsible AI in Prompt Engineering,” March 23, 2023: https://thepromptartisan.com/lesson-9-ethics-and-responsible-ai-in-prompt-engineering/.