

**THE ETHICAL DILEMMA OF ARTIFICIAL INTELLIGENCE
IN MEDICINE**

A Thesis
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
MASTER OF ARTS

by
Joseph Salvatore Capalbo
August 2024

Thesis Approval:

Nora Jones, Thesis Advisor, Urban Bioethics

ABSTRACT

Artificial Intelligence (AI) has the capability to revolutionize modern life. From humble beginnings of simple machines to current day programs capable of winning “Jeopardy!” and passing medical board exams, the applications of this maturing technology are incredibly diverse. Healthcare in particular contains many inefficiencies and opportunities for improvement for which AI programs have shown encouraging results. However, the ramifications of extensive implementation are unclear. In order to cultivate innovative technology safely, the core ethical principles of beneficence, non-maleficence, autonomy and justice must be prioritized.

TABLE OF CONTENTS

	Page
ABSTRACT.....	ii
LIST OF TABLES.....	v
LIST OF FIGURES	vi
CHAPTER	
1. INTRODUCTION	1
2. HISTORY OF AI.....	2
Early Years.....	3
AI Winters.....	5
Modern.....	5
3. CURRENT APPLICATIONS OF AI.....	7
Surveillance.....	7
Intervention.....	8
Access/Efficiency/Education	10
4. CHALLENGES OF AI.....	12
Lack of Evidence	12
Inadequate Clinical Reasoning	13
Apprehension of Change.....	14
5. ETHICAL IMPLICATIONS	16
Beneficence and Non-Maleficence	17
Autonomy	19

Justice.....	20
6. CONCLUSION.....	21
BIBLIOGRAPHY.....	22
APPENDIX	
A. PERMISSION LETTER, INCLUSION OF COPYRIGHTED MATERIAL	25

LIST OF TABLES

Table	Page
1. A systematic approach to clinical ethics.....	17

LIST OF FIGURES

Figure	Page
1. A timeline of the significant developments regarding AI in medicine.....	3

CHAPTER 1

INTRODUCTION

Modern society is on the precipice of a technological evolution with the maturation of Artificial Intelligence (AI). Organizations and entire countries are vying for supremacy. Conferences, meetings, and summits from the World Intelligence Expo in China (Jiang, 2024), to Apple Inc.'s Worldwide Developers Conference, to the keynote address of annual Internal Medicine Meeting of the American College of Physicians focus on AI's capability and its potential future.

Regarding healthcare, AI presents an opportunity to increase provider efficiency, assist with diagnostic evaluations, develop therapeutic strategies, expand access to care, and improve provider-patient relationships (Topol, 2023). However, there is limited data thus far to corroborate these expectations and the ramifications of allowing AI access to medical data is unclear. In what follows, I describe the history of AI, identify examples of its current uses and challenges, and address the ethical considerations of its implementation.

CHAPTER 2

HISTORY OF AI

While many attribute the origins of AI with the creation of robots, the modern concept can be traced to the English mathematician and computer scientist Alan Turing's 1950 publication "Computing Machinery and Intelligence" in which he described a test, later to be known as the "Turing test," to determine whether computers were capable of human intelligence (Keskinbora 2019, Kaul V 2020). A few years later in 1955, John McCarthy, an American computer scientist, was the first to document the term "artificial intelligence," defining it as "the science and engineering of making intelligent machines" (Hamet P., 2017). AI has evolved in many industries since then, but it did not enter the general public's vernacular until OpenAI released ChatGPT to the public in November 2022. To discuss the future of AI, one must appreciate its evolution (see Figure 1).

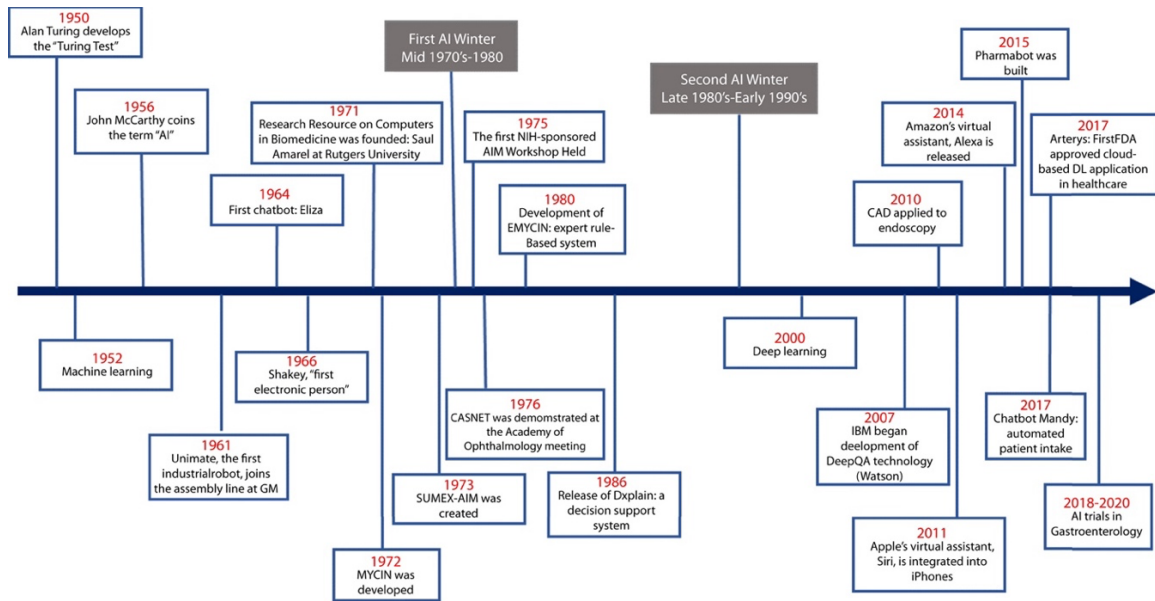


Figure 1. A timeline of the significant developments regarding AI in medicine. Reproduced from “History of artificial intelligence in medicine” by V.E.S. Kaul (2020). *Gastrointestinal Endoscopy* 92(4): 807-812. Copyright 2020 by Elsevier. Reproduced with permission from Elsevier.

Early Years

The first documentation of a “humanoid automaton” was traced back to 3rd century China when an engineer presented a human-shaped device constructed of leather and wood to the Emperor Mu of Zhou. The next significant advancement was not until the 12th century when a humanoid robot capable of striking cymbals was created by Al-Jazari, a Muslim scholar and engineer. Then were the anatomic sketches of Leonardo da Vinci, Italian Renaissance artist and engineer, in the 1490s that served as inspiration for 20th century researchers. These notes corresponded with a knight robot which would move and articulate its appendages via pulleys and cables. In 18th century, the French engineer Jacques de Vaucanson constructed the famous “Flute Player” which could performing multiple songs on a pipe. The first electronic autonomous machine was constructed by William Gray Walter, American-British neurophysiologist, in 1948. This

lead up to Alan Turing's publication and John McCarthy who would establish the field of AI in 1956 at Dartmouth College with his colleagues (Hamet P., 2017).

Thereafter, AI was focused on machines' ability to make decisions and perform tasks that previously only a human could make. General Motors was an early adopter when it employed the first industrial robot arm, Unimate, in 1961 that would follow commands and perform automated die casting. In 1964, a program called Eliza was developed by Joseph Weizenbaum to mimic human conversation using pattern matching and substitution methodology, essentially an early chatbot (Kaul V, 2020).

Although medicine did not directly utilize AI in these early years, digital data and resources were developed such as the Medical Literature Analysis and Retrieval System and the web-based search engine PubMed by the National Library of Medicine which would serve as vital building blocks for future clinical databases and medical record systems. It was not until 1976 that one of the first prototypes to demonstrate AI's use in medicine was developed by Rutgers University and presented at the Academy of Ophthalmology meeting in Las Vegas, Nevada. Their program utilized the CASNET model, a causal-associational network consisting of model-building, consultation, and a database to provide information to physicians regarding management of glaucoma based upon individual patient characteristics (Kaul V, 2020). However, this early excitement in the field of AI was followed by two "winters" characterized by reduced funding and interest.

AI Winters

The first winter in the mid to late 1970s was attributed to perceived limitations of the technology of the time. Pioneers in the field continued their pursuit. In the 1970s, the Research Resource on Computers in Biomedicine was developed at Rutgers University and Stanford University created its Medical Experimental-Artificial Intelligence in Medicine computer system which enhanced networking capabilities among clinical and biomedical researchers from several institutions. This increased collaboration led to the first National Institutes of Health-sponsored AIM workshop at Rutgers University in 1975 (Kaul V, 2020).

The second winter occurred in the late 1980s to the early 1990s due to excessive cost in development and maintenance of digital databases. However innovation continued, and the University of Massachusetts developed DXplain in 1986 as a decision support system which served as an electronic medical textbook as well as capable of generating a differential diagnosis based on symptoms (Kaul V, 2020). Other programs were developed with increasing computational powers which culminated with the program Big Blue demonstrating its advancement by defeating the world chess champion, Gary Kasparov, in May of 1997 (Hamet P., 2017).

Modern

As computer hardware and software programming improved, AI grew rapidly. In the early 2000s came the advent of “Deep Learning,” a subfield of AI in which machines can classify data, learn, and make decisions on their own based on algorithms which create an artificial neural network. It developed from “Machine Learning,” wherein programs “learn” from analyzed patterns to be applied to future scenarios but ultimately

require human input. In 2007, IBM developed Watson, an open-domain question-answering system. Its technology, called DeepQA, was able to analyze data and generate answers in a way that was easier to maintain and more cost effective as opposed to prior machine learning systems which used forward reasoning (following rules from data to conclusions), backward reasoning (following rules from conclusions to data), or hand-crafted “if-x-then-y” rules. It gained much attention when it won first place on the televised game show “Jeopardy!” in 2011 (Kaul V, 2020).

Similarly, superficial communication chatbots, such as aforementioned Eliza, evolved into meaningful conversation-based interfaces via natural language processing. This advancement led to Apple’s virtual assistant Siri in 2011 and Amazon’s Alexa in 2014. Then in 2015, Pharmabot was created to assist in medication education and Mandy, an automated patient intake process for primary care offices, was developed in 2017 (Kaul V, 2020).

With the advent of deep learning, AI systems became capable of analyzing complex algorithms, self-learning, refining risk assessment models, improving diagnostic accuracy, and modernizing workflow efficiency. This has led to a flourishing of new program development (Kaul V, 2020).

CHAPTER 3

CURRENT APPLICATIONS OF AI

AI has appropriately garnered enthusiasm from the general public for the myriads of potential uses capable of revolutionizing modern life. Healthcare in particular struggles with inefficiencies while accruing an incredible volume of data ripe for analysis. The common goal of applications in medicine is to increase the quality of care for patients through improved diagnostic accuracy, therapeutic response prediction, and anticipatory preventive medicine while decreasing medical errors and physician inefficiency (Kaul V 2020, Melillo 2024, Topol 2023). By optimizing AI to perform administrative and analytical tasks, physicians will be encouraged to spend more time with their patients, reflect more critically on the care they provide, and reduce burnout through what Dr. Topol refers to as “keyboard liberation” (DuBosar 2024, Topol 2023, Insel 2019, Melillo 2024). Additionally, through streamlined documentation, improved diagnostic yields, and refined logistics, administrative costs will reduce and health care will be more cost effective. Below are examples of current applications according to their function. This list is not intended to be exhaustive as new programs are constantly being developed, but they provide insight into the versatility and possibility of these technologies.

Surveillance

AI has the power to make necessary health maintenance screening and condition monitoring far more accessible. One of the first mobile applications of AI in medicine was “Kardia,” a device designed by AliveCor which received FDA approval in 2014 for the detection of atrial fibrillation. In years since, Apple has famously received FDA approval for their single lead ECG technology in the Apple Watch which is able to detect

atrial fibrillation and share that information with a practitioner of choice (Briganti G 2020, Kaul V 2020). A similar wearable device named Embrace received FDA approval in 2018 to detect generalized seizures with the ability to report the event to relatives and/or a health care provider. Additional wearable devices can assess gait, posture, and tremor in patients with multiple sclerosis, Parkinson disease, and Huntington disease (Briganti G, 2020).

Another area of research is attempting to identify patient characteristics through simple, inexpensive diagnostic testing to expand screening capabilities. Examples include programs that utilize a 12-lead ECG to infer sex, age (Attia Z. I., 2019), anemia (Kwon, 2020), asymptomatic left ventricular dysfunction (Attia Z. K.-J., 2019), silent arrhythmia not present at the time of recording, valvular heart disease, channelopathies, hypertrophic cardiomyopathy, and even distant conditions such as liver cirrhosis (Attia Z. I., 2021), diabetes mellitus (Kulkarni AR, 2023), and chronic kidney disease (Holmstrom L., 2023). Others assess chest radiographs to determine cardiac function, valvular disease (Ueda D., 2023), pulmonary hypertension (Liu CM, 2022), and even diabetes mellitus (Pyrros, 2023). Some have been shown to predict a patient's risk of cardiovascular disease through analysis of their electronic patient record alone (Briganti G 2020, Kaul V 2020). RETFound is an example of a program capable of reviewing retinal images to detect ocular diseases such as diabetic retinopathy and glaucoma in addition to infer the presence of heart failure, myocardial infarction, Parkinson's disease, ischemic stroke (Zhou, 2023), chronic kidney disease (Sabanayagam C, 2020), liver cancer, chronic viral hepatitis, non-alcoholic fatty liver disease, cholelithiasis, hepatic cysts (Xiao W, 2021), and estimate coronary artery calcium as well as a CT scan (Rim TH, 2021).

Other programs were created to empower patients with the tools to manage their own care. One such is a Selfie Fundus Imaging (SFI) program which was developed to enhance retinopathy screening through teaching patients to capture their own images via smartphones in lieu of traditional screening which requires trained technicians (Kumari, 2022). Wearable continuous glucose monitors allow patients with diabetes to optimize their blood sugar management through real time measurements and trends. Patients can anticipate their insulin requirement by determining whether their blood sugar level is rising or falling. The company Medtronic received FDA approval for their Guardian system which pairs a device with smartphones and have since partnered with Watson by IBM for their improved Sugar.IQ system (Briganti G, 2020).

Intervention

AI programs have also been developed to augment medical providers abilities to perform their roles. Examples include the Chinese “ENDOANGEL” used to assess bowel preparation during the withdrawal phase of a colonoscopy and the American “GI Genius” which identifies colorectal polyps via marking a live video feed during endoscopic evaluation (Kaul V, 2020). Additional projects are being developed to diagnose gastroesophageal reflux disease, atrophic gastritis, and differentiate chronic pancreatitis from pancreatic cancer while others predict outcomes of patients with gastrointestinal bleeding, esophageal cancer, inflammatory bowel disease, metastatic colorectal cancer, and esophageal squamous cell carcinoma (Briganti G, 2020). Another is the Da Vinci surgical system by Intuitive Surgical which was approved by FDA in 2000 for the use in complex, minimally invasive surgeries (Hamet P., 2017).

Aside from procedural, programs have been shown to be more accurate at interpreting pulmonary function testing and can assist in predicting the decline of glomerular filtration rate in patients with polycystic kidney disease as well as IgA nephropathy (Briganti G, 2020). Paige.ai has received FDA approval for diagnosing cancer from histopathology slides. Additional applications can quickly identify rare genetic abnormalities through rapid whole genome sequencing, discover new therapeutic targets through protein analyses, and localize DNA variants as predictors of disease or traits (Joshi 2019, Hamet P. 2017, Kaul V 2020). Avatars or psychotherapeutic programs have been developed to provide accessible kindness and reassurance for several uses such as assisting with pain management, detection of emotional disturbances, control of paranoid hallucinations, and assistance with elderly care (Hamet P., 2017).

Access/Efficiency/Education

Millions of people, both nationally and internationally, suffer from insufficient access to medical care and information. Perhaps the broadest use case for AI is to increase access for patients through more advanced resources, telehealth, and smarter search engines. Dr. Topol often refers to a story about a 6-year-old boy suffering from chronic pain and neurological symptoms. Despite evaluations from 17 physicians, his diagnosis was a mystery until his mother entered his symptoms into ChatGPT. The resulting diagnosis from that inquiry proved to be correct and the boy's condition greatly improved after the necessary surgery (Topol 2023, Melillo 2024). This is but one example of patients and families taking ownership of their care with the assistance of additional information.

Another hope for AI is to improve physician efficiency to limit the time required to perform administrative duties, documentation, and correspondingly improve care. Certain programs are already in use to perform tasks such as coding, charting, and speak-to-text to complete notation. Newer systems are even able to listen to patient encounters in the background and filter an entire conversation to identify the clinically relevant information to write an encounter note. This would greatly limit the time required to complete documentation thus allowing the physician to focus on the patient (Briganti G, 2020). As data is input electronically, ideally more simply and organized, clinical decision-making recommendations could then be provided through chart analysis thus increasing recommended screenings and optimizing workups to limit extraneous tests (Kaul V, 2020). It is theorized that this increased efficiency will prove to be cost effective as well. Health care in the U.S. is a multi-trillion-dollar industry accounting for over 15% of the country's GDP (Sun L, 2023). As the cost of care rises, the business becomes less sustainable, expanding health disparities (Insel, 2019).

Whereas AI in practice augments the physicians' tools and skills, there is opportunity to transform medical education. Such programs include the Virtual Patient Learning System (VPLS) which simulates patient encounters to develop clinical judgment and Hanover which trains students to identify the most targeted treatment options in a particular scenario. AI programs in China are even capable of passing their National Medical licensing exam (Sun L, 2023). Some universities have begun implementing curricula focused on providing digital expertise to address modern health problems and to strategize AI implementation in their future institutions (Briganti G, 2020).

CHAPTER 4

CHALLENGES OF AI

Despite the excitement surrounding AI's capabilities, many decry it as a threat to society (Insel, 2019). Much concern revolves around misunderstanding how it works, why it is being developed, and its depiction in popular culture. In 2023, the American Medical Association (AMA) announced its "principles" on AI and the White House issued an executive order on the "safe, secure, and trustworthy" development and use of AI. Both stress the importance of oversight and transparency, though fall short of announcing explicit regulation (DuBosar 2024, Executive Order 14110 2023, AMA 2023). Most concerns fall within several general insecurities: lack of evidence of benefit, inadequate clinical reasoning, and apprehension regarding implementation.

Lack of Evidence

Medicinal practice guidelines are built upon evidence. As such, a main challenge in the years to come for the broader implementation of AI is to prove clinical utility. There have been many demonstrations of AI's abilities, as discussed above, however publications documenting statistically significant improvements in clinical outcomes above the current systems are lacking. Many studies were classified into categories 4 and 5 by review boards indicating poor quality due to unreliable design or inability to be replicated. Studies suffer from such defects as selection bias and overfitting demonstrating flaws in their methodologies. Additionally, few focus on comparing the outcomes of AI, clinicians, and clinician-AI teams to demonstrate whether the technology is superior to the system it was designed to replace (Briganti G 2020, Sun L 2023). As such, one is unable to extrapolate benefit to more diverse populations. If one cannot trust

the data produced from a program, then it is worthless. More data with improved transparency into the machine learning algorithms themselves are needed to demonstrate AI's safety, efficacy, and value prior to incorporation into formal guidelines or policy (Hamet P. 2017, Matuchansky 2019, Sun L 2023).

Inadequate Clinical Reasoning

There is a consensus in the field that a medical provider, preferably a physician, must remain at the helm of developing AI programs to prevent mistakes in documentation or clinical management (DuBosar, 2024). An example from Dr. Topol is of a particular AI program diagnosing an ectopic pregnancy based upon symptoms of a 77-year old woman (Melillo, 2024). This highlights the limitation of programs to the aptitude of their designers and the data from which they are created. Programs may be subject to errors based upon race, socioeconomic class, or any number of factors unintended by the architect because their source data is inadequately diverse. Thus, for AI to take the next step to optimize its benefit, it requires more data. However, the current distribution of medical records is separated in incommunicable silos separated between each health system. The public would need to trust these organizations to share their information safely rather than the current trend of large health organizations merging with smaller ones to expand their reach (Hamet P. 2017, Attia Z. I. 2021). This again would require policy or guideline recommendations to entice health systems big and small to work together towards theoretical improvement of patient care. To ensure AI can be trusted, engineers must ensure “transparency (operations visible to user), credibility (outcomes are acceptable), auditability (efficiency can be easily measured),

reliability (AI systems perform as intended), and recoverability (manual control can be assumed if required)” (Keskinbora, 2019).

Apprehension of Change

Perhaps the most critical task will be convincing the medical community that the investment of more administrative burden, and therefore decreased efficiency, in the short term will be worth the long-term benefit. The implementation of electronic medical records is a recent precedent of the struggle of implementing new technology (Joshi, 2019). Many medical professionals remember the challenges of learning new tools and transitioning their practice as exacerbating burnout. This paired with an increased reliance on machines contributes to a concern of the “dehumanization” of medical practice (Briganti G, 2020). It is unlikely that machines will be able to provide an empathetic and compassionate environment necessary for healing (Farhud DD, 2021). This echoes the importance of focusing innovation to augment medical providers’ abilities to solve specific challenges rather than advancing technology unrestrained beyond cost-benefit limits or to create a theoretical “super intelligence” capable of surpassing human abilities. Although there is limited concern that machines will replace physicians, technology may advance to the point that it is not cost effective to employ it (Joshi 2019, Sun L 2023, Kaul V 2020, Hamet P. 2017). As pointed out in the landmark book “Crossing the Quality Chasm: A New Health System for the 21st Century,” technological advancement does not inherently equate to improvement (America, 2001). Medical professionals established in their careers may be unwilling to change their practice of reputable systems to embrace unproven products. It must also be stressed that any theoretical time accrued through increased efficiency be re-invested into the care of

those patients rather than re-allocated to an alternative task for monetary benefit (Joshi 2019, Melillo 2024).

A related concern is the lack of qualified trainers of newer technology. As many leaders in various disciplines established their practice on traditional models of medicine, highly trained teachers would be required. Would this be similar to medical equipment representatives training surgeons practice by practice or rather large session sermons given at national conferences? The former does not seem practical as most of this technology requires large data sets yet the later would require standardized guidelines endorsing particular systems from those national organizations. If there is policy support and guidelines are established, it would seem the natural place to begin introduction to new technology would be in medical education to students, residents, and younger professionals who are established in teaching settings and are more impressionable than their more senior colleagues (Sun L, 2023). There is a growing sentiment that if practitioners delay learning how to use these newer systems and technologies, then they may be at a disadvantage in years to come (DuBosar, 2024).

CHAPTER 5

ETHICAL IMPLICATIONS

In addition to the challenges listed above, one must consider the ethical implications of broad AI implementation. Ethics is an inseparable component of clinical medicine, but current moral systems are based upon interactions with other humans. How does one apply ethical concepts to non-organic intelligence? Society lacks policies and regulation for how to deal with such issues. While norms may vary by region, nation, and culture, modern bioethicists agree on four classic principles of beneficence, non-maleficence, autonomy, and justice to guide clinical decision making (Keskinbora 2019, Varkey 2021).

Jonsen et al offered a systematic approach to clinical ethics in their text *Clinical Ethics: A Practical Approach to Ethical Decisions in Clinical Medicine* as depicted in Table 1. It portrays a similar organizational formula to what a physician is classically taught when assessing a new patient, that of inquiring about chief complaints, history of presenting illness, past medical history, etc. Through this methodology, Jonsen asserts that a health care provider should begin their assessment clinically by assessing the patient's medical problems and treatment options according to the values of beneficence (act in the best interest of the patient) and non-maleficence (do no harm). Once the options and potential adverse events are determined, the patient has a right to that information to determine the correct course of action for them with respect to their autonomy. Throughout this process, one must consider the context and potential external forces at play to ensure that the outcome is fair and just. How does this formula adapt from interactions with humans to include AI (Jonsen AR 2015, Varkey 2021)?

Table 1. A systematic approach to clinical ethics	
Principles of ethics	Application of principles of ethics in patient care
Beneficence, nonmaleficence	<p><i>Clinical assessment</i></p> <p>Nature of illness (acute, chronic, reversible, terminal)? Goals of treatment?</p> <p>Treatment options and probability of success for each option?</p> <p>Adverse effects of treatment and does benefit outweigh harm?</p> <p>Effects of no medical/surgical treatment?</p> <p>If treated, plans for limiting treatment? Stopping treatment?</p>
Respect for autonomy	<p><i>Patient rights and preferences</i></p> <p>Information given to patient on benefits and risks of treatment?</p> <p>Patient understood the information and gave consent?</p> <p>Patient mentally competent? If competent, what are his/her preferences?</p> <p>If patient mentally incompetent, are patient's prior preferences known? If preferences unknown, who is the appropriate surrogate?</p>
Beneficence, nonmaleficence, respect for autonomy	<p><i>Quality of Life (QOL)</i></p> <p>Expected QOL with and without treatment?</p> <p>Deficits – physical, mental, social – may have after treatment?</p> <p>Judging QOL of patient who cannot express himself/herself?</p> <p>Who is the judge?</p> <p>Recognition of possible physician bias in judging QOL?</p> <p>Rationale to forego life-sustaining treatment(s)?</p>
Distributive justice	<p><i>External forces and context</i></p> <p>Conflicts of interest – does physician benefit financially, professionally by ordering tests, prescribing medications, seeking consultations?</p> <p>Research or educational considerations that affect clinical decisions, physician orders?</p> <p>Conflicts of interest based on religious beliefs? Legal issues?</p> <p>Conflicts of interests between organizations (clinics, hospitals), 3rd party payers?</p> <p>Public health and safety issues?</p> <p>Problems in allocation of scarce resources?</p>
Original table can be found in (Varkey, 2021), adopted from (Jonsen AR, 2015)	

Beneficence and Non-Maleficence

Beneficence and non-maleficence are often considered in tandem. Beneficence is the responsibility for health care providers to act in the best interest of their patient while non-maleficence emphasizes avoiding harm. These values represent the responsibility of

providers to weigh the benefits against burdens of interventions and to advocate for their patients' well-being while respecting their values (Keskinbora 2019, Varkey 2021). As such, they should be considered chiefly when contemplating the incorporation of AI.

As described above, AI has incredible potential to improve the medical care provided to patients. It may also greatly expand access to care, a significant concern to people of this nation and worldwide who live in underserved communities. There have been efforts in recent years to incentive physicians to practice in these settings as well to expand the field of advanced practice providers (nurse practitioners and physician assistants) who are able to practice under physician supervision with insufficient improvement (Keskinbora, 2019). To stifle AI would limit medical providers' ability to use newer tools and information to care for underserved populations, effectively limiting access to medical care and expanding social classes.

However, at what point does AI cause harm? Prioritizing AI development over privacy can lead to exploitation. Unregulated databases may introduce biases via skewed source data. Implementation of new, proprietary tools prior to traditional scrutiny and proof of benefit may have unintended consequences. "Keyboard liberation" may extrapolate generative errors or "copy-forward" inaccuracies which lead to unnecessary or even dangerous interventions. Diagnostic tools limit physicians need to perform critical analysis, essentially acting as a crutch limiting their capabilities of practicing independently. Increased productivity could convince leadership that particular positions are unnecessary, replacing jobs with programs. All these errors are devastating when caused by a health care provider but who is responsible when an AI is the source of the medical error?

The medical community and general society must develop a mechanism to debate and decide what is to be done with rapidly developing tools and information. It must be ensured that as AI advances, a “common good principle” is followed such that future programs be developed appropriately and for the benefit of humanity within ethical ideals (Sun L 2023, Joshi 2019).

Autonomy

Autonomy refers to one’s intrinsic value and power to self-determination. It has different implications in various fields, but in medicine it is referring to a patients’ ability to make choices not only with regards to their care but also to the use of their medical data. For one, the increased access to medical information and providers affords patients the means to take ownership of their care and make more informed decisions. However, how is one to respect one’s personal medical data with AI programs that rely upon data to process, learn, and improve (Keskinbora 2019, Varkey 2021, Briganti G 2020)?

Regulation regarding data protection is a recognized necessity. The Health Insurance Portability and Accountability Act (HIPAA) was passed in 1996, asserting that “a covered entity may not use or disclose protected health information, except either: (1) as the Privacy Rule permits or requires; or (2) as the individual who is the subject of the information (or the individual's personal representative) authorizes in writing” (Rezaeikhonakdar, 2023, HIPAA). This law guides medical professionals’ obligation to informed consent, truth-telling, and confidentiality. However as mentioned above, there is a dearth of regulation focused on emerging AI, raising concerns that the current laws are insufficient to protect patients’ data. For example, social networks gather large amounts of data for marketing, genetic laboratories sell customer data to pharmaceutical

companies, and malware is more commonly attacking health care agencies directly (Farhud DD, 2021). This sharing of information with third party agencies without the consent of their users has led to several lawsuits from the Federal Trade Commission (FTC) (Rezaeikhonakdar, 2023) as violations of HIPPA, yet it continues. Sharing data is a necessity for AI to progress, however, it must be performed carefully with respect to one's autonomy, even if that entails a drag on AI's innovation (Keskinbora, 2019).

Justice

Justice refers to the fair and appropriate allocation of resources. Intertwining with the values as above, it emphasizes that everyone has equal right to autonomy, beneficence, and non-maleficence regardless of their circumstances. AI brings a potential to “level the playing field” by providing access to advanced information and interventions to a greater number of people. However, opportunities tend to be stratified such that those with means benefit disproportionately. This combined with the possibility of occupations being replaced by robots may widen social gaps. It must be ensured that evolving technology remains available and attainable for all (Keskinbora 2019, Varkey 2021, Farhud DD 2021, Hamet P. 2017).

CHAPTER 6

CONCLUSION

Artificial intelligence has the potential to bring unprecedented benefits throughout society. There are many examples of promising usages, however the field is misunderstood and unregulated. It was Mary Shelley, author of the novel *Frankenstein*, who argued that science can be a destructive force that requires ethical consideration prior to implementation to avoid infringing upon fundamental rights and freedoms. Some of the preeminent minds of the time such as Stephan Hawking (theoretical physicist), Steve Wozniak (engineer, programmer, co-founder of Apple), Yann LeCun (Turing Award winning computer scientist), and Demis Hassabis (computer scientist, CEO of DeepMind Technologies) agree with this sentiment and have expressed the need for regulation (Keskinbora, 2019). It is difficult to anticipate the future of AI, but the potential benefits are too valuable and the liabilities too perilous to continue unfettered. Society must reach a consensus of regulatory standards to optimize AI's benefits while prioritizing the core ethical principles of beneficence, non-maleficence, autonomy and justice to ensure safety, efficacy, and value.

BIBLIOGRAPHY

- AMA. (2023). *AMA issues new principles for AI development, deployment & use*. Retrieved from <https://www.ama-assn.org/press-center/press-releases/ama-issues-new-principles-ai-development-deployment-use>
- America, I. o. (2001). *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington (DC): National Academies Press (US).
- Attia, Z. I. (2019). Age and Sex Estimation Using Artificial Intelligence From Standard 12-Lead ECGs. *Circulation, Arrhythmia and Electrophysiology*, 9.
- Attia, Z. I. (2021). Application of artificial intelligence to the electrocardiogram. *European heart journal*, 4717–4730.
- Attia, Z. K.-J. (2019). Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nature Medicine*, 70–74.
- Briganti G, L. M. (2020). Artificial Intelligence in Medicine: Today and Tomorrow. *Front. Med.*, 7:27.
- DuBosar, R. (2024, March). *Is AI OK for IM?* Retrieved from I.M. Matters from ACP : https://immattersacp.org/archives/2024/03/is-ai-ok-for-im.htm?utm_campaign=FY23-24_NEWS_ETOCIMMATTERS_DOMESTIC_030124_EML&utm_medium=email&utm_source=Eloqua
- Executive Order 14110. (2023). Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>, Vol. 88, No. 210.
- Farhud DD, Z. S. (2021, November). Ethical Issues of Artificial Intelligence in Medicine and Healthcare. *Iran J Public Health*, 50(11):i-v.
- Hamet P., T. J. (2017). Artificial intelligence in medicine. *Metabolism*, 69, S36-S40.
- HIPAA. (n.d.). Health Insurance Portability and Accountability Act. Pub. L. No. 104-191, (1996).
- Holmstrom L., C. M. (2023). Deep learning-based electrocardiographic screening for chronic kidney disease. *Nature Communications Medicine*, 73.
- Insel, T. R. (2019). How algorithms could bring empathy back to medicine. *Nature*, Vol.567 (7747), p.172-173.

- Jiang, B. (2024, June 24). *China's AI event World Intelligence Expo attracts few foreign firms in Tianjin amid global rift*. Retrieved from South China Morning Post: <https://www.scmp.com/tech/big-tech/article/3267795/chinas-ai-event-world-intelligence-expo-attracts-few-foreign-firms-tianjin-amid-global-rift>
- Jonsen AR, S. M. (2015). *Clinical Ethics: A Practical Approach to Ethical Decisions in Clinical Medicine, 8th Edition*. McGraw-Hill Education.
- Joshi, I. (2019). Waiting for deep medicine. *The Lancet (British edition)*, Vol.393 (10177), p.1193-1194.
- Kaul V, E. S. (2020). History of artificial intelligence in medicine. *Gastrointestinal Endoscopy*, 92(4), 807-812.
- Keskinbora, K. (2019). Medical ethics considerations on artificial intelligence. *Journal of Clinical Neuroscience*, Volume 64, Pages 277-282.
- Kulkarni AR, P. A. (2023). Machine-learning algorithm to non-invasively detect diabetes and pre-diabetes from electrocardiogram. *BMJ Innov* , 32-42.
- Kumari, S. V. (2022). Selfie fundus imaging for diabetic retinopathy screening. *Eye*, 1988–1993.
- Kwon, J. M. (2020). A deep learning algorithm to detect anaemia with ECGs: a retrospective, multicentre study. . *The Lancet*, e358–e367.
- Liu CM, S. E. (2022). Artificial Intelligence-Enabled Electrocardiogram Improves the Diagnosis and Prediction of Mortality in Patients With Pulmonary Hypertension. *JACC Asia*, 258-270.
- Matuchansky, C. (2019). Deep medicine, artificial intelligence, and the practising clinician. *The Lancet*, Vol.394 (10200), p.736-736.
- Melillo, G. (2024, April). *How AI could change health care for the better*. Retrieved from ACP Internal Medicine Meeting News: https://read.nxtbook.com/customnews_inc/american_college_physicians/imm_2024_news_fri_am_april_19/opening_session.html?utm_campaign=FY23-24_NEWS_IMDAILY_DOMESTIC_041924A_EML&utm_medium=email&utm_source=Eloqua
- Pyrros, A. B. (2023). Opportunistic detection of type 2 diabetes using deep learning from frontal chest radiographs. *Nature Communications* , 4039.
- Rezaeikhonakdar, D. (2023). AI Chatbots and Challenges of HIPAA Compliance for AI Developers and Vendors. *The Journal of Law, Medicine & Ethics*, 988–995.

- Rim TH, L. C. (2021). Deep-learning-based cardiovascular risk stratification using coronary artery calcium scores predicted from retinal photographs. *The Lancet Digital Health*, e306 - e316.
- Sabanayagam C, X. D. (2020). A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations. *The Lancet Digital Health*, e295 - e302.
- Sun L, Y. C. (2023). Artificial intelligence for healthcare and medical education: a systematic review. *Am J Transl Res*, 15(7):4820-4828.
- Topol, E. (2023, October). *Can AI catch what doctors miss?* Retrieved from TED Conferences:
https://www.ted.com/talks/eric_topol_can_ai_catch_what_doctors_miss
- Ueda D., M. T. (2023). Artificial intelligence-based model to classify cardiac functions from chest radiographs: a multi-institutional, retrospective model development and validation study. *The Lancet Digital Health*, E525-E533.
- Varkey, B. (2021). Principles of Clinical Ethics and Their Application to Practice. *Medical principles and practice : international journal of the Kuwait University, Health Science Centre*, 30(1), 17–28.
- Xiao W, H. X. (2021). Screening and identifying hepatobiliary diseases through deep learning using ocular images: a prospective, multicentre study. *The Lancet Digital Health*, e88 - e97.
- Zhou, Y. C. (2023). A foundation model for generalizable disease detection from retinal images. *Nature*, 156–163.

APPENDIX

PERMISSION LETTER FOR INCLUSION OF COPYRIGHTED MATERIAL

ELSEVIER LICENSE
TERMS AND CONDITIONS
Jul 22, 2024

This Agreement between Joseph Capalbo / Temple University ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	5834310963929
License date	Jul 22, 2024
Licensed Content	Publisher Elsevier
Licensed Content Publication	Gastrointestinal Endoscopy
Licensed Content Title	History of artificial intelligence in medicine
Licensed Content Author	Vivek Kaul,Sarah Enslin,Seth A. Gross
Licensed Content Date	Oct 1, 2020
Licensed Content Volume	92
Licensed Content Issue	4
Licensed Content Pages	6
Start Page	807
End Page	812
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	electronic

Are you the author of this Elsevier article?	No
Will you be translating?	No
Title of new work	THE ETHICAL DILEMMA OF ARTIFICIAL INTELLIGENCE IN MEDICINE
Institution name	Temple University
Expected presentation date	Apr 2025
Order reference number	105
Portions	Figure 1
The Requesting Person / Organization to Appear on the License	Joseph Capalbo / Temple University
Requestor Location	Temple University Center for Urban Bioethics 335 Poplar Street PHILADELPHIA, PA 19123 United States Attn: Temple University Center for Urban Bioethics
Publisher Tax ID	98-0397604
Total	0.00 USD
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your RightsLink account and that are available at any time at <https://myaccount.copyright.com>).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.
3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable

acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows: "Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given. The material may not be reproduced or used in any other way, including use in combination with an artificial intelligence tool (including to train an algorithm, test, process, analyse, generate output and/or develop any form of artificial intelligence tool), or to create any derivative work and/or service (including resulting from the use of artificial intelligence tools).

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written authorization of Elsevier Ltd. (Please contact Elsevier's permissions helpdesk here). No modifications can be made to any Lancet figures/tables and they must be reproduced in full.

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. **No Transfer of License:** This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. **No Amendment Except in Writing:** This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. **Objection to Contrary Terms:** Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. **Revocation:** Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply only to specific license types:

15. **Translation:** This permission is granted for non-exclusive world English rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article.

16. **Posting licensed content on any Website:** The following terms and conditions apply as follows: Licensing material from an Elsevier journal: All content posted to the web site must maintain the copyright information line on the bottom of each image; A hyper-text must be included to the Homepage of the journal from which you are licensing at <http://www.sciencedirect.com/science/journal/xxxxx> or the Elsevier homepage for books at <http://www.elsevier.com>; Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

Licensing material from an Elsevier book: A hyper-text link must be included to the Elsevier homepage at <http://www.elsevier.com> . All content posted to the web site must maintain the copyright information line on the bottom of each image.

Posting licensed content on Electronic reserve: In addition to the above the following clauses are applicable: The web site must be password-protected and made available only to bona fide students registered on a relevant course. This permission is granted for 1 year only. You may obtain a new license for future website

posting.

17. **For journal authors:** the following clauses are applicable in addition to the above:

Preprints:

A preprint is an author's own write-up of research results and analysis, it has not been peer-reviewed, nor has it had any other value added to it by a publisher (such as formatting, copyright, technical enhancement etc.).

Authors can share their preprints anywhere at any time. Preprints should not be added to or enhanced in any way in order to appear more like, or to substitute for, the final versions of articles however authors can update their preprints on arXiv or RePEc with their Accepted Author Manuscript (see below).

If accepted for publication, we encourage authors to link from the preprint to their formal publication via its DOI. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help users to find, access, cite and use the best available version. Please note that Cell Press, The Lancet and some society-owned have different preprint policies. Information on these policies is available on the journal homepage.

Accepted Author Manuscripts: An accepted author manuscript is the manuscript of an article that has been accepted for publication and which typically includes author-incorporated changes suggested during submission, peer review and editor-author communications. Authors can share their accepted author manuscript:

- Immediately
 - via their non-commercial person homepage or blog
 - by updating a preprint in arXiv or RePEc with the accepted manuscript
 - via their research institute or institutional repository for internal institutional uses or as part of an invitation-only research collaboration work-group
 - directly by providing copies to their students or to research collaborators for their personal use
 - for private scholarly sharing as part of an invitation-only work group on commercial sites with which Elsevier has an agreement
- After the embargo period:
 - Via non-commercial hosting platforms such as their institutional repository
 - Via commercial sites with which Elsevier has an agreement

In all cases accepted manuscripts should:

- link to the formal publication via its DOI
- bear a CC-BY-NC-ND license - this is easy to do
- if aggregated with other manuscripts, for example in a repository or other site, be shared in alignment with our hosting policy not be added to or enhanced in any way to appear more like, or to substitute for, the published journal article.

Published journal article (JPA): A published journal article (PJA) is the definitive final record of published research that appears or will appear in the journal and embodies all value-adding publishing activities including peer review co-ordination, copy-editing, formatting, (if relevant) pagination and online enrichment.

Policies for sharing publishing journal articles differ for subscription and gold open access articles:

Subscription Articles: If you are an author, please share a link to your article rather than the full-text. Millions of researchers have access to the formal publications on ScienceDirect, and so links will help your users to find, access, cite, and use the best available version.

Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

If you are affiliated with a library that subscribes to ScienceDirect you have additional private sharing rights for others' research accessed under that agreement. This includes use for classroom teaching and internal training at the institution (including use in course packs and courseware programs), and inclusion of the article for grant funding purposes.

Gold Open Access Articles: May be shared according to the author-selected end-user license and should contain a CrossMark logo, the end user license, and a DOI link to the formal publication on ScienceDirect.

Please refer to Elsevier's posting policy for further information.

18. **For book authors** the following clauses are applicable in addition to the above: Authors are permitted to place a brief summary of their work online only. You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version. Posting to a repository: Authors are permitted to post a summary of their chapter only in their institution's repository.

19. **Thesis/Dissertation:** If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for Proquest/UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission. Theses and dissertations which contain embedded PJAs as part of the formal submission can be posted publicly by the awarding institution with DOI links back to the formal publications on ScienceDirect.

Elsevier Open Access Terms and Conditions

You can publish open access with Elsevier in hundreds of open access journals or in nearly 2000 established subscription journals that support open access publishing. Permitted third party re-use of these open access articles is defined by the author's choice of Creative Commons user license. See our open access license policy for more information.

Terms & Conditions applicable to all Open Access articles published with Elsevier:

Any reuse of the article must not represent the author as endorsing the adaptation of the article nor should the article be modified in such a way as to damage the author's honour or reputation. If any changes have been made, such changes must be clearly indicated.

The author(s) must be appropriately credited and we ask that you include the end user license and a DOI link to the formal publication on ScienceDirect.

If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder.

Additional Terms & Conditions applicable to each Creative Commons user license:

CC BY: The CC-BY license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article and to make commercial use of the Article (including reuse and/or resale of the Article by commercial entities), provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work.

The full details of the license are available at <http://creativecommons.org/licenses/by/4.0>.

CC BY NC SA: The CC BY-NC-SA license allows users to copy, to create extracts, abstracts and new works from the Article, to alter and revise the Article, provided this is not done for commercial purposes, and that the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, indicates if changes were made and the licensor is not represented as endorsing the use made of the work. Further, any new works must be made available on the same conditions. The full details of the license are available at <http://creativecommons.org/licenses/by-nc-sa/4.0>.

CC BY NC ND: The CC BY-NC-ND license allows users to copy and distribute the Article, provided this is not done for commercial purposes and further does not permit distribution of the Article if it is changed or edited in any way, and provided the user gives appropriate credit (with a link to the formal publication through the relevant DOI), provides a link to the license, and that the licensor is not represented as endorsing the use made of the work. The full details of the license are available at <http://creativecommons.org/licenses/by-nc-nd/4.0>. Any commercial reuse of Open Access articles published with a CC BY NC SA or CC BY NC ND license requires permission from Elsevier and will be subject to a fee.

Commercial reuse includes:

- Associating advertising with the full text of the Article
- Charging fees for document delivery or access
- Article aggregation
- Systematic distribution via e-mail lists or share buttons

Posting or linking by commercial companies for use by customers of those companies.

20. Other Conditions:

v1.10

Questions? E-mail us at customercare@copyright.com.