2022

halake

[Company name]

[Date]

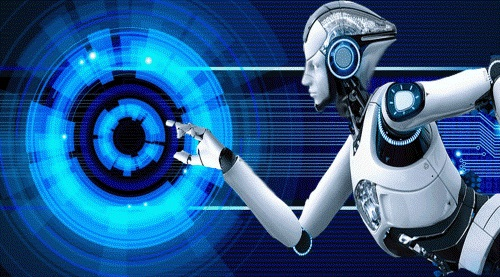
**JIMMA UNIVERSITY**

**JIMMA INSTITUTE OF TECHNOLOGY**

**FACULTY OF COMPUTING AND INFORMATICS**

**DEPARTMENT OF COMPUTER SCIENCE**

**ETHICAL AND LEGAL CHALLENGES OF ARTIFICIAL INTELLIGENCE**

****

|  |  |  |  |
| --- | --- | --- | --- |
| S.NO | FNAME | LNAME | ID NO |
| 1 | HALAKE | GUYO | RU2900/11 |
| 2 | YADENI | MISGANA | RU2909/11 |
| 3 | SHUME | ALEMAYEHU | RU2880/11 |
| 4 | YASMIN | MUSTAFA | RU3005 |

SUBMISSION TO: Mr. DESALEW

SUBMISSION DATE:4 May 8, 2022.

**#Q7.** **ETHICAL AND LEGAL CHALLENGES OF ARTIFICIAL INTELLIGENCE**

All scientists and engineers face ethical considerations of how they should act on the job, what projects should or should not be done, and how they should be handled. AI, however, seems to pose some fresh problems beyond that of, say, building bridges that don’t fall down:

• People might lose their jobs to automation.

• People might have too much (or too little) leisure time.

• People might lose their sense of being unique.

• AI systems might be used toward undesirable ends.

• The use of AI systems might result in a loss of accountability.

• The success of AI might mean the end of the human race.

* **People might lose their jobs to automation.**

The modern industrial economy has become dependent on computers in general, and select AI programs in particular. For example, much of the economy, especially in the United States, depends on the availability of consumer credit. Credit card applications, charge approvals, and fraud detection are now done by AI programs.

* **The use of AI systems might result in a loss of accountability.**

In the litigious atmosphere that prevails in the United States, legal liability becomes an important issue. When a physician relies on the judgment of a medical expert system for a diagnosis, who is at fault if the diagnosis is wrong? Fortunately, due in part to the growing influence of decision-theoretic methods in medicine, it is now accepted that negligence cannot be shown if the physician performs medical procedures that have high expected utility, even if the actual result is catastrophic for the patient.

* **The success of AI might mean the end of the human race.**

Almost any technology has the potential to cause harm in the wrong hands, but with AI and robotics, we have the new problem that the wrong hands might belong to the technology itself. Countless science fiction stories have warned about robots or robot–human cyborgs running amok. The question is whether an AI system poses a bigger risk than traditional software. We will look at **three sources of risk**.

**First**, the AI system’s state estimation may be incorrect, causing it to do the wrong

thing. For example, an autonomous car might incorrectly estimate the position of a car in the

adjacent lane, leading to an accident that might kill the occupants. More seriously, a missile

defense system might erroneously detect an attack and launch a counterattack, leading to

the death of billions. These risks are not really risks of AI systems—in both cases the same mistake could just as easily be made by a human as by a computer.

**Second**, specifying the right utility function for an AI system to maximize is not so

easy. For example, we might propose a utility function designed to minimize human suffering, expressed as an additive reward function over time

**Third,** the AI system’s learning function may cause it to evolve into a system with

unintended behavior.

* **laws of robotics**

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2. A robot must obey orders given to it by human beings, except where such orders would conflict with the First Law

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

* **Ethical challenges**

The use of AI in the clinical practice of healthcare has huge potential to transform it for the better, but it also raises ethical challenges. Those are: -

1.Informed consent to use

Health AI applications, such as imaging, diagnostics, and surgery, will transform the patient–clinician relationship. There is a need to examine under what circumstances (if at all) the principles of informed consent should be deployed in the clinical AI space. To what extent do clinicians have a responsibility to educate the patient around the complexities of AI, including the form(s) of ML used by the system, the kind of data inputs, and the possibility of biases or other shortcomings in the data that is being used? Under what circumstances must a clinician notify the patient that AI is being used at all?

These questions are especially challenging to answer in cases where the AI operates using “black-box” algorithms, which may result from noninterpretable machine-learning techniques that are very difficult for clinicians to understand fully.

**2. Safety and transparency**

Safety is one of the biggest challenges for AI in healthcare. To use one well-publicized example, IBM Watson for Oncology uses AI algorithms to assess information from patients’ medical records and help physicians explore cancer treatment options for their patients. However, it has recently come under criticism by reportedly giving “unsafe and incorrect” recommendations for cancer treatments. The problem seems to be in the training of Watson for Oncology: instead of using real patient data, the software was only trained with a few “synthetic” cancer cases, meaning they were devised by doctors at the Memorial Sloan Kettering (MSK) Cancer Center. MSK has stated that errors only occurred as part of the system testing and thus no incorrect treatment recommendation has been given to a real patient.

The recommendations of more “black-box” systems raise particular concerns. It will be a challenge to determine how transparency can be achieved in this context. Even if one could streamline the model into a simpler mathematical relationship linking symptoms and diagnosis, that process might still have sophisticated transformations beyond the skills of clinicians (and especially patients) to understand. However, perhaps there is no need to open the “black box”: It may be that at least in some cases positive results from randomized trials or other forms of testing will serve as a sufficient demonstration of the safety and effectiveness of AIs.

**3. Algorithmic fairness and biases**

AI has the capability to improve healthcare not only in high-income settings, but to democratize expertise, healthcare, and bring it to even remote areas. However, any ML system or human-trained algorithm will only be as trustworthy, effective, and fair as the data that it is trained with. AI also bears a risk for biases and thus discrimination. It is therefore vital that AI makers are aware of this risk and minimize potential biases at every stage in the process of product development. In particular, they should consider the risk for biases when deciding: -

1. which ML technologies/procedures they want to use to train the algorithms
2. what datasets (including considering their quality and diversity) they want to use for the programming.

Some of these biases may be resolved due to increased data availability and attempts to better collect data from minority populations and better specify for which populations the algorithm is or is not appropriately used. However, a remaining problem is that a variety of algorithms are sophisticated and nontransparent. In addition, as we have seen in the policing context, some companies developing software will resist disclosure and claim trade secrecy in their work. It may therefore likely be left to nongovernmental organizations to collect the data and show the biases. In cases of “black-box” algorithms, many scholars have argued that explain ability is necessary when an AI makes health recommendations, especially also to detect biases.

1. **Data privacy**

The data sharing happened for the clinical safety testing of “Streams,” an app that aims to help with the diagnosis and detection for acute kidney injury. However, patients were not properly informed about the processing of their data as part of the test. Information Commissioner’s correctly pointed out that “the price of innovation does not need to be the erosion of fundamental privacy rights”.

Although the Streams app does not use AI, this real-life example has highlighted the potential for harm to privacy rights when developing technological solutions. If patients and clinicians do not trust AIs, their successful integration into clinical practice will ultimately fail. It is fundamentally important to adequately inform patients about the processing of their data and foster an open dialog to promote trust.

* **Legal challenges**

Many of the ethical issues discussed above have legal solutions or ramifications; while there is nothing sacrosanct in our division between the two, we now shift to challenges we associate more directly with the legal system.

1. **Safety and effectiveness**

AIs are safe and effective. Stakeholders can contribute to a successful implementation of AI in clinical practice by making sure that the datasets are reliable and valid, perform software updates at regular intervals, and being transparent about their product, including shortcomings such as data biases. In addition, an adequate level of oversight is needed to ensure the safety and effectiveness of AI.

1. **Liability**

AI-based technologies also raise challenges for current liability regimes. It will be crucial to creating an optimal liability design that figures out responsibilities.

Setting the optimal liability regime depends heavily on what one thinks the “problem” is. If one is concerned that the deployment of AI-based technology in the clinical space is associated with a high risk for patients to get hurt, one might want to keep the current medical malpractice regime that attempts to meet both tort law’s two functions: (1) deterrence

(2) compensation of the victims. By contrast, if one believes that over the run of cases, reliance on AI promotes patient health, then it may be a problem if physicians prove reluctant to rely on these algorithms, especially the opaquer ones, when they remain on the hook for resulting liability. This might drive the policy-maker to a different model.

Some have proposed product liability against the makers of AI, a tort that generally entails a strict liability of the manufacturer for defects. However, there are considerable challenges to win such a claim in practice. Courts have hesitated to apply or extend product liability theories to healthcare software developers since such software is currently primarily considered as a tool to support clinicians make the final decision.

1. **Data protection and privacy**

In the world of big data, it is of pivotal importance that there are data protection laws in place that adequately protects the privacy of individuals, especially patients.

**Top Common Challenges in AI**

1. **Computing Power**

The amount of power these power-hungry algorithms use is a factor keeping most developers away. Machine Learning and Deep Learning are the stepping stones of this Artificial Intelligence, and they demand an ever-increasing number of cores and GPUs to work efficiently. There are various domains where we have ideas and knowledge to implement deep learning frameworks such as asteroid tracking, healthcare deployment, tracing of cosmic bodies, and much more.

They require a supercomputer’s computing power, and yes, supercomputers aren’t cheap. Although, due to the availability of Cloud Computing and parallel processing systems developers work on AI systems more effectively, they come at a price. Not everyone can afford that with an increase in the inflow of unprecedented amounts of data and rapidly increasing complex algorithms.

1. **Trust Deficit**

One of the most important factors that are a cause of worry for the AI is the unknown nature of how deep learning models predict the output. How a specific set of inputs can devise a solution for different kinds of problems is difficult to understand for a layman.

Many people in the world don’t even know the use or existence of Artificial Intelligence, and how it is integrated into everyday items they interact with such as smartphones, Smart TVs, Banking, and even cars.

1. **Limited Knowledge**

Although there are many places in the market where we can use Artificial Intelligence as a better alternative to the traditional systems. The real problem is the knowledge of Artificial Intelligence. Apart from technology enthusiasts, college students, and researchers, there are only a limited number of people who are aware of the potential of AI.

For example, there are many SMEs (Small and Medium Enterprises) which can have their work scheduled or learn innovative ways to increase their production, manage resources, sell and manage products online, learn and understand consumer behavior and react to the market effectively and efficiently.

1. **Human-level**

This is one of the most important challenges in AI, one that has kept researchers on edge for AI services in companies and start-ups.

For a deep learning model to perform a similar performance would require unprecedented finetuning, hyperparameter optimization, large dataset, and a well-defined and accurate algorithm, along with robust computing power, uninterrupted training on train data and testing on test data. That sounds a lot of work, and it’s actually a hundred times more difficult than it sounds.

One way you can avoid doing all the hard work is just by using a service provider, for they can train specific deep learning models using pre-trained models. They are trained on millions of images and are fine-tuned for maximum accuracy, but the real problem is that they continue to show errors and would really struggle to reach human-level performance.

1. **Data Privacy and Security**

The main factor on which all the deep and machine learning models are based on is the availability of data and resources to train them. Yes, we have data, but as this data is generated from millions of users around the globe, there are chances this data can be used for bad purposes.

Some companies have already started working innovatively to bypass these barriers. It trains the data on smart devices, and hence it is not sent back to the servers, only the trained model is sent back to the organization.

1. **The Bias Problem**

The good or bad nature of an AI system really depends on the amount of data they are trained on. Hence, the ability to gain good data is the solution to good AI systems in the future. But, in reality, the everyday data the organizations collect is poor and holds no significance of its own.

They are biased, and only somehow define the nature and specifications of a limited number of people with common interests based on religion, ethnicity, gender, community, and other racial biases. The real change can be brought only by defining some algorithms that can efficiently track these problems.

1. **Data Scarcity**

With major companies such as Google, Facebook, and Apple facing charges regarding unethical use of user data generated, various countries such as India are using stringent IT rules to restrict the flow. Thus, these companies now face the problem of using local data for developing applications for the world, and that would result in bias.

The data is a very important aspect of AI, and labeled data is used to train machines to learn and make predictions. Some companies are trying to innovate new methodologies and are focused on creating AI models that can give accurate results despite the scarcity of data. With biased information, the entire system could become flawed.

**Generally,** AI is going to be increasingly used in healthcare and hence needs to be morally accountable. Data bias needs to be avoided by using appropriate algorithms based on un-biased real time data. Diverse and inclusive programming groups and frequent audits of the algorithm, including its implementation in a system, need to be carried out. While AI may not be able to completely replace clinical judgment, it can help clinicians make better decisions.