Ethical Challenges of Artificial Intelligence & Machine Learning

Abstract:

Objective: To investigate the contributing factors and cases surrounding the concerning question, *Is it Ethical?*, in regards to Artificial Intelligence and machine learning. We briefly overview some of the broad ethical topics addressed by the Association for Computing Machinery, as well as discuss past events of ethical dilemmas involving A.I. We can put into perspective the correlations of ethical criteria that apply to both human and synthesized intelligence. We find that we are constantly on the edge of overcoming programming errors in A.I.

Keywords: Artificial Intelligence, Deepfake, Ethics, Algorithm(s)

1. Introduction

Thanks to the 'internet of things' and the growing power of A.I. we are able to collect and catalog data faster than ever before, and even more so that this data has the power to be very accurate. There are people hoping to use this data as a means to help people and improve the overall quality of life, however that is a task with a great deal of complications to execute properly. This level of accuracy has sparked a number of concerns, and in some aspects complaints, about its ability to deceive the untrained eye and in turn create problems not originally intended by its creators, and we are proving in this article whether or not that the artificial intelligence programs designed for every day people are in direct defiance of the ACM Code of Ethics.

1.1 Concise Literature Reviews

Human beings are constantly challenged with facing the task of making difficult decisions. Depending on the situation, those decisions will affect others. It is only after we have chosen to act or, in certain cases where we know the end result, do we reflect inwardly and question our actions. Imagine, leaving the decision making to a program that thinks on its own, do you think it would make the "right" decision. The time of successfully creating fully autonomous Artificial Moral Agents (AMAs) draws near and the idea of bringing these metallic beings to life is becoming normalized in the field of artificial intelligence and robotics(Tonkens, 2009). One can argue that creating a thinking machine in of itself is unethical, simply theorizing the possibility of doing such a thing presents a series of ethical challenges (Bostrom & Yudkowsky, 2014). This thought poses the question of how easily computed are ethics? Or can it even be computed?

The essential ethical challenge presented, would be to ensure that the algorithm prevents machines, such as AMA's, from harming humans. Preventing the machine from performing any actions that would prove damaging to its moral status/standing (Bostrom & Yudkowsky, 2014). The questions needing to be addressed before you can even begin coding are: "Who is writing the code?", "How will it be implemented?", "How is someone supposed to create an ethical algorithm that would make the machine more ethical than its creator?", and "What standards are both the creator and his creation being held too?" Bostrom and Yudkowsky presented the scenario of a possible error/lapse in judgement for the AI.

Suppose a bank uses a machine learning algorithm that recommends mortgage applications for approval but, it is later discovered that the algorithm is biased based on race (i.e., African American or Hispanic) which it should be blind too. Even so, statistics show that the approval rates for blacks are decreasing at a constant rate in the algorithm. It may prove rather difficult, maybe even impossible to determine why this event is happening, or if the issue can be resolved. It is said that, "When AI algorithms take on cognitive tasks with social dimensions- tasks previously performed by humans- the algorithm also inherits the social requirement." (Bostrom & Yudkowsky, 2014). The algorithm could possibly be using personal data such as the place of birth or the registered address of a customer as a factor of its discriminatory tendency. Hence, these algorithms should not only be powerful and scalable but also transparent to inspection (Bostrom & Yudkowsky, 2014).

The ACM Code of ethics states that one should "Strive to achieve high quality in both the processes and products of professional work"(Association for Computing Machinery, 2018). Transparency and avoiding harm are the two major ethical concerns when dealing with AI algorithms but, responsibility, audibility, incorruptibility, and predictability are all criteria that apply to humans performing social functions as well as the two aforementioned concerns, meaning, the same should go for any machine/program with social dimensions (Bostrom & Yudkowsky, 2014). If your program intends to replace human judgment, you should be able to produce the same logical outcome as your program including the pathways to reach your final solution. The algorithm should not be easily corrupted, nor should it be easily exploitable either, if such instances were to occur there should be someone that can be held accountable.

Another consideration with the implementation of these thinking machines is accountability."A computing professional has an additional obligation to report any signs of system risks that might result in harm. If leaders do not act to curtail or mitigate such risks, it may be necessary to "blow the whistle" to reduce potential harm. However, capricious or misguided reporting of risks can itself be harmful. Before reporting risks, a computing professional should carefully assess relevant aspects of the situation."(Association for Computing Machinery, 2018) It is essential to know

who's tasked with the responsibility of monitoring these machines, not only for a speedy resolution if an issue occurs, but also to monitor any imminent changes in the program's behavior.

Even more than that, a rising internet trend is causing great concern in the contemporary tech crowd. The art of “deepfaking” is raising concerns of ethical principles, for example, in an article “Face/Off: “DeepFake” Face Swaps and Privacy Laws” the author describes Deep fake to a concept similar of a late 90s movie Face/Off where the main character John Travolta and another character end up having each other’s faces which on the outside sounds pretty hilarious and intriguing but in reality someone parading around the internet with your face can put you or others in a compromising position (Gerstner, 2020). The article states that in early 2018 A machine learning algorithm was able to create a fake occurrence of President Barack Obama giving a speech and since then the technology has only gotten better and more accurate, being able to make it appear as though people have said things or done things that they have not actually executed in reality.

The way that deepfake works is that it takes an algorithm that uses machine learning and composites pictures from different angles of a person until it can successfully composite a near perfect grouping of pictures that began to look like a video. The caveat is, in order to possess enough pictures to make up for the many transitions in a video you would have to be a person who is notably more often exposed to cameras than the average person, which makes celebrities an ideal candidate for deepfake videos. However, in the modern age of “selfies” and social media posts theoretically if an average person posts enough “selfies” they can easily accidentally put themselves in a position to have a deepfake composite made of them.

According to Eric Gerstner, who has a vast understanding of the law, “ in its most basic form, the tort applies when one “appropriates the commercial value of a person’s identity by using without consent the person’s name, likeness, or other indicia of identity” (Gerstner, 2020). This issue is in direct violation of the ACM’s Code of Ethics sections 1.2 Avoid harm; “Well-intended actions, including those that accomplish assigned duties, may lead to harm. When that harm is unintended, those responsible are obliged to undo or mitigate the harm as much as possible. Avoiding harm begins with careful consideration of potential impacts on all those affected by decisions. When harm is an intentional part of the system, those responsible are obligated to ensure that the harm is ethically justified. In either case, ensure that all harm is minimized” (Association for Computing Machinery, 2018). There are currently little to no laws to protect citizens from any respective dangers that deepfake may provide and the creators/users of the technology have failed at their responsibility to ensure that the minimal amount of harm is being done on their end, and in code 1.6 “Respect privacy,” “Computing professionals should only use personal information for legitimate ends and without violating the rights of individuals and groups .... Merged data collections can compromise privacy features present in the original collections. Therefore, computing professionals should take special care for privacy when merging data collections” (Association for Computing Machinery, 2018). While the amount of data taken in for the original

testing to assure the program works is acceptable, it still breaks this code of ethics because it cannot be ensured that all of the information is obtained in a legitimate way.

Section 3.1 of the Code of ethics, 'Ensure that the public good is the central concern during all professional computing work' states, "People—including users, customers, colleagues, and others affected directly or indirectly—should always be the central concern in computing. The public good should always be an explicit consideration when evaluating tasks associated with research, requirements analysis, design, implementation, testing, validation, deployment, maintenance, retirement, and disposal. Computing professionals should keep this focus no matter which methodologies or techniques they use in their practice" (Association for Computing Machinery, 2018). After releasing this code, it is apparent that the creators have not put any restrictions or followed up on people using that code to ensure that it would be used for the public good, which brings up the validity of the deepfake algorithm and if it was legitimately written by computing professionals.

2. Data Analysis

There is a great deal of data on the internet, including the internet of things that we must address to give the context on the margin of error that programmers should be driven to overcome. Unbeknownst to many, the Internet of things has a far wider range than the common person would expect. Reaching deeper than mobile phones and social science data, it also deals with radio frequency identification (RFID), telecommunications, and even wireless sensor technology. It's reach and impact is so wide that it is considered one of the six "destructive civil technologies" by the US national intelligence council in the article the Internet of things: a survey the authors cite a report from the council saying "by 2025 Internet nodes may reside in everyday things food packages furniture paper documents and more" (Atzori et al., 2010, p. 2).

2.1 Methodology

2.1.1 Methodology in Comparing Technology

The previously mentioned article sets itself up as a research endeavor to try and estimate where the course of the internet is going based on current internet system's technology and how it holds up to current research being done. The only way to determine its rate of success is by opposing it to past hardware and software and looking at future potential undertakings. The article conveys how the current leading technology is in RFID since it is currently the system that satisfies all of our needs and provides us our most current strongest line of personal safety however it does acknowledge that eventually like all technology it will become obsolete. The article brings up the drawbacks of wireless sensor networks (WSN) stating the major drawbacks, "Sensor networks may consist of a very large number of nodes. This would result in obvious problems as today there is a scarce availability of IP addresses.

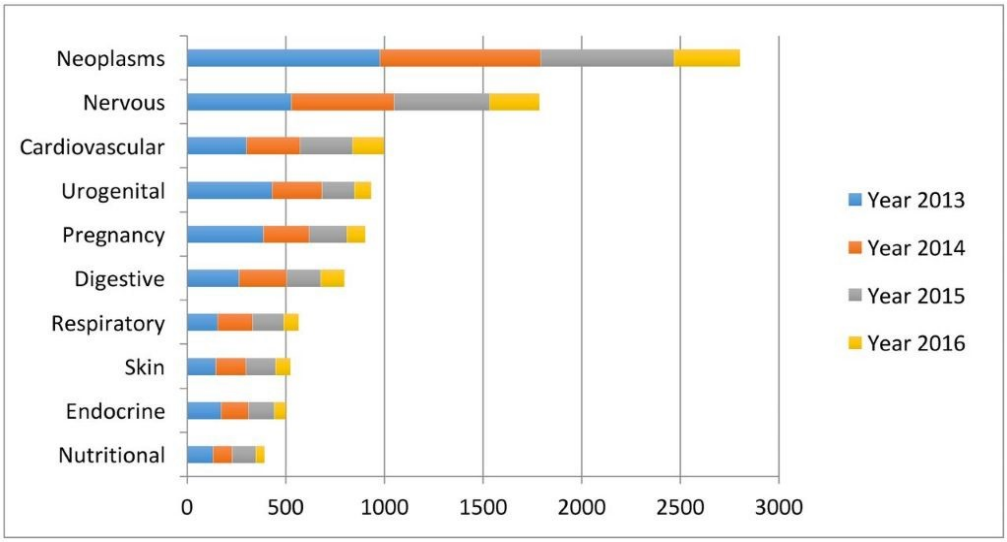
The largest physical layer packet in IEEE 802.15.4 has 127 bytes; the resulting maximum frame size at the media access control layer is 102 octets, which may further decrease based on the link layer security algorithm utilized. Such sizes are too small when compared to typical IP packet sizes. In many scenarios, sensor nodes spend a large part of their time in a sleep mode to save energy and cannot communicate during these periods. This is absolutely anomalous for IP networks. ”(Atzori et al., 2010) comparing them to the advantages of the RFID system networks (RSN) which they have described as “ RFID systems are the very small size and the very low cost. Furthermore, their lifetime is not limited by the battery duration; wireless sensor networks are the high radio coverage and the communication paradigm, which does not require the presence of a reader (communication is peer-to-peer whereas, it is asymmetric for the other types of systems); RFID sensor network are the possibility of supporting sensing, computing, and communication capabilities in a passive system. ” (Atzori et al., 2010)

After the comparison with the current RFID networks to the past wireless networks, the authors take a compelling look at possible perspectives, one being Machine-to-Machine (M2M) which is supposed to develop and maintain faster into an architecture for any and all sensor network integration and existing M2M systems. While the status of current cost affect is ongoing, it seems to show a lot of support from leading system companies according to the authors “... the integration of [Routing Over Low power and Lossy networks (ROLL)] different things into wider networks, either mobile or fixed, will allow their interconnection with the Future Internet [61]. What is worth pointing out in the cited standardization areas is the tight collaboration between standardization Institutions and other world-wide Interest Groups and Alliances. It seems that the whole industry is willing to cooperate on achieving the IoT” (Atzori et al., 2010).

2.1.2 Methodology in The Medical Field

While there are many valid and well-intentioned reasons for using artificial intelligence in the medical field considering the technological aspects can be applied in a wide variety of parameters and the implementation of these autonomous systems seem to be steadily increasing in said fields. This data was collected from several sources of existing data from a variety of articles, books, and journals accessed via online databases. Once enough data had been collected, we began cross referencing the articles to find common trends of programming oversites and ethical challenges faced when aiming to implement these machines in the different professional fields. Although AI can be an added benefit when applied in many fields, the discussion of its implementation in the medical field and the ethical concerns involved seem to be an ongoing debate. The data also made note of the concerns of trustworthiness concerning the use/misuse of personal information when these systems are implemented in areas such as human resources and customer interaction are included. Also, the majority of the data collected ranges within a 10-year time window and shows that the rate of advancement with AI and machine learning isn’t as fast as one would think.

Major concerns regarding AI and machine learning stem from the medical field, according to the article “Artificial Intelligence in HealthCare: past present and future”, “Before AI systems can be deployed in healthcare applications, they need to be ‘trained’ through data that are generated from clinical activities, such as screening, diagnosis, treatment assignment and so on, so that they can learn similar groups of subjects, associations between subject features and outcomes of interest”(Jiang F, Jiang Y, Zhi H, et al.,2017). In some cases, the implementation and use of AI



systems and machine learning are already a prominent thing in medical facilities. Disease focus, more specifically cancer, nervous system disease and cardiovascular disease are areas where research of AI implementations and the benefits that are commonly discussed.

Table 1

The leading 10 disease types considered in the artificial intelligence (AI) literature. The first vocabularies in the disease names are displayed. (Jiang F, Jiang Y, Zhi H, et al, 2017) .

As previously mentioned, AI research is picking up traction in the medical field. From the data collected, it seems the main focus is on applying machine learning techniques to complex problems, such as cancer diagnosis and kidney exchange programs. It’s suggested that it “is dependent on allowing the system to make predictions based on large amounts of patients’ personal data, by learning their own associations” (Jiang F, Jiang Y, Zhi H, et al, 2017) .

2.2 Results

2.2.1 IoT Results

To say that there are no drawbacks with the RFID systems when it comes to privacy and security would be dishonest. “The problem of data integrity has been extensively studied in all traditional computing and communication systems and some preliminary results exist for sensor networks,...However, new problems arise when RFID systems are integrated in the Internet as they spend most of the time unattended. Data can be modified by adversaries while it is stored in the node or when it traverses the network”(Atzori et al., 2010). Given this information along with the scope of the Internet of things, that leaves a lot of information and data left out in the open, and this unprotected data directly goes against the ACM code of ethics design of ethernet systems that are robustly and usably secure.

More than that, addressing that RFID systems are not perfect, but they do possess the fewest number of issues while offering the greatest amount of security for the time being is a very important step because that is documentation that technology can and will improve with time. This data gives us a general rate of how swift these changes will take place. We are working on an exponential blueprint and fighting against time to ensure that privacy is maintained. “Authentication is difficult as it usually requires appropriate authentication infrastructures and servers that achieve their goal through the exchange of appropriate messages with other nodes. In the IoT such approaches are not feasible given that passive RFID tags cannot exchange too many messages with the authentication servers. The same reasoning applies (in a less restrictive way) to the sensor nodes as well”(Atzori et al., 2010)

2.2.2 Medical Bias Results

As discussed in section 2.1.2 , the idea of implementation of autonomous systems in medicine continues to be tested. There are major hypothesized advantages to their implementation, though this comes with major concerns as well. According to “Artificial Intelligence, bias and clinical safety,” “Estimates of the impact of AI on the wider economy globally vary wildly, with a recent report suggesting a 14% effect on global gross domestic product by 2030, half of which coming from productivity improvements.

These predictions create political appetite for the rapid development of the AI industry, and healthcare is a priority area where this technology has yet to be exploited,” (Jiang F, Jiang Y, Zhi H, et al, 2017). Yes, there are advantages to machine learning in medicine but, are those advantages worth the risks? Cited above in the concise literature review, one of the major ethical concerns is transparency. If these technologies end up being exploited in corrupt ways by greedy individuals/businesses, the negative results could potentially outweigh the good. The price of medical assistance can potentially increase if these systems are sold by private companies at an

inflated price. Not only that, but, as mentioned before, these systems would rely on patient data that can be misused or affect your chances of treatment if there is an error in the algorithm of the system. For example, it was previously mentioned how the implementation can help with kidney exchange programs, “when allocating a kidney, natural features include the probability that the kidney is rejected by a particular patient, whether that patient needs the kidney urgently, etc. Even in these scenarios, identifying all the relevant features may not be easy” (Conitzer et al. 2017). Now suppose an error is found in the system showing signs of gender or racial bias for reasons unknown.

That error may cause the system to skip over your name on the list of patients in need of a kidney because the hospital database shows that kidney failure is more likely to happen in Latin females, and you happen to be both those categories, similar to the example made in the concise literature review. It may seem far-fetched but similar instances have occurred in the past with other AI systems, such as Amazon's AMZN.O; “AMZN.O machine-learning specialists uncovered a big problem: their new recruiting engine did not like women. […] The team had been building computer programs since 2014 to review job applicants’ resumes with the aim of mechanizing the search for top talent, five people familiar with the effort told Reuters. […] In effect, Amazon’s system taught itself that male candidates were preferable. It penalized resumes that included the word “women’s,” as in “women’s chess club captain” (Lauret, 2019)

3.1 Concise Summary

Through our extensive research and data we are able to show solid evidence of artificial intelligence bias. As there is no current best way to ensure that that a bias will not occur if left unattended, it ethically is too unstable to use artificial intelligence or machine learning programs without fully-extensive debugging and strenuous testing. Regarding the safety and assurance of people whose lives greatly connect with the Internet of Things should be a top priority. Albeit, the natural human desire to want the latest trend in technology is very compelling, it is imperative to realize that in our capitalistic society corporations who seek to profit from that interest will ultimately cut corners on proper testing at the cost of one's personal safety.

After reviewing and comparing the exponential rate of how rapidly our most secure technology is becoming obsolete we do not project that we will reach or surpass the optimal timeline to get ahead of the curve and provide ultimate security restrictions for everyone on all platforms. With that in mind, we foresee having artificial intelligence programs that are to be entrusted into life-saving medical procedures or self-driving smart cars sound enticing, but there is a high volume of ramifications that would have to be absorbed if these programs are integrated into modern society carelessly. We are currently unaware of a moderating AI integration board/society but it is with our best recommendation based on this research that an overseeing board of that nature would be in the best interest of the common people to further advocate for the idealistic society. It should be

discussed that this artificial intelligence “board” should run in contingency with the Association for Computing Machinery (ACM). They created the society that formed the standards that we as programmers adhere to, so we believe that it only makes sense that they continue to ensure that the rules are being maintained on a case by case basis. The same goes for any individual programmer who is committed to remaining ethical in this society they are also equally responsible for ensuring the safety of the common people.

4.1 Extended Resources

- This compelling article by Henry Ajder shows a snapshot of deepfake videos and a compilation of statistics detected by his team:
    - https://sensity.ai/deepfake-threat-intelligence-a-statistics-snapshot-from-june-2020/#:~:text=Number%20of%20deepfakes%20identified%20online&text=As%20disclosed%20in%20The%20State,7%2C964%20videos%20in%20December%202018.
- This research journal goes into what it’s like on the developers end to create a program that can make moral decisions on a human empathy/logical level:
    - http://moralai.cs.duke.edu/documents/mai_docs/moralAAAI17.pdf
- A video that discusses the poor security of the Internet of Things and how easy it is to hack into products.
    - https://www.ted.com/talks/ken_munro_internet_of_things_security
- A video, discussing the advancements being made with autonomous systems as well as, how private companies are more than willing to invest millions of dollars into the the research and development of these new technologies
    - https://www.youtube.com/watch?v=3oE88_6jAwc&t=296s
- A video, listing risks associated with AI, along with visual representation for better understanding
    - https://www.youtube.com/watch?v=1oeoosMrJz4
- This article take a look at several examples of Autonomous robots that have made significant progress in its development or has already been completed ○ https://www.youtube.com/watch?v=1oeoosMrJz4

- A journal article written by Roman V. Yampolskiy, making a case that scientists are misguided when it comes to machine ethics and how they should focus more on the safety of AI.
    - https://link.springer.com/chapter/10.1007/978-3-642-31674-6_29

- A video, where Toby Walsh (a leading researcher in AI) simplifies the relationship of ethics and AI, and also providing reasoning for the regulation of new technology

    - https://www.youtube.com/watch?v=HSsQApXQGsI
- A textbook written by both top medical experts and leading robotic data analyst that dives deep into the long running history of medical machinery.
    - https://www.researchgate.net/publication/273123956_Machine_Medical_Ethics
- A short read by Apara Venkateswaran that discusses the methods of cheating in machine learning and why it's important to strive for integrity
    - https://towardsdatascience.com/ethics-in-machine-learning-9fa5b1aadc12

5. References:

- Association for Computing Machinery. (2018). ACM Code Of Ethics And Professional Conduct. *Affirming our obligation to use our skills to benefit society* [PDF File]. New York,NY: Author. Retrieved from https://www.acm.org/binaries/content/assets/about/acm-code-of-ethics-booklet.pdf

- Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A Survey. *The International Journal of Computer and Telecommunications Networking*, 1–19. https://cs.wmich.edu/alfuqaha/spring15/cs6570/lectures/IoT-survey.pdf

- Bostrom, Nick & Yudkowsky, Eliezer (2014). The ethics of artificial intelligence. In Keith Frankish (Ed.) & William M. Ramsey (Ed.), *The Cambridge Handbook of Artificial Intelligence* (pp. 315-332). Cambridge,UK: Cambridge University Press. https://books.google.com/books?id=RYOYAwAAQBAJ&lpg=PA316&ots=A1TZugaGxx&dq=ethical%20challenges%20of%20artificial%20intelligence%20&lr&pg=PA316#v=onepage&q&f=false

- Challen, R. (2019, March 1). *Artificial intelligence, bias and clinical safety*. BMJ Quality & Safety. https://qualitysafety.bmj.com/content/28/3/231.ful 1

- Conitzer, V., Sinnott-Armstrong, W., Borg, J. S., Deng, Y., & Kramer, M. (2017). *Moral decision making frameworks for artificial intelligence*. Proceedings of the 31st AAAI Conference on Artificial Intelligence, AAAI 2017, San Francisco, CA, USA, pp. 4831–4835. [Google Scholar]

- Gerstner, Eric (2020, January). *Face/Off: "DeepFake" Face Swaps and Privacy Laws | IADC*. IADClaw.

https://www.iadclaw.org/defensecounseljournal/faceoff-deepfake-face-swaps-and-privacy-laws/

- Jiang F, Jiang Y, Zhi H*, et al* Artificial intelligence in healthcare: past, present and future*Stroke and Vascular Neurology* 2017;**2:** doi: 10.1136/svn-2017-000101 https://svn.bmj.com/content/2/4/230.full

- Lauret, Julien. (2019, August 16). *Amazon's sexist AI recruiting tool: how did it go so wrong?* Medium. https://becominghuman.ai/amazons-sexist-ai-recruiting-tool-how-did-it-go-so-wrong-e3d14816d98e

- K. (2019, August 27). *The 7 Most Pressing Ethical Issues in Artificial Intelligence*. Kambria. https://kambria.io/blog/the-7-most-pressing-ethical-issues-in-artificial-intelligence/

- Tonkens, R. (2009). A Challenge for Machine Ethics. *Minds and Machines: Journal for Artificial Intelligence, Philosophy, and Cognitive Science*, *19* (3), 421–438. https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=phl&AN=PHL2145643&site=ehost-live&scope=site&custid=ken1