Safe AI - Disaster Management

Sumanth Vullamparthi

Maharashtra Institute of Technology World Peace University, Pune

Abstract: The 'dotcom' bubble that started in the mid 1990's proved to be the start of the Internet era. The use of the Internet kept growing which led to the generation of enormous amounts of data. With the rapid expansion of cloud computing in the early 2000's enterprises could use this data and the high computational power to build smart systems or intelligent systems which we also call Artificially Intelligent systems (AI systems). This has proven to be of great strength to enterprises, as AI took over different sectors of the industry and has proved itself to be the most disruptive technology that humans ever created. But as the saying goes "With great power comes great responsibility!!!". Although AI has proven to be a very useful technology, there have been instances in the past where AI systems failed badly which caused sentiment harm, monetary harm and even harm to human life! So, with such great power at hand it becomes really important for us to create a "Safe AI ecosystem". In this paper we discuss a few case studies where AI poses danger to privacy, life and other aspects and has not performed as expected. We will also look into how organizations are keeping up with the ethical part of machine learning / AI development and research.

KEYWORDS: Artificial Intelligence, failures, humanity, data, Internet, Artificial general intelligence

INTRODUCTION

A knife can be used to cut fruits and enjoy their taste and on the other side it can also be used to cause physical harm. Considering this, we can draw a general conclusion that every product whether software product or hardware products, these are never neutral. How to use them lies in the hands of the user. So is the case with technologies. They are never neutral, its behaviour depends on how the developer who worked on it built it. Artificial Intelligence, a technology that promises to mimic the human brain is a technology that brings in a lot of potential to solve complex problems that currently humans spend time solving. In the upcoming sections we see the problems that this technology brings in and study some use cases where these issues have been looked into. The safety and security of data becomes very important in this age of technology. Even as it is said 'Data is the new gold', availability of data opens up many possibilities to build next generation products. Researchers come up with better ways of using the power of AI in the domains like health care, supply chain management, sports, EdTech and many more areas. With more research being done and new products being developed, it gets important to also look into the fall backs of this technology. Just as cyber security involves research in both cyber security and ethical hacking, AI research should also involve research in how AI systems cause problems and casualties. Therefore, in this paper we see some use cases in different domains and how the problems can be solved in these domains.

CASE STUDIES

1. AI in Healthcare

With the growing world population, it becomes highly important to address the health issues of a common man faster than it used to happen in the past. With increasing pollution, new viruses coming in, the illness rate is increasing every passing year. Artificial Intelligence has played a key role in accelerating the health care service in many ways the past few years. AI systems today help us diagnose cancer better with its effective computer vision techniques. Sickness and deadly illness could be predicted by analysing the behaviour of the human body. AI serves in the field of making medicines. It helps patients get a detailed report, predictions and recommendations based on their blood pressure reports, diabetes reports etc. These are few of the current use cases that AI serves health care. But, the scope is infinite and AI brings a lot of opportunities in this field to serve the community of patients and doctors better.

But, to reach this goal of extensive use of AI in healthcare, there are many security, ethical, trust, privacy barriers. In this case study we will look into how these barriers actually are real issues and how can enterprises overcome them. AI use cases in the healthcare domain need lots of data. This data belongs to patients and its their private data. There are many issues raised on data privacy, since this data could be used in negative ways as well. Hospitals and patients are not very comfortable in sharing their health records or data for technology advancement and research purposes. There should be campaigns to spread awareness about how AI can really help people and make them realise that these research studies can bring massive benefits in the healthcare systems which will help them and the generations to come. Gaining the trust of a common man and hospitals sharing their data becomes very important. Enterprises should make sure that their technologies are reliable and safe before they try to fetch data from hospitals or patients, because one mistake from any institute or enterprise will lead to a loss of trust in the whole AI research community. Hence, data privacy and security should be of utmost importance. Open standards should be promoted in order to create greater

incentives for data sharing between organizations and patients. There are a few ‘Data donation models’ currently existing.

Technologies like Blockchain can bring a lot of benefits in creating more secure Data donation models which would give more transparency and ownership to the owner of the data even if it is shared with organizations. Educating people on how these technologies work and how the organization is taking advantage of them to help the society becomes a key to gain the trust of people to get their data. Machine learning models have the tendency to generate a bias based on the data that is fed to it. It becomes very difficult to understand this bias before it has actually caused any issue in the market. Healthcare is not a domain where we can take the risk of disaster recovery, since every life is very precious and one wrong decision of ML models can put the organization into deep trouble. Hence, it becomes very important to understand the data and the biases in it before the models are deployed. These models should be ready to understand the ongoing changes and the increasing dangers so that they are updated with the surrounding conditions and give timely accurate solutions or predictions. Organizations should be ready for accountability. Accountability for the way they are using data of the patients and hospitals, accountability on why their AI products are working the way they are working. For this the developers of these AI products need to know the working of the model behind the “Black Box”. Hence, it is very important for developers to have a deep understanding of what they are building. Strict laws and regulations for data security and privacy, accountability, inclusiveness and transparency will help organizations expand the research in this field of AI and build more robust and effective AI systems.

2. AI in Digital Marketing

Artificial intelligence, a disruptive technology, like in any other field has itself deeply rooted in the domain of digital marketing. The opportunities it creates are tremendous as it helps advertisers create personalised and optimized ad campaigns. For example, electronic billboards can use live streaming data to publish more accurate and data based advertisements. With advanced advertising techniques like youtube ads, facebook ads taking the front seat, there are millions of consumers that are targeted on a daily basis. But, in reality many of us consume these ads without actually paying much attention to any of the content in these ads. Consider paying attention to the advertisement as a 'HIT' and otherwise as a 'MISS'. In the current situation there are many 'MISS' as compared to 'HIT'. This means that advertisers are losing a lot of money and resources on ad 'MISS'. AI helps them solve this problem by opening up doors to personalised ad optimization which tremendously brings down their expenditure in campaigns. Automated customer care is another field where AI has done great things. With the number of digital products increasing exponentially, there is a need to cater to the problems and issues faced by customers using these products. Chat bots, voice assistants do a great job of reducing the workload off the service center team.

These Services help customers on the 1st level and try their best to solve the customers' issues. Customer behaviour analysis is another sector that helps organizations to predict the customer's future behaviour and helps them design their product launch, prices, offers etc. With all the above and many more advantages of AI in this sector, there are many ethical issues that need to be kept at the highest priority while using them. One such ethical problem is job losses. For example, with AI taking the seat of customer service, many people lose their jobs who worked in customer service teams. Organizations should encourage their labourers to upskill and take up other roles rather than laying them off just because a technology like AI is capable of doing their tasks. Another obvious issue is the data privacy issue. Customers buying data, online data is being sold to these advertising companies which is not ethical at all. There should be more initiatives like the General Data Protection Regulation (GDPR) by the European Union that ensures complete protection of individual's data. Apple, a tech giant has started a great initiative

to give the powers of user's data to the user. They give an option to the user if they want to share their data for personalised ads or do, they want to keep it private. This way there is transparency and involvement of the user in the whole process. Brave Browser, another great product that promises that it does not track the activity of a user on the browser. Google chrome on the other hand tracks our behaviour and gives us ads based on it. Brave gives the authority to the user by blocking the ads by default. With Blockchain as a key component in the business model it ensures complete data privacy and immutability. Brave, an amazing ethical product, reasons this by saying 'Because what you do online is your business, not ours.' These are some of the ways organisations are taking initiatives to build ethical environments for digital marketing.

MOTIVATION

Instances where AI failed and ended up some bad/negative results

1. A robot made for grabbing automotive parts killed a man: 2015
2. Google Image Search returned racist results: 2016
3. Self driving cars indulged in deadly accidents: 2016
4. Face beautifying AI made black people look white: 2017

Need for research in Safe AI systems becomes necessary to predict and avoid the possible failures and disasters that AGI can bring in the near future.

CONCLUSION

Studying the above use cases, we conclude that there are many more problems that are going to rise in near future with respect to:

1. Data security and privacy needed for model training.

2. Trust in the model's behaviour. This trust includes trust of developers in the product, trust of government, and most importantly trust of the customers.

3. Predicting possible biases and creating all possible testing environments before deploying the model for public use.

4. Limiting the use of AI for constructive purposes and keeping this technology in safe hands. For this it becomes very important to educate the coming generation on ethical AI so that going ahead in future we are sure that this great technology is in safe hands.

There are technologies like blockchain coming up that are promising data security to the owners of the data. Combining AI with blockchain technology will help organizations to gain trust of data

owners, where there is transparency and inclusivity. Blockchain technology, in which data once stored can't be changed, will help a lot in the health care sector. Organizations should come up with better ways of gaining public trust by showing their initiatives to maintain data privacy. There should be intensive research in the field of how to monitor usage of AI. There should be cloud monitoring of AI service usage, so that destructive usage of AI can be stopped before it does high damage. We can also expect AI products that monitor this and optimize this field of monitoring constructive use of AI.

All these are some of the measures that could be taken for betterment of the ethical AI environment and we should hope for the best to happen in future even as this technology deep roots itself in almost every sector of industry.

REFERENCES

[1] Roman V. Yampolskiy,"Predicting future AI failures from historic examples", 18 October 2018.

[2] Rahaman,Mohammad Mushfequr Rahman ,"Dangers of Artificial Intelligence: 2020 - ResearchGate",(2020)

[3] Anderson, Susan & Anderson, Michael.,"The Consequences for Human Beings of Creating Ethical Robots",(2007).

[4] Tsamados, Andreas & Aggarwal, Nikita & Cowls, Josh & Morley, Jessica & Roberts, Huw & Taddeo, Mariarosaria & Floridi, Luciano.,"The ethics of algorithms: key problems and solutions.", (2021). AI & SOCIETY. 10.1007/s00146-021-01154-8.

[5] Josef Baker-Brunnbauer,"Management perspective of ethics in artificial intelligence ",16 November, 2020 - Springer

[6] Microsoft ,"Healthcare, artificial intelligence, data and ethics - A 2030 vision", December, 2018

[7] Michael Anderson, Susan Leigh Anderson - "Machine Ethics: Creating an Ethical Intelligent

Agent", December, 2007 – aaai

[8] K.Nair, R. Gupta, "Application of AI technology in modern digital marketing environment"

[9] The Past, Present and Future of AI in Marketing December, 2017, Hal Conick. [Online]. Available:https://www.ama.org/marketing-news/the-past-present-and-future- of-ai-in-marketing/. [Accessed Sept. 15, 2021]

[10] Michael J. Rigby, "Ethical Dimensions of Using Artificial Intelligence in Health Care", AMA Journal of Ethics, feb 2019. [Online]. Available: https://journalofethics.ama-assn.org/article/ethical-dimensions-using-artificial-intelligence-health-care/2019-02. [Accessed Sept. 12, 2021]

[11] Ethical and legal challenges of artificial intelligence-driven healthcare, Sara Gerke, Timo Minssen, and Glenn Cohen, Published online 2020 Jun 26, PMC. [Online].

[12] Amitai Etzioni & Oren Etzioni ," AI assisted ethics", 05 May 2016