MIT-WPU

॥ विश्वशान्तिर्ध्रुवं ध्रुवा ॥

Dr. Vishwanath Karad
MIT WORLD PEACE
UNIVERSITY | PUNE
TECHNOLOGY, RESEARCH, SOCIAL INNOVATION & PARTNERSHIPS

Seminar Report

On

Safe Ai -Disaster Management

By

Sumanth Vullamparthi

1032180622

PC30

Under the guidance of

Prof. Jayshree Aher

MIT-World Peace University (MIT-WPU)

Faculty of Engineering

School of Computer Engineering & Technology

2021-2022

MIT-WPU

Faculty of Engineering
and Technology

MIT-World Peace University (MIT-WPU)

Faculty of Engineering & Technology
School of Computer Engineering & Technology

CERTIFICATE

This is to certify that Mr. /Ms. Sumanth Vullamparthi of T.Y. B.Tech., School of Computer Engineering & Technology, Trimester – IX /X, PRN. No. 1032180622, has successfully completed seminar on

Safe AI – Disaster Management

To my satisfaction and submitted the same during the academic year 2021- 2022 towards the partial fulfillment of degree of Bachelor of Technology in School of Computer Engineering & Technology under Dr. Vishwanath Karad MIT-World Peace University, Pune.

Dr. Vrushali Kulkarni

____

Prof. Jayshree Aher

Head
School of Computer Engineering & Technology

TABLES AND ABBREVIATIONS

List Of Tables:

ABBREVIATIONS:

AI: Artificial Intelligence

ML : Machine learning

AGI : Artificial General Intelligence

ACKNOWLEDGEMENT

I would like to express my gratitude towards my Seminar Guide Prof. Jayshree Aher for her invaluable guidance and support throughout the Seminar. Her comments and motivation helped me a lot in completion of this Seminar work.

Finally, I would like to thank the School of Computer Engineering and Technology and MIT World Peace University for providing us with the platform which has led to many such work of good quality and value.

Sumanth Vullamparthi
PC30

INDEX

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

1. 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.

2. LITERATURE SURVEY

TITLE AND YEAR	PUBLICATION	AUTHORS	FINDINGS
Predicting future AI failures from historic examples: 2018	ResearchGate	Roman Yampolskiy	Cyber security has reached has come this far because of the parallel research in building more secure systems and advanced techniques of hacking systems. To make safe AI systems extensive research needs to be done in how AI systems can fail in coming future so as to avoid and prevent problems or major failures.
Dangers of Artificial Intelligence: 2020 - ResearchGate	ResearchGate	Mohammad Mushfequr Rahman	1. Presence of kill switch to abort the processes and restart with clean state in. 2. Ban aggressive use of AI in military use cases 3. Establish international monitoring system of AI accountability Hold developers and owners of the AI system for all the actions that it does.
Management perspective of ethics in artificial intelligence : 16 November, 2020 - Springer	Springer	Josef Baker Brunnbauer	1. Diversity in the developers working on the AI product// project can reduce bias 2. Management is more concentrating on how much revenue AI is generating and not on the ethical part of AI 3. AI ethical aspects should be taken care of by the manufacturers. 4. Enterprise values and AI values are 2 separate things 5. Responsibility & Data Security importance
The Consequences for Human Beings of Creating Ethical Robots - ResearchGate	ResearchGate	Susan Leigh Anderson and Michael Anderson	1. Machine Ethics - defines how machines behaviour towards human beings and other machines.

			2. Can machines change their behaviour after some time even if they are trained to behave in an ethical manner? Like we humans do 3. Machines that work in a dynamic environment where it can't be trained on all possible situations that it can go through 4. Regular updates and training on ethical basis is a must. (Treating a machine like a human infant who needs to be taught on a regular basis)
The ethics of algorithms: key problems and solutions : 2021	Springer	Andreas Tsamados, Nikita Aggarwal, Josh Cowls, Jessica Morley, Huw Roberts, Mariarosaria Taddeo & Luciano Floridi	1. Inconclusive evidence leading to results that can't be practically justified by human intervention can lead to ethical problems 2. Inscrutability of the results given by the model developed by developers themselves - Black box 3. Accountability should be a must where developers understand why the model is working the way it is working
The Consequences for Human Beings of Creating Ethical Robots	ResearchGate	Susan Leigh Anderson and Michael Anderson	1. EU and OECD - the AI governing bodies that passed rules accepted by about 42 countries. 2. AI startups have high funding so that they have enough resources to test their systems before deploying in the market 3. Courses and degrees on ethical AI have started in countries like US, UK etc so that the coming generation know how to use the technology they

			will get to work on in coming future

3. LITERATURE REVIEWS

Cyber security has reached great heights only because of the intensive research in the field of cyber security and ethical hacking. Both these work hand in hand where cyber security engineers try to build more secure systems and on the other hand ethical hackers try to find new ways to break into systems or hack systems. This helps the cyber security community to grow.

Drawing a parallel between cyber security and artificial intelligence, we can say that extensive research in failures of AI also becomes very important even as the casualties caused by AI systems increase each day. There have been many instances where AI has failed to perform the way it was intended to while building it. Stated below are some of the past instances where AI behaved weirdly.

1. Uber self-driving car fails to identify a woman with a bicycle and crashes into her leading to the death of the woman.
2. A beauty contest robot eliminates all dark people raising racism issues.
3. Robot that was designed to pick auto parts picked a man and killed him.
4. A housekeeper robot killed the house pet and cooked it.
5. Irrelevant and inappropriate ads shown to users.
6. Alexa playing adult content instead of songs

And the list goes on…

Keeping these instances in mind, we see that these AI systems that are made for serving good purposes ended up in bad situations.

There have been instances where AI has been used for negative reasons. For example, AI based drones used by terrorist groups to make targets and cause societal chaos, hackers using AI to track banking patterns and target people to loot their banks.

It is very important to consider these problems and organizations have come up with ways to deal with these problems.

1. Presence of process kill switch in AI based robots so as to avoid deadly cases.
2. Banning aggressive use of AI in the military so as to avoid unintended attacks caused due to decisions taken by AI.

3. Diversifying the developers working on an AI project so as to reduce bias in the model.
4. Rich understanding of data so as to avoid bias in the model and avoid garbage in garbage out issues.
5. Regular updates at the production level so that models adapt to the surrounding and make better decisions.

There are degree certifications that are being started in the field of AI with ethics so that people are educated on how to ethically use this great technology.

There are many startups that have brought in their products in the market. These startups should emphasize on testing of their products well before they are deployed. In order for intensive testing to be done these startups should be well funded so that they can carry out effective testing in all conditions. AI system developers should have a sense of responsibility and ownership towards the product. In order for this to happen developers should know the working behind the black box and know why the model is working the way it is working.

4. CASE STUDIES

4.1 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 health care systems which will help them and the generations to come. Gaining the trust of a common man and hospitals to share 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.

4.2 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.

5. 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 if almost every sector of industry.

6. 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]

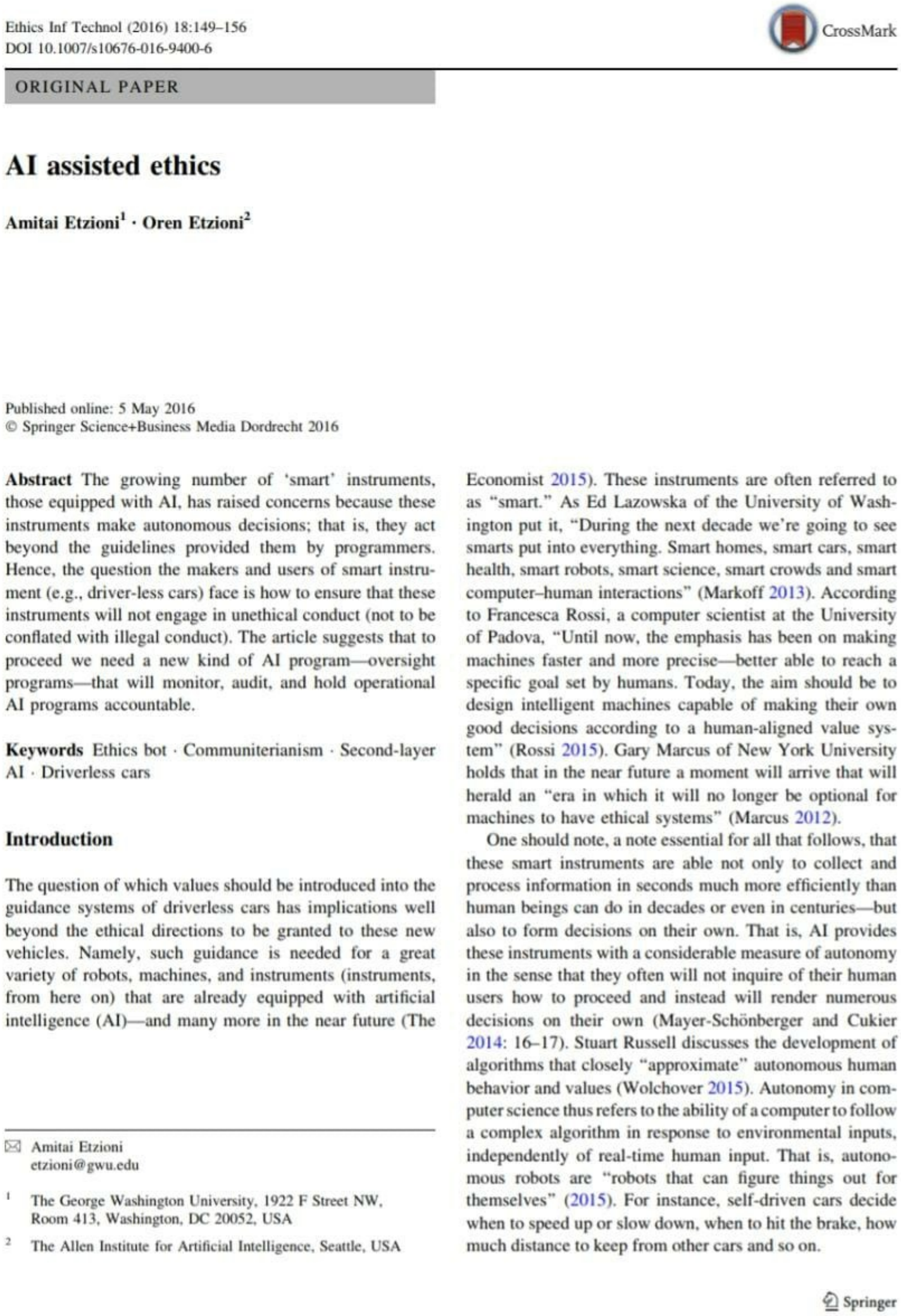
[10] Michael J. Rigby, "Ethical Dimensions of Using Artificial Intelligence in Health Care", AMA Journal of Ethics, feb 2019. [Online]. Available: https://journalofethics.ama-assn.org/article/ethical-dimensions-using-artificial-intelligence-health-care/2019-02. [Accessed Sept. 12, 2021]

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

Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7332220/. [Accessed Sept. 8, 2021]

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

BASE PAPER FIRST PAGE



Ethics Inf Technol (2016) 18:149–156
DOI 10.1007/s10676-016-9400-6

CrossMark

ORIGINAL PAPER

AI assisted ethics

Amitai Etzioni[1] · Oren Etzioni[2]

Published online: 5 May 2016

Abstract The growing number of ‘smart’ instruments, those equipped with AI, has raised concerns because these instruments make autonomous decisions; that is, they act beyond the guidelines provided them by programmers. Hence, the question the makers and users of smart instrument (e.g., driver-less cars) face is how to ensure that these instruments will not engage in unethical conduct (not to be conflated with illegal conduct). The article suggests that to proceed we need a new kind of AI program—oversight programs—that will monitor, audit, and hold operational AI programs accountable.

Keywords Ethics bot · Communiterianism · Second-layer AI · Driverless cars

Introduction

The question of which values should be introduced into the guidance systems of driverless cars has implications well beyond the ethical directions to be granted to these new vehicles. Namely, such guidance is needed for a great variety of robots, machines, and instruments (instruments, from here on) that are already equipped with artificial intelligence (AI)—and many more in the near future (The Economist 2015). These instruments are often referred to as “smart.” As Ed Lazowska of the University of Washington put it, “During the next decade we’re going to see smarts put into everything. Smart homes, smart cars, smart health, smart robots, smart science, smart crowds and smart computer–human interactions” (Markoff 2013). According to Francesca Rossi, a computer scientist at the University of Padova, “Until now, the emphasis has been on making machines faster and more precise—better able to reach a specific goal set by humans. Today, the aim should be to design intelligent machines capable of making their own good decisions according to a human-aligned value system” (Rossi 2015). Gary Marcus of New York University holds that in the near future a moment will arrive that will herald an “era in which it will no longer be optional for machines to have ethical systems” (Marcus 2012).

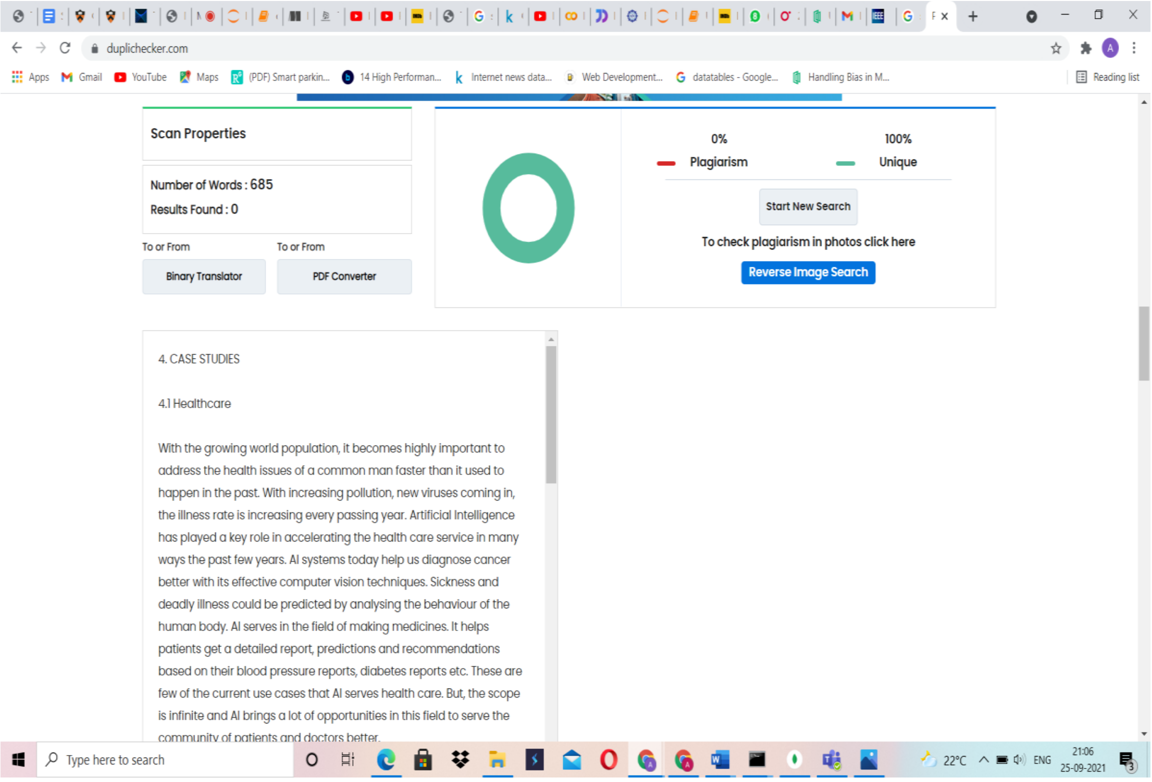
One should note, a note essential for all that follows, that these smart instruments are able not only to collect and process information in seconds much more efficiently than human beings can do in decades or even in centuries—but also to form decisions on their own. That is, AI provides these instruments with a considerable measure of autonomy in the sense that they often will not inquire of their human users how to proceed and instead will render numerous decisions on their own (Mayer-Schönberger and Cukier 2014: 16–17). Stuart Russell discusses the development of algorithms that closely “approximate” autonomous human behavior and values (Wolchover 2015). Autonomy in computer science thus refers to the ability of a computer to follow a complex algorithm in response to environmental inputs, independently of real-time human input. That is, autonomous robots are “robots that can figure things out for themselves” (2015). For instance, self-driven cars decide when to speed up or slow down, when to hit the brake, how much distance to keep from other cars and so on.

✉ Amitai Etzioni
etzioni@gwu.edu

[1] The George Washington University, 1922 F Street NW, Room 413, Washington, DC 20052, USA

[2] The Allen Institute for Artificial Intelligence, Seattle, USA

Plagiarism Check Report



Scan Properties

Number of Words : 685

Results Found : 0

0% Plagiarism

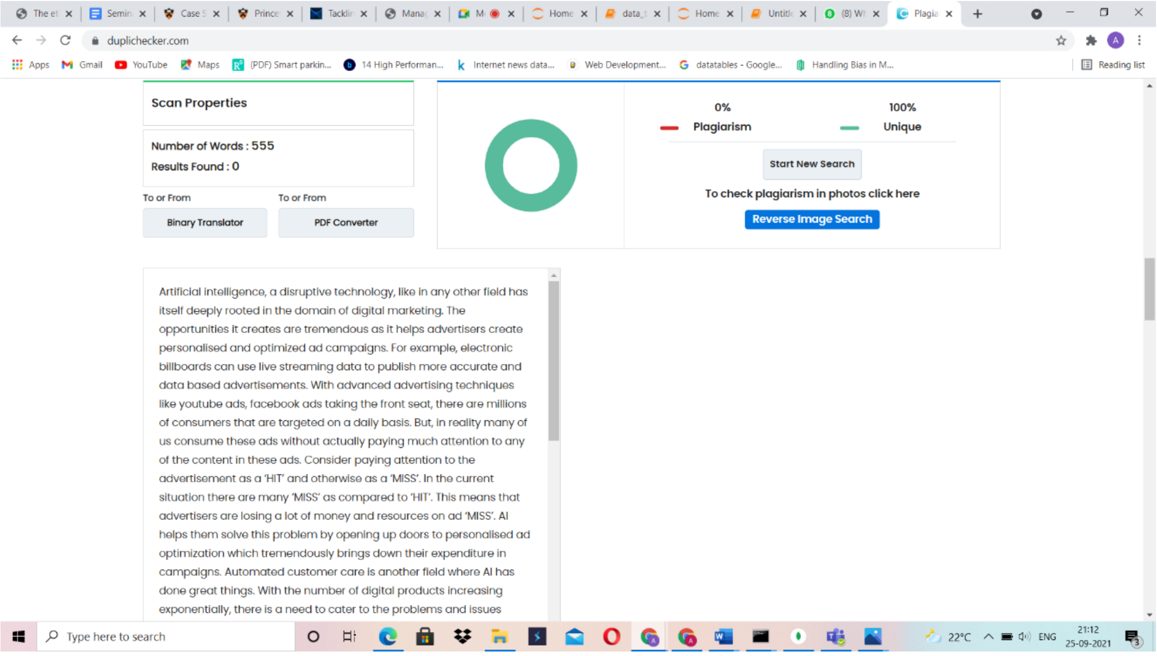
100% Unique

To check plagiarism in photos click here

4. CASE STUDIES

4.1 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.



Scan Properties

Number of Words : 555

Results Found : 0

0% Plagiarism

100% Unique

To check plagiarism in photos click here

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