EG2401 Final Report

Faculty of Engineering

Aldo
A0184544Y

Dinesh
A0184499H

Hokiewan Thezeus
A0184600M

Ritwik Jha
A0168818M

Under the guidance of:
A/P Lee Tong Heng
Professor at Faculty of Engineering

Table of Contents

1) Introduction to Artificial Intelligence

Artificial Intelligence (AI) is a system that mimics humans' intelligence in performing actions. It aims to replicate human intelligence, creativity, learning capability, and decision-making. AI is present in virtually every industry especially but not limited to healthcare, military, internet-based services and transportation.

2) Report outline

This report analyses the ethical dilemmas pertaining to the different types of Artificial intelligence discussed, exploring in detail their uses and applications. Solving these dilemmas will be of importance in critical decision making and safety of future generations. Hence, the report will explain the specific ethical dilemmas for each case as well as the current solutions, if any, and proposed solutions to these dilemmas which will be aided by line diagrams and ethical flow charts.

3) Criteria for evaluation

According to the Ethics Guidelines for trustworthy AI report by European High-Level Expert Group on AI, trustworthy AI should have three main components. They should be lawful, ethical and robust [3]. Using these three criteria as a guideline, for the purposes of this report, the following criteria will be used to evaluate each of the uses of AI presented: Privacy, Discrimination, Legal Liability and Safety and Reliability.

4) Case Study 1: Clearview AI

There are three broad areas in which AI is used on the internet. These are as follows:

1. Online Retail
2. Social Media and online content services
3. Cyber Security

In online shopping, AI is used to customize shopper's experience and tailor their search results towards items which they are more likely to be interested in. Online retailers use customers' search history to recommend products which customers have shown interest in. Examples of this include IBM Watson Assistant, Microsoft Cognitive Services and Google AI services [4]. Moreover, AI is also used by online retail to give region and country specific experiences to their customers. It also enables retailers to execute dynamic pricing. Dynamic

pricing has allowed e-commerce websites to achieve objectives such as increasing market share, increasing sales and revenue and taking advantage of trends in demand and consumption [5].

Social media and online content services such as Netflix and Spotify make use of AI to offer content (such as movies and music) recommendations, direct users to the social media trends which they are likely to be interested in--based on their past activity--and allow online marketers to have a more widespread reach to their target audiences [6].

With regards to cybersecurity, 61% of enterprises reported in a survey that they use AI to detect breach attempts [7]. AI has also shown potential in creating secured data channels through using neural networks for encryption and decryption, as a study by Google found [8]. Lastly, AI can be used by law enforcement and intelligence agencies to track suspicious online activity of individuals, using tools like facial recognition. Such tools can be used for speedily tracking down criminals on the run, or for predictive policing. This has so far been done by companies like Clearview AI and Banjo [9].

A. Case Summary

This case study is ongoing at the time of writing the report. The facts of the case thus far are as follows. Clearview AI is a US-based privately-owned company which provides facial recognition services to law-enforcement agencies [10]. Clearview has built a database by scraping photos of individuals from social media websites, including Facebook, Youtube and Twitter [10]. Facebook and Twitter spokespersons have said that such scraping is illegal and violates their user agreements, and they will take appropriate actions if they find that Clearview has violated their terms [11]. The first use of this technology was in February 2017, when Clearview AI was used to track down a shooter by the Indiana State Police [10]. Reportedly, Clearview's expanding clientele now includes FBI, Department of Homeland Security and Canadian law enforcement authorities [10].

In March 2020, Vermont Attorney General TJ Donovan filed a lawsuit against Clearview AI for violating the Vermont Consumer Data Protection Act and the Data Broker Law. The Attorney General Donovan asked the court to order Clearview to stop scraping websites for photographs [12]. The case is currently ongoing.

B. Privacy

Use of AI can have significant implications on privacy. AI inherently relies on the use of data to 'train' itself to become better at the task it is designed to execute. In Clearview's case, a facial recognition algorithm which is based on a neural network will become more and more accurate if it has access to a greater amount of facial data. Companies like Clearview AI need to source for a large amount of data to 'train' their software. How companies obtain this data determines whether there are privacy concerns regarding the data. Clearview in particular, was used a practice known as 'scraping'--which involves automated indexing of public search engines and social media websites for publicly available photographs uploaded by individuals or organizations. Clearview's CEO claims that since this data was publicly available, Clearview is only providing a catalogue of sorts by accumulating it in one search engine "in much the same way as Google's search engine" [11].

Google, however, explicitly states that "Google's Terms of Service do not allow the sending of automated queries of any sort to our system without express permission in advance from Google" [13]. Search Engine Optimization softwares regularly makes use of automated indexing to compile the most searched keywords and allow their customers' websites to maintain a competitive edge in terms of web traffic [14]. Most SEO's, however, work in collaboration with Google and hence are not voiding the terms of use or invading the privacy of individual entities. Clearview's practice of scraping on the other hand, reportedly did not come to the attention of major social media and search engine sites until late 2019, and in early 2020, Google, Youtube, Facebook and Twitter sent cease and desist letters to Clearview [15]. The main issue with Clearview's practice of scraping is that often, individuals are unaware of how to protect personal information and photographs online from being scraped. On Facebook, there is an option for users to disable their information from being available on public search engines (such as Google or Bing). Many users, however, would be unaware of this feature, and would not disable access, leading to their photographs on Facebook being publicly indexed. Even if individuals do find out about this feature, if they only disable access after their data has been scraped, it is too late, as Clearview does not take down photographs once they have been scraped, even if they are deleted at a later time from their original source [10]. Though Clearview has claimed that it is working on a feature which

allows individuals to request deletion of ‘scraped’ data, there is a degree of loss of control associated with Clearview’s method. Many users will not be aware that their data is being ‘scraped’ from their social media accounts in the first place. Moreover, Clearview may also index photographs of individuals posted by other friends, organizations or media. Access to these photographs is not controlled by the individuals themselves, causing further loss of control over privacy and anonymity.

C. Discrimination

The loss of privacy and indexing of sensitive photographs, besides their inherent invasion of individuals’ anonymity, may be used to create a profile of a person by governments or private organizations, which could be used for discriminatory practices. This has occurred before in China, where facial recognition is part of a larger overall system which assigns a ‘credit-score’ to individuals [16]. While officially, the system aims to ‘encourage socially responsible behaviour’, there are concerns that it could be used by the Chinese government to quell free speech or silence dissent against the government. While services such as Clearview do not assign a ‘credit-score’ which could be used for discrimination, it is possible that Clearview’s database can store pictures of events--political or otherwise--which a person has attended in the past to be used as evidence of political or organizational affiliations. Such a system--if misused--could severely compromise freedom of speech and individual liberty.

D. Reliability

AI software, while advanced, is not above malfunctioning. One example is Microsoft’s Tay AI. Tay AI was a chatbot launched on Twitter by Microsoft. It learned to tweet messages by learning from other Twitter users’ tweets [17]. Within 16 hours of its release, it had to be taken down, because it was flagged by many users for posting offensive messages and using vulgar language [17]. Hence, it is possible for an AI to ‘learn’ undesirable things, and act in a manner not originally intended. Similar dilemmas exist with regard to Clearview’s software. Clearview boasts a 99.6% [18] success rate for recognition, but if used on millions of individuals, it may misclassify a significant number of individuals, which could waste resources of law enforcement, or worse, lead to wrongful convictions.

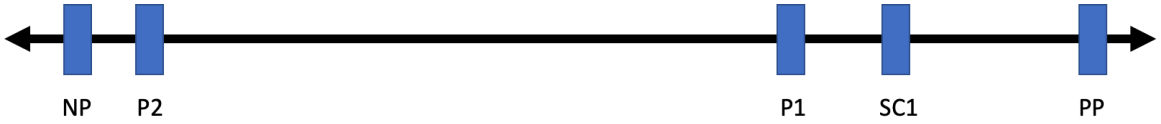
While it presently does not have any filtering features, given that it is used by law enforcement, if Clearview were to filter suspicious or potentially dangerous individuals from its database, there is no guarantee that this would be done in a manner which is completely fair. Its results might end up discriminating against individuals from certain locations, socio-economic classes, or races.

Hence, it is important to ensure that the reliability of an AI system is tested fully and well understood, before it is implemented for official use by law enforcement or government.

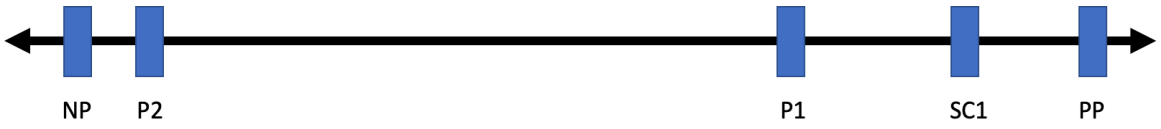
E. Ethical Analysis: Ethics Table

Action/Choice	Party 1 - Law Enforcement Agencies	Party 2
	Ethics Category - Virtue Ethics	
Allow development of Clearview AI	Virtue ethics is satisfied as Clearview enables law enforcement agencies to identify suspects much faster than would otherwise be possible. One example is when a detective in Clifton, New Jersey was able to identify a suspect in a 'matter of seconds' [10]. This allows law enforcement agencies to solve crimes quicker, saving resources and allowing public money to be channelled into other useful causes. Virtue ethics is, however, violated because law enforcement agencies have to upload sensitive photographs to Clearview's database, and the security of this database is at present 'untested.'	Virtue ethics is satisfied, as accor is using only publicly available ir that the information is protected which guarantees the right to publ Virtue ethics is not satisfied, b composed of a database which wa websites. Most of these websites practice [11]. Both Google an informing the company of the viol
Allow only Law Enforcement agencies to use Clearview	Virtue ethics is not immediately violated, but there is potential for abuse of Clearview even by law enforcers using it for legal purposes. One possible example of abuse was discussed on CBS News, which was the possibility of profiling using old images which individuals themselves were not aware of. The specific example discussed involved a hypothetical image of a person taken at a rally. The persons' attendance at the rally could be falsely used to (wrongly) deduce their political leanings or build a personal profile. Such potential uses--though within the ambit of legal use--could still compromise individuals' rights and privacy [11]. Virtue ethics is also violated because the database which Clearview has built was done by violating the terms of use of the source websites. Hence, as journalist Nicholas Thomson pointed out, this database 'may be legal, but goes against the terms of service of websites.' This constitutes a legal 'grey zone' which authorities have not fully decided upon, and as such, New Jersey's Attorney General has banned the use of Clearview for the interim [11]	Better choice of customer Legal liability is mitigated becaus Virtue ethics is satisfied. Given th authorities, limiting the use of C some extent as the use can be c invasion of individual privacy can Known permitted use of softwar bodies (NBL etc) Virtue ethics is violated, howeve that individuals within the law e technology. Moreover, in Februar by a flaw in their security systen included non-law enforcement Walmart, and the NBA [11]. Hen to law enforcement has been cont in New Jersey and Vermont.

F. Ethical Analysis: Line Diagram



Point	Ethics Line Drawing from POV of Party 1 - Clearview AI	Location from Left
NP	Clearview sells their technology to anyone who is willing to pay the asking price. No regard for privacy or anonymity of individuals.	0
PP	Clearview pivots into a non-profit, semi-government organization, adhering to strict guidelines regarding the use of facial recognition and AI. Only government, law-enforcement and government-approved use of Clearview AI is permitted. Strictly no paid use by private individuals or entities permitted.	10
P1	Clearview self-regulates and only makes itself available to law-enforcement agencies and governmental organizations. Use of Clearview is restricted to purposes of law enforcement and crime-prevention [10].	7
P2	Clearview makes itself available to anyone with the necessary connections and influence, regardless of whether they are private individuals or government organizations [11]. Ambiguity due to absence of legislation regulating use of facial recognition software is used as a grey zone to sell Clearview's capabilities to private individuals or corporate entities [11]. Ultimate implications of the services of Clearview AI is not factored into the decision-making process.	1
SC1	Clearview works closely with the government in formulating legislative framework for governing the use of Clearview and other AI in law enforcement and intelligence gathering uses.	8



Point	Ethics Line Drawing from POV of Party 2 - Government	Location from Left
NP	Introduce legislation allowing Clearview to sell their services to anyone who is willing to pay the asking price. No protection for privacy or anonymity of individuals.	0
PP	Law enforcement is able to function completely without the use of Clearview AI. There is no need to infringe upon individual privacy and anonymity.	10
P1	Law enforcement agencies use Clearview to speedily track criminals and prevent danger to the lives of innocent people [12]. No law in place which mandates Clearview to protect privacy.	7

P2	Government is aware of Clearview's violation of online platforms' terms of use, and of Clearview selling its services to private companies and individuals given the right price. Law enforcement continue to employ the services of Clearview, despite knowledge of Clearview's infringements. No legislation put forth to regulate the use of Clearview's technology.	1
SC1	Government legislates a takeover of Clearview. Clearview functions as a non-profit branch of the government, with strict guidelines regulating its use to only situations where there is grave threat to life. Individuals have the option of seeing which data is present in the database and the option to request for removal of sensitive information.	9

G. Conclusion

Clearview has much potential to be used as a force for good. At present, however, its use is unregulated, and its implication depend upon the integrity of the operators and creators. In order to ensure that individual privacy is not infringed, and prevent malicious use of this software, legal frameworks must be devised in order for software like Clearview to be put to good use. In the absence of a framework which holds either the operators or the creators accountable for the implications of the software—such as erroneous identification of a non-criminal as a criminal, or malicious use which compromises privacy of an individual—it is very difficult to ensure that software services like Clearview will be used only for the betterment of society.

5) Defence Case: DoDAAM's Automated Turrets

In military operations, humans' lives are put on the line (e.g. soldiers). Thus, some would suggest alternatives with less risk on humans' lives such as deploying drones in the military [h0]. Even further, some would suggest the use of AI for decision-making to avoid discrimination and ensure fast, accurate and consistent response to crisis. Therefore, anticipation of a future with robots dominating the military, an AI revolution in warfare and evaluation of their ethical impacts are in order.

A. Case Summary

In this section, a case study on the ethicality of AI's technology application in defence and security will be discussed. The case study revolves around automated turrets produced by DoDAAM, a defence and space company based in South Korea. Its products include

intelligent combat robots, the aEgis series. Super aEgis II, with its ability to automatically identify, track, and eliminate moving targets from great distance (and theoretically) without the need of human intervention, has been deployed to guard the border between South and North Korea. Furthermore, Super aEgis II has also been deployed to guard military air bases, power plants, airports in numerous countries around the world (e.g. in the Middle East) [h4]. According to Jungsung Park, a senior research engineer for DoDAAM, the only reason why there is still human behind Super aEgis II's trigger is due to DoDAAM clients' request for safeguards, otherwise as initially designed, it would auto-fire [h4]. To unlock the turret's firing ability, the human operators must first enter a password and then manually instruct the turret to shoot.

Some would suggest that advanced defence systems will discourage war from breaking out and promote peace in the process. Some activists support this idea and DoDAAM might have shared the same vision [h8]. Peace is a part of society's well-being. Therefore, besides fulfilling their moral duty in creating a more peaceful world, DoDAAM also complies with act-utilitarianism in the process.

IHL states that advance warning should always be delivered whenever possible before the shooting starts [h11]. DoDAAM makes sure that their turrets always instruct potential targets to turn back and leave their perimeter before they are shot. This is a precautionary measure to minimize loss of lives whenever possible and adhere to human rights to live as much as possible. However, even if such a warning is implemented, full autonomous lethal weapons still invite many criticisms [h12].

To analyse the ethicality of the case around DoDAAM's automated turrets, line diagrams and ethics tables will be used (only virtue and duty ethics will be analysed due to limited space).

B. Privacy and Discrimination

Application of AI in defence includes preventive action through surveillance such as *predictive policing* [h1]. However, as the programming and data preparation are done by humans, the resulting AI may contain implicit bias against group(s) of people. Furthermore,

cyber-attacks to AI systems (e.g. facial recognition) belonging to the officials might reveal personal data.

Gary Marcus, a cognitive scientist, stated that it is not feasible to pre-program all of the relevant (possibly ambiguous) ethics in developing an AI [h4]. Therefore, some might suggest letting the AI learn from watching how humans behave. The downside of this idea, according to a senior researcher at Oxford Martin School, is the possibility of the AI learning from a bad environment or starting to develop 'alien' behaviour. Bias in the data given might result in discriminative action such as persecution. A sophisticated AI might also be able to predict human behaviour to a certain accuracy where the distinguishing line between highly accurate predictive monitoring and privacy breaching becomes blurred.

In the future where the aEgis series are deployed to maintain order in public space, privacy and discrimination will become a much more relevant question than today's application.

C. Safety and Reliability

In terms of safety and reliability, AI (robots) is a force multiplier that increases advantage in numbers [h2]. By replacing humans on the frontline, the risk of losing lives will be minimized. South Korea shares their border with North Korea and the situation is very tense in the demilitarized zone (DMZ) [h5]. There were approximately 1000 'fracases' and around 50 serious incidents occurred in the DMZ ever since 1953. Conflict may break out anytime and soldiers are getting weird out from the long and boring guard-duty. Autonomous weapons are not affected by emotion, able to work consistently without rest, and are expendable in dealing with dangerous tasks [h2]. However, a group of academically experts in AI also press the fact that there is little scientific evidence that fully autonomous robots could, in the future, have sufficient accuracy in target identification, situational awareness, and decisions regarding appropriate use of force [h2]. This might lead to more casualties and thus decision-making should not be delegated to machines. This concern is why landmines are banned by the Ottawa treaty in 1977 [h2]. Notice that they were considered simple autonomous weapons that explode anyone who stepped on them. Instead, at the very least, *human-on-the-loop* autonomy (i.e. allowing humans to override the systems) can be implemented to balance human's judgement and AI's efficiency in decision making [h2]. It is

also recommended to obtain international agreement that ensures every fully autonomous weapon must have a function to be recalled once they are deployed to prevent more casualties when the AI (is about to) violates international war laws/treaties [h2].

D. Legal Liability

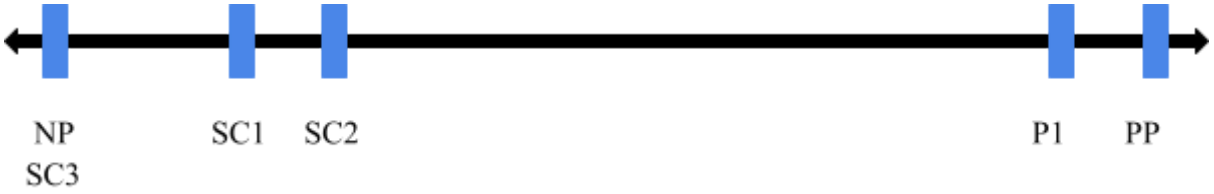
AI has no sense of ethics; it is also not exactly a citizen that has legal liability. Since AI has the capability to make their own decisions, the developers cannot be fully liable for these decisions. Currently there does not exist a law that can put AI developers liable if they cannot foresee the incidents or harm done [h3]. With clear guidelines on liability, only properly trained and tested AI/robots will ever be launched to the public and this minimizes risk of harm towards civilians. Therefore, a solution to this uncertainty regarding legal liability should start from researching laws that can ensure legal liability. A standardized quality standard should be established as a reliable quality control.

In the case of DoDAAM's turrets, IHL states that force may only be applied to lawful combatants (i.e. it is illegal to harm civilians) [h6]. As of now, the aEgis series cannot differentiate between civilians and military, moreover between allies and enemies [h4]. In order to ensure compliance with the IHL, the military should only allow fully autonomous application when DoDAAM has perfected their civilian-military-distinguishing feature.

E. Ethics Analysis: Ethics Table

Action/Choice	Party 1 - DoDAAM	Party 2 - Cust
	Ethics Category - Duty Ethics	
DoDAAM aims to make their automated turrets 'smarter', more autonomous, and receive less intervention from humans.	**DoDAAM aims to reduce the needs of putting humans on the front lines.** Less number of personnel translates to less risk of loss of lives that aligns with virtue ethics. From the perspective of preserving lives, DoDAAM has satisfied the virtue ethics and also its duty as an able party. **DoDAAM prefers the turrets to be fully autonomous and more reliable** DoDAAM wishes to serve its purpose as one of companies that specializes in developing defence systems. From the perspective of DoDAAM's management and engineers, not pursuing development is an act of duty negligence and thus this goal is aligned with their duty-ethics. In fact, DoDAAM should not only develop their turrets to be fully autonomous, but also 'smarter' as to avoid unexpected outcomes that may violate humans' rights or even violate rule-utilitarianism (misclassification of civilians as military). **DoDAAM needs to ensure competitiveness in the market** DoDAAM satisfies their duty by pursuing such developments in their products. After all, DoDAAM is a profit-driven company. Maintaining their competitiveness in the market is paramount as a part of company sustainability. By becoming the lead in the industry, DoDAAM has been satisfying their duty marvellously. **DoDAAM may prevent wars** As mentioned earlier, DoDAAM is fulfilling its moral duty in creating a more peaceful world through advanced defence systems.	**There is never enough precaution to be tak** It is within their duty-ethics to prefer the sure **Reducing exposure of soldiers to life-threat** Replacing the soldiers with autonomous turre to the soldiers' rights to live while also fulfilli **DoDAAM's turrets cannot differentiate bet** With the current limitation, granting full military's duty-ethics. However, it may be future. **Limiting the number of human soldiers on** Machines are limited to operate only within cases that require human judgement at the sce of humans, the military may fail to respond a in 2017, a defector from North Korea fled multiple times and saved by South Korean so might not have reached South Korea alive th watch from its station. From this perspective autonomous turrets to guard the border is ar thus may lead to the violation of duty-ethics.
Customers ask for safeguards to be implemented. Such as having humans on/in the loop. (i.e. humans may intervene or even decide the turrets' decision)	**DoDAAM complies with their customers' request** Jungsuk Park, a senior research engineer for DoDAAM, claimed that initially they planned the aEgis series to be fully autonomous weapons with little-to-none intervention from humans [h4]. However, in order to comply with their customers' request, they modify their products to meet the customers' specifications; the involvement of human operators. By compromising their original design of full autonomous defence systems, the engineers and management board of DoDAAM has satisfied their duty in maintaining the customers' trust and ensuring continued revenue source to sustain the company. **DoDAAM makes sure their products adhere to international law** As mentioned earlier, DoDAAM designed their turrets to adhere to the IHL and customers' requests. This fulfils their duty to be a trustful agent for their customers. If DoDAAM still wishes to include full autonomous function in their turrets, they should consider the option to pair the full autonomous function with non-lethal paralyzing weapons. Therefore, even if the turrets were to auto fire, there would be less casualties and the adversaries could be secured for further processing (i.e. judgement by humans).	**Involving humans in decision making relat** The turrets may one day hurt non-combatants ex-army ranger, shared an experience where experience when he met a spy which was ju even though it was allowed by law. Then differently. The decision to implement safegu IHL (rule-utilitarianism) and the military's du **Shoot first, ask questions later** There is a tendency for human soldiers to pr stressful situation. This weakness is eliminat self-perseverance instinct [h2]. This advant objectively. **Risk of the turrets malfunctioning** There exists a risk of cyber-attack that may In such cases, the presence of human opera conditions and deliver fast response. At th systems will one day break either due to bug according to Murphy's laws [h10]. Thus, a s will always function properly in defending the

F. Ethics Analysis: Line Diagram



Point	Ethics Line Drawing from POV of Party 1 - DoDAAM	Location from Left
NP	DoDAAM develops autonomous turrets that fire anyone within its range at will. Humans cannot interfere with its action, thus granting AI full authority to eliminate lives.	0
PP	DoDAAM develops autonomous turrets that may use non-lethal force to paralyze at will and only use lethal force where human operators instruct so. The moment the turrets notice the possible targets and before shooting, the turrets will deliver advance warning to the possible targets.	10
P1	DoDAAM installs a safeguard in aEgis series such that fires are only shot when human operators allow it (i.e. humans on the loop). The moment the turrets notice the possible targets and before shooting, the turrets will deliver advance warning to the possible targets.	9
SC1	DoDAAM develops autonomous turrets that fire anyone within its range at will. Humans can stop its action, but the turrets may shoot without the consent of the operators given the operators do not instruct otherwise after waiting for a certain length of time. However, the moment the turrets notice the possible targets and before shooting, the turrets will deliver advance warning to the possible targets.	2
SC2	DoDAAM develops autonomous turrets that can distinguish between non-combatants and lawful combatants. It will shoot lawful combatants within its range at will. Humans can stop its action, but the turrets may shoot without the consent of the operators given the operators do not instruct otherwise after waiting for a certain length of time. However, the moment the turrets notice the possible targets and before shooting, the turrets will deliver advance warning to the possible targets.	3
SC3	DoDAAM develops autonomous turrets that may use force to cripple (i.e. permanent damage) targets at will and only use lethal force where human operators instruct so. The turrets will never deliver advance warning to the possible targets. This clearly violates the IHL by dealing superfluous injury and lack of advance warning [h11] [h14].	0

Point	Ethics Line Drawing from POV of Party 2 - Customers (The Military)	Location

		from Left
NP	The military does not consider the ethical implications of installing full autonomous turrets from DoDAAM that do not require human involvement in decision making. They fully replace all soldiers with stationary aEgis series and no operators are assigned to intervene the turrets before they shoot the targets.	0
PP	The military makes sure that every shot done by the aEgis series will only result from human judgement of the situation, thus the requirement of safeguards. The aEgis series only identifies and tracks potential adversaries and human operators control the turrets from a safe distance. Besides that, the military also invest in the development of more reliable and ethical autonomous defence systems and possibly promote the use of non-lethal force for fast response in critical situations. The military also anticipates unexpected situations by placing soldiers around the border so that they can be deployed to attend to issues when necessary.	10
P1	The military makes sure that every shot done by the aEgis series will only result from human judgement of the situation, thus the requirement of safeguards. The aEgis series only identifies and tracks potential adversaries and human operators control the turrets from a safe distance. The military also anticipates unexpected situations by placing soldiers around the border so that they can be deployed to attend to issues when necessary.	9
SC1	The military requires the aEgis series they install around the border to fire at will given that after a certain period of time, no operators intervene with its decision. However, it is still possible for humans to intervene with the turrets' decision.	1
SC2	The military only installs surveillance systems to identify and track potential adversaries. All lethal weapons are not connected with DoDAAM's surveillance systems, thus requiring human operators to anticipate notice/warning from the surveillance systems and operating their weapons manually.	7
SC3	The military does not use any implementation of AI in their defence systems. All surveillance and lethal weapons use are done manually by humans. Thus, requiring more soldiers to be posted in response to higher workload.	3
SC4	The military monitors the DMZ and operates the aEgis series from a safe distance. No soldiers are stationed around the border since they do not anticipate any situation that would require human soldiers on the site. In other words, only these autonomous defence systems are physically present to guard the border.	2

G. Other ethical concerns

Paul Scharre and Robert Sparrow stated that in order to respect humans' dignity, the decisions about killing humans should only come from other humans, not machines [h13]. AI/robots do not have empathy, they lack mercy and thus ethical point of view. The European Union and a number of states in the U.S. are against letting machines judge legal penalties on a person, instead other humans should be involved in the process of decision making [h13]. If machines are not even allowed to subject persons to legal penalties, then it should be worse to let machines deem whether the persons are deserving death. Today, DoDAAM's turrets are

mostly deployed to guard borders. But it is not unimaginable that soon, autonomous defence systems might be applied for general use in guarding facilities, guarding governmental offices, or even deployed around the streets. When those days come, people will have to bear living while being in the range of fire. When it is DoDAAM's morale duty to enhance national security, it is also their duty to promote that sense of security.

H. Conclusion

The defence industry has much potential with regards to AI implementation both as force multipliers and as decision making tools. However, several promising implementations such as full automated military robots are surrounded with issues involving privacy, discrimination, safety and reliability, and legal liability. The current use of AI in defence is limited to only surveillance and recommendation making at most, while controversial decision making (e.g. using force to harm targets) is still done by human operators. This defensive approach to AI implementation is to avoid unethical outcomes in such a sensitive and vital industry. Many experts (e.g. Stephen Hawking, Elon Musk, Stuart Russell) have signed an open letter that features a potential crisis that stems from the development of AI technology in the military such as its penetration to the black market that can be used by terrorists, warlords or dictators to either control the populace, undermining nations, or persecuting a particular ethnic group. Therefore, the world should anticipate this change and respond accordingly.

6) Case study 3: Autonomous vehicles of Uber

While AI technology is still in the embryonic phase, billions of dollars is spent on research and development, accelerating AI advancements. IDC forecasts spending in Artificial intelligence technology "will increase by more than 50 percent" every year and reach around $58 billion in investments by the next year [1]. AI disrupts several industries and its impact could be felt across the transport sector. It is estimated that the AI transportation market alone will be valued at $3.5 billion by 2023 [2]. Leveraging AI in transportation improves the transportation sector safety, traffic conditions, financial expenses, productivity, employee performance and sales.

According to WinterGreen Research, by 2023, around 90 million autonomous vehicles will be on the road around the world [3]. Google and Uber have been working on the development of autonomous vehicles for public use. Implementation of AI in transport of goods in the US is expected to reduce costs by 45% as 65% of goods are transported via trucks by reduced stoppage time and delays [4]. On the other side of the world, autonomous taxis have already been implemented in Tokyo which have led to reduction in costs for services and increased accessibility to remote areas [5].

One of the key benefits of having AI technology in vehicles is the reduction of traffic accidents as well as streamlined traffic flow and reduced congestion. According to the National Safety Council, in 2017 alone there were 40,000 traffic fatalities in the U.S., with more than 90 percent of them caused by human error [6]. In 2018, traffic congestions alone cost $87 billion to Americans [7].

A. Case Summary

Uber is a US-based privately-owned company that provides transportation services which includes the new wave of autonomous vehicles. There was an incident when Uber was testing out the new self-driving capabilities of an autonomous vehicle which led to a fatality. In 2018, a lawsuit was filed against Uber for the death of Elaine Herzberg. This case has been concluded in 2019, where Uber was found not to be criminally liable by a court. Yavapai County Attorney Sheila Sullivan Polk has stated there is “no basis for criminal liability” to Uber [8]. However, the car's back up driver could still face charges due to negligence as they were watching Tv. Elaine Herzberg was crossing with her bicycle when she was hit by an Uber test vehicle. It was the first recorded case of pedestrian death due to an autonomous self-driving vehicle. Uber was not found guilty and a settlement was made to the family of Elaine for an undisclosed fee [9]. Governor Ducey authorised an “autonomous program with limited expert oversight”. It was reported that he allegedly allowed “immature” technology on to the streets and influenced the case incident conclusions [10]. As well, the public was not informed; the tests only came public knowledge long after the fact.

B. Legal liability

Legal liability is of major concern and has serious implications when it comes to AI. It has been integrated into vehicles more so than ever before and now cars can drive themselves. But this is not perfect technology and AI has made mistakes that led to casualties in the past which begs the question of where liability should be placed. In the Uber self-driving car case, Uber was not found to be criminally guilty for the fatality [8]. But it must be taken into consideration, Uber officials disabled emergency braking manoeuvres and left the control to the driver in the situation of an emergency before the test drive and eventual fatality occurred later on that day. Additionally, the radar and cameras picked up Elaine Herzberg six seconds before the collision but, according to a report from the National Transportation Safety Board, couldn't determine if she was a pedestrian [11]. There has been an outcry by the public for Uber to take responsibility for testing 'immature technology' on the public and disabling emergency braking systems of the Volvo which could have saved her life [10], [11].

Elaine Herzberg, however, was jaywalking and was not using the proper crosswalk about 100m north ahead. Signs are posted prohibiting pedestrian crossing from that area [11]. There is a case to be made that Elaine was at the wrong since she did not abide by road safety and the traffic law. It did not help the situation as the safety driver, in the Uber vehicle, was also distracted and if they had not been, it might have prevented the death of the pedestrian. The driver was watching a TV show instead of being vigilant and focusing on the road. The investigation concluded that the driver had a part to play in this very "avoidable" incident [11]. Therefore, a case was made that the driver should bear the consequence if negligence on their part is found.

The legal implications have left the state in a position left to be desired as Tempe faces a claim for $10 million from the victim's family because "the city created a dangerous situation" by not stopping people from crossing illegally [11]. Another claim against Tempe was filed against the state that accused Gov. Doug Ducey "created a dangerous situation" by advocating for Uber to test in Arizona with autonomous vehicles [11]. Gov. Ducey allegedly had 'secret' trials of the Uber self-driving cars in Arizona with little 'expert oversight' [10]. Stricter laws and regulations should have been in place which could have prevented the

situation leading to the fatality. All parties accountable should be held responsible and preventive measures need to be in place to stop this from happening in the future.

C. Safety

The automotive industry has begun applying AI in critical tasks such as self-driving cars, drones, taxis and traffic management where safety has become a major concern. Using artificial intelligence, the path of pedestrians or cyclists can be predicted, which would help in decreasing instances of traffic accidents and injuries. However, AI implementation is not as safe as it could be at this early stage. Autonomous self-driving cars could lead to drivers being distracted and not observing the happenings of the road leading to avoidable accidents. The example of this can be clearly seen in the fatality of Elaine in the Uber incident. The car, which was said to be travelling at around 38 mph, didn't slow down as it approached the pedestrian on the road [12]. This case magnifies the importance of collision avoidance systems for autonomous vehicles however the competence and focus of the driver cannot be understated [11].

Autonomous vehicle experts concluded that the car's ample array of sensors should have detected the pedestrian before she was hit [13]. Although autonomous technology has shown driverless cars are safer by eliminating human errors, it should be taken into consideration that AI technology in self-driving cars are still at its infancy. Uber, and car manufacturers alike, are always improving their technology in an attempt to minimise possible casualties but the technology can never be perfect and could cause potential harm in the present. It is important to ensure the safety regulations are met and risk assessments are done by experts. The AI system should be well tested before introduction to the public and before wide use of the application of the technology. Applying the spirit of Act-utilitarianism ethics, even with the current technology, self-driven cars are safer, in practice, than cars manually driven and better for society.

D. Security

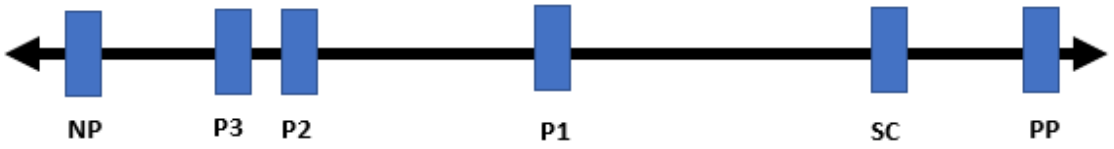
AI technology in the transportation industry is still vulnerable to cyber-attacks. The increase in demand for autonomous vehicles and increased investment year by year is paralleled by the ever-increasing concern towards security. In June 2017, Petya malware variant hacked

systems of the largest shipping company in the world which was “handling 15% of the global seaborne trade”. In June 2015, 10 flights were disrupted due to a cyber-attack against an airline’s ground computer systems at Warsaw’s Okecie airport [14]. While transportation companies know the significance of protecting data and customer security, a few corporations have experienced detrimental impacts of a cyber-attack. A demonstration of cyber-attack in self-driving cars is Tesla where one of the autonomous self-driving cars was hacked and remotely controlled to increase its speed to 85 mph in a 35-mph zone [15]. This shows AI is yet to catch up with the security threats of the modern digital age and therefore implementation should go through proper risk assessment. It is imperative to have the latest security systems safeguarding the transportation industry from cyber-attacks. A breach in security could lead to total devastation in multiple sectors not just in transportation.

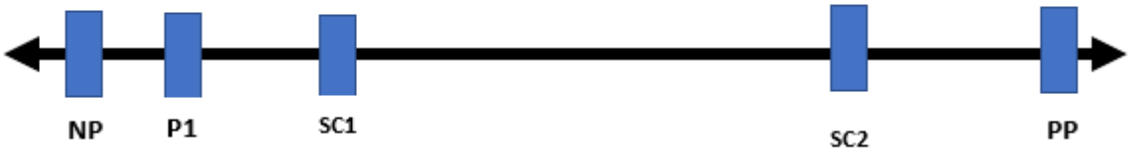
E. Ethics Analysis: Ethics Table

Action/Choice	Party 1 - Uber	Party 2 – C
	Ethics Category – Act Utilitarian	
Uber aims to make their cars autonomous with less human intervention	Act Utilitarianism ethics satisfied as Uber's mission is to provide affordable cars with increased safety measures that will have a greater good of improved public safety and better quality of life (increased safety, convenience, reduce bad decision making etc).	Act utilitarianism ethics satisfied as self-driving cars and AI technology be very supportive of Uber and worked w progression of the technology. He belie safety and supporting companies like public and so act-utilitarianism is satis "We lose tens of thousands of America by human error". He believes in Uber' error, and actually have increased high
Governor Ducey allowed Uber to conduct trials of immature technology in Arizona without public knowledge and little oversight	Act-Utilitarianism ethics satisfied since Uber complies with their customers' request. The demand for self-driving cars is very high as seen by the initial orders and craze. By conducting trials with the governor's permission, they would be able to meet and please customer demands of an earlier release date hence done for the greater good of the public Act utilitarianism ethics not satisfied as Uber autonomous system did not stop before hitting a pedestrian and possibly is faulty to which it cannot differentiate between civilians or obstacles. The incident was "definitely avoidable". The emergency brake systems were not enabled as there was a safety driver in the driver's seat to do so in case of emergency. The lack of awareness from the distracted driver led to a fatality. The worth of life is priceless and due to incompetence, a life is lost. The radar picked up Herzberg seconds before the collision but couldn't determine what she was, according to a report from the National Transportation Safety Board [11]. Act-Utilitarianism satisfied since Uber needs to ensure competitiveness in the global market. Companies like Google are at the forefront of the technology with the production of Waymo self-driving cars. For Uber to stay competitive, they need to innovate and experiment with little government hindrance. It is a private company that is money-oriented and trialling its cars with imperfect technology is a small price to pay for the potential benefits of moving faster to market with a better finished product. With increased market share, they will benefit from sizable profits and could pass their savings on to the consumer eventually benefiting for the greater good of the public.	Act-utilitarianism was not satisfied as in the state of Arizona with minim progression of tests and improvement state of road safety where "tens of th Uber would lead to "increased highway Act-utilitarianism was not satisfied si of the self-driving car in public when have led to greater losses. The auton pedestrian and possibly is faulty to w according to the National Transport should have recognised the risks invo avoidable". As well, the emergency br safety driver in the driver's seat to do been regulations in place where the te safety standard before being released have been managed and at the very lea lack of oversight from Ducey's part do and therefore Act-utilitarianism is not s

F. Ethics Analysis: Line Diagram



Point	Ethics Line Drawing from POV of Uber	Location from Left
NP	Uber develops autonomous self-driving vehicles with faulty technology that lead to the casualties and deaths of civilians. The lack of human intervention leads to drivers being distracted leading to more accidents.	**0**
PP	Uber develops autonomous vehicles that safely and efficiently transport passengers. The technology is perfected and leads to large improvements in road safety leading to lower accidents and more enjoyable ride experience. Minimal human intervention is required but can intervene during emergencies.	**10**
SC1	Uber has a safeguard mechanism which gives control to the driver when emergency situations occur where the system malfunctions. These situations are few and far between. AI technology will be able to identify possible pedestrians in the way and issue a warning to the driver well in advance giving ample time to the car. The technology being developed is trailed in a controlled environment and is tested in public only with the highest safety regulations met and risk assessments done; risks are managed for the improvement of technology and the greater good.	**8.5**
P1	Uber, without the knowledge of the public, trialled the use of the AI technology in self-driving cars	**5**
P2	Uber disabled the braking system and gave control to the driver for the case of emergency stoppages. The radar, cameras and lidar picked up Herzberg six seconds before the collision. But the computer initially couldn't determine what she was, according to a report from the National Transportation Safety Board.	**3**
P3	Uber hired a driver that was incompetent and distracted by watching a program instead of focusing on the road. They had hired a driver that had bad driving habits and put the lives of the public at serious risk resulting in the fatality of a woman. The incident was deemed as completely “avoidable” as the driver violated safety and traffic laws and guidelines [11]	**2**



Point	Ethics Line Drawing from POV of the Governor Ducey	Location from Left
NP	Governor Ducey does not consider the ethical implications of allowing Uber to trail in the state without the knowledge of the public. The high safety standards and regulations were not upheld. The lack of oversight was alarming from Governor Ducey. Ducey took incentives from Uber to freely trail in the Arizona state. Influenced the case findings and conclusion [11].	0
PP	Governor Ducey oversees the Uber trialling of self-driving cars personally. HE enforces strictest safety and road regulations. A risk assessment is done before the self-driving car is tested in the public. If the technology is imperfect, more improvements are needed to minimise risk to the public. Not only is there oversight for the technology but as well as the safety driver in the vehicle. The driver shall be hand-picked and only chosen if great competency is shown with good driving habits and strong observance to the law	10
P1	Governor Ducey does not consider the ethical implications of allowing Uber to trail in the state and does so without the knowledge of the public. The high safety standards and regulations were not upheld. The lack of oversight from Governor Ducey led to low levels of safety regulations and improper risk management. Ducey possibly took incentives from Uber to trail in the Arizona state as shown by the emails exchanged [11].	1
SC1	Governor Ducey does not consider the ethical implications of allowing Uber to trail in the state but does so with the knowledge of the public and so the public awareness of testing the self-driving car leads to pedestrians being more cautious.	3
SC2	The high safety standards and regulations were upheld. The strict oversight from Governor Ducey led to high levels of safety regulations and proper risk management. If the technology was not ready, the self-driving cars would-be put-on hold. The safety of the public is of the utmost importance.	7

G. Conclusion

The limitless potential for AI to improve the transportation industry is very easily observed. However, changes in law needs to be made to accommodate the legal liability and implications of accidents arising out of decisions made by AI. Changes in policies and regulations need to be made to ensure the public's safety and security in transportation. The AI technology might not ever be perfect but as long as it presents a safer alternative it should be considered as an option. On the surface the applications may seem all but threatening but

present and future application can raise some ethical concerns. As with any other emerging technologies, ethical and safety implications should be carefully considered before any large-scale implementation takes place. All actions and advancements come with risks but reducing the risk to as minimal as possible and acting responsibly and ethically should be the aim.

7) Case Study 4: Project Nightingale

Healthcare systems have been struggling with the increasing costs and poor outcomes [o1]. This shows that the problems in healthcare systems are like 'Pandora's Box'. They are too complex to be solved individually and manually as they generate many complicated problems. Hence, Artificial Intelligence (AI) is needed to help those responsible to overcome these problems. AI could be categorized as strong AI and weak AI [o2]. A weak AI is an AI that can only deal with specific tasks and exists currently, however, a strong AI is able to generate multiple tasks with anthropomorphic intelligence. This strong AI is usually referred to as general AI, which is the ongoing goal for many researchers nowadays to tap on as it will lead to superintelligence in the future. In healthcare systems, AI has developed new analytical tools using Big Data Analytics that can help to solve several clinical problems. Moreover, AI also has the ability to learn and adapt to new environments by utilizing complex learning algorithms to provide better user experience based on feedback given. It can also be used to predict health risks as it can analyse significantly large amounts of data [o3].

Thus, it is certain that AI is a compelling tool that likely to make a significant impact on the medical field in the future. The implementation of AI in healthcare systems has been widely used in image analysis, surgical operations, elimination of dosage errors, cybersecurity, etc [o3]. Interestingly, there are also 'consulting physician' AI robots that are programmed to have a conversation with patients and counsel them. Hence, more and more AI implementation in healthcare systems in the future will appear in no time. However, as the AI is more advanced, there are more problems on humanity and ethical issues may emerge such as unemployment, legal liability, data privacy and reliability. Morley, et. al. [o4] states that there are at least three issues on AI in healthcare systems that need to be considered: technical

possibilities and limitations; ethical, regulatory and legal framework; and governance framework. However, only ethical, regulatory and legal framework will be discussed further in this section.

A. Case Summary

Project Nightingale is a "secret" project, partnership between Google and Ascension to improve the healthcare of the U.S., that collects and uses up to 50 million personal medical information of people across 21 states [o5]. It is mentioned that this project is an AI software meant to enhance the healthcare services [o6]. The two companies have signed a Health Insurance Portability and Accountability Act (HIPAA) Business Associate Agreement (BAA) which allows patient data transfer from Ascension to Google Cloud and prohibits Google to use these data for other purposes. However, the data shared includes name, date of birth, address, family members, allergies, immunizations, radiology scans, hospitalization records, lab tests, medications and medical condition [o7]. These data sharing that includes personal data raises public and ethical concerns as there should be anonymity in the process. Moreover, these data have also been accessed by at least 150 Google employees and there is no consent taken from doctors and patients although HIPAA states that consent is not required [o7]. The United States Department of Health and Human Services (HHS) has launched an investigation into this case and it will be run by HHS' Office of Civil Rights [o8].

B. Legal Liability

Although this case study does not reflect any issue regarding data liability, this topic is still important to be discussed as new discoveries of automated medical robots are being made. *"Who is responsible when AI makes the wrong decision and leads to a fatal outcome?"* This is a common question that most people are still debating, especially in the healthcare sector. AI has been increasingly taking part in assisting medical practitioners in early diagnosis and surgical procedure in order to cut costs. However, there are still 'racist AI' that considers biased judgement/findings just like normal data. This can lead AI to make wrong decisions and errors, but who is responsible for these mistakes: doctors, AI or the provided data?

In common medical malpractice or negligence cases, *Bolam* test could be used to prove whether the doctor has violated his duty of care to a patient [o9]. However, AI can complicate this test as it introduces another opinion besides doctors' professional one. This is because hospitals may set a standardised direction in the AI system and doctors may take into consideration those AI's autonomous judgements before diagnosing the patient [o10].

Usually the results from AI healthcare systems are based on patient history, hospital transcripts and medical research. Hence, it may be difficult for AI to express clear justification for its decision as it is based on trends and calculations from the data that it has [o10]. Furthermore, in reality, doctors and hospitals actually have the decision to follow or not with AI's diagnosis and thus, can doctors and hospitals be the one that is responsible for negligence? If doctors are the one that are being blamed, they are just following the rule set by the hospitals, then can the hospitals be held accountable for AI's decision? Seeing the questions that have been asked, it looks like a cycle where the blame can be put on from AI itself to doctors, from doctors to hospitals, hospitals to AI and so on. This problem may not be solved until the end. Hence, proper legal understanding of AI and how it can lead to medical negligence still needs to be focused on before the new problems from more advanced AI occur.

C. Data privacy/Data Sharing

In this digitisation era, personal data, collected by and for AI, have raised the majority public concern on how safe their privacy will be as every step that they do will be recorded and the auto-generated results that they receive are based on AI algorithms. This shows that their data could be transparent to the world and any interested party is able to buy the data without their consent. Moreover, the world is filled with enormous data, hence, the cost of generating and distributing data has been decreasing and resulting in the growing ability of data processing. AI Researchers have been working on large amounts of data and there are ways for them to exchange data with other organizations, which can be seen from the partnership of Google and Ascension, to have enough samples to test on their study. There are big and powerful technology companies that attract talents to work in their industry, provided that non-disclosure agreement must be signed by the researchers so that the data and findings

should remain within the academia [o11]. However, there is no insight on whether Google employees have signed the agreement and make sure that the data is not shared with other third parties or combined with the data that they currently have in the database.

Therefore, to reduce this concern, the activities of data sharing should adhere to the principles that focus on maximising the health benefits of the public and avoid the violation of ethical obligation by the researchers as well as companies to conduct their study. Health Insurance Portability and Accountability Act (HIPAA) agreement, which is signed by both Google and Ascension, is one of the laws that regulates this issue, however there is a loophole in this law created in 1996. The law enables healthcare providers to disclose personal information to business associates [o5]. This raises public concerns as in this technology era, any information shared by individuals can be easily owned by others.

D. Discrimination

There are reasons why the data that is shared by Ascension by Google are not anonymous, instead the personal information is attached to health information. One of the reasons is the AI could discriminate against certain patients when it comes with different people with different ethnic groups, different gender, etc. However, Ascension has millions of patients which would minimise the bias made by the AI. Thus, there is no strong reason why Ascension does not de-identify the data that they shared to Google although it is said that the data is used solely for the development of the software [o7].

E. Ethical Analysis: Ethics Table

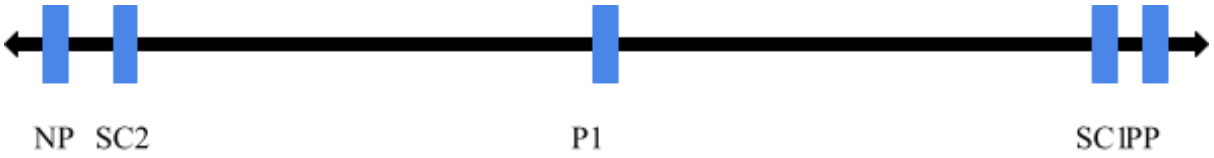
Action/Choice	Party #1: Google	
	Ethics Category -- Rights Ethics	
Ascension agreed to the agreement with Google to share the data of its patients under HIPAA BAA.	Google satisfies the right of the patients to have access to a better healthcare system if the data are being used for the sole purpose of the project. Google violates the right of the patients to protect their personal data if they do not de-identify the data that they get through the data sharing agreement [o7]. This means that they can divide the data through several parts and may use certain information of the data for other purposes.	Ascension satisfies the r information for further re they signed. Ascension violates the righ is because explicit consen under HIPAA as the pro system. Explicit consent r data are being used and h for anonymity their data [o
Google develops AI software for healthcare sector	Google satisfies the right of the public to provide a better healthcare system using AI. The software will help to suggest which treatments the patient should go and give recommendations after doctors examine him/her and input the data into the system. Through this system, additional/replacement resources could be managed to the patient's team, such as doctors and prescription.	Ascension indirectly satis health information (PHI) given that Google uses the Ascension violates the rig their approvals. Although information owner is not r Hence, HIPAA should be r

F. Ethical Analysis: Line Diagram

Google -

Point	Ethics Line-Drawing from POV of Party #1 Google	Location from Left
NP	Google uses the data from Ascension for other purposes besides the main objective that is stated in the agreement with Ascension. There is no de-identification of the data.	0
PP	Google uses the data from Ascension only for the sole purpose of the project, which is to enhance the healthcare system. There is de-identification of the data. The software seems to improve Ascension's healthcare system. All actions taken adhere to HIPAA and there is no breach of confidentiality.	10
P1	Google accesses the data for the purpose of the project. Google claims that the usage of data is still under HIPAA. There is no de-identification of the data shared by Ascension and no one knows how Google may treat the data that they get[o7].	8
SC1	The project is still ongoing, however, Google divides the data shared by Ascension and uses it for other purposes, such as combining the data that they currently have with the one from Ascension. There is a breach of confidentiality.	1
SC2	The data that Google received is anonymous and has been de-identified. Google can still develop the software, however, there is still small discrimination. Overall, the software seems to improve the healthcare system in Ascension and more people are satisfied with the services provided by Ascension.	9

Ascension -



Point	Ethics Line-Drawing from POV of Party #2 Ascension	Location from Left
NP	Ascension transfers the data to Google Cloud without consent (breach of confidentiality). The purpose and usage of the data sharing is not based on HIPAA.	0
PP	Ascension transfers the data to Google Cloud by notifying its patients about the purpose of data sharing and how the data is being stored. HIPAA should be reviewed as it was enacted in 1996 before the technology industry, thus it	10

	is outdated. Better specified regulations are specified to prevent breaching of confidentiality. Explicit consent should be asked by Ascension to its patient.	
P1	Ascension puts its patient's data to Google Cloud without notifying them about the project. Although it may seem that there is no violation of HIPAA, explicit consent is not taken [o7].	5
SC1	Ascension de-identifies the data before shares it to Google. There is no breach of confidentiality but there is no explicit consent from its patients.	9.5
SC2	Ascension shares the data without de-identifies it. No explicit consent is granted. Ascension knows that the data is used not only to develop the software but also for other purposes that Google intends to do. They do not care about the data; they only focus on getting an enhanced healthcare system developed by Google.	1

G. Conclusion

The investigation by the United States Department of Health and Human Services (HHS) is still ongoing and people, especially Americans, are still waiting for the outcomes. Privacy of each one of us is important in this digitization era. Hence, law should be made clearly to underline such problems. It is recommended that HIPAA should be revised to accommodate the issues arising from this modern era of technology.

8) Conclusion

The underlying theme in every field where AI has potential use is the absence of legal frameworks which regulate the use of AI. As a result, the liability aspect of AI is also presently unclear. If an AI were to make an erroneous judgement which led to loss of life or property, who is to bear liability, the operator/owner or the creator? These questions do not have clear answers at present. Another question which is related to this is to what extent AI can be allowed to make decisions in place of humans. As seen in Case Study 2, operators can feel uncomfortable in permitting AI to make decisions pertaining to life and death. This discomfort also exists in Case Study 1, where an AI is used to decide whether a person is the same individual as a suspect in a criminal or legal case. The interim solution to these problems has involved legal orders to stay the use of Clearview AI in the case of New Jersey. A long-term solution, however, must involve the formulation of rigorous frameworks which protect individual rights and privacy, while ensuring that accountability for the actions of AI are directed to human stakeholders. The key to successful implementation and integration of AI lies in a balanced approach and calibrated use. AI should be introduced in areas where

computer learning can benefit individuals and society. But legal frameworks must be present to ensure accountability is held ultimately by human operators.

A. References

[1] Mason, R. O., Collins, C. P., & Cox, E. L., *Four ethical issues of the information age.* Mis Quarterly, 1986.

[2] Delbridge, A. *The Macquarie dictionary*, Rev. 3rd edn. Macquarie Library, North Ryde, N.S.W, 2001.

[3] AI HLEG, Ethics Guidelines for Trustworthy Artificial Intelligence. AI HLEG, 2018.

[4] V. Grover, "4 Ways AI is Personalizing the Retail Customer Experience | MarTech Advisor," *MTA*, 2019. https://www.martechadvisor.com/articles/customer-experience-2/4-ways-ai-personalizing-retail-cx/ (accessed Mar. 28, 2020).

[5] G. Fowler, "Council Post: The Evolution Of Brick-And-Mortar Shopping With AI," *Forbes*, 2020. https://www.forbes.com/sites/forbesbusinessdevelopmentcouncil/2020/03/05/the-evolution-of-brick-and-mortar-shopping-with-ai/#3e6281a26fdf (accessed Mar. 28, 2020).

[6] M. Kaput, "What Is Artificial Intelligence for Social Media?," *Marketing Artificial Intelligence Institute*, 2019. https://www.marketingaiinstitute.com/blog/what-is-artificial-intelligence-for-social-media (accessed Mar. 28, 2020).

[7] L. Columbus, "10 Predictions How AI Will Improve Cybersecurity In 2020," *Forbes*, 2019. https://www.forbes.com/sites/louiscolumbus/2019/11/24/10-predictions-how-ai-will-improve-cybersecurity-in-2020/#16da446f6dd7 (accessed Mar. 28, 2020).

[8] M. Burges, "Google's artificial intelligence created its own encrypted messages | WIRED UK," *Wired*, 31-Oct-2016. https://www.wired.co.uk/article/google-artificial-intelligence-encryption (accessed Mar. 28, 2020).

[9] K. Quach, "Clearview said to be chasing every mugshot taken in the US over the last 15 years to paste into its facial-recog system • The Register," *The Register*, 09-Mar-2020. https://www.theregister.co.uk/2020/03/09/ai_roundup_march6/ (accessed Mar. 28, 2020).

[10] K. Hill, "The Secretive Company That Might End Privacy as We Know It - The New York Times," *New York Times*, 2020. https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html (accessed Mar. 27, 2020).

[11] B. Gillbert, "Clearview AI scraped billions of photos from social media to build a facial recognition app that can ID anyone — here's everything you need to know about the mysterious company, Business Insider - Business Insider Singapore," *Business Insider*, 2020. https://www.businessinsider.sg/what-is-clearview-ai-controversial-facial-recognition-startup-2020-3 (accessed Mar. 28, 2020).

[12] C. R. Clark, "Attorney General Donovan Sues Clearview AI for Violations of Consumer Protection Act and Data Broker Law - Office of the Vermont Attorney General," *Office of Vermont Attorney General*, 10-Mar-2020. https://ago.vermont.gov/blog/2020/03/10/attorney-general-donovan-sues-clearview-ai-for-violations-of-consumer-protection-act-and-data-broker-law/ (accessed Mar. 28, 2020).

[13] "Automated queries - Search Console Help," *Google Support*, 2020. https://support.google.com/webmasters/answer/66357?hl=en (accessed Mar. 28, 2020).

[14] H. Sharma, "Data Scraping for SEO & Analytics - Tutorial," *Optimize Smart*, 2020. https://www.optimizesmart.com/data-scraping-guide-for-seo/ (accessed Mar. 28, 2020).

[15] A. Ng and S. Musil, "Clearview AI hit with cease-and-desist from Google, Facebook over facial recognition collection - CNET," *CNet*, 2020. https://www.cnet.com/news/clearview-ai-hit-with-cease-and-desist-from-google-over-facial-recognition-collection/ (accessed Mar. 28, 2020).

[16] C. Campbell, “How China Is Using Big Data to Create a Social Credit Score | Time.com,” *Time*, 2019. https://time.com/collection/davos-2019/5502592/china-social-credit-score/ (accessed Mar. 28, 2020).

[17] O. Schwartz, “In 2016, Microsoft’s Racist Chatbot Revealed the Dangers of Online Conversation - IEEE Spectrum,” *IEEE Spectrum*, 2019. https://spectrum.ieee.org/tech-talk/artificial-intelligence/machine-learning/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation (accessed Mar. 28, 2020).

[18] “Clearview AI: Google, YouTube Venmo and LinkedIn send cease-and-desist letter to facial recognition app - CBS News,” *CBS News*, 2020. https://www.cbsnews.com/news/clearview-ai-google-youtube-send-cease-and-desist-letter-to-facial-recognition-app/ (accessed Mar. 28, 2020).

Dinesh

[1] “IDC - IT EXECUTIVE - AI,” *IDC*, 2019. [Online]. Available: https://www.idc.com/itexecutive/research/topics/ai. [Accessed: 31-Mar-2020].

[2] “AI in Transportation Market to Generate Revenue Worth $3.5 Billion by 2023,” *Prescient & Strategic Intelligence*, 2018. [Online]. Available: https://www.psmarketresearch.com/press-release/ai-in-transportation-market. [Accessed: 31-Mar-2020].

[3] A. Bolwell, “The future of transportation: How AI is helping vehicles think,” *Megatrends*, 2018. [Online]. Available: https://hpmegatrends.com/the-future-of-transportation-how-ai-is-helping-vehicles-think-c505d229876f. [Accessed: 31-Mar-2020].

[4] A. Chottani, G. Hastings, J. Murnane, and Fl. Neuhaus, “Autonomous trucks disrupt US logistics | McKinsey,” *McKinsey & Company*, 2018. [Online]. Available: https://www.mckinsey.com/industries/travel-transport-and-logistics/our-insights/distraction-or-disruption-autonomous-trucks-gain-ground-in-us-logistics. [Accessed: 31-Mar-2020].

[5] “World’s first autonomous taxi starts operating in Tokyo - Nikkei Asian Review,” *NIKKEI*, 2018. [Online]. Available: https://asia.nikkei.com/Business/Business-trends/World-s-first-autonomous-taxi-starts-operating-in-Tokyo. [Accessed: 31-Mar-2020].

[6] C. Isidore, “Self-driving cars are already really safe,” *CNN BUSINESS*, 2018. [Online]. Available: https://money.cnn.com/2018/03/21/technology/self-driving-car-safety/. [Accessed: 31-Mar-2020].

[7] “How Technology Can Fundamentally Change Your Commute,” *Forbes*, 2019. [Online]. Available: https://www.forbes.com/sites/quora/2019/09/10/how-technology-can-fundamentally-change-your-commute/#6b848c4013bd. [Accessed: 31-Mar-2020].

[8] A. Brinklow, “Uber won’t be charged in Arizona self-driving car death - Curbed SF,” *Curbed*, 2019. [Online]. Available: https://sf.curbed.com/2019/3/6/18252948/uber-autonomous-robot-death-charges-sheila-sullivan-yavapai. [Accessed: 31-Mar-2020].

[9] B. Ckerkin, “Uber Reaches Settlement with Family of Fatal Crash Victim | DMV.ORG,” *DMV*, 2018. [Online]. Available: https://www.dmv.org/articles/uber-settles-fatal-crash-suit. [Accessed: 31-Mar-2020].

[10] M. Harris, “Exclusive: Arizona governor and Uber kept self-driving program secret, emails reveal | Technology | The Guardian,” *The Guardian*, 2018. [Online]. Available: https://www.theguardian.com/technology/2018/mar/28/uber-arizona-secret-self-driving-program-governor-doug-ducey. [Accessed: 31-Mar-2020].

[11] R. Randazzo, “Uber crash death one year later: Who is involved and who is responsible?,” *az central*, 2019. [Online]. Available: https://www.azcentral.com/story/news/local/tempe/2019/03/17/uber-crash-death-who-blame-tempe-arizona-rafaela-vasquez-elaine-herzberg/3157481002/. [Accessed: 31-Mar-2020].

[12] R. Grenoble, “Software In Fatal Uber Crash Reportedly Recognized Woman, Then Ignored Her | HuffPost,” *HUFFPOST*, 2018. [Online]. Available:

https://www.huffpost.com/entry/self-driving-uber-fatal-software-at-fault_n_5af093c2e4b041fd2d296e23. [Accessed: 31-Mar-2020].

[13] C. Lago and C. Trueman, “How Singapore Is Developing Driverless Cars | CIO,” *CIO*, 2019. [Online]. Available: https://www.cio.com/article/3294207/how-singapore-is-driving-the-development-of-autonomous-vehicles.html. [Accessed: 31-Mar-2020].

[14] D. Winder, “Hackers Made Tesla Cars Autonomously Accelerate Up To 85 In A 35 Zone,” *Forbes*, 2019. [Online]. Available: https://www.forbes.com/sites/daveywinder/2020/02/19/hackers-made-tesla-cars-autonomously-accelerate-up-to-85-in-a-35-zone/#39297aba7245. [Accessed: 31-Mar-2020].

[15] S. Rao, “Cyber Security and Public sector Transportation System – mc.ai,” *Medium*, 2020. [Online]. Available: https://mc.ai/cyber-security-and-public-sector-transportation-system/. [Accessed: 31-Mar-2020].