ARUNAI ENGINEERING COLLEGE

(An Autonomous Institution)

VeluNagar,Thiruvannamalai-606603

www.arunai.org

DEPARTMENT OF ARTIFICIAL INTELLIGENCE

&

DATA SCIENCE

BACHELOR OF TECHNOLOGY

2025-2026

FIFITH SEMESTER

CCS345 – ETHICS AND AI LABORATORY

ARUNAI ENGINEERING COLLEGE

(An Autonomous Institution)

TIRUVANNAMALAI–606603

DEPARTMENT OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE

CERTIFICATE

Certified that this is a bonafide record of work done by

Name :

University Reg.No :

Semester :

Branch :

Year :

Staff-in-Charge **Head of the Department**

Submitted for the_____

Practical Examination held on_____

Internal Examiner **External Examiner**

TABLE OF CONTENT

S.No.	Date	Experiment	Page Number	Signature

EXP.NO-1 Recent case study of ethical initiatives in healthcare, autonomous vehicles and defense

1) Case study: healthcare robots

Artificial Intelligence and robotics are rapidly moving into the field of healthcare and will increasingly play roles in diagnosis and clinical treatment. For example, currently, or in the nearfuture, robots will help in the diagnosis of patients; the performance of simple surgeries; and themonitoring of patients' health and mental wellness in short and long-term care facilities. They may also provide basic physical interventions, work as companion careers, remind patients to take their medications, or help patients with their mobility. In some fundamental areas of medicine, such as medical image diagnostics, machine learning has been proven to match or even surpass our abilityto detect illnesses.

Embodied AI, or robots, are already involved in a number of functions that affect people's physical safety. In June 2005, a surgical robot at a hospital in Philadelphia malfunctioned during prostatesurgery, injuring the patient. In June 2015, a worker at a Volkswagen plant in Germany was crushed to death by a robot on the production line. In June 2016, a Tesla car operating in autopilot modecollided with a large truck, killing the car's passenger (Yadron and Tynan, 2016).

As robots become more prevalent, the potential for future harm will increase, particularly in the case of driverless cars, assistive robots and drones, which will face decisions that have real consequences for human safety and well-being. The stakes are much higher with embodied AI than with mere software, as robots have moving parts in physical space (Lin et al., 2017). Any robot with movingphysical parts poses a risk, especially to vulnerable people such as children and the elderly.

Safety

Again, perhaps the most important ethical issue arising from the growth of AI and robotics inhealthcare is that of safety and avoidance of harm. It is vital that robots should not harm people, and that they should be safe to work with. This point is especially important in areas of healthcare that deal with vulnerable people, such as the ill, elderly, and children.

Digital healthcare technologies offer the potential to improve accuracy of diagnosis and treatments, but to thoroughly establish a technology's long-termsafetyandperformance investment in clinicaltrials is required. The debilitating side-effects of vaginal mesh implants and the continued legalbattles against manufacturers (The Washington Post, 2019), stand as an example against shortcutting testing, despite the delays this introduces to innovating healthcare. Investment in clinical trials will be essential to safely implement the healthcare innovations that AI systems offer.

User understanding

The correct application of AI by a healthcare professional is important to ensure patient safety. For instance, the precise surgical robotic assistant 'the da Vinci' has proven a useful tool in minimisingsurgical recovery, but requires a trained operator (The Conversation, 2018).

A shift in the balance of skills in the medical workforce is required, and healthcare providers arepreparing to develop the digital literacy of their staff over the next two decades (NHS' Topol Review,2009). With genomics and machine learning becoming embedded in diagnoses and medical decision-making, healthcare professionals need to become digitally literate to understand each technological tool and use it appropriately. It is important for users to trust the AI presented but to be aware of each tool's strengths and weaknesses, recognising when validation is necessary. Forinstance, a generally accurate machine learning study to predict the risk of complications in patientswith pneumonia erroneously considered those with asthma to be at low risk. It reached thisconclusion because asthmatic pneumonia patients were taken directly to intensive care, and this

higher-level care circumvented complications. The inaccurate recommendation from the algorithm was thus overruled (Pulmonology Advisor, 2017).

However, it's questionable to what extent individuals need to understand how an AI system arrivedat a certain prediction in order to make autonomous and informed decisions. Even if an in-depthunderstanding of the mathematics is made obligatory, the complexity and learned nature of machine learning algorithms often prevent the ability to understand how a conclusion has been made from a dataset — a so called 'black box' (Schönberger, 2019). In such cases, one possible route

to ensure safety would be to license AI for specific medical procedures, and to 'disbar' the AI if acertain number of mistakes are made(Hart, 2018).

Data protection

Personal medical data needed for healthcare algorithms may be at risk. For instance, there areworries that data gathered by fitness trackers might be sold to third parties, such as insurance companies, who could use those data to refuse healthcare coverage (National Public Radio, 2018). Hackers are another major concern, as providing adequate security for systems accessed by a range of medical personnel is problematic (Forbes, 2018).

Pooling personal medical data is critical for machine learning algorithms to advance healthcare interventions, but gaps in information governance form a barrier against responsible and ethicaldata sharing. Clear frameworks for how healthcare staff and researchers use data, such as genomics,in a way that safeguards patient confidentiality is necessary to establish public trust and enableadvances in healthcare algorithms (NHS' Topol Review, 2009).

Legal responsibility

Although AI promises to reduce the number of medical mishaps, when issues occur, legal liabilitymust be established. If equipment can be proven to be faulty then the manufacturer is liable, but it is often tricky to establish what went wrong during a procedure and whether anyone, medical personnel or machine, is to blame. For instance, there have been lawsuits against the da Vinci surgical assistant (Mercury News, 2017), but the robot continues to be widely accepted (The Conversation, 2018).

In the case of 'black box' algorithms where it is impossible to ascertain how a conclusion is reached,it is tricky to establish negligence on the part of the algorithm's producer (Hart, 2018).

For now, AI is used as an aide for expert decisions, and so experts remain the liable party in mostcases. For instance, in the aforementioned pneumonia case, if the medical staff had relied solely on the AI and sent asthmatic pneumonia patients home without applying their specialist knowledge,then that would be a negligent act on their part (Pulmonology Advisor, 2017; International Journal of Law and Information Technology, 2019).

Soon, the omission of AI could be considered negligence. For instance, in less developed countrieswith a shortage of medical professionals, withholding AI that detects diabetic eye disease and soprevents blindness, because of a lack of ophthalmologists to sign off on a diagnosis, could beconsidered unethical (The Guardian, 2019; International Journal of Law and Information Technology, 2019).

Bias

Non-discrimination is one of the fundamental values of the EU (see Article 21 of the EU Charter ofFundamental Rights), but machine learning algorithms are trained on datasets that often have proportionally less data available about minorities, and as such can be biased (Medium, 2014). This can mean that algorithms trained to diagnose conditions are less likely to be accurate for ethnicpatients; for instance, in the dataset used to train a model for detecting skin cancer, less than 5percent of the images were from individuals with dark skin, presenting a risk of misdiagnosis forpeople of colour (The Atlantic, 2018).

To ensure the most accurate diagnoses are presented to people of all ethnicities, algorithmic biases must be identified and understood. Even with a clear understanding of model design this is adifficult task because of the aforementioned 'black box' nature of machine learning. However, various codes of conduct and initiatives

have been introduced to spot biases earlier. For instance,The Partnership on AI, an ethics-focused industry group was launched by Google, Facebook, Amazon, IBM and Microsoft (The Guardian, 2016) — although, worryingly, this board is not very diverse.

Equality of access

Digital health technologies, such as fitness trackers and insulin pumps, provide patients with theopportunity to actively participate in their own healthcare. Some hope that these technologies will help to redress health inequalities caused by poor education, unemployment, and so on. However, there is a risk that individuals who cannot afford the necessary technologies or do not have the required 'digital literacy' will be excluded, so reinforcing existing health inequalities (The Guardian, 2019).

The UK's National Health Services' Widening Digital Participation programme is one example of howa healthcare service has tried to reduce health inequalities, by helping millions of people in the UK who lack the skills to access digital health services. Programmes such as this will be critical in ensuring equality of access to healthcare, but also in increasing the data from minority groups needed to prevent the biases in healthcare algorithms discussed above.

Quality of care

'There is remarkable potential for digital healthcare technologies to improve accuracy ofdiagnoses and treatments, the efficiency of care, and workflow for healthcare professionals' (NHS' Topol Review, 2019).

If introduced with careful thought and guidelines, companion and care robots, for example, couldimprove the lives of the elderly, reducing their dependence, and creating more opportunities forsocial interaction. Imagine a home-care robot that could: remind you to take your medications; fetch items for you if you are too tired or are already in bed; perform simple cleaning tasks; and help you stay in contact with your family, friends and healthcare provider via video link.

However, questions have been raised over whether a 'cold', emotionless robot can really substitute for a human's empathetic touch. This is particularly the case in long-term caring of vulnerable and often lonely populations, who derive basic companionship from caregivers. Human interaction isparticularly important for older people, as research suggests that an extensive social network offers protection against dementia. At present, robots are far from being real companions. Although they can interact with people, and even show simulated emotions, their conversational ability is stillextremely limited, and they are no replacement for human love and attention. Some might go as far as saying that depriving the elderly of human contact is unethical, and even a form of cruelty.

And does abandoning our elderly to cold machine care objectify (degrade) them, or humancaregivers? It's vital that robots don't make elderly people feel like objects, or with even less controlover their lives thanwhenthey were dependent on humans — otherwise they may feel like they are'lumps of dead matter: to be pushed, lifted, pumped or drained, without proper reference to the fact that they are sentient beings' (Kitwood 1997).

In principle, autonomy, dignityandself-determination can all be thoroughlyrespected by a machineapplication, but it's unclear whether application of these roles in the sensitive field of medicine will be deemed acceptable. For instance, a doctor used a telepresence device to give a prognosis ofdeath to a Californian patient; unsurprisingly the patient's family were outraged by this impersonal approach to healthcare (The Independent, 2019). On the other hand, it's argued that new technologies, such as health monitoring apps, will free up staff time for more direct interactions with patients, and so potentially increase the overall quality of care (The Guardian, Press Association, Monday 11 February 2019).

Deception

A number of 'carebots' are designed for social interactions and are often touted to provide anemotional therapeutic role. For instance, care homes have found that a robotic seal pup's animal-like interactions with residents brightens their mood, decreases anxiety and actually increases the sociability of residents with their human caregivers. However, the line between reality and imagination is blurred for dementia patients, so is it

dishonest to introduce a robot as a pet andencourage a social-emotional involvement? (KALW, 2015) And if so, is if morally justifiable?

Companion robots and robotic pets could alleviate loneliness amongst older people, but this wouldrequire them believing, in some way, that a robot is a sentient being who cares about them and hasfeelings — a fundamental deception. Turkle et al. (2006) argue that 'the fact that our parents, grandparents and children might say 'I love you' to a robot who will say 'I love you' in return, doesnot feel completely comfortable; it raises questions about the kind of authenticity we require of our technology'. Wallach and Allen (2009) agree that robots designed to detect human social gestures and respond in kind all use techniques that are arguably forms of deception. For an individual tobenefit from owning a robot pet, they must continually delude themselves about the real nature of their relation with the animal. What's more, encouraging elderly people to interact with robot toys has the effect of infantilising them.

Autonomy

It's important that healthcare robots actually benefit the patients themselves, and are not just designed to reduce the care burden on the rest of society — especially in the case of care andcompanion AI. Robots could empower disabled and older people and increase their independence;in fact, given the choice, some might prefer robotic over human assistance for certain intimate taskssuch as toileting or bathing. Robots could be used to help elderly people live in their own homes forlonger, giving them greater freedom and autonomy. However, how much control, or autonomy,should a person be allowed if their mental capability is in question? If a patient asked a robot tothrow them off the balcony, should the robot carry out that command?

Liberty and privacy

As with many areas of AI technology, the privacy and dignity of users' needs to be carefullyconsidered when designing healthcare service and companion robots. Working in people's homesmeans that robots will be privy to private moments such as bathing and dressing; if these momentsare recorded, who should have access to the information, and how long should recordings be kept?The issue becomes more complicated if an elderly person's mental state deteriorates and they become confused — someone with Alzheimer's could forget that a robot was monitoring them, andcould perform acts or say things thinking that they are in the privacy of their own home. Home-care robots need to be able to balance their user's privacy and nursing needs, for example by knockingand awaiting an invitation before entering a patient's room, except in a medical emergency.

To ensure their charge's safety, robots might sometimes need to act as supervisors, restricting their freedoms. For example, a robot could be trained to intervene if the cooker was left on, or the bath was overflowing. Robots might even need to restrain elderly people from carrying out potentiallydangerous actions, such as climbing up on a chair to get somethingfrom a cupboard. Smart homeswith sensors could be used to detect that a person is attempting to leave their room, and lock the door, or call staff — but in so doing the elderly person would be imprisoned.

Moral agency

There's very exciting work where the brain can be used to control things, like maybe they've lost the use of an arm…where Ithink the realconcerns lie is with things like behavioural targeting: going straight tothe hippocampus and people pressing 'consent', like we do now, for data access'. (John Havens)

Robots do not have the capacity for ethical reflection or a moral basis for decision-making, and thus humans must currently hold ultimate control over any decision-making. An example of ethical reasoning in a robot can be found in the 2004 dystopian film 'I, Robot', where Will Smith's character disagreed with how the robots of the fictional time used cold logic to save his life over that of achild's. If more automated healthcare is pursued, then the question of moral agency will require closer attention. Ethical reasoning is being built into robots, but moral responsibility is about more thanthe application of ethics — and it is unclearwhether robots of the future will be able to handlethe complex moral issues in healthcare (Goldhill, 2016).

Trust

Larosa and Danks (2018) write that AI may affect human-human interactions and relationships within the healthcare domain, particularly that between patient and doctor, and potentially disrupt the trust we place in our

doctor.

'Psychology research shows people mistrust those who make moral decisions by calculating costsand benefits — like computers do' (The Guardian, 2017). Our distrust of robots may also come fromthe number of robots runningamok in dystopianscience fiction. News stories ofcomputer mistakes — for instance, of an image-identifying algorithm mistaking a turtle for a gun (The Verge, 2017) —alongside worries over the unknown, privacy and safety are all reasons for resistance against theuptake of AI (Global News Canada, 2016).

Firstly, doctors are explicitly certified and licensed to practice medicine, and their license indicates that they have specific skills, knowledge, and values such as 'do no harm'. If a robot replaces a doctor for a particular treatment or diagnostic task, this could potentially threaten patient-doctor trust, as the patient now needs to know whether the system is appropriately approved or 'licensed' for thefunctions it performs.

Secondly, patients trust doctors because they view them as paragons of expertise. If doctors were seen as 'mere users' of the AI, we would expect their role to be downgraded in the public's eye,undermining trust.

Thirdly, a patient's experiences with their doctor are a significant driver of trust. If a patient has an open line of communication with their doctor, and engages in conversation about care andtreatment, then the patient will trust the doctor. Inversely, if the doctor repeatedly ignores thepatient's wishes, then these actions will have a negative impact on trust. Introducing AI into thisdynamic could increase trust — if the AI reduced the likelihood of misdiagnosis, for example, orimproved patient care. However, AI could also decrease trust if the doctor delegated too much diagnostic or decision-making authority to the AI, undercutting the position of the doctor as an authority on medical matters.

As the body of evidence grows to support the therapeutic benefits for each technological approach,and as more robotic interacting systems enter the marketplace, then trust in robots is likely to increase. This has already happened for robotic healthcare systems such as the da Vinci surgicalrobotic assistant (The Guardian, 2014).

Employment replacement

As in other industries, there is a fear that emerging technologies may threaten employment (TheGuardian, 2017), for instance, there are carebots now available that can perform up to a third ofnurses' work (Tech Times, 2018). Despite these fears, the NHS' Topol Review (2009) concluded that 'these technologies will not replace healthcare professionals but will enhance them ('augment them'), giving them more time to care for patients'. The review also outlined how the UK's NHS will nurture a learning environment to ensure digitally capable employees.

2) Case study: Autonomous Vehicles

Autonomous Vehicles (AVs) are vehicles that are capable of sensing their environment and operating with little to no input from a human driver. While the idea of self-driving cars has beenaround since at least the 1920s, it is only in recent years that technology has developed to a pointwhere AVs are appearing on public roads.

According to automotive standardisation body SAE International (2018), there are six levels of driving automation:

0	No automation	An automated system may issuewarnings and/ormomentarily intervene indriving, but has no sustained vehicle control.

1	Hands on	The driver and automated system share control of the vehicle. For example, the automated system may control engine power to maintain a set speed(e.g. Cruise Control), engine and brake power to maintain and vary speed(e.g. Adaptive Cruise Control), or steering (e.g. Parking Assistance). Thedriver must be ready to retake full control at any time.
2	Hands off	The automated system takes full control of the vehicle (including accelerating, braking, and steering). However, the driver must monitor the driving and be prepared to intervene immediately at any time.
3	Eyes off	The driver can safely turn their attention away from the driving tasks (e.g. to text or watch a film) as the vehicle will handle any situations that call foran immediate response. However, the driver must still be prepared to intervene, if called upon by the AV to do so, within a timeframe specified bythe AV manufacturer.
4	Minds off	As level 3, but no driver attention is ever required for safety, meaning thedriver can safely go to sleep or leave the driver's seat.
5	Steering wheel optional	No human intervention is required at all. An example of a level 5 AV wouldbe a robotic taxi.

Some of the lower levels of automation are already well-established and on the market, while higherlevel AVs are undergoing development and testing. However, as we transition up the levels and putmore responsibility on the automated system than the human driver, a number of ethical issuesemerge.

Societal and Ethical Impacts of AVs

'We cannot build these tools saying, 'we know that humans act a certain way, we're going to kill them –here's what to do'.' (John Havens)

Public safety and the ethics of testing on public roads

At present, cars with 'assisted driving' functions are legal in most countries. Notably, some Teslamodels have an Autopilot function, which provides level 2 automation (Tesla, nd). Drivers are legallyallowed to use assisted driving functions on public roads provided they remain in charge of the vehicle at all times. However, many of these assisted driving functions have not yet been subject to independent safety certification, and as such may pose a risk to drivers and other road users. InGermany, a report published by the Ethics Commission on Automated Driving highlights that it isthe public sector's responsibility to guarantee the safety of AV systems introduced and licensed onpublic roads, and recommends that all AV driving systems be subject to official licensing and monitoring (Ethics Commision, 2017).

In addition, it has been suggested that the AV industry is entering its most dangerous phase, withcars being not yet fully autonomous but human operators not being fully engaged (Solon, 2018).The risks this poses have been brought to widespread attention following the first pedestrian fatalityinvolving an autonomous car. The tragedy took place in Arizona, USA, in May 2018, when a level 3AV being tested by Uber collided with 49-year-old Elaine Herzberg as she was walking her bike across a street one night. It was determined that Uber was 'not criminally liable' by prosecutors(Shepherdson and Somerville, 2019), and the US National Transportation Safety Board's preliminary report (NTSB, 2018), which drew no conclusions about the cause, said that all elements of the self-driving system were operating normally at the time of the crash. Uber said that the driver is relied upon to intervene and take action in situations requiring emergency braking – leading somecommentators to call out the

misleading communication to consumers around the terms 'self- driving cars' and 'autopilot' (Leggett, 2018). The accident also caused some to condemn the practice of testing AV systems on public roads as dangerous and unethical, and led Uber to temporarily suspend its self-driving programme (Bradshaw, 2018).

This issue of human safety — of both public and passenger — is emerging as a key issue concerning self-driving cars. Major companies — Nissan, Toyota, Tesla, Uber, Volkswagen — are developing autonomous vehicles capable of operating in complex, unpredictable environments without directhuman control, and capable of learning, inferring, planning and making decisions.

Self-driving vehicles could offer multiple benefits: statistics show you're almost certainly safer in a car driven by a computer than one driven by a human. They could also ease congestion in cities,reduce pollution, reduce travel and commute times, and enable people to use their time more productively. However, they won't mean the end of road traffic accidents. Even if a self-driving car has the best software and hardware available, there is still a collision risk. An autonomous car could be surprised, say by a child emerging from behind a parked vehicle, and there is always the issue of *how: how* should such cars be programmed when they must decide whose safety to prioritise?

Driverless cars may also have to choose between the safety of passengers and other road users. Saythat a car travels around a corner where a group of school children are playing; there is not enoughtime to stop, and the only way the car can avoid hitting the children is to swerve into a brick wall —endangering the passenger. Whose safety should the car prioritise: the children's', or the passenger's?

Processes and technologies for accident investigation

AVs are complex systems that often rely on advanced machine learning technologies. Several serious accidents have already occurred, including a number of fatalities involving level 2 AVs:

- In January 2016, 23-year-old Gao Yaning died when his Tesla Model S crashed into the backof a road-sweeping truck on a highway in Hebei, China. The family believe Autopilot wasengaged when the accident occurred and accuse Tesla of exaggerating the system's capabilities. Tesla state that the damage to the vehicle made it impossible to determinewhether Autopilot was engaged and, if so, whether it malfunctioned. A civil case into thecrash is ongoing, with a third-party appraiser reviewing data from the vehicle (Curtis, 2016).
- In May 2016, 40-year-old Joshua Brown died when his Tesla Model S collided with a truckwhile Autopilot was engaged in Florida, USA. An investigation by the National Highways and Transport Safety Agency found that the driver, and not Tesla, were at fault (Gibbs, 2016). However, the National Highway Traffic Safety Administration later determined that bothAutopilot and over-reliance by the motorist on Tesla's driving aids were to blame (Felton,2017).
- In March 2018, Wei Huang was killed when his Tesla Model X crashed into a highway safetybarrier in California, USA. According to Tesla, the severity of the accident was 'unprecedented'. The National Transportation Safety Board later published a report attributing the crash to an Autopilot navigation mistake. Tesla is now being sued by the victim's family (O'Kane, 2018).

Unfortunately, efforts to investigate these accidents have been stymied by the fact that standards, processes, and regulatory frameworks for investigating accidents involving AVs have not yet beendeveloped or adopted. In addition, the proprietary data logging systems currently installed in AVsmean that accident investigators rely heavily on the cooperation of manufacturers to provide criticaldata on the events leading up to an accident (Stilgoe and Winfield, 2018).

One solution is to fit all future AVs with industry standard event data recorders — a so-called 'ethicalblack box' — that independent accident investigators could access. This would mirror the modelalready in place for air accident investigations (Sample, 2017).

Near-miss accidents

At present, there is no system in place for the systematic collection of near-miss accidents. While it is possible that manufacturers are collecting this data already, they are not under any obligation to do so — or to share the data. The only exception at the moment is the US state of California, which requires all companies that are actively testing AVs on public roads to disclose the frequency atwhich human drivers were forced to take control of the vehicle for safety reasons (known as'disengagement').

In 2018, the number of disengagements by AV manufacturer varied significantly, from one disengagement for every 11,017 miles driven by Waymo AVs to one for every 1.15 miles driven byApple AVs (Hawkins, 2019). Data on these disengagements reinforces the importance of ensuringthat human safety drivers remain engaged. However, the Californian data collection process hasbeen criticised, with some claiming its ambiguous wording and lack of strict guidelines enables companies to avoid reporting certain events that could be termed near-misses.

Without access to this type of data, policymakers cannot account for the frequency and significance of near-miss accidents, or assess the steps taken by manufacturers as a result of these near-misses. Again, lessons could be learned from the model followed in air accident investigations, in which all near misses are thoroughly logged and independently investigated. Policymakers require comprehensive statistics on all accidents and near-misses in order to inform regulation.

Data privacy

It is becoming clear that manufacturers collect significant amounts of data from AVs. As these vehiclesbecome increasingly common on our roads, the question emerges: to what extent are thesedata compromising the privacy and data protection rights of drivers and passengers?

Already, datamanagement andprivacyissues have appeared, withsomeraisingconcerns about thepotential misuse of AV data for advertising purposes (Lin, 2014). Tesla have also come under fire for the unethical use of AV data logs. In an investigation by *The Guardian*, the newspaper found multipleinstances where the company shared drivers' private data with the media following crashes, without their permission, to prove that its technology was not responsible (Thielman, 2017). At the sametime, Tesla does not allow customers to see their own data logs.

One solution, proposed by the German Ethics Commission on Automated Driving, is to ensure thatthat all AV drivers be given full data sovereignty (Ethics Commission, 2017). This would allow them to control how their data is used.

Employment

The growth of AVs is likely to put certain jobs — most pertinently bus, taxi, and truck drivers — atrisk.

In the medium term, truck drivers face the greatest risk as long-distance trucks are at the forefront of AV technology (Viscelli, 2018). In 2016, the first commercial delivery of beer was made using aself-driving truck, in a journey covering 120 miles and involving no human action (Isaac, 2016). Last year saw the first fully driverless trip in a self-driving truck, with the AV travelling seven miles without a single human on board (Cannon, 2018).

Looking further forward, bus drivers are also likely to lose jobs as more and more buses becomedriverless. Numerous cities across the world have announced plans to introduce self-driving shuttles in the future, including Edinburgh (Calder, 2018), New York (BBC, 2019a) and Singapore (BBC 2017). In some places, this vision has already become a reality; the Las Vegas shuttle famously got off to a bumpy start when it was involved in a collision on its first day of operation (Park, 2017), and touristsin the small Swiss town of Neuhausen Rheinfall can now hop on a self-driving bus to visit the nearbywaterfalls (CNN, 2018). In the medium term, driverless buses will likely be limited to routes that travel along 100% dedicated bus lanes. Nonetheless, the advance of self-driving shuttles has already created tensions with organised labour and city officials in the USA (Weinberg, 2019). Last year, the Transport Workers Union of America formed a coalition in an attempt to stop autonomous busesfrom hitting the streets of Ohio (Pfleger, 2018).

Fully autonomous taxis will likely only become realistic in the long term, once AV technology hasbeen fully tested and proven at levels 4 and 5. Nonetheless, with plans to introduce self-driving taxisin London by 2021 (BBC, 2018), and an automated taxi service already available in Arizona, USA(Sage, 2019), it is easy to see why taxi drivers are uneasy.

The quality of urban environments

In the long-term, AVs have the potential to reshape our urban environment. Some of these changesmay have negative consequences for pedestrians, cyclists and locals. As driving becomes more automated, there will likely be a need for additional infrastructure (e.g. AV-only lanes). There may also be more far-reaching effects for urban planning, with automation shaping the planning of everything from traffic congestion and parking to green spaces and lobbies (Marshall and Davies,2018). The rollout of AVs will also require that 5G network coverage is extended significantly —again, something with implications for urban planning (Khosravi, 2018).
The environmental impact of self-driving cars should also be considered. While self-driving cars havethe potential to significantly reduce fuel usage and associated emissions, these savings could becounteracted by the fact that self-driving cars make it easier and more appealing to drive long distances (Worland, 2016). The impact of automation on driving behaviours should therefore not be underestimated.

Legal and ethical responsibility

From a legal perspective, who is responsible for crashes caused by robots, and how should victims be compensated (if at all) when a vehicle controlled by an algorithm causes injury? If courts cannot resolve this problem, robot manufacturers may incur unexpected costs that would discourage investment. However, if victims are not properly compensated then autonomous vehicles are unlikelyto be trusted or accepted by the public.

Robots will need to make judgement calls in conditions of uncertainty, or 'nowin' situations. However, which ethical approach or theory should a robot be programmed to follow when there's nolegal guidance? As Lin et al. explain, different approaches can generate different results, including the number of crash fatalities.

Additionally, who should choose the ethics for the autonomous vehicle — drivers, consumers, passengers, manufacturers, politicians? Loh and Loh (2017) argue that responsibility should be shared among the engineers, the driver and the autonomous driving system itself. However, Millar (2016) suggests that the use should be able to decide what ethical or behave, of doctors, who do not have the moral authority to make important decisions on end-of-life care without the informed consent of their patients, he argues that therewould be a moral outcry if engineers designed cars without either asking the driver directly for their input, or informing the user ahead of time how the car is programmed to behave in certain situations.

Ethical dilemmas in development

In 2014, the Open Roboethics initiative (ORi 2014a, 2014b) conducted a poll asking people what they thought an autonomous car in which they were a passenger should doif a child stepped out in front of the vehicle in a tunnel. The car wouldn't have time to brake and spare the child, but could swerve into the walls of the tunnel, killing the passenger. This is a spin on the classic 'trolley dilemma', where one has the option to divert a runaway trolley from a path that would hurt several people onto the path thatwould only hurt one.

36 % of participants said that they would prefer the car to swerve into the wall, saving the child; however, the majority(64 %) said they would wish to save themselves, thussacrificing the child. 44 % of participants thought that the passenger should be able to choose the car's course of action, while 33 % said that lawmakers should choose. Only 12 % said that the car's manufacturers should make the decision. These results suggest that people do not like theidea of engineers making moral decisions on their behalf.

Asking for the passenger's input in every situation would be impractical. However, Millar (2016) suggests a 'setup'procedure where people could choose their ethics settingsafter purchasing a new car. Nonetheless, choosing how thecar reacts in advance could be seen as premeditated harm, if, for example a user programmed their vehicle to always avoid vehicle collisions by swerving into cyclists. This would increase the user's accountability and liability, whilst diverting responsibility away from manufacturers.

3) Case study: Warfare and weaponisation

Although partially autonomous and intelligent systems have been used in military technology sinceat least the Second World War, advances in machine learning and AI signify a turning point in theuse of automation in warfare.

AI is already sufficiently advanced and sophisticated to be used in areas such as satellite imageryanalysis and cyber defence, but the true scope of applications has yet to be fully realised. A recent report concludes that AI technology has the potential to transform warfare to the same, or perhaps even a greater, extent than the advent of nuclear weapons, aircraft, computers and biotechnology (Allen and Chan, 2017). Some key ways in which AI will impact militaries are outlined below.Lethal autonomous weapons

As automatic and autonomous systems have become more capable, militaries have become morewilling to delegate authority to them. This is likely to continue with the widespread adoption of AI, leading to an AI inspired arms-race. The Russian Military Industrial Committee has already approved an aggressive plan whereby 30% of Russiancombat power will consist ofentirely remote-controlledand autonomous robotic platforms by 2030. Other countries are likely to set similar goals. While the United States Department of Defense has enacted restrictions on the use of autonomous and semi-autonomous systems wielding lethal force, other countries and non-state actors may not exercisesuch self-restraint.

Drone technologies

Standard military aircraft can cost more than US$100 million per unit; a high-quality quadcopterUnmanned Aerial Vehicle, however, currently costs roughly US$1,000, meaning that for the price of a single high-end aircraft, a military could acquire one million drones. Although current commercial drones have limited range, in the future they could have similar ranges to ballistic missiles, thusrendering existing platforms obsolete.

Robotic assassination

Widespread availability of low-cost, highly-capable, lethal, and autonomous robots could make targeted assassinationmore widespread and more difficult to attribute. Automatic sniping robotscouldassassinatetargets fromafar.

Mobile-robotic-Improvised Explosive Devices

As commercial robotic and autonomous vehicle technologies become widespread, some groups will leverage this to make more advanced Improvised Explosive Devices (IEDs). Currently, the technological capability to rapidly deliver explosives to a precise target from many miles away is restricted to powerful nation states. However, if long distance package delivery by drone becomes a reality, the cost of precisely delivering explosives from afar would fall from millions of dollars to thousands or even hundreds. Similarly, self-driving cars could make suicide car bombs morefrequent and devastating since they no longer require a suicidal driver.

Hallaq et al. (2017) also highlight key areas in which machine learning is likely to affect warfare. They describe an example where a Commanding Officer (CO) could employ an Intelligent Virtual Assistant(IVA) within a fluid battlefield environment that automatically scanned satellite imagery to detect specific vehicle types, helping to identify threats in advance. It could also predict the enemy's intent, and compare situational data to a stored database of hundreds of previous wargame exercises andlive engagements, providing the CO with access to a level of accumulated knowledge that wouldotherwise be impossible to accrue.

Employing AI in warfare raises several **legal and ethical questions**. One concern is that automatedweapon systems that exclude human judgment could violate International Humanitarian Law, andthreaten our fundamental right to life andthe principle of human dignity. AI could also lowerthethreshold of going to war, affecting global stability.

International Humanitarian law stipulates that any attack needs to distinguish between combatants and non-combatants, be proportional and must not target civilians or civilian objects. Also, no attack should unnecessarily aggravate the suffering of combatants. AI may be unable to fulfil theseprinciples without the involvement

of human judgment. In particular, many researchers are concerned that Lethal Autonomous Weapon Systems (LAWS) — a type of autonomous militaryrobot that can independently search for and 'engage' targets using lethal force — maynot meetthestandards set by International Humanitarian Law, as they are not able to distinguish civilians fromcombatants, andwould not be able to judge whetherthe force of the attackwas proportional giventhe civilian damage it would incur.

Amoroso and Tamburrini (2016, p. 6) argue that: '[LAWS must be] capable of respecting theprinciples of distinction and proportionality at least as well as a competent and conscientious humansoldier'. However, Lim (2019) points out that while LAWSthat fail tomeet these requirements should not be deployed, one day LAWS *will* be sophisticated enough to meet the requirements ofdistinction and proportionality. Meanwhile, Asaro (2012) argues that it doesn't matter how goodLAWS get; it is a moral requirement that only a human should initiate lethal force, and it is simplymorally wrong to delegate life or death decisions to machines.

Some argue that delegating the decision to kill a human to a machine is an infringement of basichuman dignity, as robots don't feel emotion, and can have no notion of sacrifice and what it meansto take a life. As Lim et al (2019) explain, 'a machine, bloodless and without morality or mortality, cannot fathom the significance of using force against a human being and cannot do justice to thegravity of the decision'.

Robots also have no concept of what it means to kill the 'wrong' person. 'It is only because humanscan feel the rage and agony that accompanies the killing of humans that they can understandsacrifice and the use of force against a human. Only then can they realise the 'gravity of the decision' to kill' (Johnson and Axinn 2013, p. 136).

However, others argue that there is no particular reason why being killed by a machine would be asubjectively worse, or less dignified, experience than being killed by a cruise missile strike. 'Whatmatters is whether the victim experiences a sense of humiliation in the process of getting killed.Victims being threatened with a potential bombing will not care whether the bomb is dropped bya human or a robot' (Lim et al, 2019). In addition, not all humans have the emotional capacity toconceptualise sacrifice or the relevant emotions that accompany risk. In the heat of battle, soldiers rarely have time to think about the concept of sacrifice, or generate the relevant emotions to make informed decisions each time they deploy lethal force.

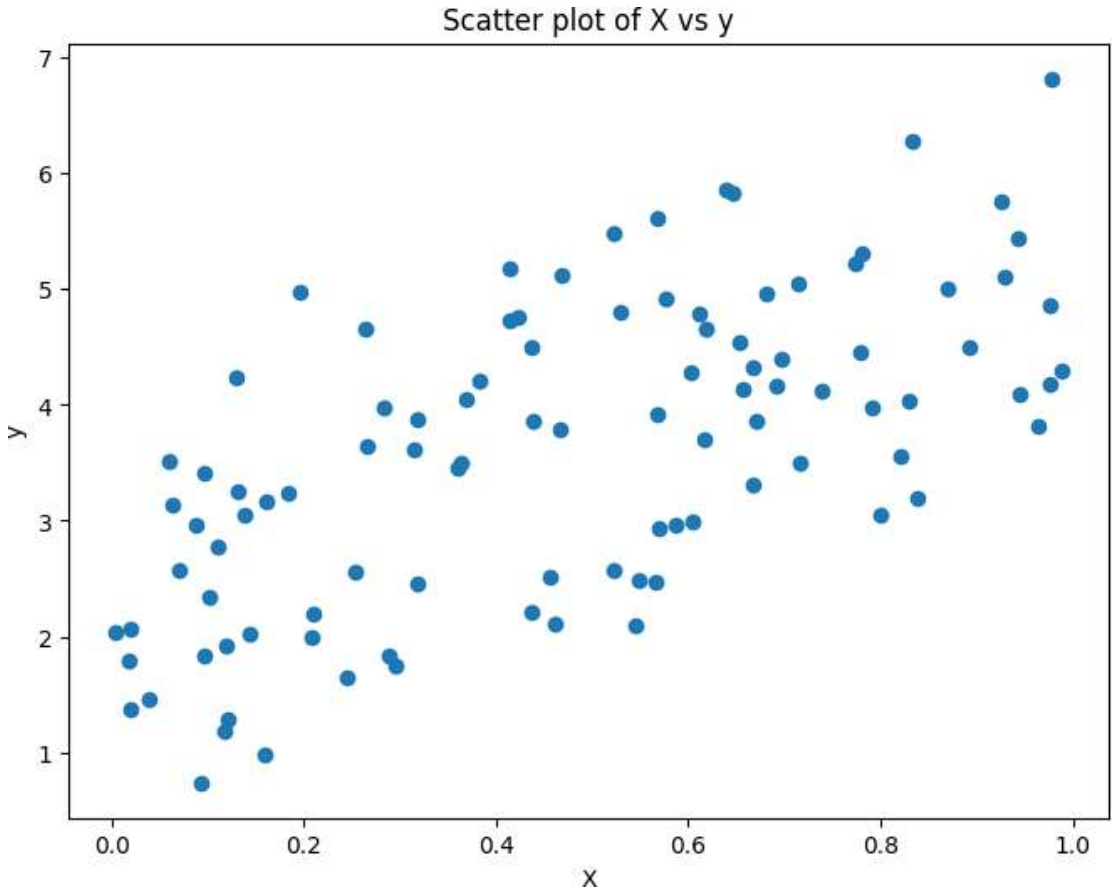
Additionally, who should be held accountable for the actions of autonomous systems — thecommander, programmer, or the operator of the system? Schmit (2013) argues that the responsibility for committing war crimes should fall on both the individual who programmed theAI, and the commander or supervisor (assuming that they knew, or should have known, theautonomous weapon system had been programmed and employed in a war crime, and that theydid nothing to stop it from happen

EXP.NO-2 Exploratory data analysis on a 2 variable linear regression model

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# Generating example data
np.random.seed(0)
X = np.random.rand(100, 1)  # Independent variable
y = 2 + 3 * X + np.random.randn(100, 1)  # Dependent variable

# Creating a DataFrame
data = pd.DataFrame(data=np.hstack([X, y]), columns=['X', 'y'])

# Scatter plot
plt.figure(figsize=(8, 6))
plt.scatter(data['X'], data['y'])
plt.title('Scatter plot of X vs y')
plt.xlabel('X')
plt.ylabel('y')
plt.show()

# Calculating correlation coefficient
correlation = data['X'].corr(data['y'])
print(f'Correlation coefficient between X and y: {correlation}')

# Fitting a linear regression model
from sklearn.linear_model import LinearRegression

model = LinearRegression()
model.fit(X, y)
```

```
# Getting model parameters
intercept = model.intercept_[0]
slope = model.coef_[0][0]
print(f'Intercept: {intercept}')
print(f'Slope: {slope}')

# Plotting the regression line
plt.figure(figsize=(8, 6))
plt.scatter(data['X'], data['y'])
plt.plot(data['X'], model.predict(X), color='red')
plt.title('Linear regression model')
plt.xlabel('X')
plt.ylabel('y')
plt.show()
```

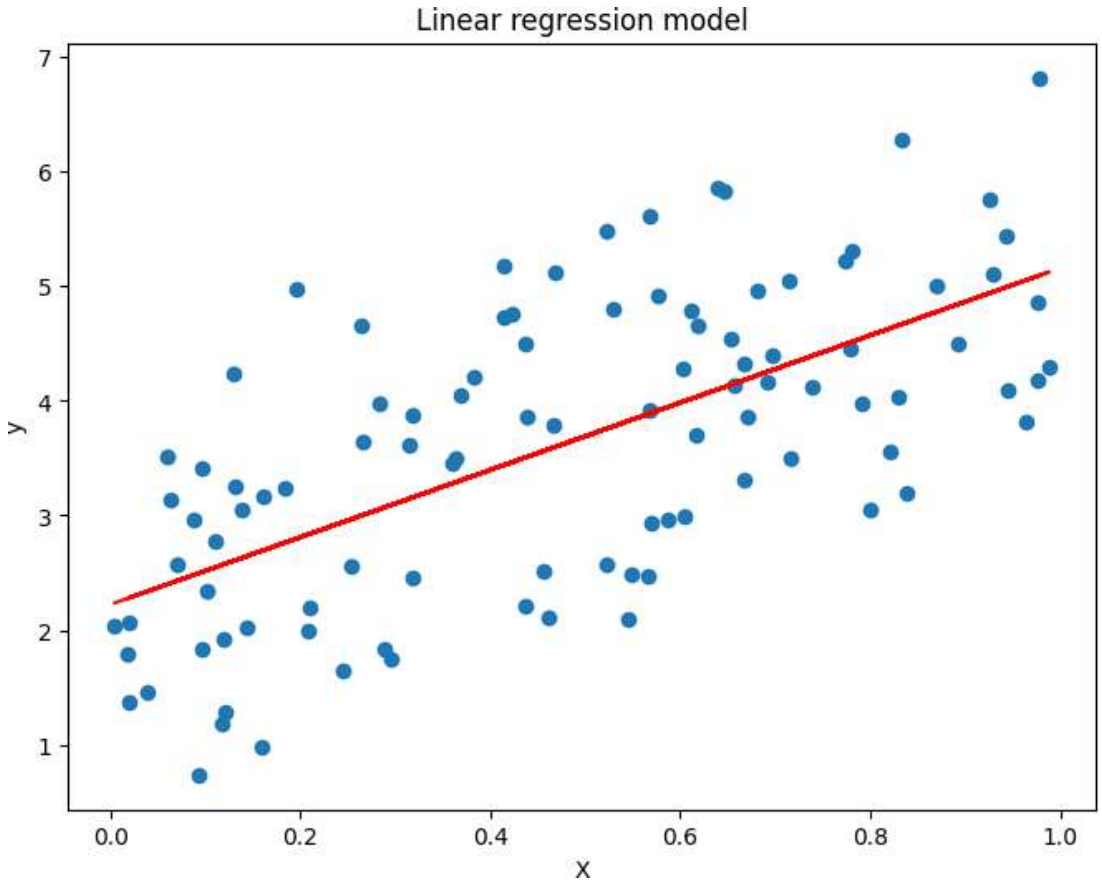
OUTPUT:



Correlation coefficient between X and y: 0.6476229996285181

Intercept: 2.2221510774472293

Slope: 2.9369350214020384

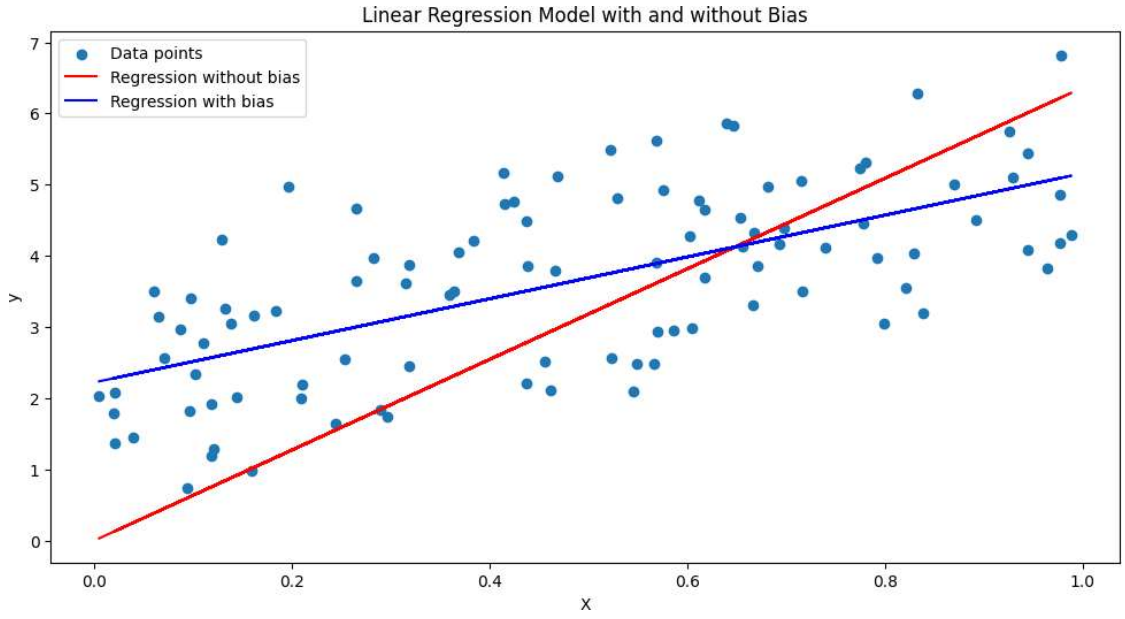

Linear regression model
7
6
5
4
3
2
1
y
0.0
0.2
0.4
0.6
0.8
1.0
X

EXP.NO-3 Experiment the regression model without a bias and with bias

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression

# Generating example data
np.random.seed(0)
X = np.random.rand(100, 1)  # Independent variable
y = 2 + 3 * X + np.random.randn(100, 1)  # Dependent variable

# Fitting a linear regression model without bias
model_no_bias = LinearRegression(fit_intercept=False)
model_no_bias.fit(X, y)

# Fitting a linear regression model with bias
model_with_bias = LinearRegression(fit_intercept=True)
model_with_bias.fit(X, y)

# Plotting the data points and regression lines
plt.figure(figsize=(12, 6))
plt.scatter(X, y, label='Data points')
plt.plot(X, model_no_bias.predict(X), color='red', label='Regression without bias')
plt.plot(X, model_with_bias.predict(X), color='blue', label='Regression with bias')
plt.legend()
plt.title('Linear Regression Model with and without Bias')
plt.xlabel('X')
plt.ylabel('y')
plt.show()

# Displaying model parameters
print("Model parameters without bias:")
```

```
print(f"Slope: {model_no_bias.coef_[0][0]}")

print("\nModel parameters with bias:")
print(f"Intercept: {model_with_bias.intercept_[0]}")
print(f"Slope: {model_with_bias.coef_[0][0]}")
```

OUTPUT:



Model parameters without bias:

Slope: 6.363033406072777

Model parameters with bias:

Intercept: 2.2221510774472293

Slope: 2.9369350214020384

EXP.NO-4 Classification of a dataset from UCI repository using a perceptron with and without bias.

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import Perceptron
from sklearn.metrics import accuracy_score

# Load the dataset from UCI repository (example with Iris dataset)
url = "https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data"
column_names = ['sepal_length', 'sepal_width', 'petal_length', 'petal_width', 'species']
data = pd.read_csv(url, names=column_names)

# Extracting features and target variable
X = data.drop('species', axis=1)
y = data['species']

# Splitting the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
# Fitting a perceptron model without bias
model_no_bias = Perceptron(fit_intercept=False)
model_no_bias.fit(X_train, y_train)
y_pred_no_bias = model_no_bias.predict(X_test)
accuracy_no_bias = accuracy_score(y_test, y_pred_no_bias)
print("Accuracy of perceptron without bias:", accuracy_no_bias)
# Fitting a perceptron model with bias
model_with_bias = Perceptron(fit_intercept=True)
model_with_bias.fit(X_train, y_train)
y_pred_with_bias = model_with_bias.predict(X_test)
accuracy_with_bias = accuracy_score(y_test, y_pred_with_bias)
print("Accuracy of perceptron with bias:", accuracy_with_bias)
```

OUTPUT:

Accuracy of perceptron without bias: 0.6

Accuracy of perceptron with bias: 0.7333333333333333

EXP.NO-5 Case study on ontology where ethics is at stake

Title: Ethical Dilemma in Healthcare Ontology

Scenario:

- In a healthcare organization, a team of data scientists and healthcare professionals is working on developing an ontology to improve patient care and treatment outcomes. The ontology aims to standardize medical terminology, facilitate data interoperability, and enhance decision-making processes.

- One day, during the development phase, the team encounters a dilemma. They realize that the ontology, if misused or misinterpreted, could potentially lead to biased decision-making, discrimination, and privacy breaches. For example, the ontology could inadvertently reinforce stereotypes, prioritize certain demographics over others, or compromise patient confidentiality.

The team faces conflicting ethical considerations:

1. On one hand, the ontology has the potential to revolutionize healthcare by enabling better data-driven decision-making, improving treatment accuracy, and enhancing patient outcomes.

2. On the other hand, the misuse of the ontology could result in ethical violations, such as discrimination based on sensitive attributes like race, gender, or socioeconomic status.

Key Stakeholders:

1. Data Scientists: Responsible for developing and maintaining the ontology.

2. Healthcare Professionals: Will use the ontology in clinical settings.

3. Patients: Directly impacted by the decisions made using the ontology.

4. Regulatory Bodies: Oversee the ethical and legal aspects of healthcare data usage.

Ethical Considerations:

1. Fairness and Bias: How can the team ensure that the ontology is unbiased and does not perpetuate systemic biases present in healthcare data?

2. Informed Consent: How should patients be informed about the use of the ontology and their data privacy rights?

3. Transparency: Should the ontology be transparent and auditable to ensure accountability and trust?

4. Accountability: Who should be held accountable for any ethical breaches related to the ontology's usage?

Resolution:

To address the ethical concerns, the team decides to:

1. Implement bias detection algorithms to identify and mitigate biases in the ontology.

2. Develop clear guidelines for patient consent, data privacy protection, and transparent communication.

3. Engage in ongoing ethical reviews and audits to monitor the ontology's impact and address any emerging issues promptly.

4. Collaborate with ethicists, patient advocates, and regulatory bodies to ensure alignment with ethical standards and legal regulations.

- By proactively addressing the ethical considerations, the team aims to harness the potential of ontology in healthcare while upholding ethical principles and safeguarding patient welfare.

- This case study highlights the complex intersection of ontology, healthcare, and ethics, emphasizing the importance of ethical awareness and responsibility in data-driven decision-making processes.

EXP.NO-6 Identification on optimization in AI affecting ethics.

The rapid advancements in Artificial Intelligence (AI) and optimization algorithms have brought about significant ethical considerations and implications. Here are some key points on how optimization in AI can impact ethics:

1. Bias and Fairness:

Optimization algorithms in AI are often trained on historical data, which can contain biases related to race, gender, or socioeconomic status. If not properly addressed, these biases can be amplified by optimization processes, leading to unfair or discriminatory outcomes. Ethical concerns arise when AI systems optimize for certain metrics at the expense of fairness and equality.

2. Transparency and Accountability:

Optimization algorithms in AI can be complex and difficult to interpret, making it challenging to understand how decisions are made. Lack of transparency can lead to accountability issues, as stakeholders may not be able to explain or challenge the outcomes produced by AI systems. Ethical considerations include the need for transparent optimization processes and mechanisms for holding AI systems accountable for their decisions.

3. Privacy and Data Protection:

Optimization in AI often involves processing large amounts of data, raising concerns about privacy and data protection. Optimization algorithms may inadvertently reveal sensitive information about individuals or groups, leading to privacy breaches. Ethical dilemmas emerge when optimizing AI systems prioritize performance over safeguarding personal data and privacy rights.

4. Manipulation and Exploitation:

Optimization algorithms can be susceptible to manipulation or exploitation by malicious actors seeking to influence outcomes for personal gain or harm. Ethical issues arise when AI systems are optimized to deceive or manipulate users, perpetuate misinformation, or engage in unethical behaviors that prioritize short-term gains over long-term societal well-being.

5. Unintended Consequences:

Optimization in AI can have unintended consequences that impact individuals, communities, or society as a whole. Ethical considerations include the need to anticipate and mitigate potential harms resulting from optimized AI systems, such as job displacement, social inequality, or loss of human autonomy. Balancing optimization goals with ethical responsibilities is crucial to minimize negative impacts.

6. Algorithmic Decision-Making:

Optimization algorithms drive decision-making processes in AI systems, influencing outcomes in various domains, including healthcare, finance, criminal justice, and social services. Ethical concerns arise when optimized algorithms make decisions that are opaque, unfair, or discriminatory, raising questions about accountability, transparency, and the potential for human oversight and intervention.

Addressing the ethical implications of optimization in AI requires a multi-faceted approach that integrates ethical principles, regulatory frameworks, stakeholder engagement, and ongoing monitoring and evaluation. By promoting ethical AI design, development, and deployment practices, we can strive to optimize AI systems responsibly and ethically, ensuring that they align with societal values and respect human rights.