Seminar Report

On

AI ETHICS:
REGULATORY
CHALLENGES
AND
OPPORTUNITIES
IN INDIA

By

Gaurav Mishra

1032221713

Under the guidance of

Dr. Chaitali Vivek Chandankhede

**School of Computer Science & Engineering
Department of Computer Engineering & Technology**

* 2024-2025   *

MIT-World Peace University (MIT-WPU)

Faculty of Engineering
School of Computer Science & Engineering
Department of Computer Engineering & Technology

<u>CERTIFICATE</u>

This is to certify that Mr. /Ms. <u>Gaurav Mishra</u> of B.Tech. CSE/AIDS/CSF, DCET, School of Computer Science & Engineering, Semester – VI, PRN. No. 1032221713, has successfully completed seminar on

AI ETHICS: REGULATORY CHALLENGES AND OPPORTUNITIES IN INDIA

to my satisfaction and submitted the same during the academic year 2024 - 2025 towards the partial fulfillment of degree of Bachelor of Technology in School of Computer Science & Engineering DCET under Dr. Vishwanath Karad MIT-World Peace University, Pune.

Dr. Balaji Patil

Seminar Guide   Program Director, DCET, SoCSE

LIST OF FIGURES & TABLES

ABBREVIATIONS

AI: Artificial Intelligence

STOA: Panel for the Future of Science and Technology

IEEE: Institute of Electrical and Electronics Engineers

 AGI: Artificial General Intelligence

ANN: Artificial Neural Network

GDPR: General Data Protection Regulation

OECD: Organisation for Economic Co-operation and Development

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my guide,Asistant Professor Dr. Chaitali Vivek Chandankhede, for their valuable guidance, encouragement, and support throughout this seminar. I would also like to thank the faculty members of MITWPU for their continuous assistance and insights. Lastly, I extend my thanks to my family and friends for their constant motivation.

Gaurav Mishra
1032221713

INDEX

ABSTRACT

The rapid advancement of Artificial Intelligence (AI) raises ethical concerns that demand immediate attention. This report presents a literature review on AI ethics, addressing fundamental principles, challenges, and frameworks. It explores fairness, accountability, transparency, privacy, and bias mitigation strategies. The research emphasizes the importance of regulatory measures, responsible AI practices, and global collaboration for ethical AI deployment. Additionally, it examines India's existing AI-related legal landscape and proposes actionable suggestions inspired by European and American regulatory models.

AI technology continues to reshape industries, governance, and social interactions, creating both opportunities and risks. While AI promises economic growth and efficiency, ethical challenges such as algorithmic bias, surveillance, loss of jobs, and autonomy concerns require comprehensive strategies. This report integrates research findings, case studies, and regulatory policies to propose a framework for ethical AI governance in India, while drawing comparisons from global AI regulations.

Keywords: AI Ethics, Fairness, Accountability, Transparency, Bias, Privacy, Regulation, Indian AI Laws, Global AI Policies, HumanCentric AI.

AI ETHICS: REGULATORY CHALLENGES AND OPPORTUNITIES IN INDIA

1. INTRODUCTION

1.1 What is AI – and what is intelligence?

AI is developing rapidly, offering many benefits but also posing ethical, legal, and social risks. This chapter introduces AI, defines intelligence, and distinguishes robots as a subset of AI.

The European Commission (2018a) defines AI as systems displaying intelligent behaviour by analyzing their environment and taking autonomous actions to achieve goals. AI may be software-based (e.g., search engines, voice assistants) or embedded in hardware (e.g., robots, autonomous cars).

Intelligence is often defined as "doing the right thing at the right time." Legg and Hunt (2007) highlight three features: (1) interaction with the environment, (2) goal achievement, and (3) adaptability. Intelligence involves learning and understanding.

A robot's environment may be a hospital, street, or home; a software AI's environment might be clinical or virtual. Human interaction is central to ethical concerns in AI.

Current AIs are "narrow"—performing specific tasks (e.g., translation, playing chess). "Artificial General Intelligence" (AGI) remains a future goal.

Machine learning enables AI to learn from data or experience. **Supervised learning** uses labeled datasets (e.g., image classification with ANNs). **Unsupervised learning** requires AI to learn patterns without labeled data, often through trial and error.

Supervised learning risks bias if training data is flawed. Unsupervised learning is slow. **Deep learning** refers to systems with many-layered ANNs and large datasets.

Importantly, AI and machine learning are not the same—many AIs do not use learning.

1.2 Definition of morality and ethics, and how that relates to AI:

Ethics are principles guiding behaviour. For instance, Kant's imperative urges us to act in ways we'd want everyone to act.

AI ethics focuses on how designers and operators can reduce harm from AI through ethical design and application. It covers short-term issues like data bias, mid-term concerns like job impact, and long-term fears like superintelligence.

AI ethics has grown from academia to global concern, as AI impacts every sector—healthcare, law, finance, and more. Numerous ethical guidelines, standards (e.g., IEEE), and AI strategies have emerged globally.

Example scenario: A bank uses an AI to approve mortgages. Though blinded to race, the algorithm disproportionately rejects black applicants. Complex models like neural networks may hide discriminatory logic, while simpler models (e.g., decision trees) are easier to audit.

AI must be **transparent**, especially in socially impactful roles. Lack of transparency and accountability can lead to unjust decisions.

Predictability is also essential—similar to legal systems relying on precedent to ensure fair expectations. AI should be **robust against manipulation**—like security systems resisting adversarial attacks.

Responsibility is another concern. When AI fails, who is accountable—developers or users? Bureaucracies may deflect blame to AI. Even with override options, users may avoid responsibility to escape blame.

In summary, AI systems replacing human social judgment must meet criteria like **responsibility, transparency, predictability, auditability, incorruptibility**, and **fairness**. These are crucial for building trust in an AI-driven society.

1.3.Organization of report

Robots and artificial intelligence (AI) come in various forms, as outlined above, each of which raises a different range of ethical concerns. These are outlined in Chapter 2: Mapping the main ethical dilemmas and moral questions associated with the deployment of AI. This chapter explores in particular: Social impacts: this section considers the potential impact of AI on the labour market and economy and how different demographic groups might be affected. It addresses questions of inequality and the risk that AI will further concentrate power and wealth in the hands of the few. Issues related to privacy, human rights and dignity are addressed as are risks that AI will perpetuate the biases, intended or otherwise, of existing social systems or their creators. This section also raises questions about the impact of AI technologies on democracy, suggesting that these technologies may operate for the benefit of state-controlled economies. Psychological impacts: what impacts might arise from human-robot relationships? How might we address dependency and deception? Should we consider whether robots deserve to be given the status of 'personhood' and what are the legal and moral implications of doing so? Financial system impacts: potential impacts of AI on financial systems are considered, including risks of manipulation and collusion and the need to build in accountability. Legal system impacts: there are a number of ways in which AI could affect the legal system, including: questions relating to crime, such as liability if an AI is used for criminal activities, and the extent to which AI might support criminal activities such as drug trafficking. In situations where an AI is involved in personal injury, such as in a collision involving an autonomous vehicle, then questions arise around the legal approach to claims (whether it is a case of negligence, which is usually the basis for claims involving vehicular accidents, or product liability). Environmental impacts: increasing use of AIs comes with increased use of natural resources, increased energy demands and waste disposal issues. However, AIs could improve the way we manage waste and resources, leading to environmental benefits. Impacts on trust: society relies on trust. For AI to take on tasks, such as surgery, the public will need to trust the technology. Trust includes aspects such as fairness (that AI will be impartial), transparency (that we will be able to understand how an AI arrived at a particular decision), accountability (someone can be held accountable for mistakes made by AI) and control (how we might 'shut down' an AI that becomes too powerful).

1.4. Artificial General Intelligence

AI ETHICS: REGULATORY CHALLENGES AND OPPORTUNITIES IN INDIA

There is broad consensus that current AI falls short of true human-level intelligence, despite excelling in narrow tasks like chess. Once AI achieves a task, it's often no longer seen as "intelligent." This points to a missing quality: generality. Artificial General Intelligence (AGI) refers to AI capable of operating across diverse domains, not just one.

Modern AI systems like Deep Blue are highly specialized—they outperform humans in one domain but cannot adapt beyond it. Unlike animals, which have fixed skill sets, humans can learn across tasks—building both hives and dams, for example. AGI aspires to this general capability.

AGI poses unique challenges. While traditional machines (like toasters or reactors) operate in narrowly defined contexts, AGI must function in unpredictable and novel environments. A toaster doesn't know it's toasting—it just follows design instructions. If misused (e.g., cloth instead of bread), it could cause harm. Task-specific AIs already go beyond this model, operating in vast, dynamic decision spaces that even their creators can't fully predict.

Deep Blue, for instance, didn't rely on pre-programmed moves; it evaluated future outcomes better than its human creators could, sacrificing predictability for performance. This highlights a core issue with AGI: safety and behavior may not be locally predictable, even with perfect programming.

Humans act toward long-term goals (like feeding themselves) across countless contexts Nature never directly prepared us for—like going to the Moon. AGI must also reason about consequences far beyond immediate actions. Specifying good AGI behavior requires non-local criteria—like ensuring outcomes are not harmful to humans—which demand foresight and ethical reasoning from the AI.

Just as we can't predict how a radically designed aircraft will fly safely without understanding its principles, we can't guarantee AGI safety without verifying how it reasons ethically. Assurances of safety must be based on verifiable, trustworthy reasoning—not mere optimism.

Constructing trustworthy AGI will require a shift in engineering: from checking behavior to understanding intention and ethical cognition. This makes AGI ethics fundamentally different from that of simpler technologies:

- The AI's specific behavior may be unpredictable, even if safe.

- Safety must be verified by understanding what the AI is trying to achieve.

- Ethical thinking must itself become a design component.

1.5. Impact On Society

1.5.1. The labour market :

People have been concerned about the displacement of workers by technology for centuries. Automation, and then mechanisation, computing, and more recently AI and robotics have been predicted to destroy jobs and create irreversible damage to the labour market. Leontief (1983), observing the dramatic improvements in the processing power of computer chips, worried that people would be replaced by machines, just as horses were made obsolete by the invention of internal combustion engines. In the past, however, automation has often substituted for human labour in the short term, but has led to the creation of jobs in the long term (Autor, 2015). Nevertheless, there is widespread concern that artificial intelligence and associated technologies could create mass unemployment during the next two decades. One recent paper concluded that new information technologies will put 'a substantial share of employment, across a wide range of occupations, at risk in the near future' (Frey and Osborne, 2013). AI is already widespread in finance, space exploration, advanced manufacturing, transportation, energy development and healthcare. Unmanned vehicles and autonomous drones are also performing functions that previously required human intervention. We have already seen the impact of automation on 'blue-collar' jobs; however, as computers become more sophisticated, creative, and versatile, more jobs will be affected by technology and more positions made obsolete.

Impact on economic growth and productivity ,Economists are generally enthusiastic about the prospects of AI on economic growth. Robotics added an estimated 0.4 percentage points of annual GDP growth and labour productivity for 17 countries between 1993 and 2007, which is of a similar magnitude to the impact of the introduction of steam engines on growth in the United Kingdom (Graetz and Michaels, 2015). Impact on the workforce It is hard to quantify the effect that robots, AI and sensors will have on the workforce because we are in the early stages of the technology revolution. Economists also disagree on the relative impact of AI and robotics. One study asked 1,896 experts about the impact of emerging technologies; 48 percent believed that robots and digital agents would displace significant numbers of both 'blue' and 'white' collar workers, with many expressing concern that this would lead to vast increases in income inequality, large numbers of unemployable people, and breakdowns in the social order (Smith and Anderson, 2014). However, the other half of the experts who responded to this survey (52%) expected that technology would not displace more jobs than it created by 2025. Those experts believed that although many jobs currently performed by humans will be substantially taken over by robots or digital agents, they have faith that human ingenuity will create new jobs, industries, and ways to make a living.

AI ETHICS: REGULATORY CHALLENGES AND OPPORTUNITIES IN INDIA

Ford (2009) issues an equally strong warning, and argues that: 'as technology accelerates, machine automation may ultimately penetrate the economy to the extent that wages no longer provide the bulk of consumers with adequate discretionary income and confidence in the future. If this issue is not addressed, the result will be a downward economic spiral'. He warns that 'at some point in the future — it might be many years or decades from now — machines will be able to do the jobs of a large percentage of the 'average' people in our population, and these people will not be able to find new jobs'. However, some economists dispute these claims, saying that although many jobs will be lost through technological improvements, new ones will be created. According to these individuals, the job gains and losses will even out over the long run. 'There may be fewer people sorting items in a warehouse because machines can do that better than humans. But jobs analysing big data, mining information, and managing data sharing networks will be created' (West, 2018).

If AI led to economic growth, it could create demand for jobs throughout the economy, including in ways that are not directly linked to technology. For example, the share of workers in leisure and hospitality sectors could increase if household incomes rose, enabling people to afford more meals out and travel (Furman and Seamans, 2018). Regardless, it is clear that a range of sectors will be affected. Frey and Osborne (2013) calculate that there is a high probability that 47 percent of U.S. workers will see their jobs become automated over the next 20 years. According to their analysis, telemarketers, title examiners, hand sewers, mathematical technicians, insurance underwriters, watch repairers, cargo agents, tax preparers, photographic process workers, new accounts clerks, library technicians, and data-entry specialists have a 99 percent chance of having their jobs computerised. At the other end of the spectrum, recreational therapists, mechanic supervisors, emergency management directors, mental health social workers, audiologists, occupational therapists, health care social workers, oral surgeons, firefighter supervisors and dieticians have less than a one percent chance of this. In a further study, the team surveyed 156 academic and industry experts in machine learning, robotics and intelligent systems, and asked them what tasks they believed could currently be automated (Duckworth et al., 2019). They found that work that is clerical, repetitive, precise, and perceptual can increasingly be automated, while work that is more creative, dynamic, and human oriented tends to be less 'automatable'. Worryingly, eight times as much work fell between 'mostly' and 'completely' automatable than between 'mostly not' and 'not at all' automatable, when weighted by employment. Activities classified as 'reasoning and decision making' and 'coordinating, developing, managing, and advising' were less likely than others to be automatable, while 'administering', 'information and data processing' and 'performing complex and technical activities' were likely to be more so. Overall the model predicted very high automation potential for office, administrative support, and sales occupations, which together employ about 38 million people in the U.S. Also at high risk of automation were physical processes such as production, farming, fishing and forestry, and transportation and material moving, which employ about 20 million people in total. In contrast, occupations that were robust to automation included education, legal, community service, arts, and media occupations, and to a lesser extent, management, business, and financial occupations. Unsurprisingly, the study found that occupations with the highest salaries and levels of education tend to be the least amenable to automation. However, even this does not guarantee that an

occupation's activities cannot be automated. As the authors point out, air traffic controllers earn about US$125,000 a year, but it is thought that their tasks could largely be automated. In contrast, preschool teachers and teaching assistants earn under $30,000 a year, yet their roles are not thought to be amenable to automation.

1.5.2. Labour-market discrimination: effects on different demographics :

The impacts of these sizeable changes will not be felt equally by all members of society. Different demographics will be affected to varying extents, and some are more at risk than others from emerging technologies. Those with few technical skills or specialty trades will face the most difficulties (UK Commission for Employment and Skills, 2014). Young people entering the labour market will also be disproportionately affected, since they are at the beginning of their careers and they will be the first generation to work alongside AI (Biavaschi et al., 2013). Even though many young people have time to acquire relevant expertise, few gain training in science, technology, engineering, and math (STEM) fields, limiting their ability to withstand employment alterations. According to the U.S. Department of Education (2014), there will be a 14 percent increase in STEM jobs between 2010 and 2020 — but 'only 16 percent of American high school seniors are proficient in mathematics and interested in a STEM career'. Women may also be disproportionately affected, as more women work in caregiving positions — one of the sectors likely to be affected by robots. Due to discrimination, prejudice and lack of training, minorities and poor people already suffer high levels of unemployment: without high-skill training, it will be more difficult for them to adapt to a new economy. Many of these individuals also lack access to high-speed Internet, which limits their ability to access education, training and employment (Robinson et al., 2015). Special Eurobarometer survey 460 identified that EU residents have a largely positive response to the increasing use of digital technology, considering it to improve society, the economy, and their quality of life, and that most also consider themselves competent enough to make use of this technology in various aspects of their life and work (European Commission, 2017). However, crucially, this attitude varied by age, location, and educational background — a finding that is central to the issue of how AI will affect different demographics and the potential issues arising around the 'digital divide'. For instance, young men with high levels of education are the most likely to hold positive views about digitisation and the use of robots — and are also the most likely to have taken some form of protective measure relating to their online privacy and security (thus placing them at lower risk in this area). These kinds of socio-demographic patterns highlight a key area of concern in the increasing development and implementation of AI if nobody is to be disadvantaged or left behind (European Commission, 2017). Consequences 'When we're talking about 'AI for good', we need to define what 'good' means. Currently, the key performance indicators we look to are framed around GDP. Not to say it's evil, but it's about measuring productivity and exponential profits'. (John Havens) It is possible that AI and robotic technologies could exacerbate existing social and economic divisions, via putting current job classes at risk, eliminating jobs, causing mass unemployment in automatable job sectors. Discrimination may also be an issue, with young people potentially being disproportionately affected, alongside those without high-skill training.

1.6 Inequality

'The biggest question around AI is inequality, which isn't normally included in the debate about AI ethics. It is an ethical issue, but it's mostly an issue of politics – who benefits from AI?' (Jack Stilgoe) AI and robotics technology are expected to allow companies to streamline their businesses, making them more efficient and more productive. However, some argue that this will come at the expense of their human workforces. This will inevitably mean that revenues will be split across fewer people, increasing social inequalities. Consequently, individuals who hold ownership in AI-driven companies are set to benefit disproportionately.

Concentration of power among elites 'Does AI have to increase inequality? Could you design systems that target, for example, the needs of the poorest people? If AI was being used to further benefit rich people more than it benefits poor people, which it looks likely to be, or more troublingly, put undue pressure on already particularly marginalised people, then what might we do about that? Is that an appropriate use of AI?' (Jack Stilgoe) Nemitz (2018) writes that it would be 'naive' to ignore that AI will concentrate power in the hands of a few digital internet giants, as 'the reality of how [most societies] use the Internet and what the Internet delivers to them is shaped by a few mega corporations...the development of AI is dominated exactly by these mega corporations and their dependent ecosystems'. The accumulation of technological, economic and political power in the hands ofthe top five players – Google, Facebook, Microsoft, Apple and Amazon – affords them undue influence in areas of society relevant to opinionbuilding in democracies: governments, legislators, civil society, political parties, schools and education, journalism and journalism education and — most importantly — science and research. In particular, Nemitz is concerned that investigations into the impact of new technologies like AI on human rights, democracy and the rule of law may be hampered by the power of tech corporations, who are not only shaping the development and deployment of AI, but also the debate on its regulation. Nemitz identifies several areas in which tech giants exert power:

1.   Financial. Not only can the top five players afford to invest heavily in political and societal influence, they can also afford to buy new ideas and start-ups in the area of AI, or indeed any other area of interest to their business model — something they are indeed doing.

2.   Public discourse. Tech corporations control the infrastructures through which public discourse takes place. Sites like Facebook and Google increasingly become the main, or even only, source of political information for citizens, especially the younger generation, to the detriment of the fourth estate. The vast majority of advertising revenue now also goes to Google and Facebook, removing the main income of newspapers and rendering investigative journalism unaffordable.

3.   Collecting personal data. These corporations collect personal data for profit, and profile people based on their behaviour (both online and offline). They know more about us than ourselves or our friends — and they are using and making available this information for profit, surveillance, security and election campaigns.

4.Privacy, human rights and dignity

AI will have profound impacts on privacy in the next decade. The privacy and dignity of AI users must be carefully considered when designing service, care and companion robots, as working in people's homes means they will be privy to intensely private moments (such as bathing and dressing). However, other aspects of AI will also affect privacy. Smith (2018), President of Microsoft, recently remarked: '[Intelligent 3] technology raises issues that go to the heart of fundamental human rights protections like privacy and freedom of expression. These issues heighten responsibility for tech companies that create these products. In our view, they also call for thoughtful government regulation and for the development of norms around acceptable uses.

2. Privacy and Data Rights:

AI technologies like Intelligent Personal Assistants (IPAs) – Amazon Echo, Google Home, and Siri – raise privacy concerns as they continuously listen and collect user data. Surveys show users fear hacking (68.63%) and data misuse, though many still use these devices due to convenience.

Big Data and AI have eroded anonymity, as algorithms can analyze vast records to reveal identities and preferences. Facial recognition, online behavior, and social media patterns can predict personality, political views, and emotional states, sometimes turning harmless data into sensitive insights.

Machine learning needs large datasets, but users have limited control over how their data is used. While regulations like the EU's GDPR protect personal data, they often exclude anonymous aggregated data used in model training. Worse, trained models can sometimes be reverse-engineered to reveal personal information.

Ethical concerns arise over individuals' rights to control data in trained models. Issues include consent, data removal, and being informed about how models are used. Scholars advocate rights such as accessing, erasing, and tracking trained models using personal data.

Human Rights

AI affects democracy and dignity. Predictive analytics may expose individuals' political beliefs, leading to targeted manipulation, discrimination, or persecution. Such profiling could result in bullying, job loss, or worse in authoritarian regimes.

Surveillance

AI-enabled surveillance—via drones, facial recognition, and smart devices—is expanding into homes and public spaces, raising deep privacy concerns. These systems may monitor behavior without consent and be used by governments for control. In countries like China, AI surveillance has reportedly contributed to mass detentions based on ethnicity or beliefs.

In India, AI is used by law enforcement to digitize criminal records and monitor behavior with drones. While aimed at safety, these measures risk infringing on privacy and are often implemented without safeguards or oversight.

Freedom of Speech

AI is used to combat fake news and hate speech, but tools like sentiment analysis risk misinterpreting context, leading to unjust censorship. In India, such tools are used to remove content and monitor social media. Surveillance also deters free expression, causing self-censorship among citizens.

Democracy

AI can undermine democratic processes through targeted misinformation. During elections in the US and UK, bots spread biased content to sway opinions. Campaigns now use AI to micro-target voters, manipulating behavior using personal data (e.g., Cambridge Analytica).

Automated accounts and cyber troops are used in many countries to influence online discourse, suppress dissent, and create false impressions of public support. AI-driven propaganda challenges the integrity of democratic elections and public trust.

News Bubbles and Echo Chambers

AI-powered news feeds personalize content but reduce exposure to diverse opinions, creating filter bubbles. This contributes to societal polarization, as individuals are only shown viewpoints aligned with their beliefs—undermining public dialogue essential to democracy.

The End of Democracies?

Some argue AI may favor authoritarian regimes. Unlike democracies, such states can use unrestricted access to vast datasets to boost AI development. With no strong privacy protections (e.g., in China), these governments may gain a technological edge while disregarding individual rights. The convenience and efficiency of AI may lead societies to accept such models, prioritizing control over freedom.

Who is Responsible?

A major ethical and legal challenge is determining who is responsible for AI actions—developers, users, manufacturers, or the AI itself. In self-learning systems, where behavior evolves beyond original programming, accountability becomes complex.

The European Parliament attributes responsibility to human actors but acknowledges challenges with adaptive systems. The idea of distributed agency suggests AI actions result from the combined input of many stakeholders, requiring shared responsibility.

In military or autonomous systems, the responsibility may shift depending on design, deployment, or context. When AI behaves unpredictably, a "responsibility gap" emerges—complicating legal and moral accountability. New frameworks are needed to address these challenges.

Tort Law

Tort law deals with harm caused by one party to another, relevant to AI-related injuries. In the case of autonomous vehicles, liability traditionally falls under negligence (driver fault) or product liability (manufacturer fault).

With self-driving cars, software errors may lead to accidents. If the defect could have been avoided through a better design, it may be considered unsafe under the risk-utility test. However, these cases are complex, needing expert analysis of algorithms and sensor data, making litigation costly and difficult.

This legal shift from straightforward traffic liability to complex product liability cases may hinder victims' ability to receive timely compensation.

3. International ethical initiatives

While official regulation remains scarce, many independent initiatives have been launched internationally to explore these – and other – ethical quandaries. The initiatives explored in this section are outlined in Table 2.1 and will be studied in light of the associated harms and concerns they aim to understand and mitigate.

Table 3.1(International ethical initiatives)

Initiative	Location	Key issues tackled	Publications	Sources of funding
The Institute for Ethics in Artificial Intelligence	Germany	Human-centric engineering and a focus on the cultural and social anchoring of rapid advances in AI, covering disciplines including philosophy, ethics, sociology, and political science.		Initial (2019) funding grant from Facebook ($7.5 million over five years).
The Institute for Ethical AI & Machine Learning	United Kingdom	The Institute aims to empower all from individuals to entire nations to develop AI, based on eight principles for responsible machine learning: these concern the maintenance of human control, appropriate redress for AI impact, evaluation of bias, explicability, transparency, reproducibility, mitigation of the effect of AI automation on workers, accuracy, cost, privacy, trust, and security.		unknown
The Institute for Ethical Artificial Intelligence in Education	United Kingdom	The potential threats to young people and education of the rapid growth of new AI technology, and ensuring the ethical development of AI-led EdTech.		unknown
The Future of Life Institute	United States	Ensuring that the development of AI is beneficial to humankind, with a focus on safety and existential risk: autonomous weapons arms race, human control of AI, and the potential dangers of advanced 'general/strong' or super-intelligent AI.	*'Asilomar AI Principles'*	Private. Top donors: Elon Musk (SpaceX and Tesla), Jaan Tallinn (Skype), Matt Wage (financial trader), Nisan Stiennon (software engineer), Sam Harris, George Godula (tech entrepreneur), and Jacob Trefethen (Harvard).
The Association for Computing Machinery	United States	The transparency, usability, security, accessibility, accountability, and digital inclusiveness of computers and networks, in terms of research, development, and implementation.	Statements on: algorithmic transparency and accountability (January 2017), computing and network security (May 2017), the Internet of Things (June 2017), accessibility, usability, and digital inclusiveness (September 2017),	unknown

The Foundation for Responsible Robotics	The Netherlands	Responsible robotics (in terms of design, development, use, regulation, and implementation). Proactively taking stock of the issues that accompany technological innovation, and the impact these will have on societal values such as safety, security, privacy, and well-being.		unknown
AI4People	Belgium	The social impacts of AI, and the founding principles, policies, and practices upon which to build a 'good AI society'.	*'Ethical Framework for a Good AI Society'*	Atomium— European Institute for Science, Media and Democracy. Some funding was provided to the project's Scientific Committee Chair from the Engineering and Physical Sciences Research Council.
The Ethics and Governance of Artificial Intelligence Initiative	United States	Seeks to ensure that technologies of automation and machine learning are researched, developed, and deployed in a way which vindicate social values of fairness, human autonomy, and justice.		The Harvard Berkman Klein Center and the MIT Media Lab. Supported by The Miami Foundation (fiscal sponsorship), Knight Foundation, Luminate, Red Hoffman, and the William and Flora Hewlett Foundation.
Saidot: Enabling responsible AI ecosystems	Finland	Helping companies, governments, and organisations develop and deploy responsible AI ecosystems, to deliver transparent, accountable, trustworthy AI services. Enabling organisations to develop human-centric AI, with a focus on increasing the levels of trust and accountability in AI ecosystems. The platform offers software and algorithmic systems that can 'validate [an] intelligence system's trustworthiness' (Saidot, 2019)		
euRobotics	Europe	Maintaining and extending European talent and progress in robotics – AI industrialisation and economic impact.		European Commission

4.India's Approach to leadership in AI

AI refers to the ability of machines to perform cognitive tasks like thinking, perceiving, learning, problem solving and decision making. Initially conceived as a technology that could mimic human intelligence, AI has evolved in ways that far exceed its original conception. With incredible advances made in data collection, processing and computation power, intelligent systems can now be deployed to take over a variety of tasks, enable connectivity and enhance productivity. As AI's capabilities have dramatically expanded, so have its utility in a growing number of fields. The truly transformative nature of the technology, yet the nascent stage of its adoption worldwide, provides India with an opportunity to define its own brand of AI leadership. #AIforAll - the brand proposed for India implies inclusive technology leadership, where the full potential of AI is realised in pursuance of the country's unique needs and aspirations. The strategy should strive to leverage AI for economic growth, social development and inclusive growth, and finally as a "Garage" for emerging and developing economies. While AI has the potential to provide large incremental value to a wide range of sectors, adoption till date has been driven primarily from a commercial perspective. Technology disruptions like AI are once-in-a generation phenomenon, and hence large-scale adoption strategies, especially national strategies, need to strike a balance between narrow definitions of financial impact and the greater good. NITI Aayog has decided to focus on five sectors that are envisioned to benefit the most from AI in solving societal needs: a) Healthcare: increased access and affordability of quality healthcare, b) Agriculture: enhanced

farmers' income, increased farm productivity and reduction of wastage, c) Education: improved access and quality of education, d) Smart Cities and Infrastructure: efficient and connectivity for the burgeoning urban population, and e) Smart Mobility and Transportation: smarter and safer modes of transportation and better traffic and congestion problems. To truly reap the benefits of deploying AI at scale, the report identifies the following barriers that need to be addressed in order to achieve the goals of #AIforAll: a) Lack of broad based expertise in research and application of AI, b) Absence of enabling data ecosystems – access to intelligent data, c) High resource cost and low awareness for adoption of AI, d) Privacy and security, including a lack of formal regulations around anonymisation of data, and e) Absence of collaborative approach to adoption and application of AI. Superior research capabilities have been the cornerstone of leadership aspirations in emerging technologies and effectively realising the growth potential requires expertise in both core and applied research. Despite indications of recent positive efforts in this aspect of technology, AI research in India is still in its infancy and requires large scale concerted and collaborative interventions. The paper proposes a two-tiered structure to address India's AI research aspirations: a) Centre of Research Excellence (CORE) focused on developing better understanding of existing core research and pushing technology frontiers through creation of new knowledge; b) International Centers of Transformational AI (ICTAI) with a mandate of developing and deploying application-based research. Private sector collaboration is envisioned to be a key aspect of ICTAIs.

The research capabilities are proposed to be complemented by an umbrella organisation responsible for providing direction to research efforts through analysis of socio-economic indicators, studying global advancements, and encouraging international collaboration. Pursuing "moonshot research projects" through specialised teams, development of a dedicated supranational agency to channel research in solving big, audacious problems of AI – "CERN for AI", and developing common computing and other related infrastructure for AI are other key components research suggested. As technology increasingly disrupts the nature of jobs and shifts the benchmarks of technological aptitude, skilling and reskilling of workforce forms an integral part of our approach to adopting AI. There is an emergent need for reskilling the existing workforce and developing future talent in accordance with the changing needs of the job market. This could be done via the adoption of decentralised teaching mechanisms working in collaboration with the private sector and educational institutions to prescribe certification with value. Furthermore, promotion of job creation in new areas, like data annotation needs to be identified and promoted, as these would have the potential of absorbing a large portion of the workforce that may find itself redundant due to increasing automation. Adoption of AI across the value chain viz. startups, private sector, PSUs and government entities, will truly unlock the potential by creating a virtuous cycle of supply and demand. The barriers to AI development and deployment can effectively be addressed by adopting the marketplace model – one that enables market discovery of not only the price but also of different approaches that are best suited to achieve the desired results. A three-pronged, formal marketplace could be created focusing on data collection and aggregation, data annotation and deployable models. There could be a common platform called the National AI Marketplace (NAIM). Furthermore, for accelerated adoption of a highly collaborative technology like AI, the government has to play the critical role

of a catalyst in supporting partnerships, providing access to infrastructure, fostering innovation through research and creating the demand by seeking solutions for addressing various governmental needs. As AI-based solutions permeate the way we live and do business, questions on ethics, privacy and security will also emerge. Most discussions on ethical considerations of AI are a derivation of the FAT framework (Fairness, Accountability and Transparency). A consortium of Ethics Councils at each Centre of Research Excellence can be set up and it would be expected that all COREs adhere to standard practice while developing AI technology and products. Data is one of the primary drivers of AI solutions, and thus appropriate handling of data, ensuring privacy and security is of prime importance. Challenges include data usage without consent, risk of identification of individuals through data, data selection bias and the resulting discrimination of AI models, and asymmetry in data aggregation. The paper suggests establishing data protection frameworks and sectorial regulatory frameworks, and promotion of adoption of international standards. In order for India to ride the AI innovation wave, a robust intellectual property framework is required. Despite a number of government initiatives in strengthening the IP regime, challenges remain, especially in respect of applying stringent and narrowly focused patent laws to AI applications – given the unique nature of AI solution development. The importance of data to development of useful models is one such example. To tackle these issues, establishment of IP facilitation centers to help bridge the gap between practitioners and AI developers, and adequate training of IP granting authorities, judiciary and tribunals is suggested. The AI strategy is aimed at primarily guiding an inevitable wave of change for quicker and better impact. The AI ecosystem is rapidly evolving and taking societies into uncharted territory. For now, we can begin to ask some of the big questions that each society must answer for itself: are we ready to manage data ethically? How do we bridge the digital divide? Which innovations are worthy of public funds and partnerships? Bringing these questions into the open is the most important step in ensuring that AI advances create a better society. There has been tremendous activity concerning AI policy in different countries over the past couple of years. Governments in USA, UK, France, Japan and China have released their policy and strategy papers relating to AI. In order to establish a leadership role, it is important for India to take the plunge and start by releasing a Strategy Paper to initiate the roll out of an ambitious programme that would ensure for India its rightful place in this transformational era.

4.1. Identifying priority areas for India's efforts in Artificial Intelligence :

A national AI strategy needs to be premised on a framework which is adapted to India's unique needs and aspirations, while at the same time, is capable of achieving the country's full potential of leveraging AI developments. Such a framework could be seen as an aggregation of the following three distinct, yet inter-related components: a) Opportunity: the economic impact of AI for India b) AI for Greater Good: social development and inclusive growth c) AI Garage for
40% of the world: solution provider of choice for the emerging and developing economies (exChina) across the globe Opportunity: the economic impact of Artificial Intelligence for India
AI is emerging as a new factor of production, augmenting the traditional factors of production viz. labor, capital and innovation and technological changes captured in total factor productivity.

AI ETHICS: REGULATORY CHALLENGES AND OPPORTUNITIES IN INDIA

AI has the potential to overcome the physical limitations of capital and labour, and open up new sources of value and growth. From an economic impact perspective, AI has the potential to drive growth through enabling:

(a)   intelligent automation i.e. ability to automate complex physical world tasks that require adaptability and agility across industries,

(b)   labour and capital augmentation: enabling humans to focus on parts of their role that add the most value, complementing human capabilities and improving capital efficiency, and

(c) innovation diffusion i.e. propelling innovations as it diffuses through the economy.

AI innovations in one sector will have positive consequences in another, as industry sectors are interdependent based on value chain. Economic value is expected to be created from the new goods, services and innovations that AI will enable. Accenture, in its recent AI research reports5 , provides a framework for evaluating the economic impact of AI for select G20 countries and estimates AI to boost India's annual growth rate by 1.3 percentage points by 2035.
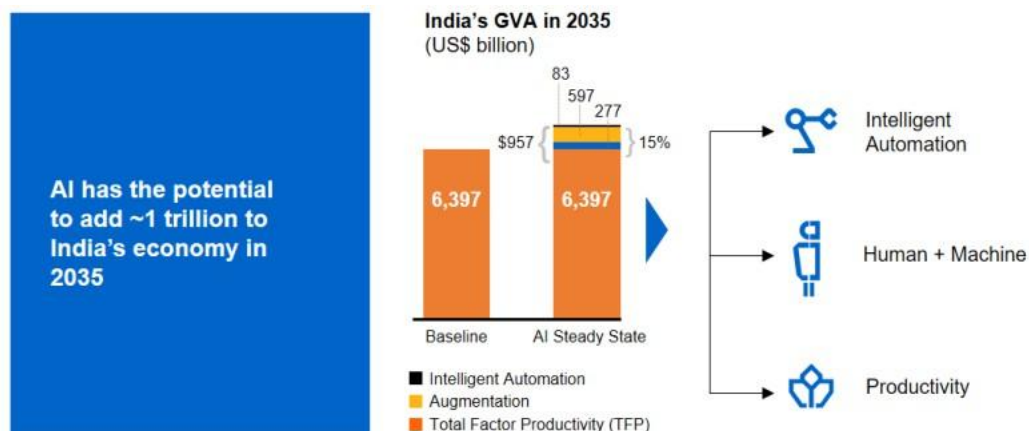
Figure 4.1(source : Accenture)

Artificial Intelligence has the potential to provide large incremental value to a wide range of sectors globally, and is expected to be the key source of competitive advantage for firms. a) Healthcare: Application of AI in healthcare can help address issues of high barriers to access to healthcare facilities, particularly in rural areas that suffer from poor connectivity and limited supply of healthcare professionals. This can be achieved through implementation of use cases such as AI driven diagnostics, personalised treatment, early identification of potential pandemics, and imaging diagnostics, among others. b) Agriculture: AI holds the promise of driving a food revolution and meeting the increased demand for food (global need to produce 50% more food and cater to an additional 2 billion people by 2050 as compared to today). It also has the potential

to address challenges such as inadequate demand prediction, lack of assured irrigation, and overuse / misuse of pesticides and fertilisers. Some use cases include improvement in crop yield through real time advisory, advanced detection of pest attacks, and prediction of crop prices to inform sowing practices. c) Smart Mobility, including Transports and Logistics: Potential use cases in this domain include autonomous fleets for ride sharing, semi-autonomous features such as driver assist, and predictive engine monitoring and maintenance. Other areas that AI can impact include autonomous trucking and delivery, and improved traffic management. d) Retail: The retail sector has been one of the early adopters of AI solutions, with applications such as improving user experience by providing personalised suggestions, preference-based browsing and image-based product search. Other use cases include customer demand anticipation, improved inventory management, and efficient delivery management. e) Manufacturing: Manufacturing industry is expected to be one of the biggest beneficiaries of AI based solutions, thus enabling 'Factory of the Future' through flexible and adaptable technical systems to automate processes and machinery to respond to unfamiliar or unexpected situations by making smart decisions. Impact areas include engineering (AI for R&D efforts), supply chain management (demand forecasting), production (AI can achieve cost reduction and increase efficiency), maintenance (predictive maintenance and increased asset utilisation), quality assurance (e.g. vision systems with machine learning algorithms to identify defects and deviations in product features), and in-plant logistics and warehousing. f) Energy: Potential use cases in the energy sector include energy system modelling and forecasting to decrease unpredictability and increase efficiency in power balancing and usage. In renewable energy systems, AI can enable storage of energy through intelligent grids enabled by smart meters, and also improve the reliability and affordability of photovoltaic energy. Similar to the manufacturing sector, AI may also be deployed for predictive maintenance of grid infrastructure. g) Smart Cities: Integration of AI in newly developed smart cities and infrastructure could also help meet the demands of a rapidly urbanising population and providing them with enhanced quality of life. Potential use cases include traffic control to reduce congestion and enhanced security through improved crowd management. h) Education and Skilling: AI can potentially solve for quality and access issues observed in the Indian education sector. Potential use cases include augmenting and enhancing the learning experience through personalised learning, automating and expediting administrative tasks, and predicting the need for student intervention to reduce dropouts or recommend vocational training.

4.2. Current AI Laws in India and Global Suggestions:

India does not yet have a dedicated AI law as of March 2025. Instead, it relies on existing frameworks, policies, and recommendations that only partially address AI-related issues.

AI ETHICS: REGULATORY CHALLENGES AND OPPORTUNITIES IN INDIA

Absence of Specific AI Laws
India lacks clear legal provisions for AI governance, creating uncertainty in areas like liability, accountability, and ethics.

Digital Personal Data Protection Act (DPDPA) 2023
The DPDPA regulates personal data use, influencing AI training datasets. It strengthens user rights but does not explicitly address AI-related governance, leaving loopholes in managing automated decision-making.

Information Technology Act, 2000 (IT Act)
The IT Act addresses data breaches and privacy but lacks clarity on AI ethics, algorithmic accountability, and liability in cases of AI-related harm.

NITI Aayog's National Strategy for AI (2018)
The #AIForAll strategy promotes ethical AI in sectors like healthcare and education but is not legally binding, limiting its impact.

MeitY Initiatives
MeitY has published white papers on AI ethics and data governance but has not enacted binding regulations. The lack of a centralized AI authority hampers enforcement.

TRAI Recommendations
TRAI has proposed a risk-based AI classification system and suggested creating an AI regulatory body—Artificial Intelligence and Data Authority of India (AIDAI). These ideas are yet to be implemented.

5. Proposed AI Ethics Law for India: A Multi-Stakeholder Approach

India needs a structured AI ethics law involving collaboration between the government, academia, industry, and civil society.

1. Establishment of a Central AI Regulatory Authority
AIDAI should oversee AI regulation, with MeitY coordinating with NITI Aayog and DoT. Ethics review committees must assess high-risk AI systems in sectors like healthcare and law enforcement.

2. Ethical Guidelines for AI Development and Deployment
AI systems must ensure transparency and explainability. Institutions like IITs should develop bias detection standards. Privacy laws like DPDPA need to be strengthened. Critical AI decisions should require human oversight.

3. Legal and Accountability Framework
AI liability laws must hold developers and companies responsible for harm. Guidelines should prevent AI from violating intellectual property rights. Whistleblowers exposing unethical AI practices should be protected.

4. Industry Participation and Self-Regulation
Companies like TCS, Infosys, and Wipro should create internal ethical AI guidelines. Startups working on responsible AI should be incentivized. Public-private partnerships must drive ethical AI development.

5. AI Ethics Curriculum and Public Awareness
AI ethics should be included in IIT and IIM courses. Public consultations should precede lawmaking. Government-led programs must promote AI literacy among citizens.

Conclusion: Towards a Balanced and Responsible AI Ecosystem

India must adopt an enforceable AI ethics law that promotes innovation while protecting citizens' rights. By learning from global regulations and adapting to India's socio-economic needs, the country can build a responsible and inclusive AI future.

6.LITERATURE SURVEY :

Artificial Intelligence (AI) has made significant strides in recent years, impacting various sectors, including healthcare, education, finance, and governance. The ongoing debates surrounding the ethics of AI, its regulation, and its impact on the workforce are crucial in shaping the future of this technology. The following literature review explores the critical dimensions of AI, focusing on ethics, societal impact, and policy frameworks.

1. AI Ethics and Initiatives

AI's ethical implications are increasingly coming under scrutiny, particularly as it becomes integrated into everyday life. Bird et al. (2020) examine the various ethical issues raised by AI technologies, emphasizing the need for ethical frameworks to guide the development of AI systems. The European Parliamentary Research Service (EPRS) report highlights ongoing initiatives and the role of governance in ensuring that AI is developed and used responsibly. As AI becomes more pervasive, questions about privacy, accountability, bias, and transparency are becoming more pressing.

2. The Impact of AI on Employment

The impact of AI on employment and labor markets is a central theme in the ongoing discourse surrounding its development. Frey and Osborne (2013) in their study on the future of employment argue that a large proportion of jobs, particularly those involving routine tasks, are susceptible to automation. Their analysis, based on the likelihood of job computerization, raises

concerns about the displacement of workers and the potential creation of a global underclass, as discussed by Gray and Suri (2019). As AI systems evolve, the workforce may face unprecedented challenges, necessitating policies that prioritize reskilling and transition plans for displaced workers.

3. AI Regulations and Policy Frameworks

The regulation of AI is a crucial step in ensuring that its development aligns with societal values and objectives. In the EU, the proposal for a regulation on AI, as laid out by the European Commission (2021), is a comprehensive effort to create a legal framework for AI technologies. The proposal emphasizes safety, transparency, and accountability in the development of AI systems, with a focus on high-risk applications. In the United States, the White House's executive order (2023) further emphasizes the need for safe, secure, and trustworthy AI, promoting ethical guidelines and fostering research in AI safety.

The U.S. National Institute of Standards and Technology (NIST) also released its AI Risk Management Framework (2023), which provides guidelines for managing risks associated with AI systems. This framework offers a systematic approach to the identification, assessment, and mitigation of risks in AI development, contributing to more responsible AI deployment.

In India, the National Strategy for Artificial Intelligence (2018) by NITI Aayog focuses on the potential for AI to drive national development, emphasizing the need for a strategic and inclusive approach to AI adoption. This document calls for policies that promote innovation while ensuring that AI applications are ethically sound and contribute positively to society.

4. Defining Intelligence and AI

The definition of intelligence itself has been debated for decades. Legg and Hutter (2007) provide an extensive collection of definitions of intelligence, which are critical in understanding the underlying principles of AI. These definitions not only help to clarify the capabilities of AI systems but also offer insights into how these systems should be developed and assessed in terms of their cognitive functions.

5. AI and the Future of Humanity

The ethical challenges associated with AI are not limited to its societal impact but also extend to the relationship between humans and machines. Havens (2016) in *Heartificial Intelligence* explores the intersection of AI and human values, arguing that embracing our humanity is key to maximizing the benefits of AI while minimizing its risks. The book provides a philosophical perspective on AI, stressing the need to balance technological advancement with human dignity and empathy.

6. Statistical and Machine Learning Foundations

Understanding the technical foundations of AI is essential to navigating its broader implications. Hastie, Tibshirani, and Friedman (2001) in *The Elements of Statistical Learning* present a comprehensive guide to statistical learning and machine learning techniques that power AI

systems. This text lays the groundwork for understanding the algorithms that drive AI, highlighting their potential and limitations. It is crucial for both developers and policymakers to understand these technical aspects to make informed decisions about AI regulation and deployment.

7. RESEARCH COMPONENT

Introduction

Artificial Intelligence (AI) is reshaping India's industries, from healthcare to agriculture, promising economic growth and societal benefits. However, as of March 2025, India's lack of a dedicated AI ethics law leaves citizens vulnerable to risks like algorithmic bias, privacy breaches, and job displacement. This blog explores India's regulatory gaps and proposes a hybrid framework inspired by European and American models to ensure ethical AI deployment.

The Current State of AI Ethics in India

India's #AIForAll strategy by NITI Aayog [6] aims to leverage AI for inclusive growth, targeting sectors like healthcare and education. Yet, without legal enforceability, it remains aspirational. The Digital Personal Data Protection Act (DPDPA) 2023 regulates data use but lacks AI-specific provisions, while the Information Technology Act, 2000, falls short on accountability for AIdriven decisions. TRAI's proposed Artificial Intelligence and Data Authority of India (AIDAI) offers hope, but its delay highlights a regulatory vacuum [Section 6.1]. Meanwhile, AI applications—like facial recognition in law enforcement—raise privacy and surveillance concerns without safeguards.

Lessons from Global Leaders

Europe's AI Act (2023) [3] offers a risk-based approach, banning high-risk practices (e.g., mass surveillance) and mandating oversight for critical AI systems. Its GDPR [1, p. 32] ensures robust data protection, a model India could adapt to strengthen DPDPA. In the U.S., the Executive Order on AI (2023) [4] sets ethical guidelines for federal use, while the NIST AI Risk Management Framework [5] encourages industry self-regulation. These approaches balance innovation with accountability, offering India a blueprint.

A Proposed Path Forward

India needs a multi-stakeholder AI ethics law led by AIDAI, integrating:

- Transparency: Mandate explainable AI in critical sectors (e.g., healthcare diagnostics).

- Bias Mitigation: Collaborate with IITs to develop standardized bias detection tools.

- Accountability: Introduce EU-inspired liability clauses for AI harm.

- Industry Role: Incentivize TCS and Infosys to adopt self-regulatory codes.

- Public Awareness: Launch campaigns to educate citizens on AI rights, drawing from EU public consultations.

Conclusion

India stands at a crossroads in the AI era. By blending Europe's regulatory rigor with America's innovation-friendly flexibility, India can foster an ethical AI ecosystem that benefits all. Policymakers, industry, and citizens must act now to shape a future where AI serves humanity responsibly.

Target Audience: Policymakers, tech professionals, and the Indian public.

8. REFERENCES :

[1]   E. Bird, J. Fox-Skelly, N. Jenner, R. Larbey, E. Weitkamp, and A. Winfield, "The ethics of artificial intelligence: Issues and initiatives," European Parliamentary Research Service, STOA, PE 634.452, Mar. 2020.

[2]   C. B. Frey and M. A. Osborne, "The future of employment: How susceptible are jobs to computerisation?" Oxford Martin Programme, Oxford, U.K., 2013.

[3]   European Commission, "Proposal for a Regulation on Artificial Intelligence (AI Act)," COM(2021) 206 final, Brussels, Belgium, Apr. 2021.

[4]   The White House, "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," Washington, DC, USA, Oct. 2023.

[5]   National Institute of Standards and Technology (NIST), "AI Risk Management Framework," U.S. Department of Commerce, Gaithersburg, MD, USA, Jan. 2023.

[6]   NITI Aayog, "National Strategy for Artificial Intelligence: #AIForAll," Government of India, New Delhi, India, Jun. 2018.

[7]   S. Legg and M. Hutter, "A collection of definitions of intelligence," in Proc. Frontiers Artificial Intell. Appl., vol. 157, pp. 17–24, Jun. 2007.

[8]   M. L. Gray and S. Suri, "Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass," Houghton Mifflin Harcourt, Boston, MA, USA, 2019.

[9]   T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction," Springer, New York, NY, USA, 2001.