Volume 10, Issue 6, June – 2025 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1665

The Artificial Intelligence Revolution: Ethical Frontiers, Impact and Regulation

Aahana Jain[1]

[1] Greenwood High International School

Publication Date: 2025/06/30

Abstract: Artificial Intelligence (AI) has rapidly transformed industries and redefined global economic and social landscapes. This paper explores the multifaceted impact of AI systems across key sectors such as healthcare, finance and retail and the economic implications of AI – including its potential to contribute trillions to the global GDP and its dual role in creating and displacing jobs. Additionally, this paper critically addresses the core ethical issues emerging from AI integration such as bias and fairness, transparency and explainability, sustainability, and misuse and weaponization. It also analyzes ongoing efforts to solve these issues and various case studies in order to highlight the importance of the ethical use of AI. As AI systems continue to evolve, ensuring that their progress aligns with ethical standards and human values remains one of the most pressing challenges of our time.

Keywords: *Artificial Intelligence, AI Ethics, Algorithmic Bias, Environmental Impact, Transparency, Data Privacy, Labour Market Impact, AI Policies.*

How to Cite: Aahana Jain (2025). The Artificial Intelligence Revolution: Ethical Frontiers, Impact and Regulation, *International Journal of Innovative Science and Research Technology* 10(6), https://doi.org/10.38124/ijisrt/25jun1665

I. INTRODUCTION

In recent decades, the rapid proliferation of artificial intelligence (AI) and machine learning (ML) has significantly impacted sectors such as healthcare, finance, transportation and law, and has ushered an era of technological innovation. Artificial intelligence refers to the general ability of computers to emulate human thought or systems capable of performing tasks that require human-like intelligence. AI encompasses machine learning, deep learning, neural networks, natural language processing and more. Machine learning refers to the algorithms that enable systems to make predictions or decisions without explicit programming. [1]

However, as these technologies have evolved and have become integral to societal functions, they have raised various pressing ethical concerns. Issues such as algorithmic bias, lack of transparency and data privacy violations have become increasingly evident, especially in applications such as medicine and predictive policing. When such systems operate on biased data or operate opaquely, they risk perpetuating systemic inequalities and undermining public trust. Furthermore, AI raises the question of accountability when it leads to harm, as responsibility is often spread over different groups of people including developers and users. The impact of AI on employment, its use in surveillance and its potential for misuse and weaponization further complicate this ethical landscape.

Ultimately, the true measure of progress of AI systems will not just lie in how intelligent our systems become, but in how wisely and justly we choose to use them - by embedding conscience into the architecture of the future.

II. HISTORY AND CONTEXT

- The first mention of artificial intelligence can be traced back to 1726 in Jonathan Swift's novel 'Gulliver's Travels,' which anticipates the concept of algorithmic text generation through the machine The Engine – a large mechanical contraption used to assist scholars in generating new ideas. [2][3]
- In 1914, the Spanish engineer Leonardo Torres y Quevedo demonstrated the first chess-playing machine El Ajedrecista that operated using electromagnets and was fully automated.
- Next, in 1921, Czech playwright Karel Capek coined the term robot in his science fiction play 'Rossum's Universal Robots'
- In 1950, Alan Turing published the paper "Computing Machinery and Intelligence," which proposed a test for

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1665

machine intelligence called the Imitation Game or the Turing test.

- In 1951, Marvin Minsky and Dean Edmunds built the first artificial neural network, called Stochastic Neural Analog Reinforcement Calculator (SNARC) - one of the earliest efforts to simulate human brain learning processes using reinforcement learning.
- In 1955, John McCarthy held a workshop at Dartmouth wherein the term artificial intelligence was first coined. In 1958, he created LISP (list processing), the first programming language for AI research. [2]
- In 1959, Arthur Samuel coined the term machine learning. James L. Adams created the Standford Cart in 1961 – one of the first autonomous vehicles – that was independently able to navigate through a room full of chairs. In 1966, Joseph Weizenbaum created the first 'chatterbot', named ELIZA, which used natural language processing (NLP) to converse with humans. [2]
- In 1986, Ernst Dickmann and his team at Bundeswehr University of Munich created the first driverless car that could drive up to speeds of 55mph on a clear road.
- From 2006 onwards, companies such as Netflix and Facebook started using AI as part of their advertising and user experience algorithms, and in 2011 Apple's virtual assistant Siri was launched.
- In 2020 OpenAI started beta testing GPT-3, one of the most sophisticated AI models to date. Moreover, DeepMind's AlphaFold 2 made a breakthrough by accurately predicting the 3D structure of proteins from their amino acid sequences. In 2021, DALL-E was launched, which is capable of generating highly detailed images from textual descriptions. In 2022, ChatGPT was launched and in 2024 Sora was launched– a model capable of generating videos from text descriptions. [3]
- Now, AI and ML algorithms are being increasingly adopted into multiple fields including criminal sanctions, loan offerings, healthcare, recruitment, finance, transportation and more. This has raised multiple concerns regarding the ethical use of AI in order to steer its trajectory towards responsible and justified outcomes.

III. IMPACT OF AI

- *Sector Specific Impact:*

- *Healthcare:*

AI is transforming the field of medicine and healthcare. It was valued at $16.61 billion in the global healthcare market in 2024, comprising the value of total products and services sold. [7] It can accurately analyze X-rays and CT scans, improve the speed and accuracy of diagnosis, identify diseases like osteoporosis and cancer, analyze vast amounts of genomic and other data and create customized treatment plans through predictive analysis. Moreover, it can help with remote patient care, identify trends to detect irregular patterns, detect frauds and can enhance the management of medical records. [5][6] AI was used during the Covid-19 pandemic in order to remove virus-related misinformation on social media. [7]

- *Retail and E-commerce:*

AI is revolutionizing the landscape of retail and E-commerce by enabling personalized shopping experiences through analyzing customer behaviour, purchase history and offering tailored product recommendations. Moreover, AI enables dynamic pricing optimization by analyzing market conditions, competitor pricing and consumer demand. It also helps in inventory management, demand forecasting, visual searches, image recognition, customer segmentation, customer service through AI chatbots, stock management, and so on. [6]

- *Banking and Finance:*

AI is being leveraged in financial and banking operations for fraud detection and prevention by analyzing transaction patterns, and for credit scoring and risk assessment through ML algorithms. Moreover, through technologies such as Optical Character Recognition (OCR), deep learning and Natural Language Processing (NLP), AI systems can accurately scan documents, extract data and enhance decision making processes. It can also help with portfolio and debt management, and financial report generation. [6]

- *Transportation and Logistics:*

In the field of transportation, AI has paved the way for self-driving or autonomous vehicles, inventory management and efficient space utilization across warehouses. It also helps in resource management, route optimizations and prevention of the bullwhip effect, among other uses.

- *Entertainment and Media:*

This industry has embraced AI in order to enhance user experiences and generate personalized content. It contributes in game design and storytelling by improving non-Player Characters (NPCs) and analyzing extensive datasets to create captivating narratives. Moreover, it is used in content recommendation and for editing movies. For example, IBM's Watson was used for producing the trailer for the film 'Morgan'. Further, it is used to create targeted advertising and is employed by various social media platforms such as Facebook, Instagram, Snapchat, etc. in order to deliver personalized products and services to their users. AI is being increasingly used in the field of music, both for composition and lyrics generation. [6]

- *Education:*

In the education sector, AI has transformed how we learn. It provides a platform for personalized learning and automated grading. Moreover, it can analyze student performance, identify areas for improvement and automate tasks such as grading and report management.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1665

The impact of AI on other sectors such as hospitality and information technology is also similar. Though this has led to some positive outcomes, it has caused the displacement and transition of various jobs.

➢ *Impact of Ai on the Global Economy and Employment:*

Artificial intelligence is rapidly transforming the global economy, reshaping the future of our work. Generative AI could inject $2.6-$4.4 trillion annually into the global economy – a value almost equivalent to UK's GDP in 2021 which was about $3.1 trillion. [8] This would give a boost of about 15%-40% to the $11-$17 trillion of economic value that non-generative artificial intelligence and analytics is estimated to proffer. It is also estimated that AI could drive a 7% increase (approximately $7 trillion) in global GDP and cause a 1.5% rise in productivity growth over a period of 10 years. [10] Most of this value – about 75% - would come from customer operations, marketing and sales, software engineering, and research and development. Estimates suggest that 0-30% of the hours worked globally could be automated by 2030 and that current generative AI and other technologies have the potential to automate work activities that take up about 60-70% of workers' time. [8][9] Generative AI has the potential to cause a labour productivity growth of 0.1-0.6% annually through 2040, depending on the rate at which technology is adopted and workers are reemployed into other activities.[8]

However, it is estimated that by 2030, 14% of employees could be forced to change their careers and that AI may replace 300 million jobs. [11] It is approximated that 75-375million people may need to switch occupational categories or garner new skills and about 400-800 million individuals could be displaced by automation by 2030. This occupational change suggests that a large number of people would need to learn new skills or shift occupations in the coming years. Despite this, while each new technological wave inevitably displaces some jobs, history shows that the creation of new roles and industries typically offsets these losses over time. A report estimates that 250-280 million jobs could be generated from the impact of rising incomes on consumer goods, with an additional 50-85 million jobs generated from higher spending on health and education.[9]

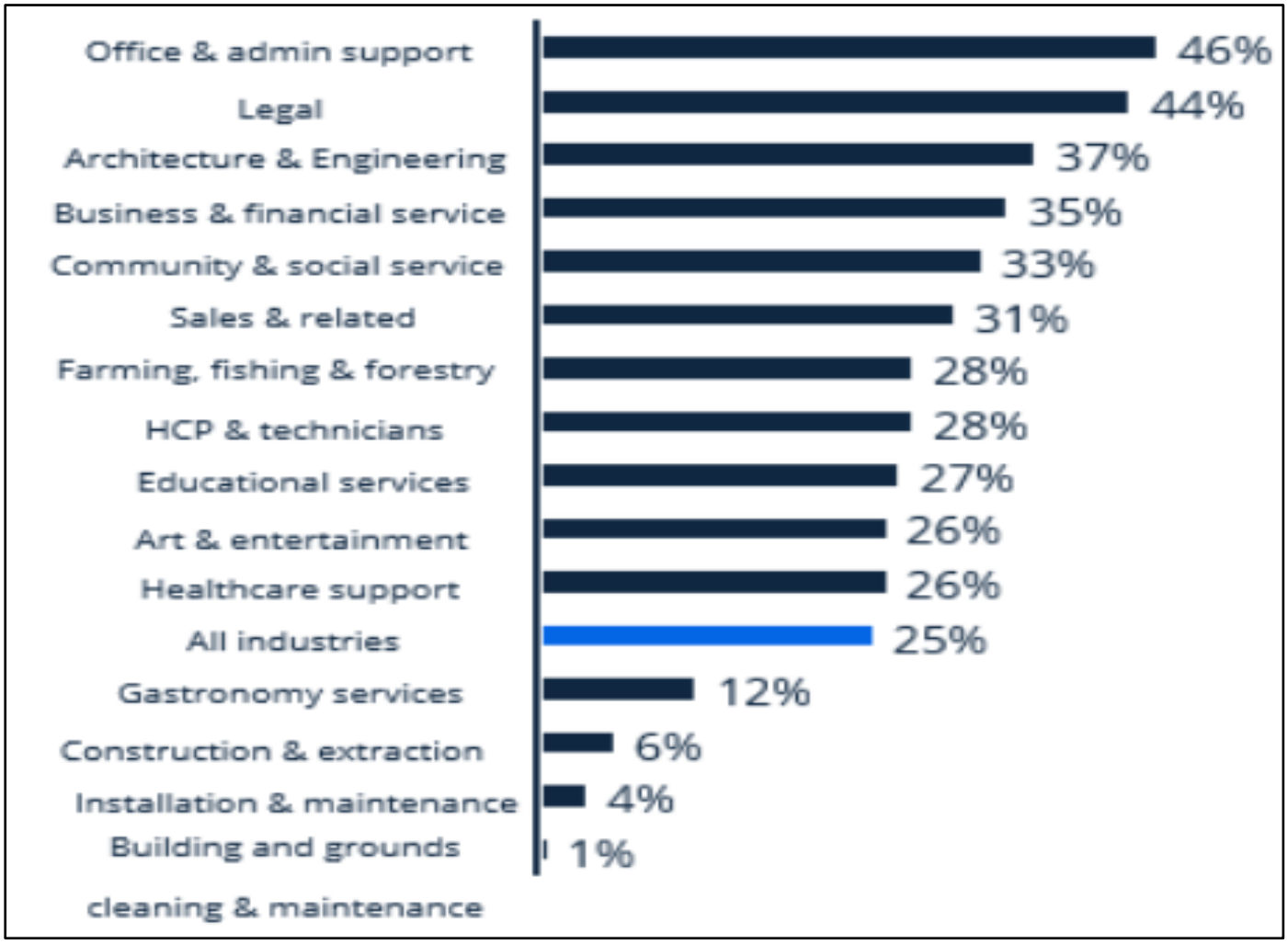

Fig 1 Tasks that could be automated by AI in the U.S. and Europe

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1665

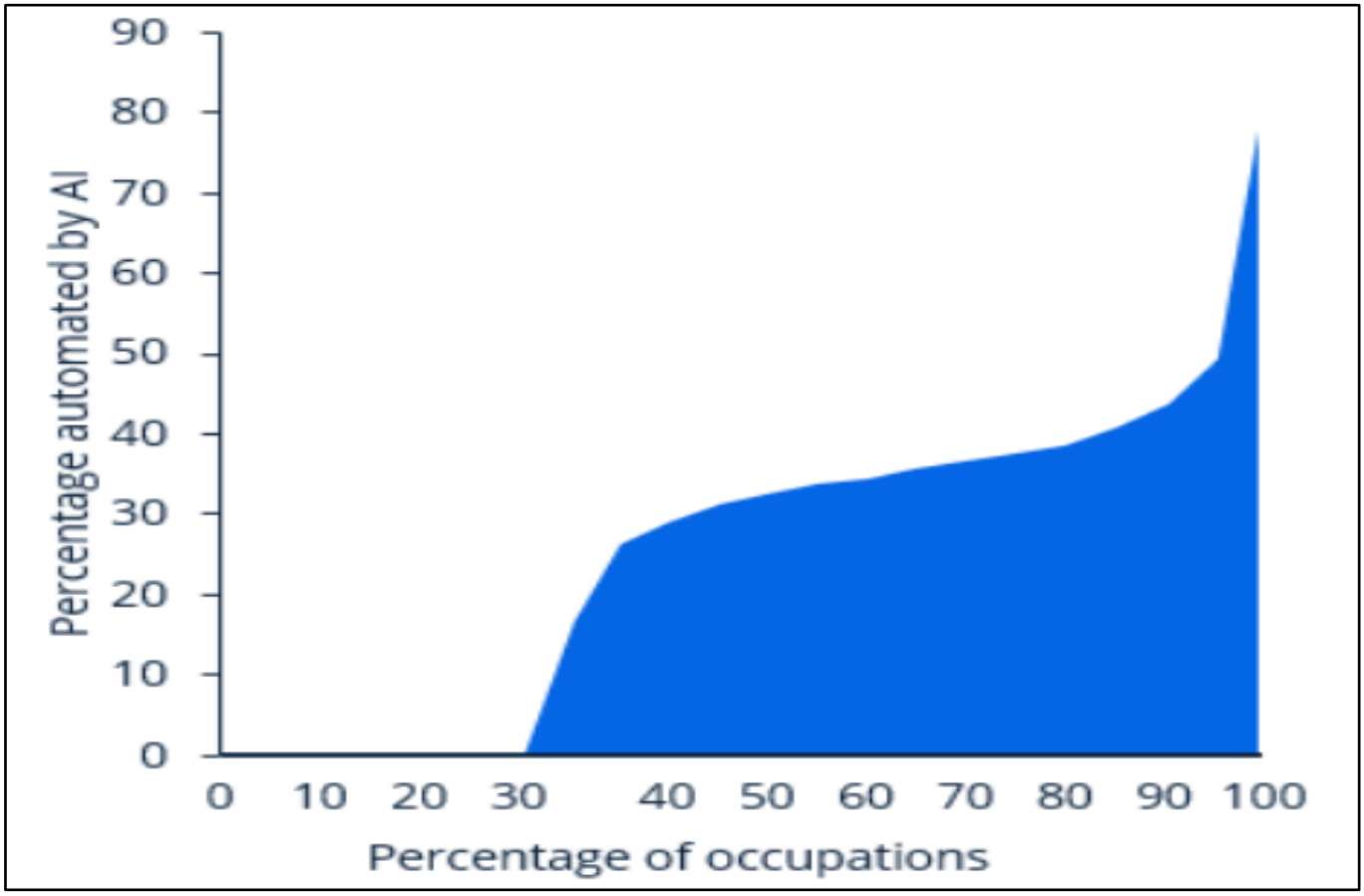

Fig 2 Share of Current Occupational Workload Exposed to Automation by AI

IV. CORE ETHICAL ISSUES

As AI and ML systems reshape various industries and redefine the nature of work, they also raise profound ethical concerns. Their growing role in autonomous decision making, including in medical diagnoses and self-driving vehicles, forces society to confront questions about fairness, accountability, transparency and responsibility in the use of intelligent systems. The primary ethical issues include:

➢ *Fairness and Bias:*

Fairness and bias is one of the most significant ethical concerns with regards to AI and ML systems. Algorithmic bias in AI occurs when two data sets are not considered equal, which could arise due to biased assumptions in the AI algorithm development process or due to built-in prejudices in the training data. [13] Machine learning algorithms learn from historical data, and if that data contains biases, these algorithms can perpetuate and even exacerbate biases, amplifying societal inequalities and resulting in discriminatory outcomes. [14][15]

For example, facial recognition systems have been found to have higher error rates for people with darker skin tones, resulting in discrimination and privacy violations. Specifically, in experiments involving Contrastive Language-Image Pre-training (CLIP), images of black people were misclassified as non-human at more than twice the rate of any other race. Moreover, AI systems misunderstood black speakers, particularly black men, twice as often as white speakers. [13][14] Such biases can lead to unfair hiring processes that favour candidates from specific backgrounds while excluding others, which undermines diversity; reinforce stereotypes and marginalize certain communities; and perpetuate racial, gender, or socioeconomic discrimination when making decisions related to criminal justice, healthcare and loan practices. [15]

- *Mitigation:*

Such biases violate individual rights and erode our trust in technology. Preventing such a bias requires the use of diverse and representative datasets, comprehensive audits of datasets by multiple teams, rigorous testing, and implementing fairness aware algorithms. [14][16]

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1665

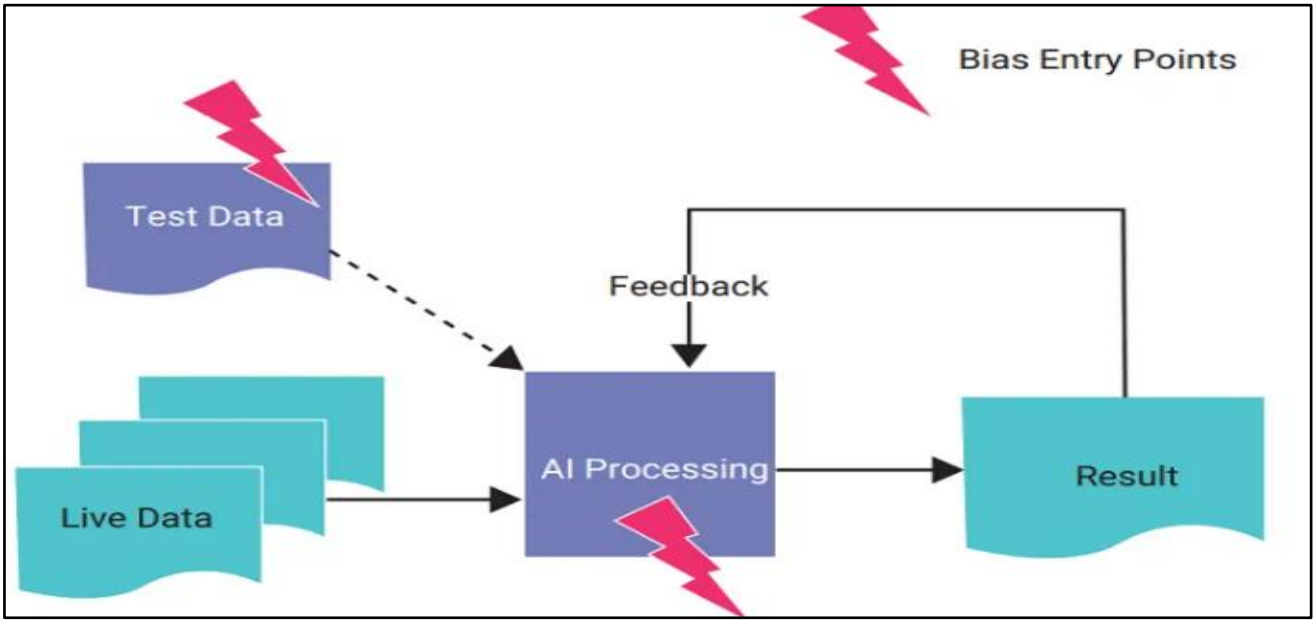

Fig 3 Bias in AI systems

➢ *Privacy and Data security:*

Training AI models requires vast amounts of data, and the collection and utilization of personal information by such systems poses significant risks, which raises questions about data privacy and informed consent. Unauthorized access, data breaches or misuse of sensitive information (which includes names, addresses, health records, financial records, and so on) can have severe consequences for organizations and individuals as they can lead to identity theft or even financial fraud. Moreover, AI used for surveillance without proper regulation can lead to invasive tracking and profiling of individuals.

- *Mitigation:*

For this reason, it is essential to establish clear guidelines for data collection and ensuring that individuals are aware of and consent to how their data is used. Strong encryption, access controls, regular security assessments, transparent data usage policies, and adherence to privacy regulations such as the General Data Protection Regulation (GDPR) are important to ensure safe data usage. [14][16]

➢ *Transparency and Explainability:*

AI and ML systems often operate as black boxes. This means that the internal workings and the decision-making processes are opaque and are not easily understood, even by its creators. This makes it difficult to understands how these systems arrive at their output and make decisions. This can lead to mistrust and can even be problematic in certain situations, such as applications involving healthcare or finance, where the reasoning behind a certain decision is crucial. [14]

- *Mitigation:*

Explainable AI aims to make AI systems more transparent and understandable to users and stakeholders. This not only ensures trust, but also allows for the identification and rectification of potential biases and errors. To address this concern, researchers are working on developing more interpretable AI models and methods for explaining AI decisions. Moreover, promoting open-source AI development can also help. [14][16]

➢ *Accountability and Responsibility:*

The primary question regarding accountability and responsibility is that whether the developer, the organization deploying the system, the users, or the AI itself is responsible when the AI system makes a harmful decision.

- *Mitigation:*

Establishing legal and ethical accountability frameworks is essential for defining liabilities and ensuring that developers and organizations take appropriate measures to prevent any form of harm caused by AI. [14][16]

➢ *Job Displacement and Economic Impact:*

Though AI is leading to increased productivity and the creation of new opportunities in certain sectors, it has led to the displacement of jobs in multiple industries. This uneven distribution of the benefits of AI can also lead to an increase in economic inequalities.

- *Mitigation:*

Addressing this issue would require investments and policies for retraining and upskilling programs for affected workers, social safety nets, support for displaced workers, fostering innovation in emerging industries, equitable access to AI technologies and inclusive economic policies. [16]

➢ *Weaponization and Misuse:*

The intended or unintended misuse of AI can have severe security, ethical and humanitarian consequences. Weaponization of AI is one of the biggest threats looming

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1665

over the international community as these technologies would not be limited by the same barriers as human soldiers. They could traverse all kinds of terrain and engage in atypical warfare such as in space or cyberspace. One of the biggest threats is the lethal autonomous weapons system (LAWS) which creates complex security issues.[17] In 2023 the US Department of Defense (DoD) defined "AWS as being capable of once activated, to select and engage targets without further intervention from a human operator." [18] Essentially, such systems can search, aim and attack automatically, according to programmed instructions.[17] Examples of automated weapons systems include the Israeli Iron Dome, the German MANTIS, the Swedish LEDS-150, and the UK's Taranis drone that is expected to be fully operational by 2030. Countries like the U.S. and Russia are also developing robotic tanks that can operate autonomously or be remotely controlled. [18][19] The proliferation of such intelligent weapons brings with it the prospect of an arms race and the development of 'killer robots,' and possibilities that terrorist groups or militant organizations may get their hands on such weapons. Geopolitical competition over AI supremacy is intensifying, with nations like China, US and Saudi Arabia investing billions is autonomous weaponry and AI, raising the risk of warfare driven by machines.[18] Additionally, current technology is also at the risk of getting hacked, changing the intended function of such weapons. [17]

Despite these issues, the international community has not reached a consensus in regulations regarding AWS. The UN's Group of Governmental Experts (GGE) had drafted a report in 2023 emphasizing the need for human control and developmental guidelines; however, these were deemed as a minimum standard and a comprehensive legal framework is still lacking. [18]

Apart from this, the risk of AI being used for cyberattacks poses a serious danger to society. It can be used to enhance cyberattacks, making them faster, more targeted and more difficult to detect. AI can be employed for automated phishing, malware generation, and more. Moreover, deepfakes pose serious ethical issues through the creation of hyper-realistic, yet fake, audios and videos. They can be used to spread misinformation, manipulate political narratives, undermine trust in media, violate privacy and dignity, and much more.[24] On the other hand, it can be used for cybersecurity through pattern recognition, real-time monitoring, autonomous mitigation and image-matching technology in order to prevent terrorist content on websites. [19][20]

➢ *Autonomy and Control:*

Human oversight is crucial to ensure that AI systems operate within ethical boundaries and do not cause unintended harm. This involves setting limits on the decision-making capabilities of AI systems and ensuring human intervention in critical situations.

- *Mitigation:*

Approaches such as human-in-the-loop and robust monitoring systems can help mitigate this issue. [16]

➢ *Environmental Impact:*

AI presents a serious environmental issue due to its resource intensiveness. The supply chain associated with AI systems and their continued maintenance carries a significant environmental cost. [21]

Most large-scale AI deployments are housed in data centres, including those operated by cloud-service providers. These data centres require large amounts of energy to power, water for construction, cooling and operation, and release excessive amounts of greenhouse gases into the atmosphere.

- It is estimated that AI-related infrastructure may soon consume more water than Denmark, a county with over 6 million people. [22]
- The energy required by data centres comes from the burning of fossil fuels which produce greenhouse gases like methane and carbon dioxide that cause global warming. Moreover, GPUs (Graphic Processing Units) are used for the training of AI models and processing related data. These GPUs run on electricity, which again requires burning fossil fuels. The biggest amount of energy is required for training the AI system, which requires days or even months of feeding data into a GPU. Some studies indicate that training a model to understand and process human language produces 6,26,155lb of CO_2 over a course of 3.5 days. This is the environmental impact equivalent to the lifetime of 5 cars.[23] Further, the training of GPT-3 on a 500-billion-word database produced around 550 tons of CO_2, equivalent to flying 33 times between Australia and the UK.[25] The International Energy Agency even reported that a request made through ChatGPT consumes 10 times more electricity than a google search.[22]
- Apart from this, the microchips that power AI require the use of rare earth elements, most of which are mined in environmentally destructive ways. Additionally, data centres produce electronic waste which contains toxic metals like mercury and lead.[22] Such heavy metals can enter the food chain through soil and water sources, and their biomagnification can lead to serious ailments in humans and animals, including nervous system damages and cancer.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1665

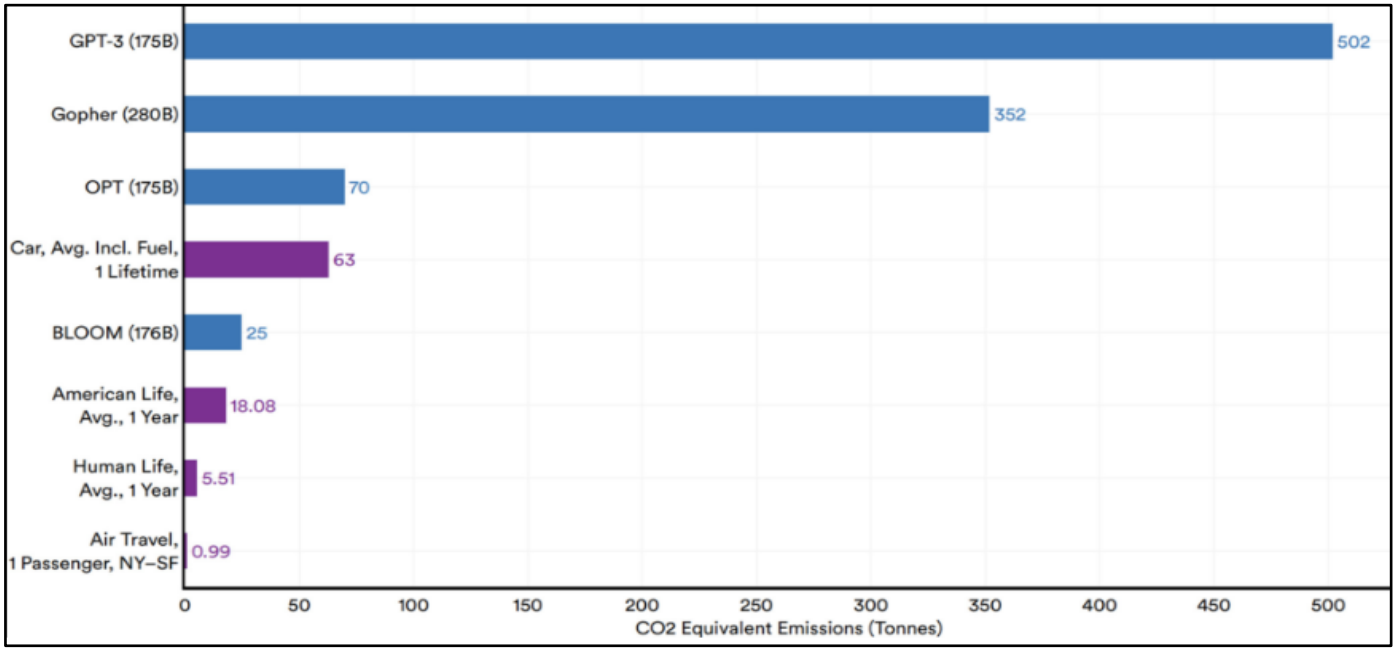

Fig 4 CO_2 Equivalent Emissions for Training ML Models (Blue) vs Real Life Cases (Purple)

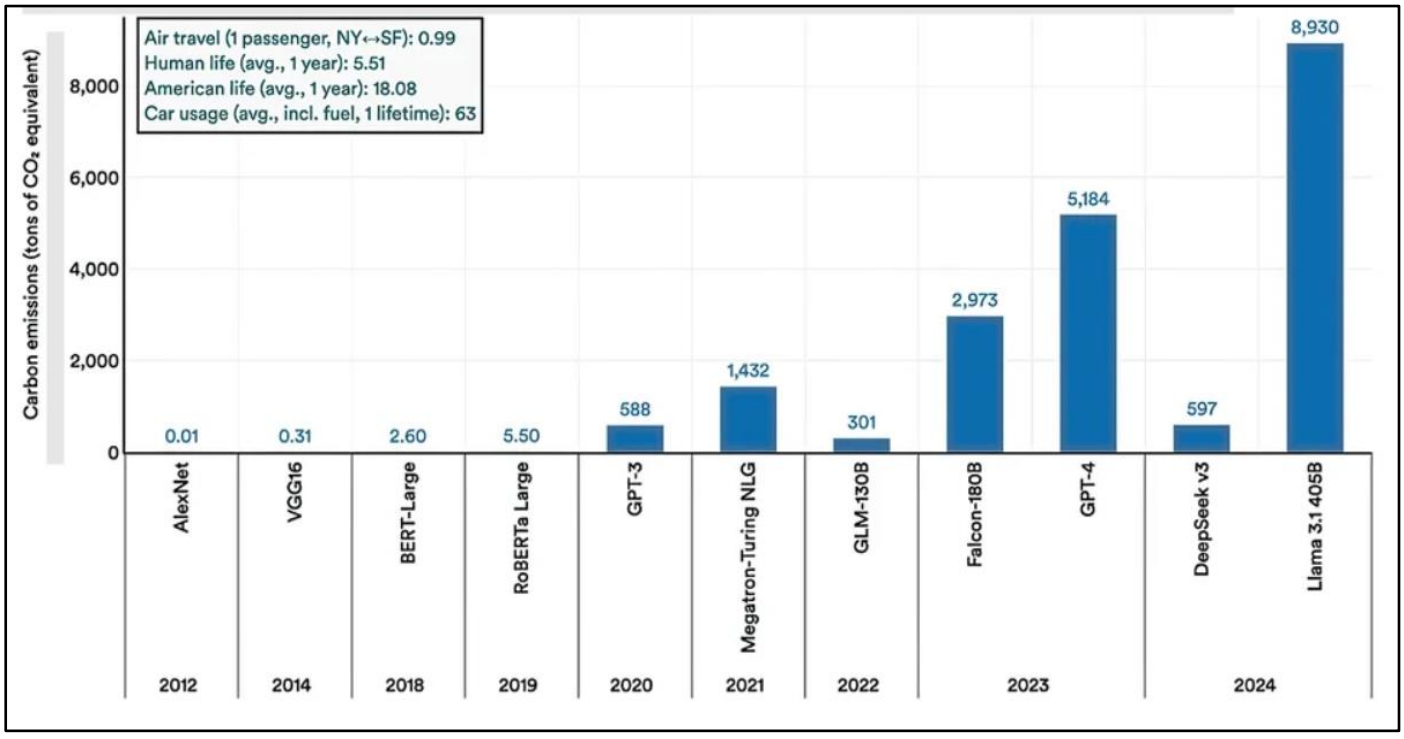

Fig 5 Estimated Carbon Emissions from Training Select AI Models

➢ *Mitigation:*

As the use of AI becomes more ubiquitous, several environmentally focused applications have begun to emerge. For example, the AI for Good movement focuses on the ways in which AI can be leveraged for achieving the UN Sustainable Development Goals, many of which focus on the environment aspect of sustainability. [21] Environmental concerns regarding AI led to the birth of a new term – Green AI. This comprises both Green in AI and Green by AI. Green by AI aims to reduce greenhouse gas emissions by enhancing efficiency across other sectors such as agriculture, transportation, etc. For instance, computer vision technologies can detect gas leaks in pipes to reduce emissions from fossil fuels. ML algorithms can also optimize heating, lighting, etc. by analyzing data from building. On the other hand, green in AI, is an energy efficient AI, with a low carbon footprint, better quality data and logical transparency. This can include data centre optimization and efficient GPUs to reduce greenhouse gas emissions. Neuromorphic computing is another emerging area that aims to create energy efficient computing systems. [25]

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1665

V. REGULATION AND GOVERNANCE

Regulations on AI and ML based systems can mitigate many of the risks associated with AI; ensure that AI-ML is developed and implemented in ways that are fair, transparent and respectful to human rights; promote the adoption of appropriate guidelines and standards; provide assurance that these technologies are safe, reliable and responsible; ensure that its benefits are shared globally; and advocate for sustainable AI practices. [27] Some global and region-specific regulations are mentioned below.

➢ *Global Initiatives:*

- *OECD AI Principles:*

The Organization for Economic Co-operation and Development's (OECD) AI principles are the first intergovernmental standards on AI and were initially adopted in 2019, with some amendments made in 2024. To date, 47 countries have committed to and endorse these principles. [28][29]

Its value-based principles for AI include:

✓ Inclusive growth, sustainable development and well-being
✓ Human rights and democratic values, including fairness and privacy
✓ Transparency and explainability
✓ Robustness Security and Safety
✓ Accountability

The OECD also issued 5 recommendations to policy makers: encouraging governments and individuals to invest in AI research and development, free of inappropriate bias; fostering an inclusive, trustworthy and sustainable AI-enabling ecosystem; establishing policy frameworks that promote AI, while ensuring accountability; collaborating with stakeholders to build human capacity and prepare for labour market transformations; and strengthen international cooperation. [30]

- *UNESCO Recommendations on the Ethics of AI:*

The UNESCO produced the first ever global standard on AI ethics, called the 'Recommendation on the Ethics of Artificial Intelligence,' in November 2021, and it is applicable to all 194 member states of UNESCO.

The recommendation is built upon four core values:

✓ Respect, protection and promotion of human right and fundamental freedoms and human dignity
✓ Living in peaceful, just and interconnected societies
✓ Ensuring diversity and inclusiveness
✓ Environment and ecosystem flourishing

The framework also outlines 10 core principles centred on a human-rights approach to AI. These emphasize ensuring human oversight; promoting public understanding of AI through open and accessible education, media, etc.; the promotion of safety, security, inclusivity, fairness, justice and non-discrimination by AI actors; assessing AI technologies against their impact on sustainability; ensuring transparency, explainability, auditability and traceability in AI systems; privacy and data protection; and participation of stakeholders and respect for internation and national laws.[31]

The recommendation provides 11 policy areas for action.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1665

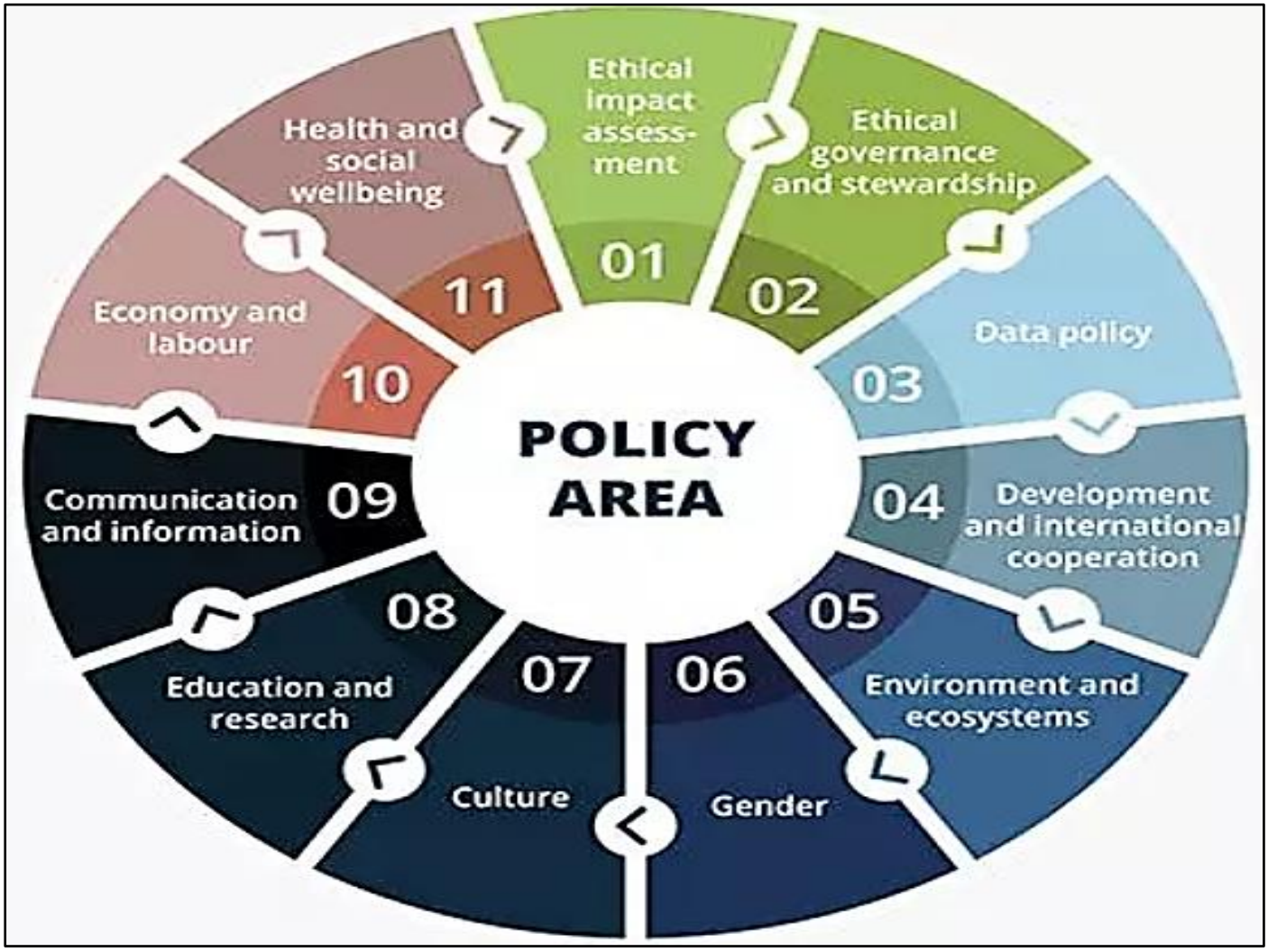

Fig 6 11 areas for policy action

Apart from this, the UNESCO has developed two practical methodologies for effective implementation:

- ✓ Readiness Assessment methodology (RAM): helps assess whether member states are prepared to effectively implement the recommendation.
- ✓ Ethical Impact Assessment (EIA): a process that helps AI project teams and stakeholders identify and assess the impacts of an AI system. [31]

- *The Hiroshima AI Process Comprehensive Policy Framework (HAP):*

The HAP was launched by the G7 (an informal forum of 7 major advanced economies – US, UK, Japan, Italy, Germany, France, Canada) under Japan's presidency in May 2023, with the aim to promote safe, secure and trustworthy AI. It operates in cooperation with international organizations such as the OECD and the GPAI (Global partnership on AI). The framework includes key elements such as the International Code of Conduct, the International Guiding Principles, project-based cooperation on AI and the OECD's report towards a G7 common understanding of Generative AI. It aims to govern AI in a way that upholds democratic values, fairness, accountability, transparency and safety. It also seeks to encourage openness, inclusivity and fairness in AI related discussions, and foster stakeholder and international collaboration. [32][33]

➢ *COUNTRY SPECIFIC REGULATIONS:*

- *The European Union:*

✓ *The EU AI Act:*

This is the world's first comprehensive legal framework for AI. It was approved by the European parliament on March 13th, 2024 and is expected to be fully applicable in the second half of 2027. It follows a risk-based approach, classifying AI systems into four tiers – unacceptable risk (banned) like real-time remote biometric identification for law enforcement in public spaces, high risk (strictly regulated) like AI solutions for the administration of justice or AI safety components in critical infrastructure, limited risk (transparency obligations) like AI chatbots, and minimal risk (no specific requirements) like AI-enabled video games.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1665

High risk AI systems must adhere to strict obligations such as risk management, human oversight and data governance. Non-compliance can result in fines of up to 35 million euros or 7% of global turnover. All businesses operating within or interacting with the EU market must comply with these regulations. [34][35]

✓ *General Data Protection Regulation (GDPR):*

While the GDPR is not an AI specific regulation, it directly impacts AI systems using personal data, including how it is collected, processed and stored. It sets out mandatory rules for how organizations must use personal data in an integrity friendly way and levies harsh fines for non-compliance of privacy and security standards. Though it was drafted and passed in the EU, it imposes obligations on all organizations as long as they target or collect data related to people in the EU. It mandates transparency, consent and accountability in data handling and gives individuals the rights over automated decision making. [36][37]

- *The United States:*

The US pursues a decentralized regulatory framework for AI, that is, most regulatory policies are focused on sectoral levels. The lack of a nationalized AI law posits that the oversight and regulation of AI falls on existing agencies. For example, the Federal Trade Commission (FTC) targets the issue of consumer protections and seeks to apply fair and transparent business practices in the field. Similarly, the National Highway Traffic Safety Administration (NHTSA) regulates the safety aspect of AI technologies in autonomous cars. [38]

✓ *California's Generative AI Training Data Transparency Act (AB 2013):*

This act was signed into law on September 28, 2024, and it takes effect on January 1, 2026. It is the first law in the US to mandate the disclosure of training data for generative AI systems. This law applies to any entity that develops, modifies, or provides generative AI systems that have been made accessible to the Californian public since January 1, 2022.

It requires developers to publish a high-level summary of their training datasets including the copyright and ownership status, descriptions of data types, cleansing and processing methods, the dates of collection and first use, and the personal information content. This act raises transparency and accountability in AI development. [34]

✓ *California Consumer Privacy Act (CCPA):*

This is California's data privacy law that previously did not directly address the use of AI or automated decision-making technology (ADMT). The creation of the CPRA (California Privacy Rights Act) led to the creation of an agency (CPPA) that issued draft regulations about consumers' rights to access information about and opt out of automated decisions. The draft regulations under the CCPA that apply to AI and ADMT aim to enhance transparency and accountability. They apply to for-profit organizations that make significant decisions using AI (like employment, healthcare, loans) or conduct extensive profiling, and require them to conduct risk assessments. They must give consumers pre-use notices, opt-out options and explanations of how decisions would impact individuals. [39]

✓ *Colorado Senate Bill 24-205:*

This is a regulation aimed at protecting residents from algorithmic discrimination in high-risk AI systems – those that make decisions in areas such as employment, housing, healthcare, education, etc. It is set to take effect on February 1, 2026. It requires developers and deployers of AI systems to prioritize transparency, risk management and consumer rights, so that such systems are used ethically and without bias in decisions that significantly affect individuals' lives.

Developers of AI systems must exercise reasonable care to prevent algorithmic discrimination and must provide deployers with information such as data sources, system limitation, and so on. Deployers must implement risk management frameworks consistent with standards such as the NIST AI RMF, conduct impact assessments and ensure users are informed when such systems are used. [34]

Apart from this the Texas Responsible AI Governance Act (TRAIGA) is a regulatory framework designed to govern the use, deployment and development of AI systems in Texas. In addition to state-level regulations, the US Senate introduced the Artificial Intelligence Research, Innovation and Accountability Act, which seeks to establish federal guidelines for transparency, risk assessment and accountability in generative AI, high-impact and critical-impact AI systems.

- *The United Kingdom:*

The UK has not framed a comprehensive AI regulation. Instead, it has opted for a cross-sector, outcome-based framework for regulating AI that is marked by 5 core principles. These are safety, security and robustness, appropriate transparency and explainability, fairness, accountability and governance, and contestability and redress. It follows a pro-innovation approach that puts AI oversight into the hands of existing regulators who will implement frameworks in their own sectors by applying existing laws and issuing supplementary guidance.[40] Bodies such as the AI Security Institute will provide further tools and guidance for organizations.

Moreover, the Bletchley Declaration on AI Safety that was launched at the UK-hosted AI Safety Summit marked a global consensus on AI safety. It focused on the risks of advanced AI systems, especially frontier models; enhancing the scientific understanding of these risks; and cross-country policies to address these risks. It emphasized the dual-use nature of AI – its transformative potential and its risks – and

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1665

advocated for AI safety, shared responsibility among nations and development of global standards and oversight mechanisms.[41]

- *China:*

China has a strong stance on regulatory oversight and takes a centralized approach. However, instead of regulating AI broadly, it deals with different AI advancements separately. Some of its regulations are:

✓ *Interim Measures for Managing Generative AI Services:*

These rules were jointly issued by the Cyberspace Administration of China (CAC) and six other ministries, and came into effect on August 15, 2023. They apply to all AI content services, including text, picture, audio and videos, that are accessible to the Chinese public. It requires model providers use training data from legal sources and obtain user content for personal data; label AI-generated content and adhere to content moderation and accuracy standards; ensure no subversion of core socialist values; establish user complaint channels; conduct security assessments and file algorithms with the CAC if they have the potential of influencing public opinion. [42][43]

✓ *The Administrative Provisions on Deep Synthesis in Internet-based Information Services:*

These came into effect on January 10, 2023. They impose strict requirements on service providers to ensure data security, transparency and data management. Providers must strengthen data management and transparency by complying with data protections laws like the Data Security Law and the Personal Information Protection Law, and implement real-identity authentication systems. They must also establish guidelines and processes for identifying and dealing with false or damaging information created using deep synthesis technology. Moreover, it is mandatory to label any information generated using deep synthesis technologies and conduct security assessments for tools involving biometric or sensitive information related to national or public interest.[44]

VI. CASE STUDIES

There are various real-world scenarios that depict the urgent need for ethical AI governance with responsible AI design, transparency, human oversight, accountability and responsibility.

➢ *NYC AI Chatbot encourages business owners to break the Law:*

In march 2024, it was reported that the Microsoft-powered chatbot named 'MyCity' was giving entrepreneurs incorrect information that would lead them to break the law. This chatbot was intended to help provide New Yorkers with information on starting and operating a business in the city. However, it falsely claimed that owners could cut off their workers' tips, serve food that had been nibbled by rodents and much more. It also claimed that landlords could discriminate based on source of income. This case highlights the risk and potential harm of deploying AI in public-facing government services without proper human oversight. [45]

➢ *Air Canada pays for chatbot lies:*

In February 2024, Air Canada, the largest Airline in Canada, was ordered to pay for damages to a passenger caused by incorrect information given by its virtual assistant. Its chatbot gave a passenger incorrect information regarding bereavement fares. Following its advice, when the passenger submitted refund claims after the purchase of his ticket, the airline tuned him down saying that bereavement fares could not be claimed after ticket purchase. Subsequently a tribunal was held and they were required to pay the passenger CA$812.02 in damages. [45]

➢ *ChatGPT Hallucinates Court Cases:*

In a New York federal court filing, an attorney had used ChatGPT in order to find precedent to support a case filed by an Avianca employee. However, at least six of the cases submitted did not exist and included false names, docket numbers and more. As a result, a $5000 fine was imposed on him. This case demonstrates the dangers of uncritically relying on generative AI in high-stakes applications and highlights concerns about trust and reliability in AI applications. [45]

➢ *Amazon's discriminatory AI hiring tool:*

In 2014, Amazon started working on an AI-powered recruiting software in order to help its HR department screen applications for the best candidates. However, this project was scraped in 2018. The model was trained on 10 years-worth of resumes submitted to Amazon and rated candidates from 1-5. However, due to the training data having higher male apllications, the system penalized applications with the word 'women's', and so was also less likely to recommend applicants from women's colleges. This case is significant as it highlights the risk of bias in training data and how they can perpetuate existing inequalities, and demonstrates the importance of human oversight, regulation and diversity in the development of AI systems. [45]

These cases underscore that the risks associated with AI and ML based systems are not theoretical and that without robust oversight, AI can amplify harm just as easily as it can drive progress. They reveal the urgent need for transparent, accountable and human-centred AI so that technological advancement does not come at the cost of justice, safety and human dignity.

VII. CONCLUSION

As artificial intelligence and machine learning systems continue to advance and integrate into critical aspects of society—from healthcare and education to warfare and governance—the need for robust, ethically grounded, and

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1665

globally coordinated regulation becomes increasingly urgent. While numerous frameworks and policies, such as the EU AI Act and the OECD Principles, demonstrate progress, gaps remain in enforcement, alignment, and adaptability across the globe. Effective AI governance must not only ensure safety, transparency, and fairness, but also safeguard human dignity, privacy, and rights. Ultimately, through collaborative global effort rooted in shared ethical principles, we can ensure that AI evolves to uplift humanity and becomes a force for responsible progress and collective good.

ACKNOWLEDGEMENT

I would like to express my sincere gratitude towards my parents for their valuable guidance, insightful feedback and unwavering support throughout my journey of writing this paper. I would also like to thank my advisor Mr. Prabhat Kumar Tiwari for his constant support.

REFERENCES

[1]. https://ai.engineering.columbia.edu/ai-vs-machine-learning/
[2]. https://www.tableau.com/data-insights/ai/history
[3]. https://www.ibm.com/think/topics/history-of-artificial-intelligence
[4]. https://www.sciencedirect.com/science/article/pii/S0893395224002667
[5]. https://www.chitkara.edu.in/blogs/the-impact-of-artificial-intelligence-on-various-industries/
[6]. https://www.leewayhertz.com/ai-use-cases-and-applications/
[7]. https://www.usa.edu/blog/how-ai-is-revolutionizing-healthcare/
[8]. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier
[9]. https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages
[10]. https://www.goldmansachs.com/insights/articles/generative-ai-could-raise-global-gdp-by-7-percent
[11]. https://explodingtopics.com/blog/ai-replacing-jobs
[12]. https://www.statista.com/site/insights-compass-ai-future-ai-work
[13]. https://www.isaca.org/resources/isaca-journal/issues/2022/volume-4/bias-and-ethical-concerns-in-machine-learning
[14]. https://www.intelegain.com/ethical-considerations-in-ai-machine-learning/
[15]. https://www.dataversity.net/top-ethical-issues-with-ai-and-machine-learning/
[16]. https://www.geeksforgeeks.org/artificial-intelligence/top-9-ethical-issues-in-artificial-intelligence/
[17]. https://www.yu.edu/sites/default/files/inline-files/DISEC%20Weaponization%20of%20Artificial%20Intelligence.pdf
[18]. https://gjia.georgetown.edu/2024/07/12/war-artificial-intelligence-and-the-future-of-conflict/
[19]. https://bernardmarr.com/weaponizing-artificial-intelligence-the-scary-prospect-of-ai-enabled-terrorism/
[20]. https://www.sciencedirect.com/science/article/pii/S1877050924014492
[21]. https://www.ey.com/en_nl/insights/climate-change-sustainability-services/ai-and-sustainability-opportunities-challenges-and-impact
[22]. http://unep.org/news-and-stories/story/ai-has-environmental-problem-heres-what-world-can-do-about
[23]. https://www.scu.edu/environmental-ethics/resources/ai-and-the-ethics-of-energy-efficiency/
[24]. https://www.orfonline.org/expert-speak/debating-the-ethics-of-deepfakes
[25]. https://students.bowdoin.edu/bowdoin-science-journal/csci-tech/ai-save-or-ruin-the-environment/
[26]. https://spectrum.ieee.org/ai-index-2025
[27]. https://www.sciencedirect.com/science/article/pii/S0893395224001893
[28]. https://oecd.ai/en/ai-principles
[29]. https://www.spiceworks.com/tech/artificial-intelligence/articles/ai-regulations-around-the-world/
[30]. https://www.oecd.org/en/topics/sub-issues/ai-principles.html
[31]. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics
[32]. https://www.japan.go.jp/kizuna/2024/02/hiroshima_ai_process.html
[33]. https://www.drishtiias.com/daily-updates/daily-news-analysis/the-hiroshima-ai-process-for-global-ai-governance
[34]. https://www.modulos.ai/global-ai-compliance-guide/#key-ai-regulations-around-the-world
[35]. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
[36]. https://www.gdprsummary.com/gdpr-summary/
[37]. https://gdpr.eu/what-is-gdpr/
[38]. https://www.spiceworks.com/tech/artificial-intelligence/articles/ai-regulations-around-the-world/
[39]. https://www.ibm.com/think/news/ccpa-ai-automation-regulations
[40]. https://www.deloitte.com/uk/en/Industries/financial-services/blogs/the-uks-framework-for-ai-regulation.html
[41]. https://www.insightsonindia.com/2023/11/03/bletchley-declaration/
[42]. https://chambers.com/legal-trends/the-legal-requirements-of-generative-ai-in-china

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jun1665

[43]. https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-china
[44]. https://www.china-briefing.com/news/china-to-regulate-deep-synthesis-deep-fake-technology-starting-january-2023/
[45]. https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html