Regulate To Safeguard: A Critical Analysis of the European Union's First of Its Kind AI Act and the Indian Scenario

Sudha Hegde*

Arundhati**

INTRODUCTION:

One of the predicaments of 21st-century humans is whether to freely allow scientists and industry to develop artificial intelligence devices or to lay down some restrictions in their path to engineer superintelligence. Over recent years, both academia and the AI industry have hit a brick wall in their quest for an answer. Another bewildering aspect is that academia and AI enthusiasts are in splits when it comes to predicting any existential threats or risks from AI at its infancy stage. However, even in its infancy, AI-driven technologies have divulged a litany of risks. Machines infused with Artificial Intelligence have already been accomplishing tasks once deemed unattainable by beating world champions in games like chess, and Chinese board game Go, Jeopardy, to driving cars, proving mathematical theorems, and analyzing complex images and big data (Russell et al., 2010). Even though the initial algorithms were written by humans, now AI technology has developed the capacity to learn from its actions and subsequently modify its algorithms. This self-learning and self-modifying trait of AI raises concern, but the most bothersome angle to this problem is that it is just a piece of the puzzle. AI experts have also been alarming the world about the challenges of deploying AI tools for human advancement without deterring basic human rights like freedom, equality, and privacy (Grosz & Stone, 2018).

As AI is not completely autonomous and is susceptible to human manipulation, one more concern is about how the politically and financially powerful organizations are going to use it. The unholy alliance of misuse of technology by humans and algorithmic biases serves as a ticking time bomb threatening to detonate social trust, justice, peace and unity. The present civilization is no doubt a result of human intelligence. Humans, through the application of reasoning, problem-solving and creativity created complex social structures, governance models and art forms. The addition of artificial intelligence to human cognition will have a profound impact on social, economic, medical, scientific, financial, and military spheres some

* Assistant Professor of Journalism, Government First Grade College, Kalasa, Karnataka, India.

** Dean and Associate Professor of Law, Karnataka State Higher Education Academy, Dharwad, India.

of which may warrant concern. Autonomous weapons, mass surveillance, algorithmic bias, employment disruptions leading to inequality and cybersecurity are the major risk areas.

Weaponization of AI

Flynn Coleman who calls the present time, an intelligent machine age, opines that human cognitive superiority is ending with self-evolving machines. He warns that humans are occupied with predicting the outcomes of the AI revolution rather than putting measures in place to make humanitarian ideals central to any technology (Coleman, 2020). He also makes a notable observation concerning AI governance that political leaders are under informed to make effective laws to control AI research and development.

In 2023, the UN General Assembly adopted a resolution to emphasize the responsible development of lethal automated weapons and acknowledged the serious risks and challenges posed by such weapons including the risk of proliferation to non-state actors (First Committee, 2023). Another potential risk with automated weapon systems is they can be used in a manner where the deployer is untraceable by making it impossible to fix the accountability (Sabatini, 2023).

In military establishments across the world, already the competition to develop advanced AI weapons is underway. The US military's MQ-9 Reaper drone, which can carry out airstrikes with minimal human intervention, the US Navy's MQ-25 Stingray, which is an autonomous refueling drone, and the Israeli Harpy drone which destroys enemy radars are all examples of automated weapons. Along with this, intelligent and command systems are also developed by various military giants for effective warzone coordination like the US Army's Integrated Battle Command System, Lockheed Martin's Legion autonomous combat vehicle, and the Chinese People's Liberation Army (PLA) Sharp Sword system. Chinese military has carried out simulated airstrikes on Taiwan using its Sword AI system in April 2023 after Taiwan’s state head met the US leaders against the wishes of Chinese leaders (Lau, 2023).

As the AI weaponization of armed forces could destabilize traditional deterrence mechanisms and dynamics, they can pose risks to global stability and peace, especially to the less advanced countries (Dahab Gilan, 2018) . This makes it pertinent to bring AI weaponization under some sort of regulation. When it comes to algorithmic bias, there are several examples where biased machine-learning algorithms have given unfair outcomes. One such example is social media bots and algorithms. To provide appropriate content to users according to their preferences algorithms embedded in social media platforms are forming echo

chambers and filter bubbles which support the prevailing views and social biases of users (Cinelli et al., 2021).

Facial Recognition and Deepfake

When it comes to the application of AI in facial recognition tools, studies have shown that AI tools used for facial recognition produce higher false positives for certain ethnicities and races. One of the reasons attributed to this error is the information fed to train the algorithm. As the training data was not as diverse for people of certain ethnicities like black people or African natives as it was for lighter people, it produced higher false positives and false negatives leading to misidentification and illegitimate access (Livingston, 2020). Even though facial recognition technology is advancing by leaps and bounds and being deployed by both state and non-state actors for a variety of purposes, research studies still opine that they are more likely to produce wrong results in African and Nordic faces due to the lack of datasets (Martin, 2022).

Deepfake videos have become a constant headache for Indian celebrities. From prime minister to Bollywood actors have been bearing the brunt of the misuse of this technology. Recently, deepfake videos of Aamir Khan and Ranveer Singh supporting a particular political party emerged on the internet and went viral. As India is voting for the Loksabha and considering the huge supporter base these actors enjoy, such videos may influence voters (‘DCP Nalawade Issues Appeal after Aamir Khan and Ranveer Singh File FIRs against Deepfake Videos’, 2024) . The word "deepfake" is a mashup of "deep learning" and "fake," capturing the essence of the manipulated videos and images.

Coined by an anonymous Reddit user in late 2017, deepfakes use deep learning techniques to create startlingly realistic forgeries. This user specifically employed deep learning to swap faces in pornographic videos, raising ethical concerns about the potential misuse of this technology (Rana et al., 2022). When created with malicious motives such fabricated videos may harm the reputation of the targeted person, malign his image, and spread misinformation or abusive content. Some sections of the Indian Penal Code and the Indian IT Act, 2000 have provisions to deal with cybercrimes like deepfakes and impose punishment up to 3 years. However, it is gainful if the EU AI Act advocates any new measures in this regard to safeguard the privacy and dignity of individuals.

Social Scoring Systems

The Social Credit System can be traced to the ideas of Confucianism and which idealizes social order and individual contribution (Szczotka, 2022). Chinese Qing Dynasty had

also implemented a meritocratic system. In 1958 China implemented a Hukou household registration system to track the movements of individuals (Chan, 2019). In the 21st century, both state and non-state players have been collecting social scoring data related to financial history, online activity, buying preferences, and political inclination for different objectives. Proponents of social scoring systems argue that they help governments select worthy beneficiaries for services and promote good citizen practices. However, on the flip side, such systems raise privacy concerns, discrimination, and manipulation by the data holder.
Hence, scholars are calling for transparency in datasets, parameters used in social scoring, and ethical frameworks to rule out exclusion and bias (Raz & Minari, 2023).

The specter of autonomous weapon systems raising the chilling possibility of uncontrolled warfare to societal biases embedded in facial recognition software and social scoring posing a threat of privacy erosion, a comprehensive regulatory framework was the need of the hour. Furthermore, the burgeoning use of AI in medicine necessitates safeguards against algorithmic bias that could impact diagnoses and treatments. The EU's AI Act, with its emphasis on risk assessment, human oversight, and transparency, arrived at a critical juncture. This paper seeks to analyze the present Act to evaluate whether the Act addresses these challenges and ensures the responsible development and deployment of AI that serves humanity, not the other way around.

THE EUROPEAN UNION'S AI ACT, 2023:

The European Union approved the AI Act to regulate the use of Artificial Intelligence across the European Union after a three-day marathon negotiation with the European Commission, Council, and Parliament, on December 8, 2023 (Corragio, 2023).

The Artificial Intelligence Act passed by the EU is the first of its kind wide-ranging AI law which is an official framework on AI, and addresses the risks of AI and strives to make the AI technology human-centric. It endeavours to set forth precise demands and responsibilities for developers and users of AI, guaranteeing the conscientious utilization of this technology. Notably, the regulation strives to ease the bureaucratic and economic pressures on enterprises, with a specific focus on alleviating burdens for small and medium-sized businesses (SMEs) (European Commission, 2024).

The AI Act stands as a pivotal component amidst a broader spectrum of policy endeavours directed at nurturing the progression of trustworthy AI. This includes the comprehensive Artificial Intelligence Innovation Package and the Coordinated Plan on AI. Unified, these initiatives strive to safeguard the rights and welfare of individuals and

enterprises operating within the AI sphere. Moreover, they aim to fortify commitment, investment, and innovative ventures across the European Union's AI landscape. Positioned as the inaugural worldwide legal framework for AI, the AI Act endeavours to propel the evolution of dependable AI both domestically within Europe and on a global scale. Central to its mission is the assurance that AI systems uphold core rights, adhere to safety protocols, and uphold ethical standards, while also addressing the risks posed by highly influential and consequential AI models. (European Commission, 2024).

RESEARCH QUESTIONS:

1. What criteria does the European Union employ to classify extensive AI applications?
2. What ethical guidelines are proposed for the development of AI?
3. What provisions from the EU's regulations could India adapt to its own legislation?

RESEARCH OBJECTIVES:

1. To identify and analyze the specific parameters utilized by the European Union to categorize diverse AI applications.
2. To examine and synthesize the suggested measures and ethical guidelines proposed by experts and regulatory bodies for the development of AI.
3. To explore potential strategies and recommendations for India to adapt relevant provisions from the EU's AI regulations to enhance its own legislative framework.

RESEARCH METHODOLOGY:

The present study uses a qualitative method to carry out a legal analysis. The European Union AI Act was examined for intent, implications and effectiveness. This study helps to understand the strengths, weaknesses, and potential impacts of the AI Act, particularly in areas like the weaponization of AI, social scoring, and deep fakes by dissecting the various provisions of the act.

KEY POINTS OF AI ACT

In delineating Artificial Intelligence (AI) systems, the European Union, within the AI Act (2023, Section 3(1)), defines them as entities imbued with a certain degree of autonomy, utilizing either machine or human-provided data to accomplish defined objectives. Employing

machine learning or logic-based methodologies, these systems generate outputs influencing their surrounding environment.

The bedrock of the AI Act lies in its classification system, which gauges the potential risk posed by AI technologies to individual health and safety or fundamental rights. This regulatory schema stratifies AI systems into four risk levels: Unacceptable Risk, High Risk, Low Risk, and Minimal Risk (WEF, 2023).

Unacceptable Risk entails AI systems contravening EU values or posing threats to fundamental rights, such as those manipulating individuals, exploiting vulnerabilities, or causing physical or psychological harm (Art. 5 EU AI Act). The Act further outlaws AI-driven social scoring by public entities, as well as the deployment of real-time and remote biometric identification systems in publicly accessible spaces for law enforcement purposes. There are ongoing deliberations to expand this prohibition to encompass predictive policing and to prohibit all AI systems facilitating emotion recognition and biometric surveillance (Hoffman, 2023).

High Risk refers to AI systems carrying significant risks to individual health, safety, or fundamental rights. This encompasses systems utilized in critical infrastructure management, education, employment, law enforcement, migration management, and the administration of justice and democratic processes. Providers of high-risk systems must adhere to various mandatory requirements, including risk management and impact assessment, data governance, transparency, human oversight, and cybersecurity before market introduction or deployment.
AI system's deemed Low or Minimal Risk are subjected to specific transparency obligations. This category encompasses systems interacting with humans, such as chatbots and motion recognition systems, as well as those generating or manipulating image, audio, or video content like deepfakes. Users must be duly informed when engaging with such systems. Notably, based on parliamentary suggestions, bots and deepfakes may be classified as high-risk (Wörsdörfer, 2023).

In terms of application scope, the AI Act primarily pertains to providers and deployers introducing AI systems and GPAI models to the EU market or utilizing them within the EU, provided they are established or situated within the EU. It extends its jurisdiction to providers or deployers of AI systems based in third countries, whose systems' output is employed within the EU.

However, the AI Act excludes AI systems purposed for military, defense, or national security aims, regardless of whether they are placed on the market, deployed, or utilized by public or private entities, from its ambit. Likewise, the Act does not encompass AI systems and

models, alongside their output, developed and deployed solely for scientific research and development purposes (Madiega, 2024).

DISCUSSION

AI experts have expressed different opinions about the provisions of the act. Many critics have hailed the different rules for dissimilar risk levels (Veale & Zuiderveen Borgesius, 2021). The EU's AI Act has been met with debate concerning its potential effects on innovation and responsible AI development. Critics argue the Act's detailed requirements could create a compliance burden for smaller businesses and startups, potentially stifling innovation and hindering European competitiveness (Moran, 2024). Additionally, concerns surround the Act's categorization of AI systems into risk levels, with some arguing it may be too broad or lack the necessary nuance to capture the complexities of various AI applications. Furthermore, a lack of alignment with global AI standards could create barriers to international collaboration (Wörsdörfer, 2023).

Beyond these concerns, some argue the Act lacks key elements for robust regulation. Notably, individuals cannot lodge complaints against AI systems, hindering mechanisms for redress. The Act also fails to adequately address sustainability risks associated with AI, such as the environmental impact of data centers and electronic waste. Finally, the Act's implementation timeline, while aiming to be future-proof, might not account for the rapid pace of AI development, potentially hindering its long-term effectiveness and creating a fragmented global regulatory landscape.

When it comes to social scoring the Act, effectively bans categorizing people on their social behaviour, and personal characters and predicting their future choices. The act bans social scoring systems being adopted by both state and private agencies because such systems pose a threat to fundamental rights to privacy and equality. However, in the case of Deepfakes, the act considers it as of limited risks and imposes only transparency obligations like labelling the content as AI-generated. But, deep fake videos of celebrities endorsing illegal products and extreme ideology have been rising and spreading misinformation around the world. With the many countries and States in India going for elections in coming days, it is a challenge to prevent the use of deep fake technology for political gains. Hence, the categorization of deep fake as a limited risk needs to be debated considering the misinformation capability of this technology.

Another interesting element in the Act is that even though the EU advocates for the regulation or prohibition of lethal autonomous weapons systems, the EU Artificial Intelligence Act 2023 does not deal with this subject directly. While the AI Act aims to protect fundamental rights and ensure transparency, human rights organisations argue that it does not go far enough. They suggest that the act should more robustly tackle the potential harms of AI, including invasive surveillance and social scoring systems, which could undermine personal freedoms and privacy.

THE CURRENT AI REGULATORY LANDSCAPE IN INDIA.

As of now, India does not have a dedicated, standalone law exclusively governing artificial intelligence (AI). Instead, the country has adopted a policy-based, sector-specific, and ethical guideline-driven regulatory approach. Key initiatives such as NITI Aayog's *National Strategy for Artificial Intelligence* (2018) and *Principles for Responsible AI* (2021) form the backbone of India's AI governance framework. These documents emphasize the importance of ethical considerations like inclusivity, transparency, safety, privacy, and accountability while promoting AI in critical sectors including healthcare, agriculture, education, smart cities, and mobility. Additionally, the 2023 *Digital Personal Data Protection Act* plays a pivotal role in regulating the use of personal data in AI systems by setting standards for data processing, consent, and the rights of individuals, thereby addressing one of the foundational concerns associated with AI technologies.

India's regulatory strategy is characterized as "pro-innovation," aiming to strike a balance between encouraging technological advancement and mitigating potential risks. Rather than enforcing rigid AI-specific legislation, the government has preferred issuing sector-specific guidelines that provide direction without stifling innovation. For example, in the financial sector, SEBI's 2019 circular requires disclosure of AI and machine learning usage by market participants, enhancing transparency and oversight. In healthcare, the National Digital Health Mission has introduced standards to ensure the reliability and safety of AI-driven systems, including proper data handling and diagnostic validation. These sectoral regulations, combined with overarching ethical frameworks, reflect India's cautious yet forward-looking stance on AI governance, focusing on adaptability, accountability, and trust-building across stakeholders (Singh, 2025).

WHAT DOES THIS ACT MEAN FOR INDIA?

The development of the EU's Artificial Intelligence Act offers India a valuable model framework for drafting its regulations to govern and control AI technology. India, with its burgeoning population and vast technological potential, stands to benefit significantly from establishing robust laws in this area. By doing so, India can not only secure its position in the global AI sphere but also facilitate technology exchange with other nations.

India has already witnessed substantial growth in sectors like fintech and healthcare, thanks to technological advancements that offer cost-effective and efficient alternatives to traditional methods. However, to fully harness the potential of AI and protect against potential exploitation, India needs strong laws and regulations, especially concerning privacy and data protection.

India, while preparing its own AI act can make two distinct choices, firstly it can root its AI governance based on consumer rights and ethics. The second choice is to lower the constraints for the service providers in terms of data collection, transparency, and commercial consumption to promote a competitive AI market. When it comes to China, its AI regulations are tilted at protecting the interests of the government while the US is still stringent to protect privacy. One more distinct challenge India faces is its mounting digital divide despite increasing internet penetration. Hence, Indian regulators should promote a collaborative approach in AI development to ensure equity and access. And regulators also need to keep the linguistic, cultural, social and geographical differences of the country in mind. The huge income disparity of the country also needs to be deliberated to rule out the possibility of very few at the helm of economic power misusing AI for their mere economic gains at the cost of the underprivileged majority.

Even though the territorial jurisdiction of the EU Act is the European Union, the act also applies to any AI service provider who is placing the service on the EU market. India can borrow from this provision and make its Indigenous Act applicable to any AI service provider regardless of nationality. The current Indian IT Act of 2020 is considered outdated and inadequate in addressing the complexities of AI. The Digital Personal Data Protection Bill, 2022, while a step in the right direction, still has shortcomings. Therefore, enacting a robust AI Regulation Act will not only fill the existing gaps but also demonstrate India's commitment to shaping the future of AI responsibly and ethically (Yati, 2023).

7. CONCLUSION

Artificial intelligence (AI) technology has emerged as a highly promising innovation in recent years, offering cost-effective solutions and identifying anomalies that traditional methods may overlook. However, along with its benefits, AI also poses new risks and threats. The EU's Artificial Intelligence Act recognizes the broad impact of AI across society and aims to regulate its applications accordingly. And it has come at a time when the world is extremely polarised with armed conflicts between major countries and the need of the hour is to put technologies for humane use. India can draw inspiration from the EU's approach to AI regulation and develop its framework to govern and manage AI technology keeping the paramount benefit of its vast population. One approach India could consider is the use of a regulatory sandbox, which provides a controlled environment for testing and scaling AI-based business models. This approach can accelerate the development, deployment, and commercialization of AI technologies while ensuring that risks and ethical concerns are adequately addressed.

Developing a robust regulatory framework is crucial to ensure the responsible development and deployment of AI while protecting the rights of individuals and the broader interests of society. This might involve formulating specific AI regulations, strengthening international collaborations, enhancing technical capacity within regulatory bodies, and embedding ethical principles into the core of regulatory policies. Striking the right balance between regulation and innovation remains a significant challenge. While excessive regulation can hinder technological growth, while under regulation may lead to ethical concerns and public harm. Therefore, it is imperative for policymakers to design regulations that support innovation while maintaining strong ethical safeguards. This requires ongoing dialogue and collaboration among government bodies, industry leaders, and civil society. Given the fast-paced evolution of AI technologies, India’s regulatory strategy must be agile and responsive. By prioritizing ethics, security, and inclusive stakeholder engagement, India can develop a regulatory environment that encourages innovation while ensuring accountability and public trust.

REFERENCES:

1. Chan, K. W. (2019). China's hukou system at 60: Continuity and reform. In R. Yep, J. Wang, & T. Johnson (Eds.), *Handbook on Urban Development in China*. Edward Elgar Publishing. https://doi.org/10.4337/9781786431639.00011
2. Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, *118*(9), e2023301118. https://doi.org/10.1073/pnas.2023301118
3. Coleman, F. (2020). *A Human Algorithm: How Artificial Intelligence is Redefining Who We Are*. Melville House UK.
4. Dahab Gilan, O. (2018). *The weaponization of artificial intelligence (AI) and its implications on the security dilemma between states: Could it create a situation similar to 'mutually assured destruction' (MAD)* [The American University in Cairo]. https://fount.aucegypt.edu/cgi/viewcontent.cgi?article=1807&context=etds
5. DCP Nalawade issues an appeal after Aamir Khan and Ranveer Singh file FIRs against deepfake videos. (2024, April 24). *The Times of India*. https://timesofindia.indiatimes.com/entertainment/hindi/bollywood/news/dcp-nalawade-issues-appeal-after-aamir-khan-and-ranveer-singh-file-firs-against-deepfake-videos/articleshow/109566538.cms
6. European Commission. (2024, April 4). *AI Act | Shaping Europe's digital future*. European Commission. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
7. First Committee. (2023). *First Committee of the United Nations General Assembly* [Resolution on Lethal Autonomous Weapons]. https://meetings.unoda.org/ga-c1/general-assembly-first-committee-seventy-eighth-session-2023
8. Grosz, B. J., & Stone, P. (2018). A century-long commitment to assessing artificial intelligence and its impact on society. *Communications of the ACM*, *61*(12), 68–73. https://doi.org/10.1145/3198470
9. Lau, J. (2023, April 9). China's PLA launches simulated precision strikes on Taiwan as 'Joint Sword' drills enter second day. *South China Morning Post*. https://www.scmp.com/news/china/military/article/3216461/chinese-military-launches-simulated-precision-strikes-taiwan-joint-sharp-sword-drills-enter-day-2
10. Livingston, M. (2020). Preventing Racial Bias in Federal AI. *Journal of Science Policy & Governance*, *16*(02). https://doi.org/10.38126/JSPG160205
11. Martin, K. (2022). *Ethics of data and analytics: Concepts and cases* (First edition). CRC Press/Tyalor & Francis Group.

12. Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake Detection: A Systematic Literature Review. *IEEE Access*, *10*, 25494–25513. https://doi.org/10.1109/ACCESS.2022.3154404
13. Russell, S. J., Norvig, P., & Davis, E. (2010). *Artificial intelligence: A modern approach* (3rd ed). Prentice Hall. https://people.engr.tamu.edu/guni/csce421/files/AI_Russell_Norvig.pdf
14. Sabatini, C. (Ed.). (2023). *Reclaiming human rights in a changing world order*. Brookings Institution Press.
15. Singh, A. (2025, February 26). *The AI regulatory landscape in India: What to know*. AZoRobotics. https://www.azorobotics.com/Article.aspx?ArticleID=742
16. Szczotka, P. (2022, October 24). Influence of Confucianism on the Chinese Political System: A CasSingh, A. (2025, February 26). *The AI regulatory landscape in India: What to know*. AZoRobotics.e of Social Credit System and Socialist Core Values. *Institute of New Europe*. https://ine.org.pl/en/influence-of-confucianism-on-the-chinese-political-system-a-case-of-social-credit-system-and-socialist-core-values/
17. Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, *22*(4), 97–112. https://doi.org/10.9785/cri-2021-220402
18. Yati, S. (2023, October 29). Critical Analysis of European Union's AI Draft Policy: A Step Towards Restricting Existential Threat. *RSRR*. https://www.rsrr.in/post/critical-analysis-of-european-union-s-ai-draft-policy-a-step-towards-restricting-existential-threat