Available online at http://www.ijims.com

ISSN - (Print): 2519 – 7908 ; ISSN - (Electronic): 2348 – 0343

IF:4.335; Index Copernicus (IC) Value: 60.59; Peer-reviewed Journal (Meets the UGC norms)

Unveiling the Ethical Challenges of Artificial Intelligence

Shikha Gupta[1], Rama Bansal[2] , Arundhati[3] ,Ms. Kishori Ravi Shankar[4]

[1, 2] Computer Science Deptt., Shaheed Sukhdev College of Business Studies, University of Delhi, Delhi, India
[3] Management and Law Deptt., Shaheed Sukhdev College of Business Studies, University of Delhi, Delhi, India
[4] School of Journalism and Communication, O.P Jindal Global University, Sonipat, India

Abstract

Artificial intelligence (AI) is touted to be one of the most path-breaking technological advancements of contemporary times. It is seen as capable of transforming all aspects of human life, ranging from healthcare to finance to education to crime prevention. AI basically refers to a machine's capacity for carrying out tasks that otherwise required human intelligence. Mainly, this involves identifying problems, devising solutions and making decisions based on pattern detection. The capacity to do these and more promises greater efficiency and productivity. Nevertheless, there are significant challenges on the path to widespread AI adoption, ranging from workplace adaptation to ethical issues surrounding misuse and replication of social biases. We argue that these challenges can be overcome through multidisciplinary dialogues among the multiple stakeholders who stand to be affected by AI usage, development and adoption.

Keywords: AI, ethics, challenge, technology

1. Technical challenges

The technical obstacles encountered in the application of AI encompass a range of issues stemming from the intricate characteristics of AI systems and their interactions with data, algorithms, and the surrounding context. Presented below are key technical challenges.

1.1 Data Bias:

Data bias presents a pervasive obstacle, as it can occur at any stage during the data lifecycle: data collection, processing, and algorithmic training utilized. Data bias refers to training data used to assemble AI systems' inadequacy of representation or balance from which the trained model makes imbalanced conclusions. To understand and identify the vulnerabilities of biased data the following are some characteristics of data collection, preprocessing, and algorithms to take into consideration (Figure 1).

Data Collection:

Sampling Bias: This occurs when a specific segment of population omits while collecting data. For example, if a survey has been conducted using any online platform, it might not approach individuals without internet access.

Selection Bias: This type of bias may identify when an imbalance dataset is there i.e. when a certain type or category of data has been preferred than others.

Response Bias: This happens when individuals provide not appropriate responses due to inaccurate judgement or perception or misunderstanding of survey questions.

Data Processing:

Preprocessing Bias: This identifies during preprocessing and data cleaning. For example, if outliers are removed without desirable reasons or if missing data has been imputed in a biased manner.

Feature Selection Bias: This occurs by selecting improper attributes of data that cannot accurately identify the complete population.

Normalization Bias: Normalization can magnify biases which can be available in the data, if applied without due diligence. For example, normalizing data using historical patterns could continue with present inequalities.

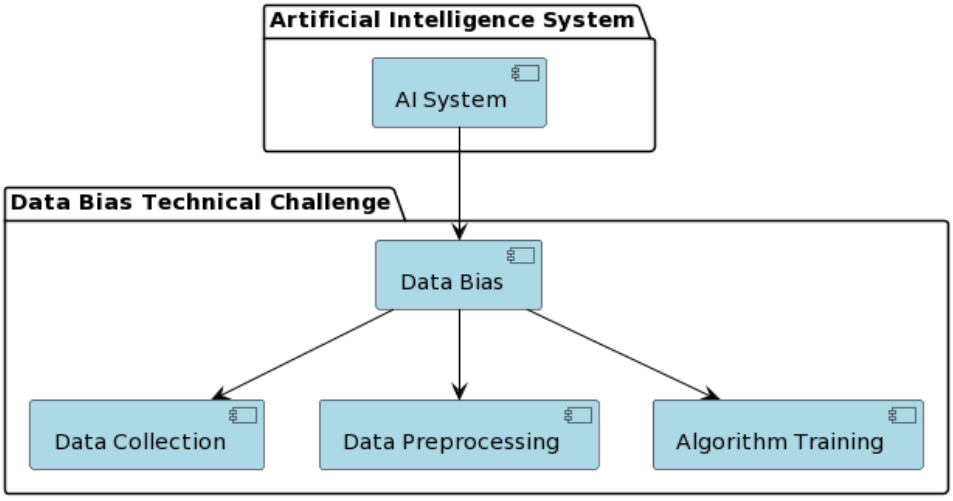

Figure 1: Data Bias

Algorithm Training:

Labeling Bias: Occurs when the symbols assigned to training data sets are subjective in nature or reflecting underlying biases. For example, in image recognition, if labels are inappropriate to identify demographics i.e. disproportionately represents the labeled image.

Algorithmic Bias: This occurs when the training dataset is not accurately addressed to train the machine. This may lead to inappropriate outcomes using discriminatory decision-making systems.

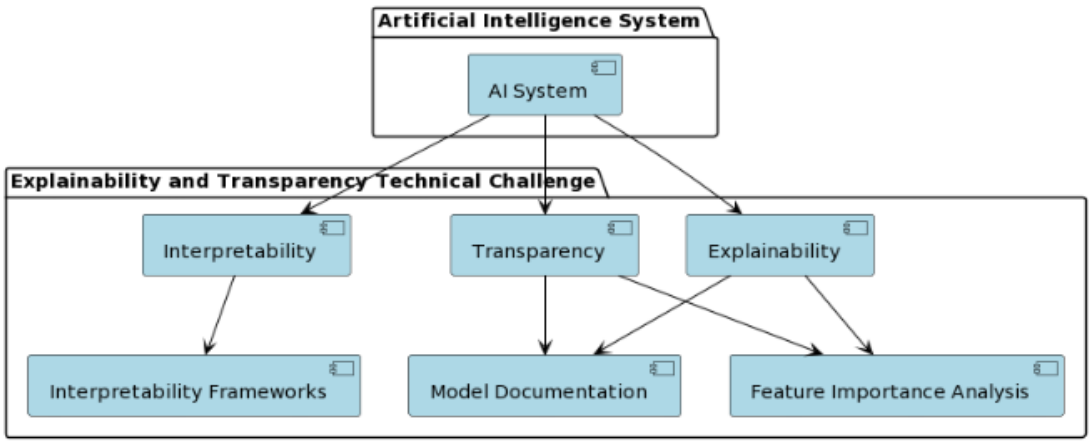

Figure 2: Explainability and Transparency

1.2. Explainability and Transparency:

Explainability and transparency (Figure 2) denote the ease of recognising and interpreting the decision-making pathways of AI systems. Easy and clear interpretation of AI algorithms can help stakeholders in making informed decisions, and identify errors in the system, if any. Techniques for enhancing explainability and transparency include documenting models and interpretability frameworks. Interpretability refers to how easily humans can decode the inner mechanisms of an AI system, grasping its different features, and understanding how a change in an input variable may result in a change in the output variable. Transparency requires full disclosure of sources, strategies and the architecture of the AI system. In addition, complete and clear explanation of the context of system use and opportunities for user interaction can also enhance transparency and explainability.

1.3. Algorithmic Fairness:

One of the key issues to be addressed to build ethical AI systems is to ensure algorithmic fairness (Figure 3). This involves recognising that algorithms are often biased, and end up mirroring social stereotypes and biases. Thus, it is important to recognise and quantify the biases that may crop up during the construction of an algorithm. This can be done by analysing the algorithm for any discriminatory patterns on the grounds of gender, race, religion, sexuality, colour, etc. Post identification, bias mitigation strategies such as bias-aware records sampling, function engineering, fairness-conscious regularisation or antagonistic debiasing may be employed to weed out the biases. Using fairness-embedded or adaptable constraints can also help the system respond to societal inequities as they change over time.

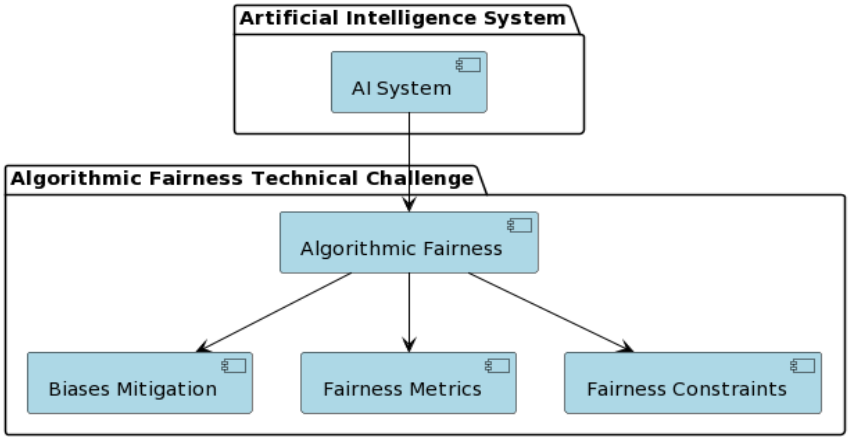

Figure 3: Algorithmic Fairness

1.4. Robustness to Adversarial Attacks:

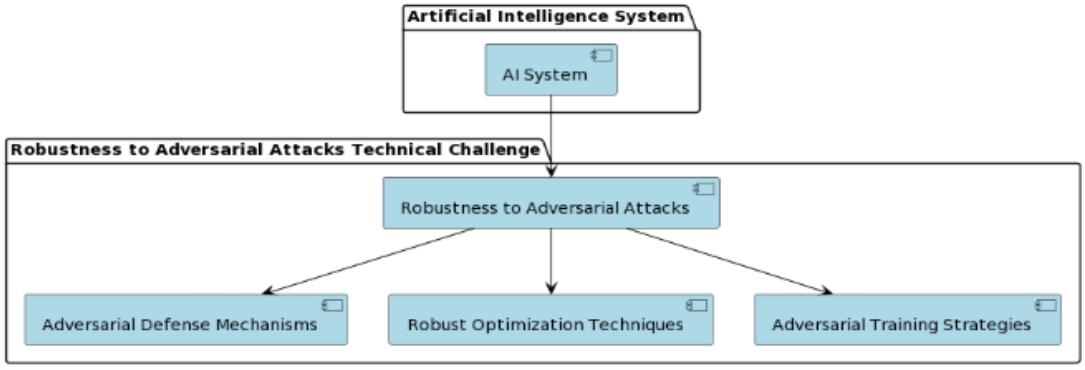

Figure 4: Robustness to Adversarial Attacks

Any AI system is vulnerable to adversarial attacks from external sources, which can result in the manipulation of algorithms and other internal mechanisms, thereby impacting the performance of the system (Figure 4). Protection from adversarial attacks involves a two-pronged approach: (i) robust optimisation and (ii) adversarial training. The former involves fashioning the system in such a way that the final output is minimally impacted by an external attack, i.e. the internal systems can deliver the required outputs despite the attack and are only minimally or peripherally affected. The latter involves training the AI system to actively recognise and thwart adversarial attacks by building the system's capacity to recognise adversaries.

1.5. Security and Privacy Concerns:

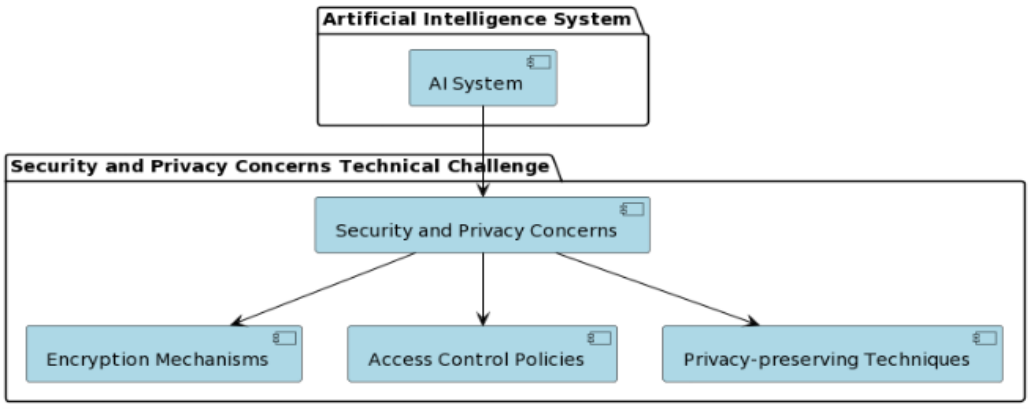

Figure 5: Security and Privacy Concerns

Of all the issues raised with AI systems, security and privacy concerns are of the highest importance (Figure 5). Both are closely related as inadequate security can lead to breach of privacy. An insecure system is more vulnerable to data breaches and leaks, thereby compromising the privacy of data owners. Moreover, this data can be potentially misused by misrepresenting individuals through identity theft, or targeting individuals into revealing information about themselves by using threats and forcing consent. Thus, challenges related to security and privacy concerns are interlinked and can be addressed by identifying the multiple stakeholders involved and designing protocols that view privacy as an inviolable right.

2. Organizational challenges

Organizational challenges play a significant role in successful implementation of Artificial Intelligence within businesses and institutions. These challenges often stem from internal dynamics, resource constraints, and the need for strategic alignment. Some key organizational challenges of AI are:

2.1. Change Management:

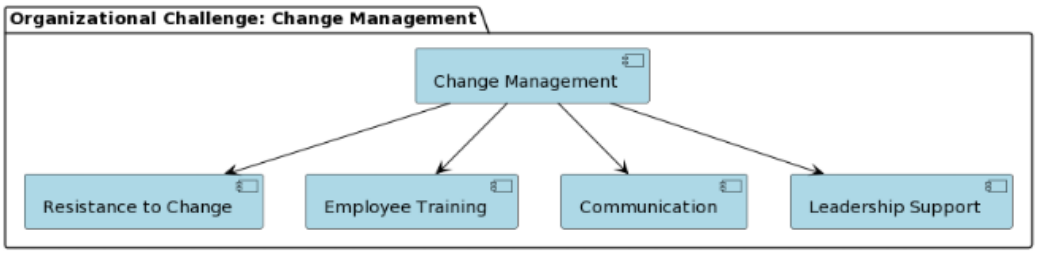

Figure 6: Change Management

This refers to managing the changes that occur at a workplace in the aftermath of the implementation of AI-based systems in different aspects of the workflow (Figure 6). This may involve integrating AI into everyday work, such as for inventory management or automated communication systems. This may require significant employee adjustment, including learning how to operate or oversee new systems. In some cases, the use of AI systems may also cause insecurity among employees regarding their employment status. Effective change management requires timely training of employees and leadership to facilitate the ease of using AI systems efficiently.

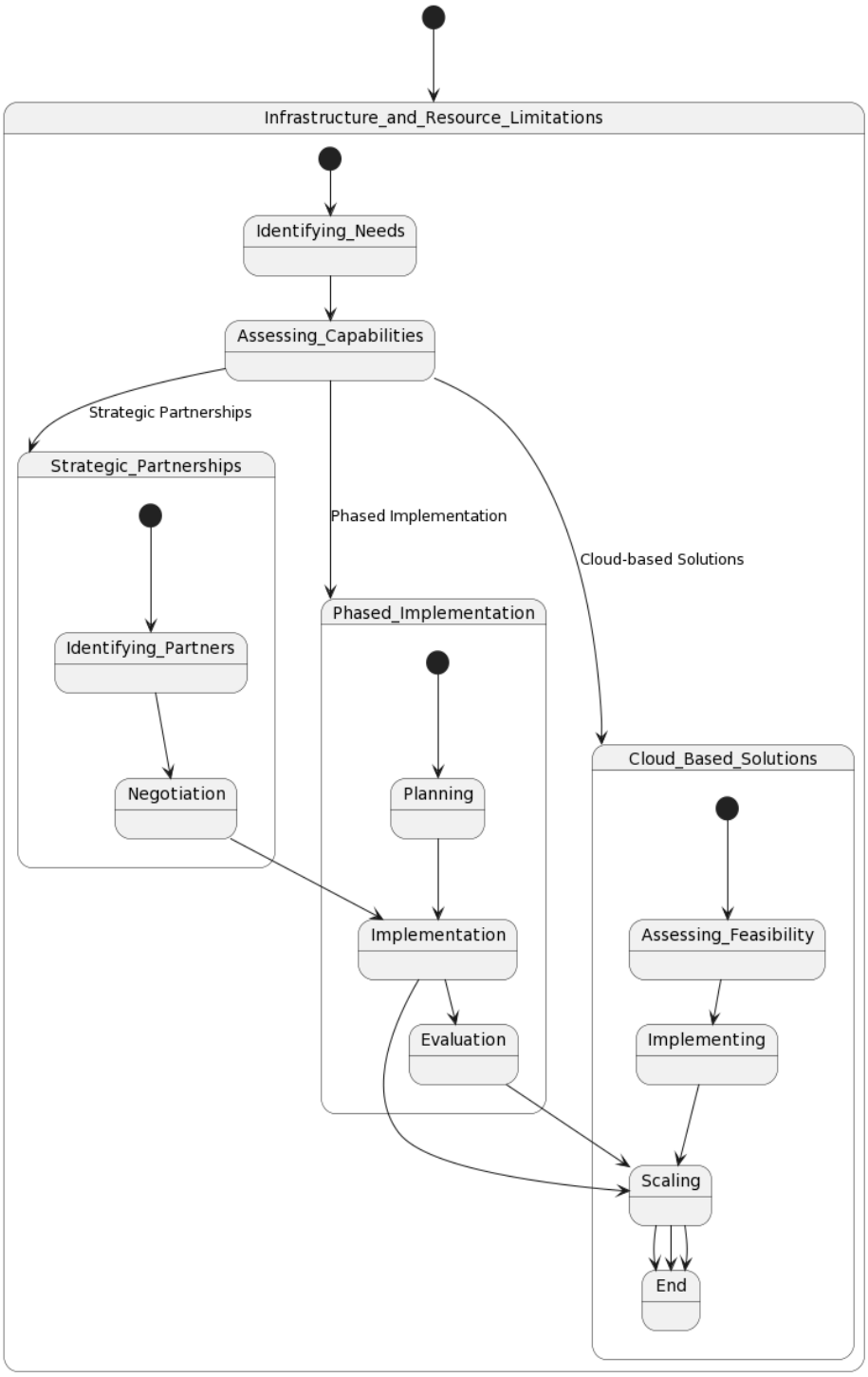

Figure 7: Infrastructure and Resource Limitations

2.2. Infrastructure and Resource Limitations:

Implementation of AI-based systems at the workplace require appropriate infrastructure, including but not limited to computational technology and access to different hardware and software (Figure 7). This, in turn, is dependent on the ability of a company to invest in new

infrastructures or updating existing infrastructures. The initial steps for bringing novel infrastructure are identifying needs (to assess what kinds of AI support a company needs) and assessing capabilities (to ascertain the kind of investments a company is capable of making). This is followed by developing strategic partnerships for bringing in appropriate AI solutions, and a phase-by-phase implementation of the same.

2.3. Lack of Skilled Personnel:

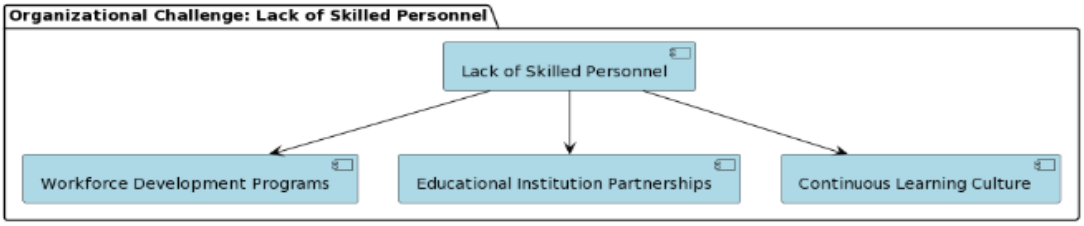

Figure 8: Lack of Skilled Personnel

One of the challenges with respect to the adoption of AI systems is the lack of skilled personnel who understand AI systems and how to deploy them (Figure 8). This challenge is particularly pronounced when a new AI-based system is adopted in an existing industry/workplace with an entrenched workforce. There are several ways to navigate these challenges. Training programmes can be conducted for existing employees to educate them about adapting to and using AI systems. Partnerships can also be developed with educational institutions for facilitating the hiring of graduates who have studied the design and use of AI systems.

2.4. Misalignment of AI with Organizational Goals:

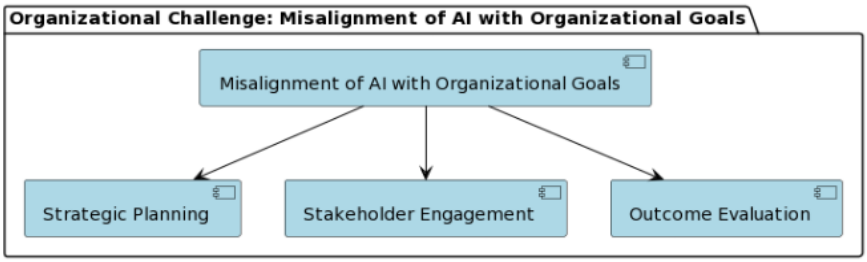

Figure 9: Misalignment of AI with Organizational Goals

While most technological innovations promise more efficient systems, not all advancements are appropriate for a particular industry. The adoption of AI may backfire if AI systems are not properly aligned with organizational goals of a firm. In order to avoid this pitfall, a company needs to engage in strategic planning (to determine where and how AI systems are needed and can be deployed), stakeholder engagement (to bring all stakeholders on the same page and ensure they understand how the adoption of AI benefits them and the company) and outcome evaluation (to assess how useful the adoption of the AI system has been) (Figure 9).

3. Social and Ethical challenges

Social and ethical challenges surrounding AI encompass a range of concerns related to its impact on society, individuals, and the broader ethical implications of AI systems.

3.1. Job Displacement:

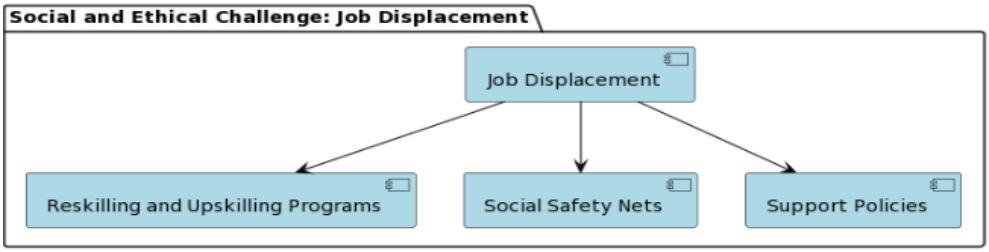

Figure 10: Job Displacement

One of the most disruptive consequences of AI adoption is job displacement (Figure 10), i.e. rendering redundant many jobs that people would do, which would now be taken over by AI. Automation has always resulted in the transformation of jobs, and AI is yet another avatar of that. However, the path ahead lies not in shunning AI, but in re-skilling people through lifelong education and supportive policies for unemployment relief till they find appropriate employment. Moreover, AI development and deployment is also likely to generate new avenues for employment that people can be trained for.

3.2. Societal Inequities Exacerbated by AI:

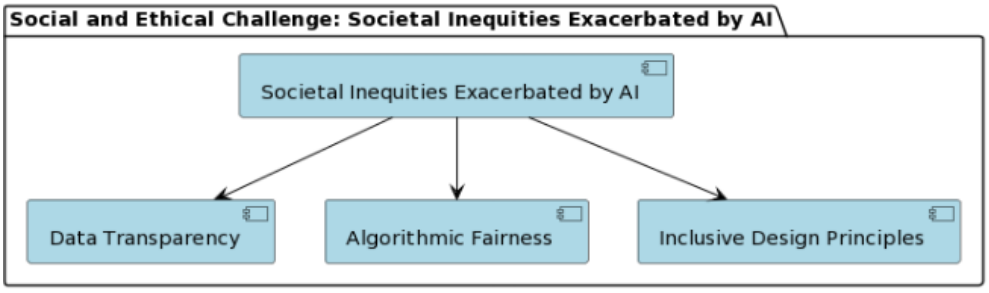

Figure 11: Societal Inequities Exacerbated by AI

AI systems depend on algorithms for functioning and many of the biases in these algorithms affect the output of the systems themselves. Algorithms often depend on large datasets for effective pattern detection and prediction, and when datasets are homogenous and do not account for human diversity and replicate the existing social inequities (Figure 11), AI systems do the same. For instance, datasets are often built on information about men, failing to capture the differences about women and sexual minorities. Thus, even the solutions generated by AI systems may be biased towards men more than women, exacerbating existing social divides. This has implications for everything from healthcare to crime prevention, where biases can either make AI more beneficial for one group at the expense of another, or make some groups more vulnerable to surveillance and stereotyping.

3.3. Ethical Dilemmas Surrounding Autonomous Systems:

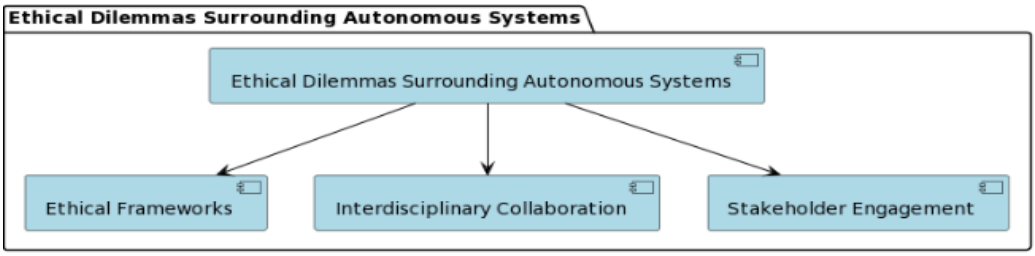

Figure 12: Ethical Dilemmas Surrounding Autonomous Systems

While AI-driven automated systems may bring the promise of complete independence from human intervention, these pose ethical dilemmas. For instance, pinning responsibility for consequences is difficult when there is no human agent involved. In case of a self-driving car, it would not be possible to hold the AI system accountable in case of an accident, and ensuring justice would also be impossible. Similar questions have also been raised with regard to the use of robots for various human functions. The only way out of ethical dilemmas (Figure 12) is interdisciplinary collaboration to detect issues early and resolve them through dialogues among all stakeholders.

3.4. Potential for Misuse and Unintended Consequences:

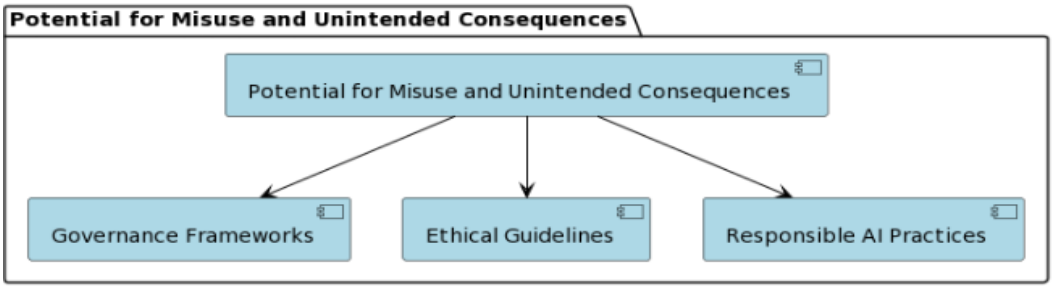

Figure 13: Potential for Misuse and Unintended Consequences

Another issue related to AI systems is the potential for misuse and unintended consequences (Figure 13). Certain technologies such as facial recognition or biometrics may be misused for data theft, impersonation and other criminal activities. There are also many AI-related technologies that may be used in contexts that the creators never thought of, resulting in unintended and maybe harmful consequences. The only way to mitigate this challenge is by designing robust frameworks and regulations for governance and remedies in case of any problems.

4. Regulatory and Legal challenges:

Regulatory and legal challenges surrounding AI encompass a range of issues related to the development, deployment, and use of AI technologies. These challenges arise from the complex nature of AI systems, their interactions with data, and the need to address ethical, privacy, and accountability concerns. Regulatory and legal challenges of AI:

4.1. Lack of Clear Legal Frameworks for AI:

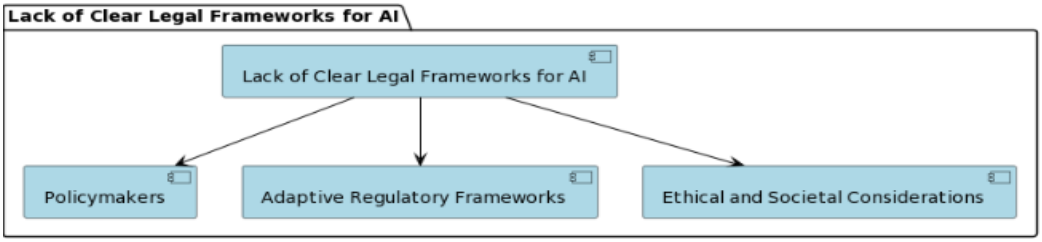

Figure 14: Lack of Clear Legal Frameworks for AI

Since many AI technologies are at their nascent stages, and widespread adoption of AI systems across industries and spheres is still a distant possibility, legal and regulatory frameworks are yet to catch up (Figure 14). Even with existing AI systems, there is a lack of clear regulations as the full spectrum of the consequences of AI systems are yet to be known. It is imperative that policymakers notice this gap and draft laws by examining existing frameworks for governing technologies. This will require engaging with stakeholders from multiple spheres ranging from data scientists, coders and civil society activists to comprehensively understand how the law can be used to govern technologies effectively.

4.2. Data Privacy Regulations:

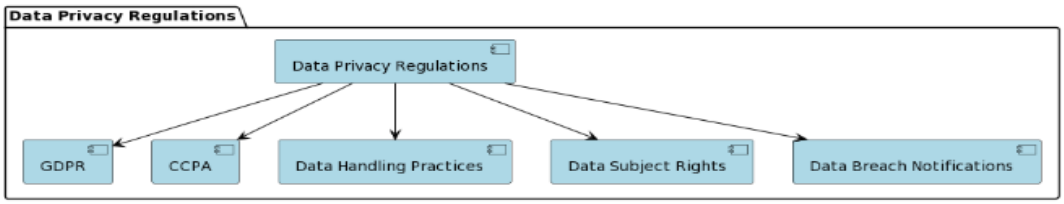

Figure 15: Data Privacy Regulations

One of the areas related to AI where some progress has been with respect to governance and legal regulation is data privacy (Figure 15). Concerns about the privacy of one's personal (or even official) data and its security are being taken seriously by lawmakers worldwide and both Europe and the United States of America have enacted laws for the same. The General Data Protection Regulation, enacted in 2018, is a privacy law relevant to businesses handling private records of individuals within the European Union (EU) and European Economic Area (EEA). It imposes rigorous duties on facts controllers and processors, including acquiring explicit consent for statistics processing, ensuring information security measures, and permitting people to correct and delete private information. Compliance failure can result in fines, highlighting the regulation's emphasis on accountability. Similarly, the California Consumer Privacy Act in the USA enhances California citizens' management over their non-public data and imposes duties on businesses dealing with such data.

4.3. Intellectual Property Concerns:

Ownership rights are essential in IP law, and defining the rightful owners of creations or innovations is a crucial aspect of such laws. In AI, possession rights extend to algorithms, datasets, software, and different intellectual property in development. Clear ownership guidelines are vital to avoid disputes and ensure appropriate recognition and compensation for owners. License agreements and effective IP enforcement mechanisms are critical for deterring infringement and protecting IP rights (Figure 16). Copyright regulation grants can also be used

to recognise the distinctive rights of creators of authentic works, which includes AI-generated content material.

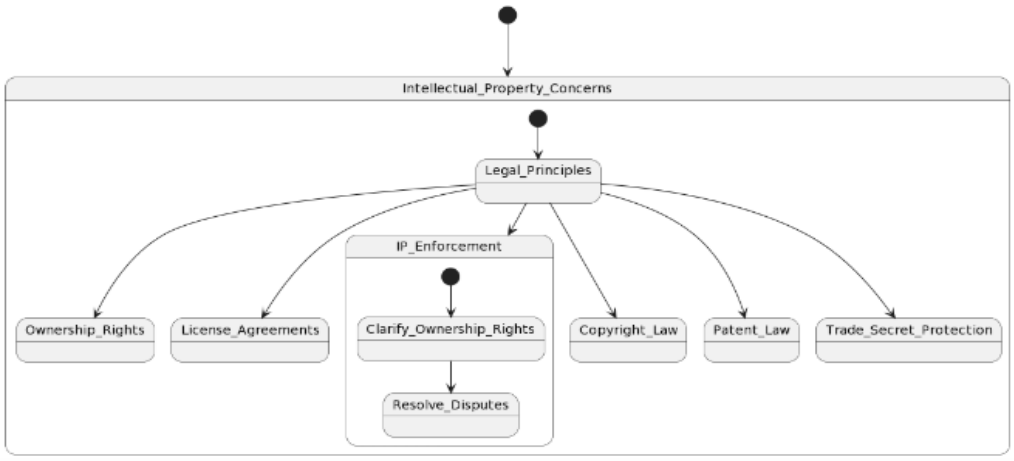

Figure 16: Intellectual Property Concerns

4.4. AI in the Contemporary Indian Context

Regulatory frameworks for AI in contemporary India are a fairly recent phenomenon. AI for All was initiated in 2018 by NITI Aayog. The purpose was to serve as an inclusive approach to AI (NITI Aayog 2018). AI innovation and deployment in areas of healthcare, agriculture, education and transportation were decided to be critical for the nation. As per the strategy's recommendations, the legislative framework for data protection and cybersecurity was also framed.

In February 2021, NITI Aayog drafted certain principles for the responsible use of AI in continuation of this AI strategy (NITI Aayog 2022). The ethical considerations surrounding AI solutions are categorised into system and societal considerations. While system considerations primarily address decision-making principles, fair inclusion of beneficiaries and accountability, the societal considerations concentrate on automation's impact on job creation and employment. The outreaching principles for the responsible governance of AI systems are safety, reliability, inclusivity, non-discrimination, equality, privacy and security, transparency, accountability and protection and reinforcement of human values.

The Digital Personal Data Protection Act, 2023 governs the processing of digital personal data in India, irrespective of its original format and it can be utilised to tackle some of the privacy issues related to AI platforms. India currently lacks some specific laws addressing generative AI and AI-related crimes. However, at present, various provisions within existing legislation offer both civil and criminal remedies. For example, Section 66E of IT Act, 2000 addresses crimes related to deep fakes (which are digitally manipulated media including videos, audio and images created using AI), specifically privacy violations that are punishable by imprisonment up to 3 years or a fine of Rs. 2 lakh.

Conclusion:

In conclusion, it can be said that AI, like many of its predecessor technologically advanced systems, also poses significant challenges for its large scale adoption. There are both practical and ethical issues with respect to the adoption of AI, and these vary according to the sector

(healthcare, media, police, etc) in which these systems are to be used. None of the problems have simplistic solutions, but need to be framed from an interdisciplinary standpoint, involving all stakeholders such as end users, developers, policymakers and others. The legal and regulatory framework to govern the use and development of AI systems is still at its initial stages, given that the potential long-term consequences of adopting these systems are as yet unknown. The Indian government is now in the process of framing rules for the use of AI across sectors, and is modifying existing acts and legal statutes to account for the changes ushered in by the widespread use of AI. It would bode well for different countries to collaborate to unitedly acknowledge the challenges brought in by these new systems, and devise solutions to ensure that they are efficiently and effectively used.

Acknowledgements

We thank Saumya, Gunika, Suhani and Rahul, students of Bachelor of Science program at Shaheed Sukhdev College of Business Studies, University of Delhi for their participation.

References

Béjean, M., Brabet, J., Mollona, E., & Vercher-Chaptal, C. (Eds.). (2024). Disruptive Digitalisation and Platforms: Risks and Opportunities of the Great Transformation of Politics, Socio-economic Models, Work, and Education. Taylor & Francis.

Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. WW Norton & Company.

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. Science, 356(6334), 183-186.

Dafoe, A., & Russell, S. (2020). Yes, We Are Worried About the Existential Risk of Artificial Intelligence. In Artificial Intelligence Safety and Security (pp. 3-23). CRC Press.

Elliott, A. (Ed.). (2021). *The Routledge social science handbook of AI*. Routledge.

Etzioni, O., & Etzioni, O. (2019). Toward robust and verified AI: Specification testing, robust training, and formal verification. arXiv preprint arXiv:1910.12536.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1).

Jha, P. K., Polcumpally, A. T., & Saigal, V. (Eds.). (2024). Emerging digital technologies and india's security sector: AI, blockchain, and quantum communications. Taylor & Francis.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389-399.

Jobin, A., Ienca, M., Vayena, E., & Goodman, K. W. (2019). Artificial intelligence: The global landscape of ethics guidelines. Ethics and Information Technology, 1-29.

Konya, A., & Nematzadeh, P. (2024). Recent applications of AI to environmental disciplines: A review. *Science of The Total Environment*, *906*, 167705.

Marcus, G., Davis, E., & Mahadevan, S. (2014). Rebooting AI: Building artificial intelligence we can trust. AAAI Magazine, 35(4), 22-29.

Menon, S. T. (2024). Going beyond conscientiousness to task pursuit orientation: Exploring an individual difference variable with potential implications for professional achievement and remote work. *Computers in Human Behavior Reports*, *13*, 100357.

Mhlanga, D. (2023). Responsible industry 4.0: A framework for human-centered artificial Intelligence. Taylor & Francis.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 2053951716679679.

Nadeem, A. (2023). Gender Bias in AI: Examination of Contributing Factors and Mitigating Strategies (Doctoral dissertation).

NITI Aayog. (2018). National Strategy for Artificial Intelligence: #AIForAll. https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf.

NITI Aayog. (2022). Responsible AI #AIForAll, Adopthttps://www.niti.gov.in/sites/default/files/2022-11/Ai_for_All_2022_02112022_0.pdfing the Framework: A Use Case Approach on Facial Recognition Strategy.

Author's biography

Dr. Shikha Gupta is currently working as Professor (Computer Science) and Head of Department (Computer Science) at Shaheed Sukhdev College of Business Studies, University of Delhi. She is a member of Skill Enhancement Council of University of Delhi under New Education Policy (NEP) 2020. With more than 25 years of experience spanning industry and academia, her recent interests include Bioinformatics, Educational Technology, Social network analysis, and Process mining. She has published and presented more than 30 research papers in journals and international conferences. Author and editor of books, she has been invited as a reviewer and editorial board member for journals, and technical committee member for international conferences.

Dr. Rama Bansal has over eleven years of teaching experience in Computer Science, during this period she has had the opportunity to teach a diverse range of subjects like Machine learning, Data Mining, Full Stack Development, Advance Java etc. During her studies, she stood the first Rank in MCA in college and Ranked 5th in the University. Additionally, She also excelled in various programming competitions by standing first during my academic tenure. She has also conducted numerous workshops on Web Development using React, Machine Learning and Data Mining at various colleges and has published research papers in various academic journals.

Ms. Kishori Ravi Shankar Assistant Professor at Shaheed Sukhdev College of Business Studies, teaching Law and Management related subjects.

Dr. Arundhati teaches at School of Journalism and Communication, O.P Jindal Global University, Sonipat, India.