International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Determining due diligence principles to enable safe harbour protection for Generative Artificial Intelligence

Kavya Mohan[1], *Jansi S*[2]

ABSTRACT :

Safe harbour protection under Information Technology Act, 2000 currently shields intermediaries from liability for third-party content. The proposed Digital India Act acknowledges diverse intermediaries including AI, yet it also raises a question of whether to grant safe harbour to AI, leaving a legal gap, in immunity for content generated by AI. This research paper focuses on developing key requirements for determining due diligence that provide safe harbor for GAI. This paper adopts a doctrinal approach and studies the current literature, government strategy documents and judicial decisions. For this purpose, reference has been made to OECD principles and foreign laws like the California Generative Artificial Intelligence Accountability Act, Canada's principle for generative AI technology, Brazil, Law No. 21 of 2020 and the EU AI Act's specific provisions regarding transparency, accountability, privacy, fairness in GAI services. The GAI system should inform users about their interaction and provide explanations for AI results. Documentation and accountability are crucial, along with AI audits and human supervision. Privacy principles should be upheld, following the Digital Personal Data Protection Act and The Information Technology (Reasonable Security Practices) Rules, 2011. Fairness in AI systems requires human oversight, reporting, grievance redressal, risk assessments, and model reskilling. Due diligence can be ensured if a GAI model adheres to the principles of transparency, accountability, privacy and fairness. The Digital India Act will be a key legislation in bringing clarity of law regarding safe harbour and due diligence to AI. It can include the elements provided in the paper for establishing sufficient safeguards to the users of GAI.

Keywords: Generative Artificial Intelligence, Transparency, Accountability, Privacy, Fairness.

Chapter I

Background :

Safe harbor provision for intermediaries in India has been evolving and judiciary has played a major role in it. The definition of intermediaries has proved to be too narrow once again with the development of AI. Generative AI like Open AI's chat gpt, has become very popular among the public. The data collection, data usage, content delivered, and content modified by GAI, are regulated by the existing laws meant to deal with traditional intermediaries, and the DPDP Act does not explicitly extend to AI resulting in a lack of clarity in the due diligence to be performed by the GAI companies. The proposed Digital India Act will provide clarity in the future. This paper is an attempt at defining the essentials to ensure transparency, accountability, privacy and fairness in order to fulfill requirements of due diligence for GAI.

Chapter II

Research methodology :

This paper adopts a doctrinal approach and studies the current literature, government strategy documents and judicial decisions on due diligence by AI. Reference has been made to the OECD principles and foreign laws like the California Generative Artificial Intelligence Accountability Act, Canada's principle for generative AI technology, Brazil, Law No. 21 of 2020 and the EU AI Act's specific provisions regarding transparency, accountability, privacy, fairness in GAI services. Brazil being a developing country is a valuable reference for Indian setup as it has adopted the principles specifically for AI, which is absent in India. California is having laws exclusively for GAI. EU AI includes GAI in the category of general purpose AI and classifies it based on risks. It also has some exceptions for open and free AI. The *Principles for Responsible, Trustworthy and Privacy-Protective Generative AI Technologies of Canada*, is referred to understand key principles regarding openness, accountability, privacy regulations of GAI.

Research problem :

Safe harbor protection currently shields intermediaries from liability for third-party content. The proposed Digital India Act acknowledges diverse intermediaries including AI, yet it also raises a question of whether to grant safe harbour to AI, leaving a legal gap, in immunity for content generated by AI. This research paper focuses on granting safe harbour protection for Generative AI as it is advancing rapidly, but global laws to regulate it are still developing.

Research objective

1. To develop key requirements for establishing due diligence principles that provide safe harbor for GAI.
2. To analyze existing laws and regulatory standards, that govern transparency, accountability, privacy, fairness in GAI.

Research question

1. Whether safe harbor protection be extended to Generative AI?
2. What are the essential elements to establish due diligence for GAI?

Literature Review :

Intermediaries are supposed to be mere conduits of information and GAI is not covered under the definition, in light of the decision in Myspace Inc. v. Super Cassettes Industries Limited, as it also modifies the data, unlike search engines[1]. Therefore the traditional notions of intermediaries cannot be extended to GAI.

There is a rising threat of toxic information generated and circulated by AI and different states are dealing with it differently. All countries face the dilemma of regulation and creation of conducive environment for AI's development. Also governance of AI include dealing with competition issues, intellectual property issues, privacy issues etc. The issue is further complicated as startups funded by venture capitalists are risk averse. However, Philipp Hacker, recommends all countries to have mandatory regulation of AI to counter toxic information[2].

With regard to the situation in India, there is a lack of clarity in the definition and scope of liability of AI. The Intermediary guidelines are requiring proactive monitoring. The judiciary has played a major role in clarifying the duties of intermediaries. now that the differences between active and passive intermediaries are decreasing, there is a requirement for clarity[3].

Advisories issued under the Information and Technology Act, 2000 ("IT Act") and the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 ("IT Rules"), does not provide legal clarity but had created confusion to the tech companies. The issuance of multiple advisories, subsequent clarifications and withdrawal of some advisories creates a very uncertain legal environment for tech companies[4]. A legislation which clearly demarcates the duties and the obligations of the users and the companies is needed.

The Proposed Digital India Act[5] will address these issues and cover within its ambit the new technologies like cloud computing, artificial intelligence (AI), blockchain, and the Internet of Things (IoT).

Existing legal framework in India regarding safe harbour under the Information Technology Act

The reading of Sections 79 and 2(w) of the Information Technology Act provides that intermediaries are those persons that receive, store, and transmit third-party information on behalf of any other person without having actual or construed knowledge of the content. The safe harbour provision was initially limited to Network Service Providers (NSP). The scope of NSP has been ever expanding since then. It is now being seen as synonymous to the term "intermediary" and includes telecom service providers, network service providers, Internet service providers, web-hosting service providers, search engines, online payment sites, online auction sites, online market places and cyber cafes[6]

In 2008-09, the definition of an intermediary was expanded from NSPs, who performed primary functions. The amendment to Section 79 provided safe harbour and clarified that 'knowledge' refers only to actual knowledge. This reflects **a role-based liability framework rather than a proactive approach**. There is some redundancy in the terms used under Section 79[7].

In 2008, prior to the amendment, the Bazee.com[8] case imposed strict liability for not expeditiously removing pornographic content, although no time limit was specified, unlike now. In Christian Louboutin v. Nakul Bajaj[9], it was held that safe harbour is only available to passive intermediaries, and a website facilitating buying and selling would not qualify. An online service provider is an intermediary only if it acts as a mere conduit. Active involvement equates to actual knowledge.

[1] *Generative AI & Disruption: Emerging Legal and Ethical Challenges*, Nishith Desai Assocs., https://www.nishithdesai.com (last visited Oct. 8, 2024)

[2] Philipp Hacker, *AI Regulation in Europe: From the AI Act to Future Regulatory Challenges*, 27 Harv. J.L. & Tech. 1 (2023).

[3] Indranath Gupta & Lakshmi Srinivasan, *Evolving Scope of Intermediary Liability in India*, 10 J. Int'l Media & Ent. L. 67 (2023).

[4] Navigating AI Regulation in India: Unpacking the MeitY Advisory on AI in a Global Context, ELP Law, https://elplaw.in/ai-regulation-india/ (last visited Oct. 6, 2024).

[5] Vidhi Centre for Legal Policy, *Explained: The Digital India Act 2023* (Oct. 9, 2023), https://vidhilegalpolicy.in/blog/explained-the-digital-india-act-2023/.

[6] Vakul Sharma, Information Technology Law and Practice, 6ed, chapter 27.

[7] Indranath Gupta & Lakshmi Srinivasan (2023) Evolving scope of intermediary liability in India, International Review of Law, Computers & Technology, 37:3, 294-324, DOI: 10.1080/13600869.2022.2164838

[8] Avnish Bajaj vs State, (2005) 3 CompLJ 364 Del

[9] Christian Louboutin Sas vs Nakul Bajaj & Ors , AIR ONLINE 2018 DEL 1962

However, this was overturned in Amway v. Amazon[10]. Courts and guidelines have significantly contributed to the interpretation of intermediary liability. With advancements, distinguishing between active and passive intermediaries has become increasingly difficult—what was once relevant may no longer be.

The MEITY introduced the intermediary guidelines in 2021 to include Social media intermediaries, OTT platforms, and digital news agencies, to engage in proactive monitoring including tracking the first originator and disabling access to information. However inclusion of AI into regulatory framework is a work in progress under the Digital India Act.

Proposed Digital India Act

With a steep rise in the number of users and kinds of intermediaries, the MEITY has come up with the proposed Digital India Act, 2023, to replace the IT Act, 2000. The proposal aims at removing the redundancies in the IT Act and upholding the constitutional rights in cyber space. Regarding Safe harbour protection, the proposal recognises multiple intermediaries like (OTT, AI, gaming, ecommerce, digital media, social media) but is not conclusive as to whether AI being an intermediary qualifies for safe harbour provision.

Safe harbour to AI

The nature of GAI is such that it is neither an active nor a passive intermediary. Hence bringing it within the ambit of Sec 79 has peculiar challenges. According to S .79(3), an intermediary does not

(i) Initiate the transmission,

(ii) Select the receiver of the transmission, and

(iii) Select or modify the information contained in the transmission. Hence an intermediary should be a passive actor and have no role in altering the information transmitted.

However,

1. An AI is the originator of the information as it creates a customised response for the prompt of the user.
2. An AI does select addressee in certain situations like content delivery AI, social media marketing AI[11].
3. An AI modifies the information to customise to the user's prompts.

The conditions are similar in USA where section 230 of the Communications Decency Act of 1996, of the US (a corresponding provision in the like of secion 79 of the IT Act), is also unclear. Courts have not yet decided whether or how Section 230 may be used as a defense against claims based on outputs from recently released generative AI products. Generative AI products are not all the same, are likely to continue to evolve, and can rely on data and inputs from diverse sources, Section 230 analysis may lead to different outcomes in different cases. There is a need to look closely at how the particular AI product at issue generates an output and what aspect of the output the plaintiff alleges to be illegal[12].

Due Diligence framework for fixing liability

Due diligence protects the companies from facing unjustified liability for content hosted by it. The position of GAI is peculiar as the policy makers are facing a dillemma over balancing the interests of the users and of the companies. Operators of GAI may argue that the third-party information on which the programme has been trained on, is merely being presented to the user after an automated process. Hence the safe harbor provision must be adapted to the changing complexities of intermediaries. The advisories issued by Meity recently and its withdrawal indicates the haze in the control of GAI[13]. The due diligence of GAI and the intermediaries that use them should be based on the following principles

Transparency

Transparency means making people understand how an AI system is created, trained, and operates, in the appropriate application area, which enable the users to make informed decisions[14] and also how the system generates content with the information given by the user[15]. Transparency also means delivering useful information as well as clarity on what information is being supplied and why. This transparency allows users to decide whether the

[10] Amway v. Amazon, CM APPL. 32954/2019

[11] *Personalization in Conversational AI*, SN Computer Science, https://link.springer.com/article/10.1007/s42979-023-00120-1 (last visited Oct. 6, 2024).

[12] Kathy A. Goforth, *Generative AI Will Break the Internet: Beyond Section 230* (Harvard J. Law Tech, 2024),https://jolt.law.harvard.edu/assets/digestImages/Generative-AI-Will-Break-the-Internet_-Beyond-Section-230-Final-Review.pdf.

[13] *Navigating AI Regulation in India: Unpacking the MeitY Advisory on AI in a Global Context*, ELP Law, https://elplaw.in/ai-regulation-india/ (last visited Oct. 6, 2024).

[14] OECD, *AI Principles* (last visited Oct. 14, 2024), https://oecd.ai/en/dashboards/ai-principles/P7.

[15] Rick Cai, *Comparing Transparency Requirements: from Global Legislative Efforts on Generative AI*, The Digital Constitutionalist (Oct. 11, 2023), https://digi-con.org/comparing-transparency-requirements-from-global-legislative-efforts-on-generative-ai/.

outcome is reliable[16] or not. Thus, transparency does not generally include the publication[17] of the source or other private code, or the sharing of confidential datasets, which may be too technically difficult to be useful in comprehending an outcome.

Laws and advisory guidelines of different countries emphasize the need for transparency in generative AI, while this paper focuses on the elements that need to be fulfilled to ensure transparency.

1. **Informing users:** It is important to clearly inform users[18] that they are not engaging with a human but GAI model. This information to the users be conveyed[19] in an understandable form and be available before, during and after the use of such system[20]. Consent popups[21] or other such tools be used to inform users about reliability of generated content. Article 52 of the EU AI Act[22] also requires that users must be notified when interacting with AI systems, ensuring transparency, particularly in high-risk applications.

2. **Explain:** Explainability[23] refers to giving persons affected by an AI system's output with easily understandable information that allows them to challenge the outcome. Developers and other AI actors also benefit by being able to identify issues, de-bug the system and learn more about the problem[24]. When AI actors explain an outcome, they may provide the main factors in a decision, data/input sources, processes, and logic that led to the specific outcome.

Accountability

The AI ecosystem consists of four key actors[25]. First, there are the suppliers of AI knowledge, who provide essential inputs for the development of AI systems. Second, the designers and developers are actively involved in creating, deploying, and operating these technologies. Third, the AI systems utilized by the end-users for various applications and services. Lastly, the people impacted by the consequences and results of AI technology are known as stakeholders. Each actor should have a clear obligation to ensure that laws and regulations are properly followed. The advisory guidelines[26] of India emphasizes accountability by mandating transparency, informed consent, and responsible use of generative AI technologies, ensuring that developers acknowledge and address the potential limitations of their systems. The requirement for unique identifiers[27] makes it clear who is accountable for the outputs generated by the AI systems.

1. **Documentation:** Whether the AI system is developed internally or by an external vendor, documentation and logs need to accompany the system throughout the supply chain. Each participant ranging from the developer to the vendor to the deployer should carry out their own risk assessments and document the measures taken to manage those risks.
2. **Traceability:** Traceability means keeping track of details from start to finish. It is the need to record detailed information[28] about specific elements or components of an AI system, like input data or models it relies on. It is crucial for effective audits of the system. In essence, transparency refers to the provision of information and disclosures related to an AI system, whereas traceability[29] pertains to the ability to track the components of the AI system before, during, and after its deployment. The traceability checklist[30] is given under the data science toolkit. The advisory guidelines[31] of India enhance traceability by requiring that all information created, generated, or modified through AI

[16] China Cyberspace Administration, *Generative Artificial Intelligence Service Management Provisions* (July 13, 2023), http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm.

[17] Ibid., at 14

[18] **S.B. 896**, 2023-2024 Reg. Sess., § 2(d) (Cal. 2024), https://www.derechosdigitales.org/wp-content/uploads/Brazil-Bill-Law-of-No-21-of-2020-EN.pdf.

[19]Brazil, Law No. 21 of 2020, art. 5(v), https://www.derechosdigitales.org/wp-content/uploads/Brazil-Bill-Law-of-No-21-of-2020-EN.pdf.

[20] Office of the Privacy Commissioner of Canada, *Principles for Responsible, Trustworthy and Privacy-Protective Generative AI Technologies* (Dec. 7, 2023), https://www.priv.gc.ca/en/privacy-topics/technology/artificial-intelligence/gd_principles_ai/

[21] Ministry of Electronics and Information Technology, *Advisory on Artificial Intelligence* (Mar. 15, 2024), https://www.meity.gov.in/writereaddata/files/Advisory%2015March%202024.pdf.

[22] Regulation (EU) 2021/1231 of the European Parliament and of the Council of 15 June 2021, *Artificial Intelligence Act*, art. 52, 2021 O.J. (L 231) 1, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

[23] Ibid., at 14 OECD.

[24] OECD, *Artificial Intelligence and the Digital Economy: A Policy Perspective* (2023), https://www.oecd-ilibrary.org/docserver/2448f04b-en.pdf?expires=1728904777&id=id&accname=guest&checksum=61DE9B045F1D7F03CFDFEA4C830D0A7C.

[25] ibid., at 24.

[26] Ministry of Electronics and Information Technology, *Advisory on Artificial Intelligence* (Mar. 15, 2024), https://www.meity.gov.in/writereaddata/files/Advisory%2015March%202024.pdf.

[27] Ministry of Electronics and Information Technology, *Advisory on Artificial Intelligence* (Mar. 15, 2024), guideline 3, https://www.meity.gov.in/writereaddata/files/Advisory%2015March%202024.pdf.

[28] Ibid., at 24

[29] Ibid., at 24

[30] IDB-OECD (2021), Responsible use of AI for public policy: Data science toolkit, https://publications.iadb.org/publications/english/document/Responsible-use-of-AI-forpublic-policy-Data-science-toolkit.pdf.

[31] Ministry of Electronics and Information Technology, *Advisory on Artificial Intelligence*, guideline 3 (Mar. 15, 2024), https://www.meity.gov.in/writereaddata/files/Advisory%2015March%202024.pdf.

systems be labeled with unique metadata. This means any output generated by generative AI can be traced back to its source, holding developers and intermediaries accountable for the content their systems generate.

3. **Audits to ensure compliance:** Auditing AI systems after development[32] plays an important role in verifying their proper functioning and ensuring that appropriate risk assessment and mitigation strategies are in place. It's essential to regularly check, revisit, and reassess[33] accountability measures, including bias testing and evaluations, as both generative AI systems and AI regulations continue to evolve.

   European Commission[34] outlines methodologies for evaluating AI performance, ethics, and compliance with regulations, emphasizing the importance of human oversight in AI auditing processes.

   Developers must ensure that their systems adhere to existing legal frameworks[35], thereby increasing scrutiny over how these technologies are built and deployed. Developers[36] and operators must ensure compliance with relevant laws. Developers of generative AI are required not only to adhere to current laws but also to implement best practices and international standards. This compels them to establish strong risk management strategies to assess and enhance their AI systems continuously.
4. **Human oversight:** Human oversight is essential for managing the unpredictability of AI systems, especially when it comes to generating content. Under Article 14 of the EU AI Act, requires developers of high-risk AI systems, like generative AI, to provide clear, understandable guidelines about how their systems work. Additionally, they should equip, human oversight team with tools that allow them to step in and intervene when necessary[37]. This is important to keep AI system under control in order to prevent generating harmful outcomes or unintended outcomes.

Privacy

In India, privacy jurisprudence is driven by the DPDP Act and the landmark judgement of KS Puttaswamy. The DPDP Act provides for the processing of digital personal data in a manner that recognises both the right of individuals to protect their personal data and the need to process such personal data for lawful purposes[38]. AI is not specifically mentioned in the recently passed DPDP act.

1. **Data minimization:** According to Open AI, the model is trained on, '(1) information that is publicly available on the internet, (2) information that we license from third parties, and (3) information that our users or our human trainers provide'[39]. The privacy policy that the users agree to, contains the details about the information collected and the purposes for which it is being used. This data is the foundation for the fundamental operation of machine learning and artificial intelligence models as LLM's are trained on such data collected. It does not violate the rights of its users because it uses the data that is already available in the public. Section 3(c)(ii) of the DPDP Act, provides that personal data made or caused to be made publicly available by the Data Principal or by any other individual who is required by law to do so is exempt from the application of the Act's requirements.
2. **Consent:** Consent should be specific, unambiguous, freely given, and one should be able to withdraw it with the same ease as when it was given, according to Art.6 GDPR. In India, Section 4 and 5 of the Digital Personal Data protection Act deal with user consent when a data intermediary processes personal data. The user consent is sine qua non for lawful collection and processing of data. Consent should include informing the users the different ways in which their information shall be used and and about what they could exercise a control over their data.
3. **Need for classification as personal data and otherwise:** The data of users that is publicly available is not classified and includes personal data[40]. Rule 5 of The Information Technology (Reasonable Security Practices and Procedures and sensitive personal information) Rules, 2011, mandates getting the consent of the user before collection of sensitive personal data and provides for storage limitation, purpose limitation, withdrawal of consent and autonomy of the data subject to amend the information. Generative AI companies cannot distinguish personal or sensitive personal data due to the massive volumes of the data being processed. Such Sensitive personal information collected like login credentials and payment information is shared to affiliates, vendors and service providers, law enforcement, and parties involved in Transactions subject to limitations laid down by the local laws. Hence the local laws must clearly define the duties of intermediaries in the manner of processing of data.
4. **Right to erase (Art.17 of GDPR)** Unlike traditional data aggregators, data fiduciaries, GAI running on big data has no mechanism for unlearning and the Right to be forgotten takes a new dimension with the rapid development of AI. A possible solution for this issue is the option of temporary chats as provided by Chat gpt. However, such protections and safeguards are only available to vigilant and informed

[32] Ibid., at 24.

[33] Office of the Privacy Commissioner of Canada, *Principles for Responsible, Trustworthy and Privacy-Protective Generative AI Technologies* (Dec. 7, 2023), https://www.priv.gc.ca/en/privacy-topics/technology/artificial-intelligence/gd_principles_ai/.

[34] European Commission, Auditing Artificial Intelligence: Key Insights and Proposals (2020), https://ec.europa.eu/futurium/en/system/files/ged/auditing-artificial-intelligence.pdf.

[35] Cal. S.B. 896, 2023-2024 Reg. Sess. (Cal. 2024), https://digitaldemocracy.calmatters.org/bills/ca_202320240sb896.

[36] Brazil, Law No. 21 of 2020, art. 6(vi), https://www.derechosdigitales.org/wp-content/uploads/Brazil-Bill-Law-of-No-21-of-2020-EN.pdf.

[37] Caitlin Andrews, *EU AI Act Shines Light on Human Oversight Needs*, IAPP (June 12, 2024), https://iapp.org/news/a/eu-ai-act-shines-light-on-human-oversight-needs.

[38] Preamble, Digital Personal Data Protection Act, No. 22 of 2023, India.

[39] *How ChatGPT and Our Language Models Are Developed*, OpenAI Help Center, https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed (last visited Oct. 8, 2024).

[40] *The Impact of the DPDP Act on Artificial Intelligence and Machine Learning*, Tsaaro Blog, https://tsaaro.com/blogs/the-impact-of-the-dpdp-act-on-artificial-intelligence-and-machine-learning/ (last visited Oct. 8, 2024).

users, as the default settings do not provide them. Consequently, current models operate on a self-regulatory basis within the limitations of local laws. Therefore, it is high time to strengthen the local laws of India.

Fairness

Content created by GAI tools may contain prejudices or stereotypes, or it may be discriminatory or not representative (e.g., biases relating to various and intersecting identity characteristics such as gender, color, and ethnicity). These biases are frequently caused by the vast volumes of internet data, used to train many generative models[41]. For instance, training data may not contain viewpoints that are less common in the data or that have surfaced since the model was trained, and it is likely to represent prevailing historical biases. This affects net neutrality. Following strategies can be used to develop a robust AI system free from bias.

1. **Risk assessment and treatment:**

The OECD principles on AI[42], provides approaches to treat bias and discrimination at every stage of the AI development and deployment process. This includes technical and process strategies like having an inclusive group of members in the planning and designing team, making the model stakeholder centric etc. Risk assessment has to start right from the planning stage and continue throughout. The responses of AI should be assessed through self regulation and through reporting systems. For instance, in the UK, any system that influences the behaviour of humans are classified as high risk systems and are mandated to follow multi pronged regulatory compliance directives.

2. **Reporting systems**

The reporting of lapses in fairness by AI, and subsequent action like arbitration, mediation, settlement, apology and correction measure will aid in the development of a fair AI system.

Grievance redressal and management are crucial for the protection of human interests. The IT Act in India mandates the significant social media intermediaries to have a resident grievance officer and a grievance redressal mechanism; And automated tools should be deployed to identify and remove harmful content which shall be periodically reviewed to identify any bias or discrimination in its functioning. However this is not explicitly extended to AI systems. All intermediaries including AI are directed to not host any information that promotes bias or discrimination that threatens the electoral process[43] as per the Advisory of MEITY.

3. **Human in the loop:**

Human centrality[44] emphasises on protection of fundamental rights while AI deals with human interests. Such human interference is needed to protect the rule of law and democratic values. Rule 4(8) of the Intermediary guidelines and digital media ethics Code, 2021 mandates SSMI's to have a human oversight of the automated tools used to remove violative content.

4. **Re-skilling and up-skilling the AI system.**

The loopholes identified through the above processes should be corrected by retraining the data set[45], removing unnecessary data and narrowing the data used for machine learning[46],treating poisoned data[47], oversampling minority views and undersampling majority views[48], etc,.

Finding :

1. Safe harbour though necessary for GAI, cannot be extended under the present IT Act. Therefore a separate law catering to the peculiarities of GAI is needed.
2. Due diligence can be ensured if an AI model adheres to the principles of transparency, accountability, privacy and fairness. Such compliance will also be in line with the OECD Principles and the present standards set by different countries around the world for AI.

Discussion :

- In order to **maintain transparency**, the GAI system should notify users that they are interacting with GAI, and when users are impacted by the results of AI, the GAI actors should provide an explanation of how the results arrived.
- Appropriate documentation is necessary in order to trace back its source and **ensure accountability**. AI audits will be carried out to make sure GAI is abiding by the law and human supervision is also necessary to prevent unintended consequences.

[41] Canada, *Guide to the Responsible Use of Generative AI*, https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/guide-use-generative-ai.html#toc-5 (last visited Oct. 8, 2024).

[42] OECD (2022), Rationale for the OECD AI Principle on "Human-centred values and fairness", https://oecd.ai/en/dashboards/ai-principles/P6 (accessed on 9 April 2022).

[43] Ministry of Electronics and Information Technology, Advisory on Artificial Intelligence (Mar. 15, 2024), https://www.meity.gov.in/writereaddata/files/Advisory%2015March%202024.pdf.

[44] Art 5, (II), Projeto de Lei No. 21, de 2020, Senado Federal, Brasil.

[45] Schwartz, R. et al. (2021), Proposal for Identifying and Managing Bias in Artificial Intelligence (SP 1270),https://www.nist.gov/artificial-intelligence/proposal-identifying-and-managing- bias-artificial-intelligence-sp-1270.

[46] Spracklen, L. (2021), Sparse Models are Fast Models: Improving DNN Inference Performance by over 10X.

[47]Wang, Y. et al. (2020), Generalizing from a few examples: A survey on few-shot learning, https://arxiv.org/abs/1904.05046.

[48] Iosifidis, V. and E. Ntoutsi (2018), "Dealing with bias via data augmentation in supervised learning senarios", Jo Bates Paul D. Clough Robert Jäschke, https://www.bibsonomy.org/bibtex/2631924ce7d73cd8e3bb6477e84d408fa/entoutsi.

- To **ensure privacy**, principles of data minimisation and consent in line with the sprit of Digital personal Data protection Act should be upheld; and the The Information Technology (Reasonable Security Practices And Procedures And Sensitive Personal Data Or Information) Rules, 2011 which provides Rules for purpose limitation, and storage and deletion of sensitive personal data within a timeframe should be adhered to.
- **Ensuring fairness** in AI systems is crucial which requires intervention at every stage of the AI model's development and deployment. This includes human oversight, reporting and grievance redressal mechanisms, periodic risk assessments and even reskilling and upskilling the model.

Scope and Limitation :

1. Only Generative AI is taken into consideration
2. The technical aspects of AI, Machine learning and algorithms are beyond the scope of the paper.
3. Safety, security, robustness of AI and their potential risk management processes still need to be addressed in detail.
4. The elements of due diligence determined in this paper, transparency, accountability, fairness, and privacy, are not exhaustive.

Chapter III

Conclusion :

We have attempted to draft a due diligence framework; however more protection in the areas of risk assessment and security of GAI have to be concretised. There is a need to improve handling and protection of personal data by GAI, to bolster user trust and establishing trustworthy AI systems. Policies must enforce these practices to promote accountability, human centrality, non discrimination, neutrality and transparency. Likewise the Digital India Act can cover aspects inspired from the legal systems around the world and develop a solid law for AI ensuring protection of human rights at all stages of AI usage.

Chapter IV

Recommendations and future directions

1. The Digital India Act will be a key legislation in bringing clarity of law regarding safe harbour and due diligence to AI. It can include the elements provided in the paper for establishing sufficient safeguards to the users of GAI.
2. For continuous monitoring of compliance, civil society groups and watchdog agencies may undertake independent assessment and rating of the protections given by companies to its users.