TRILEGAL

APPROACH PAPER ON

REGULATION OF GEN AI IN INDIA

June 2025

Contents

1 Executive Summary

Artificial Intelligence (**AI**), particularly Generative AI (**Gen AI**), has so rapidly permeated into all facets of modern society that its governance has become the subject of vigorous debate. As India considers the regulatory path most appropriate for its unique circumstances, it must learn from the approaches other countries have taken. But, rather than be bound by their choices, it should look to find a path that meets its specific requirements, and in doing so, potentially develop a framework that could be a model for others. This paper sets out the broad considerations that India ought to consider when regulating Gen AI:

- Artificial intelligence will be transformational for India's development.

  "Due to its socio-economic situation, if India wants to take full advantage of AI and reach those presently underserved, we need to develop an approach to the regulation of Gen AI that is different from that being followed in the more developed countries of the world."

- Most countries have chosen to focus their AI regulations on mitigating risk. Some have assumed that since it is a new technology, the risks of AI need to be separately stipulated in new legislation. This, however, overlooks the fact that, in almost all instances, especially where AI is used to generate content, existing laws are more than capable of addressing the new concerns that AI raises. Instead, India should look at how to use these existing laws to enforce against any potential harms, and close any enforcement gaps. Further, it is in India's interests to remain nimble when it comes to the regulation of Gen AI, and where industry self-regulation can help, it should be encouraged.

  India should not enact new legislation to address Gen AI risk, where existing laws can address the potential harm that could arise. Instead, India should look at how to use these existing laws for enforcement. Where industry self-regulation can help, it should be encouraged.

- At the same time, laws can also help to enable AI innovation and adoption. Many of the laws in the rulebook today were drafted before AI came into existence. As a result, the stipulations contained in them may not enable all that Gen AI has to offer, or place India in a good competitive position *vis-a-vis* other major economies.

  To unlock the full potential of Gen AI, we may need to amend some of our existing legal frameworks that inadvertently constrain the operation of AI systems or leave India at a comparative disadvantage relative to other major economies.

- Since Gen AI systems are fundamentally probabilistic and non-deterministic, they operate differently from conventional products. We currently take a binary approach to liability – when a product fails to perform as advertised, we hold the person responsible for it liable. This approach makes little sense, given that the probabilistic nature of AI is a feature, not a bug.

  We need a new liability standard for AI – one where persons responsible have the chance to remedy faults based on a notice and remediation framework, and are made liable only if they fail to do so or if they are negligent. Highly sensitive content, such as content that poses Chemical, Biological, Radiological and Nuclear risks, Child Sexual Abuse Material and Non-Consensual Intimate Images, could be exemptions to this rule.

2 Introduction

Artificial Intelligence (**AI**) is a term used to describe the computational technologies that simulate human learning, comprehension, problem-solving, decision-making, creativity, and autonomy. Recent improvements in deploying deep neural networks on highly performant hardware have resulted in outcomes that have begun to approach human intelligence. There have been calls to regulate this new technology because of the perceived risk that harm could be visited on society if it is allowed to proceed unchecked. This is particularly the case in the context of Generative AI Models (**Gen AI model**), where AI models have the capability to generate content of various modalities such as language, image, video and audio (**Gen AI**).

A prominent example of Gen AI is Large Language Models (**LLM**) – the transformer-based technology that lies at the cutting edge of this revolution – which can understand the relationships between words and phrases from sequences of text. By representing words as multi-dimensional vectors and assigning them weights, they can predict the next word in a sentence while staying true to patterns of grammar, semantics, and other conceptual relationships. Similarly, convolutional neural networks enable Gen AI Models to develop *'computer vision'*, where they can learn the hierarchy of features in images and videos (e.g., simple features such as edges, corners, textures and, thereafter, complex patterns and shapes). These technologies have enabled Gen AI Models to be used in a variety of downstream applications and systems (**Gen AI System**) to undertake a variety of activities.

Gen AI Systems can create original content (text, images, videos, etc.) in response to user prompts that are, in many instances, very similar to that generated by humans. Concerns have been expressed that these responses could result in harm. Countries around the world have been trying to put in place regulations to mitigate the risks of Gen AI. While some have enacted formal legislation, others have merely issued guidance on the possible consequences.

As India considers its options, we must approach this assessment from our own unique perspective. As much as we need to learn from what other countries have done, we must ensure that our approach to regulating Gen AI aligns with our national objectives. As concerned as we need to be about the harms that can result, the risk mitigation we deploy should align with the benefits we stand to gain.

2.1 The Dilemma of Regulating Gen AI

Gen AI Systems offer us in India an unprecedented opportunity to leapfrog the expected trajectory of our economic and social development. They provide novel solutions to various challenges in healthcare, financial inclusion, and education, to name a few. In each of these areas, narrowly designed Gen AI use cases can bridge the development divide and help foster inclusive growth. Enabling bespoke services capable of being delivered at population-scale will allow us to narrowly target interventions to those who need them the most.

That said, there are risks to using Gen AI Systems. Gen AI can be used to generate misleading content of such high quality that even sophisticated users might be misled into believing it is true. Thanks to its capabilities, it will make it easier for laypersons to exploit the vulnerabilities of computer systems that threaten the security of the national critical infrastructure that we rely on. While information has always been available (e.g., textbooks or through web search), AI systems have increased the accessibility of such information, making knowledge that would otherwise have been hard to obtain and appreciate more easily understood. This could also apply to information that may be hazardous, becoming more accessible to malefactors.

Indian policymakers have to resolve this complex dilemma. They need to safeguard society from the many risks that Gen AI presents. At the same time, they need to ensure that we are well-positioned to avail ourselves of its benefits. This might mean that India will have to adopt an approach more broadly tolerant of risk than developed countries. In the developed world, AI is a luxury that can be foregone if the danger it poses is too high. In India, AI is a potential avenue for the poorest among us to avail of the benefits society offers.

2.2 MeITy Report on AI Governance

On 6 January 2025, the Ministry of Electronics and Information Technology (**MeITy**) published a report that sets out recommendations on how AI should be governed in India. It proposes a whole-of-government approach and the establishment of an empowered mechanism to coordinate AI Governance. It suggests that the government set up a

permanent secretariat tasked with providing technical advice and an AI incident reporting database to mitigate repeated bad outcomes. It encourages industry engagement with a view to securing voluntary commitments on transparency and leans into the use of technology measures to address AI risks.

While the recommendations are sound, there is a lot of discussion about the risks of harm and the need to encourage fairness, transparency and openness. While none of these principles can be ignored, they must only be adopted after carefully considering India's social and economic situation. Given that AI offers India the opportunity to leapfrog several stages of its development, in several circumstances, the real risk of AI is that we will not use it enough. Perhaps with this in mind, the MeITy has followed up this report with further initiatives under the IndiaAI mission, such as launching the AIKosha portal (for datasets and models), AI Compute Portal (for dissemination of compute resources), and the AI Competency Framework for Public Sector Officials and iGOT-AI Mission Karmayogi (for government and civil services), to name a few. The MeITy is also keen on the development of an indigenous foundation model for India and has selected a startup to develop India's first sovereign LLM.

While we do not disagree with the findings and the recommendations of the MeITy report, there are several issues it has not addressed in relation to Gen AI. This paper attempts to fill in those gaps by offering a set of policy considerations that the Government ought to give credence to in enacting a regulatory framework for Gen AI in India. However, this paper is in no way attempting to cover all aspects relating to Gen AI regulation.

2.3 Scope and Structure of the Paper

India does not yet have a comprehensive approach to regulating Gen AI.[1] That said, Gen AI is already being used in the country in various ways. Rather than unquestioningly adopting the regulatory models developed economies use, we need to define our own strategy for regulating Gen AI in India – one that draws on global best practices but enables all India hopes to achieve.

This paper presents a framework India can use to develop such a strategy. It argues that our attempts at risk mitigation must always be balanced against the benefits we stand to gain from using AI. As a result, India may need to develop a bespoke set of trade-offs to suit its unique circumstances.

The scope of this paper is to address the regulation of providers, developers and deployers of Gen AI Models for content generated by these Gen AI Models.[2] Simply put, the provider of a Gen AI Model is the entity that makes available the Gen AI Model in the Indian market for use by third parties, including as part of a Gen AI System. The provider is typically also the developer of the Gen AI Model, but this may not always be the case. A deployer is the entity that uses the Gen AI Model as part of a Gen AI System under its own authority for its operations.[3]

In relation to Gen AI, the paper does not assess principles of liability concerning secondary harms arising out of the use of Gen AI Systems in specific contexts, or harms that arise during the training or development of Gen AI Models. It does not propose principles of liability applicable to end users of the Gen AI System and any harms that may arise

---

1 While the Government of India has issued various advisories – relating to misinformation arising out of AI-generated deepfakes (MeITy Advisory dated 26 December 2023 on '*Due diligence by Intermediaries and Grievance Reporting Mechanism, under the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 – regarding*' and generally in relation to AI models, requiring intermediaries and platforms to undertake certain due diligence obligations (MeITy Advisory dated 15 March 2024 on '*Due Diligence by Intermediaries/Platforms under the information Technology Act, 2000 and the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021'*), these attempts at establishing a regulatory framework have confused more than they have clarified. More recently, the MeITy constituted an advisory group, chaired by the Principal Scientific Advisor, that produced a report titled *'Report on AI Governance Guidelines Development'* that provides actionable recommendations for AI governance in India.

2 This paper distinguishes between first order harms of Gen AI Models and secondary harms arising from the use of Gen AI Models. First order harms arise from the outputs, i.e., content generated by Gen AI Models in response to the prompts inputted to it. The harm that arises in this case is a result of such prompt and such output and does not involve any subsequent use or applications of the output. For example, hate speech generated by a Gen AI Model in response to a harmless prompt from a user is a first order harm. A second order harm on the other hand is where the output generated by a Gen AI Model is subsequently used or applied in a specific scenario by the user to cause harm. For instance, where a user prompts a Gen AI Model to generate malware code which the user thereafter injects into a computer resource. First order harms are in the nature of *'content'* harms and are the focus of this paper.

3 To illustrate this by way of an example, if A has developed an LLM called *'XYZ'* and a chatbot based on XYZ called *'XYZ-Chat'* and makes it available to third parties on the market, A would be considered the provider of the Gen AI Model XYZ and the Gen AI System XYZ-Chat. If B, an e-commerce entity, licenses XYZ-Chat from A and integrates it into its customer support application *'HelpEcom AI'*, then it would be considered the deployer of HelpEcom AI. A would also step into the shoes of the deployer of XYZ-Chat where it is being integrated within the chatbot HelpEcom AI to users under its own authority.

from their use of the Gen AI System towards specific ends. It also does not address AI systems such as agentic AI or kinetic systems, which function at a level of complexity higher than Gen AI Models.
The rest of the paper has been structured as follows:

- In *Section 2,* we analyse regulations from around the world to evaluate the approach other countries have taken to regulating AI. This analysis provides us with a list of questions India would need to answer to develop its regulatory frameworks.

- In *Section 3,* we discuss the risks that Gen AI poses. We evaluate whether India needs to enact a new dedicated regulation to mitigate these risks or whether its existing regulations are broad enough to suffice.

- In *Section 4,* we evaluate the impediments that existing regulatory frameworks impose on the development of Gen AI and ask whether amendments are required to promote AI's growth in India.

- In *Section 5,* we examine how our traditional approach to product liability could affect the growth of Gen AI and question whether this essentially binary framework is suited to the probabilistic way in which Gen AI works.

- In *Section 6*, we summarise the paper's findings and set out our recommendations as to the approach the Indian government should adopt in developing a regulatory framework for Gen AI.

3 International Approaches to AI Regulation

Countries have already started to regulate AI. While the European Union (**EU**) and China have introduced new laws, others have issued guidance and other soft law instruments. In addition, AI organisations and industry bodies have put in place self-regulatory measures and public safety testing regimes to help mitigate the risk of harm.

These measures vary in focus, orientation, and approach. They seek to classify AI in different ways to arrive at the outcomes they suggest. An analysis of these approaches will help arrive at the key regulatory considerations India should assess to formulate its own approach to Gen AI regulation.

3.1 Other Jurisdictions

Let us first understand how certain key jurisdictions, both developed and developing, have gone about governing AI.

3.1.1 The European Union

The EU has enacted a full-fledged legislation in 2024 – the EU Artificial Intelligence Act (**EU AI Act**)– that applies to AI systems placed on the market, put into service, or used in the EU, as well as general-purpose AI models placed in the EU market.

If an AI model has high impact capabilities, determined by technical tools and benchmarks, or if it has similar capabilities or impact as decided by the European Commission (having regard to several criteria, including parameters, dataset size, etc.), it is considered to have systemic risk. If the AI model uses a large amount of computation for its training, it is assumed to have high impact capabilities. Based on such classifications, the EU AI Act imposes certain obligations on providers of general-purpose AI models. It also imposes obligations on the providers, importers, distributors, and deployers of AI systems based on the level of risk they pose – classifying them as either prohibited, high-risk, limited or minimal-risk AI. It imposes strict penalties for non-compliance, with fines ranging from €35 million or 7% of global turnover to €7.5 million or 1.5 % of turnover.

The EU AI Act is focused on the protection of fundamental rights of individuals within the EU. To do that, it imposes strict obligations with regard to how these models should behave and the transparency and reporting requirements of organisations that are making these models available for use. While it is a horizontal regulation, it interacts with and sometimes defers to existing sectoral legislation, and it does not require all AI applications to comply with the same obligations. The requirements are tiered according to risk, with the most stringent rules applying only to high-risk systems.

The EU AI Act has established a new regulator – the European AI Office – which is responsible for implementing and enforcing the Act and providing guidance to businesses and organisations on compliance with the law.

3.1.2 China

China has enacted different regulations to address specific manifestations of AI. These include regulations on algorithmic recommendations, deep-fakes (referred to as deep synthesis information services) and generative AI services. Within these domains, the regulations stipulate what AI companies can do, the reviews and assessments they must conduct and other stipulations as to how outputs must be generated. While China, like the EU, has enacted laws, they are more vertically focused than the EU.

The Cyberspace Administration of China (**CAC**) has considerable latitude in interpreting these regulations. This will most likely align with the priorities of the Chinese government – to ensure that AI companies respect social mores and adhere to the political direction and orientation of the government.
China has also created a national algorithm registry to which organisations must submit reports of the algorithms their systems use. The scope of the registry is continuously being updated to include various AI models and will likely continue to evolve as AI develops.

3.1.3 United States of America

AI governance in the United States of America (**US**) is in a state of flux.

Under the erstwhile Biden administration, the US White House had issued an Executive Order on AI and secured Voluntary Commitments from the AI industry. The focus of these measures had been to address concerns such as fraud, bias, disinformation, and risks to national security, and set out guiding principles that address safety, consumer protection, workers' protection, and responsible government use.

The Executive Order imposed reporting requirements on so-called *'dual use foundation models'* that require a quantity of computing power greater than 10^{26} integer or floating-point operations per second (**FLOPS**). Training compute was accordingly used as a proxy for dangerous capabilities.

The Donald Trump administration, on 23 January 2025, passed an order revoking the 2023 Biden Executive Order and calling for departments and agencies to revise or rescind actions taken pursuant to it. The new policy of the US under the Trump administration is to *'sustain and enhance America's global AI dominance in order to promote human flourishing, economic competitiveness, and national security'*. While a proposed AI Action Plan is expected to elaborate on these policies, it appears that the new federal US focus is more on the enablement, rather than regulation, of AI. At the state level, however, several legislatures are considering AI legislation.

3.1.4 United Kingdom

The United Kingdom (**UK**) had proposed a pro-innovation approach to regulating AI *via* a White Paper on AI Regulation in 2023, in which it outlined five cross-sectoral principles that UK's existing regulators would interpret and apply within their remits. The new Labour government in 2024 has since announced an AI Opportunities Action Plan in January 2025 to encourage innovation, secure access to compute, create a national data library and empower the UK's AI Safety Institute (since February 2025, the AI Security Institute).

The Action Plan also notes that well-designed and implemented regulation may be required to protect UK citizens from the most significant risks presented by AI, but that this must be done without blocking the path towards AI's transformative potential. To that end, an AI safety bill is being considered to statutorily impose obligations on frontier model developers, including model testing prior to release.

3.1.5 Singapore

Singapore does not currently have a dedicated horizontal regulation on AI. Instead, the Infocomm Media Development Authority has set up the AI Verify Foundation to enable the development of AI benchmarking and voluntary testing tools. In 2024, the foundation published a Model AI Governance Framework for Generative AI to set out certain voluntary principles for a trusted AI ecosystem in Singapore.

On the principle of accountability, this voluntary framework proposes allocating responsibility *ex ante* based on the level of control that each stakeholder in the development chain has, similar to responsibility sharing in the cloud industry. This may necessitate a consideration of different model types. The framework suggests supporting this with *ex post* safety nets for end-users, such as indemnity and insurance.

Other aspects of the voluntary framework include baseline safety practices during development, disclosures (similar to food labels), safety evaluations (the AI Verify Foundation suggested an initial set of standardised model safety evaluations for LLMs), incident reporting and testing.

3.1.6 Brazil

The Federal Senate of Brazil, in December 2024, approved a new bill setting out the framework for the development, use and governance of AI systems. Similar to the EU, the law proposes a risk-based classification of AI systems based on a preliminary assessment by system developers, distributors and application developers.
Excessive risk systems are prohibited, whereas high-risk systems are subject to algorithmic impact assessment, human supervision and transparency measures. Individuals and groups affected by AI systems also have certain rights, such as to clear, accessible information about the use of AI in their interactions with such systems, to request review by humans of automated decisions in certain circumstances, and certain non-discrimination rights.

3.1.7 Japan

The Japanese government has passed an Act on the Promotion of Research and Development and the Utilisation of AI-Related Technologies, focused primarily on the promotion of AI and establishment of an AI Strategy Headquarters. Japan's approach to AI governance has been to rely on its existing system of laws and certain soft laws published by various ministries. Pertinent among these are the AI Guidelines for Business Ver1.0 issued by the Ministry of Economy, Trade and Industry in 2024, which proposes non-binding principles for AI developers, AI providers and AI business users. For example, AI developers ought to assess the potential impact of their AI in advance and take measures to address the same, whereas AI business users must comply with guidelines set by AI providers.

These examples provide us with various approaches from which to derive the regulatory considerations India needs to evaluate.

3.2 Regulatory Considerations

The following key considerations need to be examined:

- What should be the regulatory approach towards content generated by Gen AI?
- How should Gen AI Models/Systems be classified?
- How should the risks of content generated by GenAI be addressed?

Each of them has been examined in detail below.

3.2.1 Regulatory Approach

What sort of regulation should we enact? Should we enact a new law or rely on existing provisions?

Centralised Omnibus Law: The EU AI Act is an omnibus law that takes a broad horizontal approach to regulating AI. It requires all AI applications and general-purpose AI models to adhere to the regulatory framework it describes, regardless of the sector in which it is being applied or the use to which it is put. It creates a single regulatory authority responsible for the enforcement and governance of AI that will look into AI-related matters regardless of the sector to which it is applied. Brazil has adopted a similar approach to AI regulation.

The challenge with a one-size-fits-all regulatory strategy is that it comes at the cost of flexibility and agility. By casting regulatory principles in stone, it will be unable to properly account for sectoral variations and the future evolution of this dynamic technology. A case in point are the DeepSeek LLMs, which are estimated to have been trained below the EU AI Act compute threshold of 10^{25} FLOPS (i.e., the threshold to be classified as a general-purpose AI model with systemic risks).

Vertical Domain-Specific Legislation: China has implemented laws regulating different types of AI, including separate ones for generative AI and algorithmic regulations. This aligns with its regulatory interests to ensure that the development of AI is aligned with the interests of the State. Domain-specific laws allow the Chinese government to control each of these manifestations of AI so that no matter which sector it is applied to, it will perform as the State intends.

This is far more fine-grained than the EU's omnibus approach. While it is well suited to the Chinese command and control structure, with the CAC exerting control over all domains in which Gen AI is deployed, it will struggle to address sector-specific harms that could arise.

Guidance: In the US, the erstwhile Biden administration had chosen to offer guidance rather than issuing regulations, with the Trump administration revoking the Biden presidential order on AI. In the UK, sectoral regulators may continue to determine the likely effect of AI in their domains and decide whether to enact new subordinate legislation or merely be more thoughtful about how existing regulations should be extended to cover the new harms caused by AI. Japan has also so far relied on a light-touch approach focused on existing laws, and Singapore similarly takes a voluntary and non-regulatory approach to AI governance.

This approach is more flexible. Since these stipulations are not hardcoded into law, they allow for sectoral variations in application and, as a result, are more adaptable to short-term technology evolutions.

Self-Regulation: Companies and industry associations have also come together to develop self-regulatory frameworks for AI. These range from those established by AI developers[4] to special AI-focused initiatives of government bodies of which the industry is a part.[5] There are also efforts to put in place voluntary commitments that companies can sign up to.[6]

These efforts at self-regulation rely on the companies themselves determining what their commitments should be. Given that the constraints they impose on themselves are likely guided by commercial self-interest, this approach may not provide adequate comfort.

3.2.2 Classification

Do we need to identify an appropriate metric that can be used to classify Gen AI Models to stipulate a *de minimis* threshold above which regulations apply? If so, what criteria should we adopt? The EU AI Act, in determining if a general-purpose AI model has high impact capabilities, takes into account factors such as number parameters, tokens, amount of compute, modalities and benchmarks.

As much as highly capable systems are more likely to have a higher risk of harm than less capable ones, imposing higher obligations based solely on their capabilities will prevent us from taking full advantage of all these systems offer. The Indian ecosystem of deployers must be allowed to leverage highly capable models made available on an open-source/open-weight basis, even if they do not have the deep pockets to stand behind all the risks of such highly capable models. At the same time, the foundation model developer must not be held to task for the deployment-level harms they cannot foresee. Moreover, even less performant models can, in the wrong hands, be used to cause harm.

In developed countries, AI is a luxury that will augment their already advanced lifestyles. In this context, it makes sense to have different mechanisms by which risk can be mitigated even if, as a result, they will be forced to forego some of the benefits of AI.
To developing countries like India, AI systems are transformational, offering solutions for problems that would otherwise have taken decades to solve. To us, being able to use AI systems is existential, not optional. That being the case, the broad imposition of increased regulatory burden on more capable models would be an unnecessary curtailment of the benefits that could be derived from Gen AI. The regulatory approach towards other types of AI models and systems may similarly need to be reassessed (although that is outside the scope of this paper).

3.2.3 Risks

The EU AI Act and Brazil have listed various applications of AI systems and ascribed a level of risk to these uses. It is assumed that whenever AI is used in a particular sector (e.g., critical infrastructure, education, border control) or for a particular purpose (e.g., biometrics, employment), it will be *prima facie* harmful. As a result, entire sectors could be denied the benefits of AI. What's more, a range of use cases may never be developed because the regulatory burden of doing so will be too great for any entrepreneur to bear.

In developing countries like India, risks of Gen AI should be measured based on outcomes rather than broadly imposed on specified sectors and applications. This will allow the benefits of Gen AI to reach the largest cross-section of society. Furthermore, to evaluate whether new laws need to be enacted to address Gen AI risks, we should first assess the extent to which existing laws address the harms that need to be mitigated.

---

4 One such example is the Frontier Model Forum established by Anthropic, Google, Microsoft and OpenAI.

5 AWS, Dell, Google, IBM, Microsoft, OpenAI, Red Hat, Adobe, Meta, among others, are members of the AI Verify Foundation in Singapore.

6 Amazon, Anthropic, Google, Inflection, Meta, Microsoft, OpenAI, Adobe, Cohere, IBM, Nvidia, Palantir, Salesforce, Scale AI, and Stability had signed up to voluntary commitments in the US under the Biden administration, although the status of the same in the Trump administration is unclear. Similar efforts are underway in India in the form of an AI Promise initiated by People+AI.

India needs to adopt the most flexible and agile regulatory approach to account for the evolution of technology. India needs to balance its approach to risk mitigation against the benefits that will accrue from using AI.

Doing this will most likely diverge from the approach of developed countries in terms of both the classification of Gen AI Models/Systems and the definition of risk. That being the case, how should India approach the regulation of Gen AI?

In this context, three questions arise.

- *Do we need to enact new regulations to mitigate Gen AI risks?*
- *Do we need new regulations to unlock all that Gen AI has to offer?*
- *Do we need to think differently about the liability of Gen AI systems?*

The subsequent sections of this paper will address each of these questions.

4 Regulations to Mitigate Gen AI Risks

In almost every instance, what is identified as a new Gen AI risk is just a new manifestation of an existing risk that has been either amplified or transformed by AI. The question we need to ask is whether we need a new law or if existing regulations are broad enough to cover the new risks of Gen AI.

In this section, we will examine the regulation of Gen AI under two broad categories:

Covered Harms: the risk of content harms already recognised within our existing legal frameworks, but which Gen AI either exacerbates or makes easier to commit.

New Harms: the risk of content harms that were previously not possible but which, on account of the new capabilities unlocked by Gen AI, have now come into being.

4.1 Covered Harms

Gen AI Systems can be used to generate audio and video content that is indistinguishable from reality. Advanced image generation models can create images that look like real photographs. Voice models can not only imitate the way people speak, but can also capture intonation and mannerisms with such fidelity that it seems as if the person concerned is actually speaking. Systems that deploy such Gen AI Models can manipulate reality to the point where our trust in the truth vanishes. In the hands of malicious actors, this could result in the generation of false, yet utterly believable, information that manipulates public opinion. This sort of fake content can be tailored to inflame base emotions by amplifying false news, conspiracy theories, and propaganda in more persuasive ways than traditionally fact-checked sources. In the hands of malicious actors, all this could be used for defamation, to compromise the political process or call into question the integrity of institutions.

AI is not harmful in and of itself. However, it offers easier (often cheaper) ways to cause harm. It is not as if fake content was not being created before Gen AI. However, the fact that high-quality videos can be generated in minutes with just a prompt – without needing to commission studios and artists – considerably reduces the friction in generating fake news. As a result, it is much easier than ever to produce misleading information, create computer viruses, or disturb the peace. While not the risks of AI systems themselves, these harms are more easily caused by the availability of Gen AI.

Given the widespread availability of these technologies, it has (indicatively) become easier to do the following:

Impersonation: Gen AI can be used to impersonate a real person by generating believable text that falsely attributes statements to them. Speech AI can clone their voice and speaking mannerisms, effectively putting words into their mouths. AI systems can be used to generate video content that portrays real persons doing things that never happened.

Impersonation is covered under the Information Technology Act, 2000[7] *and is a crime under the Bharatiya Nyaya Sanhita, 2023.*[8]

Misleading Advertising: Gen AI can be used to generate false information about a product or service –advertisements that suggest it performs in ways that it cannot. Since AI content is highly believable, consumers could be misled into thinking this is true.

The making of false or misleading advertising is covered under the Consumer Protection Act, 2019.[9]

7 Section 66D of the Information Technology Act, 2000 makes it a punishable offence to, by means of any communication device or computer resource, cheat by personation.

8 Section 319 of the Bharatiya Nyaya Sanhita, 2023 makes it an offence to cheat by pretending to be some other person, or by knowingly substituting one person for another, or representing that he or any other person is a person other than he or such other person really is.

9 Section 89 of the Consumer Protection Act, 2019 makes false or misleading advertising a punishable offence.

Election Interference: Gen AI can be used to generate content that presents political parties or candidates in poor light. It can be used to engender feelings of enmity or hatred between different classes of citizens for political advantage. This can unfairly influence the outcomes of elections.

Specific provisions of the Representation of the People Act, 1951 address this.[10]

Forgery: Image generation AI could be used to generate believably false documents and electronic records. In the context of a commercial transaction, this could mislead parties into entering a contract or performing actions, believing the false document to be true.

This sort of forgery would be a crime under the Bharatiya Nyaya Sanhita, 2023.[11]

Fraudulent Contracts: Gen AI could be used to induce a contracting party to accept the terms of a contract by misleading them into believing that someone has endorsed the transaction or is part of a product or service that it is not.

Contracts obtained by fraud are voidable under the Indian Contract Act, 1872.[12]

Disrupting the Peace: AI-generated fake news can be used to inflame tensions between communities by portraying individuals from different castes, religions or communities as doing something they did not.

This is a crime under the Bharatiya Nyaya Sanhita, 2023.[13]

Defamation: Gen AI can be used to generate content that makes it appear as if a person did or said something that they did not. This can negatively impact a person's reputation.

The act of making an imputation intending to harm a person's reputation is a crime under the Bharatiya Nyaya Sanhita, 2023[14] *regardless of the technology used to do so*.

Bias: Gen AI Systems have been shown to demonstrate bias when flawed human decisions in their training data influence the inferences they make.

All entities that are 'State' within the meaning of Article 12 of the Indian Constitution have a constitutional obligation to act fairly. Even though this does not extend to private parties, discrimination against certain types of persons by private parties – such as in relation to scheduled castes and tribes,[15] *disabled persons,*[16] *persons with mental disabilities,*[17] *etc. – is prohibited under the law.*

---

10 Section 125 of the Representation of the People Act, 1951 makes it an offence to, in connection with an election, promote on grounds of religion, race, caste, community or language, feelings of enmity or hatred, between different classes of the citizens of India. Section 171 makes it an offence to voluntarily interfere or attempt to interfere with the free exercise of any electoral right. Section 174 makes it an offence to use undue influence or impersonate anyone else during an election. Section 175 makes it an offence to make or publish any statement purporting to be a statement of fact which is false and which he either knows or believes to be false or does not believe to be true, in relation to the personal character or conduct of any candidate.

11 Section 336 of the Bharatiya Nyaya Sanhita, 2023 makes it an offence to make a false document or false electronic record with intent to cause damage or injury to the public or to any person.

12 Section 19 of the Indian Contract Act, 1872 states that where consent to an agreement is obtained by fraud or misrepresentation, the agreement is a contract voidable at the option of the party whose consent was so caused.

13 Section 196 of the Bharatiya Nyaya Sanhita, 2023 makes it an offence to, through the use of electronic communication, promote disharmony or feelings of enmity, hatred or ill-will between different religious, racial, language or regional groups or castes or communities. Similarly, Section 353 of the Bharatiya Nyaya Sanhita, 2023 makes it an offence to publish or circulate any statement, false information, rumour, or report, including through electronic means, with intent to cause fear or alarm to the public, or to any section of the public whereby any person may be induced to commit an offence against the State or against the public tranquillity or is likely to incite any class or community of persons to commit any offence against any other class or community.

14 Section 356 of the Bharatiya Nyaya Sanhita, 2023 makes it an offence to make or publish in any manner, any imputation concerning any person intending to harm the reputation of that person.

15 Section 3, The Scheduled Castes and the Scheduled Tribes (Prevention of Atrocities) Act, 1989.

16 Section 16 and 20, Rights of Persons with Disabilities Act, 2016.

17 Section 21 and 108, The Mental Healthcare Act, 2017.

Copyright: Although the very nature of Gen AI systems is that they ought to create a new product or output, in some cases, they could generate content that is a verbatim copy of the content created by a human creator.

Copyright law treats any such actions as an infringement.[18] Tools like output filters can help restrict substantially similar outputs.

In India, we already have laws that can address these risks of Gen AI. For the most part, these laws are drafted in terms broad enough to address the harms Gen AI could cause. There is no need to enact a new law merely because these risks are being caused using a new technology.

That said, we should educate regulators and law enforcement agencies on how Gen AI technology may make it possible for perpetrators (bad actors) to commit harm in new ways. They should be trained in new forensic and investigative techniques so that they can learn when Gen AI systems are being misused to cause harm and secure the evidence required to prosecute them.

4.2 New Harms

Since AI is a new technology, it could create new harms that may not be adequately addressed by existing law. Listed below are a couple of examples of some new harms that could result from the widespread availability of Gen AI.

4.2.1 Intellectual Property Rights

Since Gen AI Models depend on the data they are trained on, artists whose content has contributed to the training dataset may request appropriate revenue-sharing mechanisms, allowing them to partake in the revenue these models earn. While some companies have promised to pay rightsholders a fee when their content is used, it is unclear how this will be implemented, given the multi-epochal training cycle of neural networks.

Copyright law only protects the specific expression of ideas in original works of authorship (not the ideas underlying the work), and it is only in these scenarios where AI-generated output copies existing works of art that a situation of infringement arises. However, there may also be cases where AI generates content remarkably similar, in look and style, to the works of existing artists, even if it is not substantially similar to any specific artwork. While this is likely not copyright infringement under Section 51 of the Copyright Act, the widespread generation of AI artwork in the style of an artist could diminish the market for her creations and impact the artist's ability to monetise her talent.

However, rather than widening the scope of copyright infringement, the aforementioned aspects can instead be addressed through contractual relationships between relevant parties (such as revenue sharing arrangements) and appropriate and proportional protection granted to the intellectual property holder over training data (discussed below in section 5.1).

4.2.2 Ease of Accessibility of Information

Gen AI Systems enable access to information that may be otherwise hard for individuals unfamiliar with the subject to find, interpret, contextualise or understand on their own. They can present this information in plain and simple language. This goes beyond merely finding and presenting information that has been published online – it could also extend to the creation of new content (synthesising old content) that is customised per the user's requirement to ensure they understand what they have requested. This has created an unprecedented intersection between those who know how to use information to cause harm and those who have no moral compunctions doing so. This has given rise to concerns that criminal elements will be able to build bombs, make chemical or biological weapons, and 3D-print handguns far more easily than is currently possible.

While Gen AI might make this information easier to access, the information has always existed. It is only when it is used to cause harm that a crime or other violation takes place. In such an event, only the person who uses the information should be held liable. That said, we should consider measures to mitigate this risk.

[18] Section 51 of the Indian Copyright Act, 1957.

4.3 Evaluating Harm

We may be able to address the risks of Gen AI better if we know of their existence early enough to mitigate them. Since AI is being used around the world, evidence of new risks could arise anywhere. AI researchers would benefit from a system that highlights risks as and when they arise.

AI Safety Institutes (**AISI**) are government-backed organisations that identify, assess, and mitigate potential risks associated with advanced AI models and systems, particularly those with highly capable general-purpose capabilities. Once established around the world, they will be able to share information with each other so that the harms that manifest in one jurisdiction can be quickly addressed and mitigated in another.

AISIs perform empirical evaluations of AI models and systems to develop standards for safety. They establish evaluation benchmarks against which new AI models can be tested. Based on these efforts, they can lay down guidelines and best practices for the safe development and deployment of Gen AI technologies. AISIs allow us to address the risks of AI preemptively. This will limit the number of people affected and allow us to take preemptive action to mitigate the consequences.

On 31 January 2025, the MeitY announced the establishment of the IndiaAI Safety Institute for Responsible AI Innovation, which is to be set up on a hub and spoke model with various research and academic institutions and private sector entities taking up projects under the IndiaAI mission. India should further empower the IndiaAI Safety Institute - not only to assess the implications of AI models and systems within its jurisdiction, but also to benefit from being a part of the network of AISIs worldwide so that we can receive early warning indications of new threats.

Most AI-generated content risks are adequately covered by existing law. There is no need to enact new laws simply because harms already covered under existing regulations can be caused by new technology. That said, to the extent that Gen AI makes possible new risks that otherwise might not have occurred, existing laws may need to be amended to address these new harms.

In addition to putting in place legal and regulatory frameworks that specify how harms caused by Gen AI can be addressed, it is just as important to have institutional frameworks designed to identify Gen AI risks expeditiously to ensure that they can be mitigated. The IndiaAI Safety Institute can be empowered, similar to other safety institutes across the world, in this regard.

5 Regulations to Enable Gen AI

While media attention is focussed on the competition between frontier models to see which of them will be the first to achieve artificial general intelligence, for India, the real value of Gen AI lies in a multitude of narrow use cases that will allow us to deliver services to meet the many requirements of a diverse population.
Given the accessibility issues in India, including those of diverse languages and illiteracy (linguistic and digital), Gen AI can play a significant role in providing access. The voice-enabled payment system launched by National Payments Corporation of India and *Jugalbandi* (which deals with information about subsidies, benefits, and government schemes) are good examples of this. By deploying Gen AI on top of our digital infrastructure, India will be able to make digital public infrastructure more accessible.

To do this, we need first to assess what impediments, if any, lie in our way so that we can amend them appropriately, so that existing regulations do not prevent us from using Gen AI to its full potential.

5.1 Copyright

Most LLMs are rich in Western knowledge but lack the local context of Indian conditions. This is because local data is often under-represented in the training datasets on which the model was trained. To remedy this, foundational models need to be post-trained and fine-tuned using datasets rich in local context. In the process of compiling these datasets, there is a possibility that some copyrighted works may be included. Under copyright law, the author's permission will be required even to *'copy'* this into the training dataset.

Around the world, specific exemptions have been introduced into law to facilitate this. These take the form of:

- Fair use exceptions, typically limited to scholarship, research, comment and other similar grounds (as has been recently examined by the US Copyright Office for AI).
- Express copyright exemption for text and data mining (**TDM**) or computational data analysis as is available in the EU and Singapore.

While Section 52 of the Indian Copyright Act lists several exemptions for fair dealing, these exemptions may not provide sufficient certainty to AI developers to would give them the confidence to train LLMs in India.

To provide abundant clarity, India should consider amending the Indian Copyright Act to provide an express exemption for TDM or computational data analysis. This will permit developers to use the large database of contents, including copyrighted data, for training purposes without violating the copyright owners' rights, so long as they have put appropriate technical mechanisms in place to ensure that the output of the Gen AI Model does not result in copyright infringement.

At the same time, it is important to ensure that the labour, creativity, and investment put in by traditional content-generation industries are respected. Therefore, such an exemption could also be made subject to the safeguard that the first copy obtained for such training must have been lawfully accessed.

5.2 Data Protection

It is possible that some personal information may have crept into training datasets that have been scraped off the internet. While the Digital Personal Data Protection Act, 2023 (**DPDP Act**) does not apply to personal data that has been made or caused to be made publicly available by the person to whom it pertains, there are numerous other means by which such data could find its way online. The mere act of training the model using this data would result in it being processed. Since this would occur without the consent of the person to whom it pertains (and without any other legitimate ground for processing), this would amount to a breach of data protection law. If there is a risk that developers will be held liable under data protection law, they will refrain from using these datasets for training – which could have a chilling effect on the entire industry.
Further, the current public data exemption under the DPDP Act is narrow and limited to only those public data which has been made public by the data principal. Since it is almost impossible for a developer to determine whether, in fact, the data principal has put out the data in public, it is necessary to look at a broader exemption for the purpose of AI training.

It might be necessary to amend our data protection law to allow developers to process the personal data contained in training datasets for the sole purpose of training a Gen AI Model.[19] One way might be to clarify that such processing will be permitted for other legitimate uses, provided the data fiduciary balances its legitimate interests against the individual's interests, rights, and freedoms.

A number of laws apply to various aspects of the Gen AI lifecycle. Since many of these laws were enacted before Gen AI came into being, the way they have been framed does not account for how Gen AI operates. That being the case, AI companies risk being held liable for violating various provisions of these laws.

Considering this, even the risk that such accusations could be brought against them could have a chilling effect on innovation in Gen AI. If we want to encourage the use of Gen AI in a broad range of sectors, we need to make appropriate amendments to existing laws to give certainty and encourage innovation.

19 The Digital Personal Data Protection Act, 2023 does exempt processing carried out for *inter alia* research purposes from most of the obligations under the Act (including procuring consent) if such processing is not used to make a decision about the data principal. While one could argue that this could be read broadly to cover training of AI models, it would be far better for a clearer exemption to be set out in the law.

6 Liability

In Section 4, we discussed the risks of AI-generated content and concluded that, for the most part, existing Indian laws that define criminal, civil, or tortious wrongs adequately address most of the new concerns that Gen AI gives rise to, and there is no need for new horizontal AI regulation at this point.

That said, GenAI is a fundamentally non-deterministic technology, and traditional liability regimes will need to evolve to appropriately account for that uncertainty. It will not be appropriate for those responsible for Gen AI to face legal action under traditional liability regimes applicable to media-based applications – i.e., content laws. Given this, we have proposed some policy considerations which may be relevant for the assessment of such a liability regime.

6.1 Assessing Liability

When there is undesirable conduct in society, a legal system must assess how best to deal with this conduct. Where society wants to penalise these, some conduct is classified as criminal offences (for instance, theft), others as civil wrongs (for instance, breach of a certain statute like data protection law) and, in some countries in the world, as a third class of harm called tortious harms (such as negligence).

Depending on the classification adopted, legal systems then utilise various tests to assess the person responsible for causing this harm. Further, an assessment is also required as to the degree of responsibility the person bears with respect to the conduct. Once the person responsible for the harm and the degree of responsibility is established, the appropriate legal consequences are affixed to the extent of the responsibility.

We currently view liability, for the most part, as a binary exercise of determining whether a person is liable or not based on the application of such tests and other principles and mitigating factors. The laws that affix liability ultimately stipulate these circumstances as simply and clearly as possible to ensure that those responsible can take the necessary actions to ensure harm does not arise in the first place.

While this might generally work in the context of actions which are predictable, such as those within a factory or in the course of an ordinary commercial business – it is difficult to apply to the regulation of Gen AI. This is because causation in most instances presupposes that the carrying out of the action that caused the legal wrong is within the control of the entity that carries out that action. In the case of a Gen AI Model, it is difficult to predict the output generated by an LLM as the technology itself is non-deterministic.

Globally, there are several examples of laws which provide for exemption from liability, for instance, Good Samaritan laws which support emergency assistance, carrier liability exemptions, online intermediary safe harbour exemptions, etc. The Indian legal system has already experimented with a limited alternative to the above theory of liability, in its application of the concept of safe harbour for online intermediaries. There is a recognition that while the provision of a platform by online intermediaries may lead to certain legal wrongs being committed (such as the defamation of an individual by content posted by third parties), the intermediary has no control over the specific conduct that led to the illegal action.

As online intermediaries sometimes amplify some of these harms caused by third parties, intermediary regulations have evolved by providing them with a safe harbour, but with the condition that they take down illegal or prohibited content once brought to their notice. In other words, a requirement to prevent further amplification of harm, which in fact is at times the only functional role played by an intermediary.

India has gone a step further by introducing further checks and balances on the conduct of intermediaries for them to avail of safe harbour including some requirements such as (a) accountability (through grievance redressal mechanisms), (b) user awareness through terms and conditions, and policies, and (c) implementing of technological measures, such as filters, etc. These measures have been introduced from the perspective that while the intermediary has no role in the creation of the illegal content itself, it could play a role in amplifying its further transmission.

6.2 AI Liability

Gen AI Systems generate outputs in response to prompts from users. Since the act of generating these responses is largely non-deterministic, applying a degree of responsibility that has been thus far applied in a deterministic liability regime may not always be compatible with the way these Gen AI Models work.

In this regard, it is possible to argue that, at a high level, the developer has control over the universe of outcomes an AI model can produce, e.g., by deciding the datasets on which the AI model learns, controlling the extent to which retrieval-augmented generation is built into the Gen AI Model, or controlling the model's temperature, i.e., the randomness of the outputs. However, by and large, these AI systems are – as a feature, and not a bug – designed to be creative and adaptable to their specific deployment contexts. If it is almost impossible to predict to a degree of legal certainty, *ex-ante*, what these systems will do in each and every instance, a deterministic liability approach may not be an effective means of applying AI liability.

It is very difficult for providers of Gen AI Models to foresee all possible outcomes of the application of Gen AI. This makes it next to impossible to identify all the possible safeguards that a person of ordinary prudence in the AI value chain should take in using this probabilistic technology. One alternative is to heavily constrain the way the model functions so that the outputs generated do not stray across the boundary of what is acceptable. This, however, will stifle innovation and restrict the growth of Gen AI technology and its use cases.

Gen AI Systems are so useful to us precisely because of the unexpected yet highly relevant connections they make. If we force them to operate within rigid constraints to comply with our traditionally deterministic approach to liability, they will no longer be able to make these unexpected connections that we are relying on them to offer. This will be made worse by the fact that AI developers, out of fear of being held liable, will limit how Gen AI Models operate to avoid any risk of being held liable.

Therefore, when applied to Gen AI Models, it may not be appropriate to hold providers of Gen AI Models/Systems directly liable for any and all harm they cause. Doing so would have such a chilling effect on AI innovation that it would force us to forego the many benefits AI systems offer.

That said, simply because Gen AI Models/Systems are probabilistic does not mean that responsible entities should be absolved of all liability to mitigate harm from their downstream uses.

6.3 Safe Harbour from Liability

We propose that, rather than holding Gen AI Models and Gen AI Systems, as well as deployers of such systems (**Responsible Entity**) liable for harm, they should only be found liable if – in a *post facto* proceeding where harm has actually occurred – it can be demonstrated that they acted negligently towards the underlying legal obligation to prevent such harm (i.e., duty of care) with the knowledge that their action or inaction may lead to such harm.

For instance, if a Gen AI Model/System is aimed and marketed as such for the purpose of creating deep fakes violating intellectual property rights, there is both knowledge of the harm and negligence in the duty of care, which could be attributed to the Responsible Entity. Similarly, if a Gen AI Model/System generated illegal content and no measures were taken by them to rectify the same, it may be demonstrated that the Responsible Entities were negligent in taking the necessary measures/due diligence to prevent it.

6.3.1 Due diligence to meet the duty of care

The actual due diligence required to be conducted by the Responsible Entities will differ depending on the unlawful content involved and should be determined based on the context and foreseeability. For instance, the due diligence measures required for content involving nude imagery should be different from those involving a violation of intellectual property rights.

Many Responsible Entities currently implement due diligence measures, such as implementation of filters, setting of temperatures and weights, red-teaming etc., as an example of duty of care. Given that it is a nascent technology, there should not be a clear stipulation of specific measures which apply to all scenarios, but such measures should instead be allowed to develop over time.

The extent and nature of such measures would also depend on the role played by the specific Responsible Entity in the AI value chain (e.g., developers will be expected to take different measures than deployers) as well as the mode of release of the Gen AI Model/System (e.g., providers of closed/API access controlled Gen AI Models may be expected to undertake greater measures than providers of open-source/open-weight Gen AI Models).

6.3.2 Causation

Existing principles of causation would continue to apply, that is to say, the nexus will need to be established between the acts or omissions of the Responsible Entity and the harmful outputs generated by the Gen AI Model/System. However, due regard must also be given to other factors such as the deployer's acts, the end-user's prompts and the end-user's acts in the context of ultimate use by the user. With respect to the end users of the Gen AI Model/System, their liability may continue to be determined under the provisions of existing laws depending on their roles in contributing to the harm (e.g., through the prompts inputted and context of use).

6.3.3 The notice and remediation approach

The safe harbour should also be subject to a notice and remediation approach, similar to the approach which exists in the context of intermediary safe harbour. Where a Responsible Entity is brought to notice regarding an unlawful outcome of a Gen AI Model or System, it should be required to take measures to prevent the Gen AI Model or System from continuing to generate the same or similar harmful outcome. This will help prevent continuing non-compliances of the same type and ensure that remediation of the harm is done immediately.

Responsible Entities are best placed to do what needs to be done in order to set them right and determine what measures are appropriate. As such, by importing such standards, the Responsible Entity is allowed the opportunity to rectify the harms so that they do not recur. Rather than holding them liable for every incident that has been observed, it would be far more useful to give them a chance to set their systems right and only hold them liable if, as mentioned above, they acted negligently towards the underlying legal obligation to prevent such harm (i.e., duty of care) with the knowledge that their action or inaction may lead to such harm.

This graduated approach to liability recognises the probabilistic nature of Gen AI Models/Systems and, in the interests of encouraging innovation, gives Responsible Entities an opportunity to address the harms caused when a model crosses the line. It does not, however, take them completely off the hook, as it is designed to hold them liable where they deliberately disregard safety measures or are reckless.

6.3.4 Exceptions

In certain circumstances, Gen AI Systems and the knowledge they make available could have significant adverse effects. Given the high risk involved with such content, the following harms could be considered as an exception to the general safe harbour approach.

Chemical, Biological, Radioactive and Nuclear Risks: In certain circumstances, Gen AI Systems and the knowledge they make available could have dangerous consequences. This category of risks, commonly referred to as Chemical, Biological, Radioactive and Nuclear (**CBRN**) risks, are where a Gen AI System may generate content knowhow, processes and information that can aid a user to build CBRN weapons or cause CBRN harm (i.e., information that goes beyond a mere chemistry textbook explanation of fission/fusion or a simple web search on IEDs). Due to their hazardous nature, they may have to be placed on a different pedestal from all others. Indian law already recognises alternative forms of liability that apply to dangerous or hazardous activity (strict and absolute liability). It should be possible to apply these standards to the CBRN risks of Gen AI. This would mean that Responsible Entities would be liable for any CBRN risk outputs generated by the Gen AI Models/Systems

Highly sensitive harms: There could be other high-risk harms as well, for instance, child sexual abuse material (**CSAM**) and non-consensual intimate imagery (**NCII**). Due to the seriousness of the harm, by and large, it can be argued that all Responsible Entities ought to be able to foresee and prevent the generation of such outputs. It is due to the heightened sensitivity of the content that there is an overriding public interest in requiring Responsible Entities to be extra careful in ensuring that such harms do not arise.

In the above cases, the Responsible Entities have a duty of care to all downstream AI entities (such as other providers, deployers and end users) and may be required to adopt heightened measures to mitigate the risks of such

foreseeable harms. Where such harm arises, and it is demonstrated in a *post facto* proceeding that the Responsible Entity failed to take such measures, the question of liability may arise.

Given that Gen AI is a fundamentally probabilistic and non-deterministic technology, it should not be subject to our existing binary liability framework. What is required is a graduated approach – one that gives those responsible enough opportunity to amend the way in which their products function and only holds them liable if they wilfully disregard their obligations or are grossly negligent. This may be supplemented with a notice and remediation obligation, where if the Responsible Entity is brought to notice of a harm, they take measures to ensure that it is mitigated. Only for CBRN harms, or harms with a higher sensitivity, such as CSAM and NCII, should a higher standard of care be rigorously enforced in every instance.

At present, no such legal framework exists. Accordingly, we may need to enact suitable legislation to give legal effect to this alternate approach to liability to account for the use of probabilistic systems such as Gen AI.

7 Recommendations for Gen AI Regulation in India

Given the development challenges it faces, Gen AI offers India an unprecedented opportunity to leapfrog its current development trajectory. It can be used to deliver to those currently overlooked and underserved the benefits that are currently not available to them.

However, in order to achieve this outcome, we will need to ensure that our approach to regulating AI is aligned with these objectives. While we should safeguard ourselves against the risk of harm, we need to ensure that in doing so, we do not deny ourselves the benefits that AI has to offer. We need to recognize that Gen AI Models/Systems fundamentally differ from other frameworks currently the subject of regulation.

We can now set out our recommendations on how Gen AI should be regulated in India.

7.1 Focus

India's regulatory efforts should be oriented towards maximising the benefits of Gen AI for the country. To do this, we must make various trade-offs against the risk of harm. In all such instances, we need to maximise the opportunity for Gen AI to enable innovation, taking, if necessary, a more aggressive approach to risk if it will allow us to achieve our developmental goals.

7.2 Approach

India should not enact an omnibus Gen AI regulation. This would hard-code the regulations into law, making it difficult to adapt as Gen AI evolves. Instead, we should provide guidance to sectoral regulators that will educate them on how they should think to address the new challenges that Gen AI brings. It is in India's interests to remain nimble when it comes to the regulation of Gen AI, and where industry self-regulation can help, it should be encouraged.

7.3 Mitigation of Harms

India needs to adopt a systematic approach to assessing whether regulation is required in order to mitigate the harms caused by Gen AI:

- *First, we need to assess whether existing regulations adequately cover the harm caused by Gen AI outputs. If they do, then we need to evaluate the legislative text to ensure that these existing regulations have been framed broadly enough to address the risks that Gen AI poses. If both criteria have been met, there is no need to enact a new regulation for Gen AI.*
- *Where regulations exist but the legislative text does not adequately address the risks that Gen AI poses, we should evaluate whether it would suffice to issue sub-legislative guidance to regulators and law enforcement authorities that can be used to inform their actions in relation to Gen AI-related harms. This will be the nimblest way to approach a fast-evolving technology.*
- *It is only when none of the approaches listed above are adequate to address the harms of Gen AI that we should consider legislative amendments.*

Rather than blindly adopting the regulatory formulations other jurisdictions have put in place, India must develop its own approach from first principles – looking to derive one that best suits its unique circumstances. Since we heavily rely on Gen AI to provide benefits that, but for Gen AI, would have taken decades to permeate into the farthest reaches of the country, we need to have a different risk-reward trade-off compared to other countries.

7.4 Enabling Regulation

India should introduce amendments to existing legislation that will offer some form of protection to those in the Gen AI development ecosystem from the strict implementation of copyright, data protection and other laws that currently constrain the way in which Gen AI development can proceed. This could take the form of an express exemption set out in the text of the law that allows them to proceed subject to the fulfilment of certain clearly described conditions.

In all such instances, it is important to ensure that all such exemptions are framed so that the exemptions granted do not come at the cost of the rights these laws were supposed to grant and the protections they were supposed to provide.

7.5 Liability

Since the very nature of Gen AI makes it impossible to ensure that it performs in a predictable way 100% of the time, the liability frameworks we use in relation to these systems must appropriately account for that uncertainty. Therefore, we recommend a safe harbour for Responsible Entities provided that they are not negligent and implement a notice and remediation measure. The exemption to this will be for high-risk harms that arise from content related to CBRN risks, CSAM and NCII.

The information contained in this may address a specific requirement for the recipient and must not be construed as advertisement or solicitation in any form or manner.

For more information,

please visit our website
www.trilegal.com