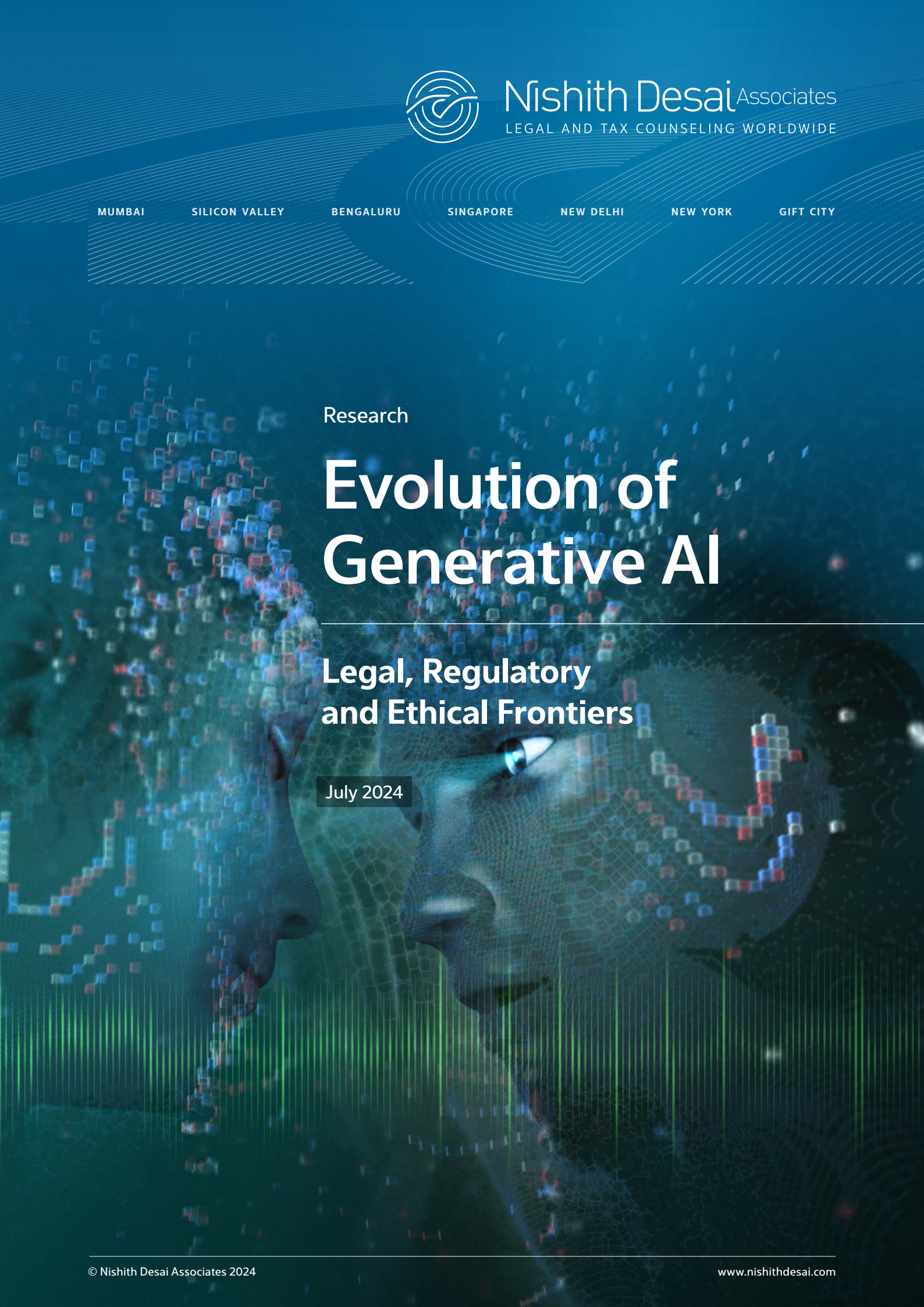
Nishith Desai Associates
LEGAL AND TAX COUNSELING WORLDWIDE

MUMBAI SILICON VALLEY BENGALURU SINGAPORE NEW DELHI NEW YORK GIFT CITY



Research

Evolution of Generative AI

Legal, Regulatory and Ethical Frontiers

July 2024

Research

Evolution of Generative AI

Legal, Regulatory and Ethical Frontiers

July 2024

DMS Code: 30219.2

FT INNOVATIVE LAWYERS 2019
FINANCIAL TIMES
ASIA-PACIFIC WINNER

Ranked as the 'Most Innovative Indian Law Firm' in the prestigious FT Innovative Lawyers Asia Pacific Awards for multiple years. Also ranked amongst the 'Most Innovative Asia Pacific Law Firm' in these elite Financial Times Innovation rankings.

Chambers AND PARTNERS

IFLR1000

The LEGAL 500

BENCHMARK LITIGATION

asialaw

WWL

ASIAN LEGAL BUSINESS

INDIA BUSINESS LAW JOURNAL

INDIA BUSINESS LAW JOURNAL THE A LIST 2023 2024 LAW ASIA

INDIA BUSINESS LAW JOURNAL DEALMAKER 2023 LAW ASIA

Mergermarket An Acuris company

ITR WORLD TAX

iel ELITE

BUSINESS TODAY

THE NATIONAL LAW REVIEW

equaliFI

IDEX LEGAL AWARDS

ILA

LEGAL ERA BY THE PEOPLE. FOR THE PEOPLE. OF THE PEOPLE

LawWiser .com

Disclaimer

Contact

For any help or assistance please email us on **concierge@nishithdesai.com**
or visit us at **www.nishithdesai.com**.

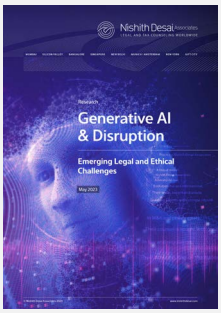
Acknowledgements

Huzefa Tavawalla
huzefa.tavawalla@nishithdesai.com

Tanisha Khanna
tanisha.khanna@nishithdesai.com

Karishma Karthik
karishma.karthik@nishithdesai.com

Rhythm Vijayvargiya
rhythm.vijayvargiya@nishithdesai.com



May 2023

Generative AI & Disruption

Emerging Legal and Ethical Challenges

Please also see our previously published paper on this topic, in which we examined some of the key use cases of GAI from a legal and regulatory perspective.

Contents

Introduction

As a result of the fast-paced advancement in technology, Artificial Intelligence ("**AI**") has emerged as a powerful tool for content creation in the last few decades. AI is essentially the method of making machines mimic human intelligence, and Generative AI ("**GAI**") is its branch that uses machine learning technology for the generation of new content.[1]

GAI uses unsupervised learning algorithms to create new audio, visual, and textual media based on user prompts.[2] There are supervised learning algorithms as well that are used to train models on labeled data. Such machine-learning techniques can be used to pre-train generative models to grasp underlying structures and relationships within the data, for fine-tuning models, and conditional generation.[3] One of the key advantages of GAI is its autonomous functioning, which allows it to process large quantities of data quickly and efficiently, enabling faster content creation.[4] Several major industries like healthcare, finance, media and entertainment, automotive, and education have adopted GAI, leveraging its capabilities for various applications. These include marketing and sales, manufacturing and industrial operations, software engineering, healthcare and life sciences, and education.[5] With its ability to generate novel content from user prompts, GAI has become an attractive option for businesses seeking to produce unique and engaging content.

Unlike other AI models that primarily classify, analyze, or act on existing data, GAI goes beyond and creates new content every time it is prompted.[6] This makes GAI a valuable tool for content creators, as it can provide them with an endless stream of fresh ideas and material. Additionally, with the increase in computing power and the development of more efficient deep learning networks, GAI models can now generate novel content in every new trial even from the same prompts.[7]

There is a myriad of legal issues to consider before businesses adopt GAI. While India has emerged as one of the nations with the highest adoption rate of GAI tools,[8] a spectrum of legal implications on the usage of GAI across sectors have also arisen, requiring immediate consideration. Given the booming start-up sector in India and the increased enthusiasm of businesses to develop in-house GAI tools, the issue of ownership of intellectual property in content generated by such tools, allocation of authorship, and potential of infringement and liability for the same also need to be studied. There are also concerns about decision-making, bias, prejudice, and stereotyping arising out of the usage of such tools. Further, the liability for unlawful content generated by AI tools is an imperative issue in India, given its diversity and representation of various cultural groups. There have been many instances of content generated by AI tools in the absence of sufficient guard rails to moderate against offensive, discriminatory, or politically charged content resulting in public outrage.[9]

1 See: https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai, last accessed June 24, 2024.

2 See: https://www.leewayhertz.com/generative-ai-use-cases-and-applications, last accessed June 24, 2024.

3 See: https://www.geeksforgeeks.org/can-i-use-unsupervised-learning-followed-by-supervised-learning/?ref=ml_lbp, last accessed June 24, 2024.

4 See: https://www.thedigitalspeaker.com/what-is-generative-ai-how-disrupt-society, last accessed June 24, 2024.

5 See: https://www.suntecindia.com/blog/generative-ai-adoption-by-industries-trends-and-statistics, last accessed June 12, 2024); https://www.snaplogic.com/blog/adopting-generative-ai-industry-function-landscape, last accessed June 24, 2024.

6 See: https://arxiv.org/pdf/2301.04655.pdf, last accessed June 24, 2024.

7 See: https://arxiv.org/pdf/2301.04655.pdf, last accessed June 24, 2024.

8 See: https://www.businesstoday.in/technology/news/story/india-leads-asia-pacific-region-in-generative-ai-adoption-report-429976-2024-05-17, last accessed June 24, 2024.

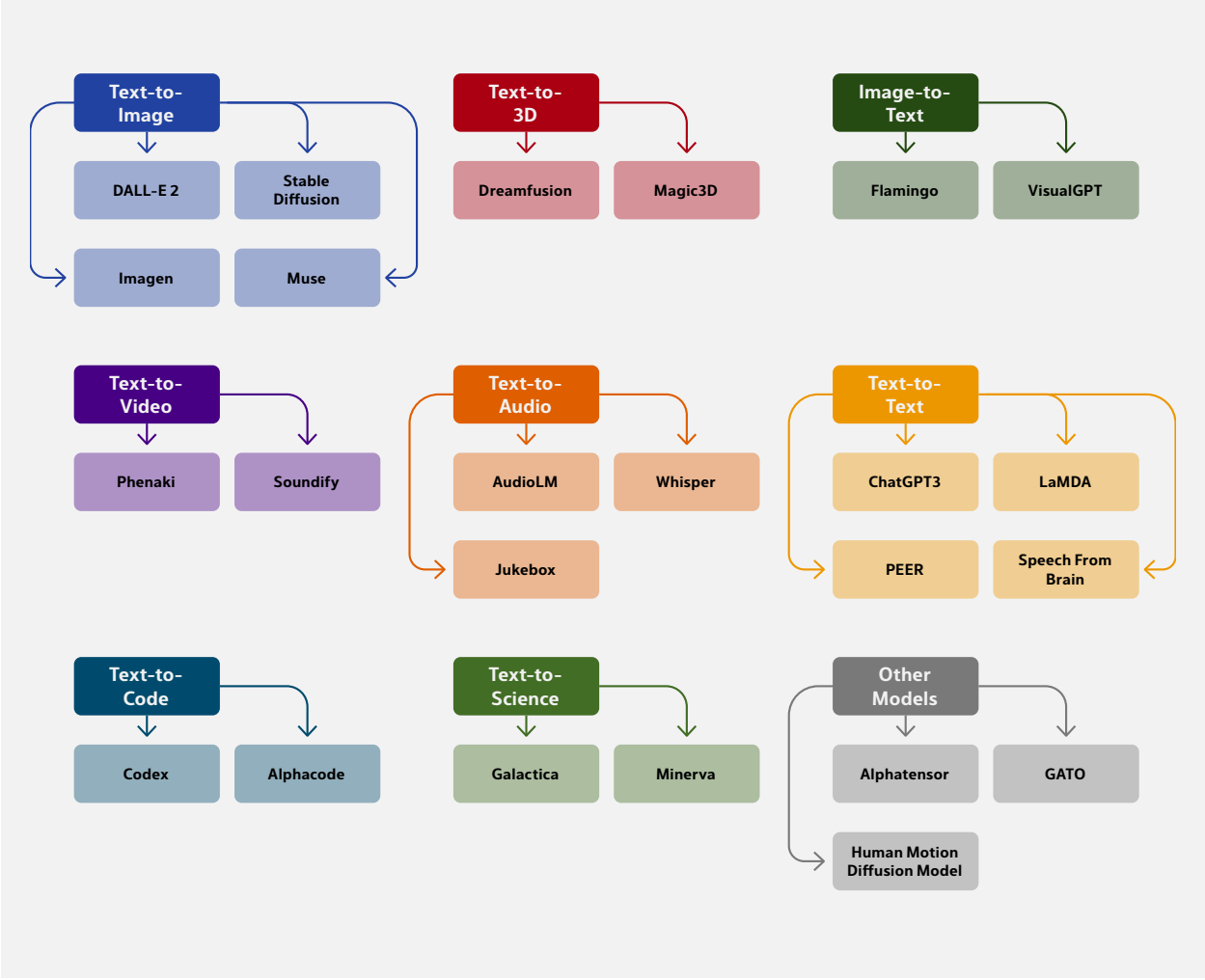
9 See: https://www.dnaindia.com/analysis/report-dna-special-why-artificial-intelligence-tech-chat-gpt-has-been-labeled-as-anti-hindu-3017614, last accessed June 24, 2024.

Further, as India moves towards a culture of data protection, with the milestone enactment of the DPDPA (discussed below), the collection and processing of archives of personal data which form part of the training data of AI models also raises many legal and operational issues. In this paper, we discuss such issues in detail and bring out the key concerns from a legal perspective. We have s systematically analyzed the types of content generated through GAI, its use cases and benefits along with the key legal concerns this technology brings with its advancement. We have also discussed the way forward to accommodate GAI keeping in mind the legal and ethical considerations attached to it.

Content Generation through Generative Artificial Intelligence

GAI models can be trained to generate novel content in an array of forms, like text, images, videos, audio, codes, and 3D models. The GAI platforms used to generate such content learn from a large amount of input data, allowing them to produce results that are often indistinguishable from human-created content.

Below is a flowchart that depicts how different inputs generate different kinds of content on GAI-based platforms:



Source: Brizuela & Garrido, ChatGPT is not all you need. A State of the Art Review of large GAI models [1]

1 See: https://arxiv.org/pdf/2301.04655.pdf, last accessed June 24, 2024.

We have also explained various forms of the content generated through GAI below:

A. Text

There exist several text-to-text GAI models like ChatGPT,[2] T5,[3] and BART,[4] which can generate a text output to a text-based input. GAI models can generate coherent and human-like text with unprecedented fluency and accuracy. It has a wide range of capabilities ranging from prose and poetry writing, search, text summarization, and image and code generation.

Large Language Models ("**LLM**")are mathematical models of the statistical distributions of phrases, words, parts of words, and even individual characters, over a corpus of human-generated text. Thus, for example, when a user prompts "Knock Knock", the LLM deciphers that there is a very high probability of "Who's there?" following the initial prompt LLM do not understand what is being asked from them and what result they produce; they merely generate the most statistically likely sequence of terms.

Although LLM may not have the originality of thought or creativity, the results are compelling enough to be that of a human author. They have a solid potential to be used as a collaborative tool in many sectors, such as research and education, programming and design, customer service, therapy, clinical documentation, spam detection, HR processes, coding, legal sector, sentiment analysis, etc.

B. Images

Various platforms like Dall-E,[5] Midjourney,[6] and Artbreeder,[7] use GAI to generate images based on textual descriptions entered by the users. Dall-E is one of the platforms that combines elements of natural language and image processing, to generate novel images based on textual prompts. The 'text-to-image' generator's brilliance lies in its capability to combine distinct and unrelated concepts in semantically plausible ways. These models have shown a remarkable 'understanding' of visual concepts (e.g., macroscopic, cross-sectional, or overexposed pictures), places (photo of a teddy bear in Times Square), and time (photo of computers in the 1950s). It can also reproduce images in a particular style (sketch or cartoon) or emulate the style of a particular artist.[8]

The GAI models majorly use transformer architecture and are trained to first recognize the relationship between natural language and image concepts. These relationships are 'encoded' in a mathematical language understandable by machines. The encodings represent the similarity/dissimilarity between image/caption pairs.[9]

2 See: https://openai.com/index/chatgpt, last accessed June 24, 2024.

3 See: https://research.google/blog/exploring-transfer-learning-with-t5-the-text-to-text-transfer-transformer, last accessed June 24, 2024.

4 See: https://scontent-bom2-1.xx.fbcdn.net/v/t39.8562-6/240738846_578931553279271_1785451284662340239_n.pdf?_nc_cat=101&ccb=1-7&_nc_sid=e280be&_nc_ohc=Th2kRrsQiK0Q7kNvgG6ifWM&_nc_ht=scontent-bom2-1.xx&oh=00_AYB6S99_V7FBu3x7QFpWwctyJqsyQ2U_Pyyq_WzW_HrACw&oe=66595D4A, last accessed June 24, 2024.

5 See: https://openai.com/index/dall-e, last accessed June 24, 2024.

6 See: https://www.midjourney.com/home, last accessed June 24, 2024.

7 See: https://www.artbreeder.com/create/composer, last accessed June 24, 2024.

8 See: https://daleonai.com/dalle-5-mins, last accessed June 24, 2024.

9 See: https://www.assemblyai.com/blog/how-dall-e-2-actually-works, last accessed June 24, 2024.

The next step involves reversing the encodings to produce images on text prompts. It is important to mention that these models do not simply reproduce the original encoded image but reproduce the *salient features* of that image. Thus, it can produce novel images from the same text prompt.[10]

C. Sounds and Audios

Various GAI tools and platforms like Jukebox,[11] Replica Studios,[12] and Soundraw[13] generate sound/music compositions through textual prompts that define genres, styles, instruments, structures, etc. with an option to customize such compositions for the required purposes. Lately, a high number of users have commenced using these platforms to generate compositions[14] with an aim to commercialise the same. Users also utilise these GAI platforms to generate sounds/audios/music for audiobooks, podcasts, commercials, games, etc. as these platforms allow users to contextualise the purpose of the compositions in the text-based prompts.

In November 2022, a human and an AI platform were seen performing a musical piece in collaboration, and GAI made it possible for the AI platform to improvise on the pitch, rhythm, and timbre like a human being would.[15]

GAI can be used to produce sounds and audio through the process of audio synthesis. Audio synthesis involves training a neural network to generate audio data based on patterns in existing audio samples; the network can then be used to create new audio that is similar in style and content to the original data.[16] Another way of generating sound through GAI is producing audio directly from mathematical equations that model sound waves.[17] These processes have the potential to produce highly realistic audio that sounds like it was created by a human.

There have been models that were designed to generate realistic speech and music by processing audio inputs. The aim behind such models was to move a step ahead from the text to audio models. These models relied on semantic and acoustic tokens accordingly to capture the details of the audio inputs which could result in high-quality synthesis and output.[18]

D. Videos

GAI is also very widely used for the generation of videos through a process called video synthesis. Similar to audio synthesis, this process relies upon the training of a neural network to generate results based on the patterns in video samples.

10 See: https://www.assemblyai.com/blog/how-dall-e-2-actually-works, last accessed June 24, 2024.

11 See: https://openai.com/index/jukebox, last accessed June 24, 2024.

12 See: https://www.replicastudios.com, last accessed June 24, 2024.

13 See: https://soundraw.io, last accessed June 24, 2024.

14 See: https://time.com/6340294/ai-transform-music-2023, last accessed June 24, 2024.

15 See: https://www.wired.co.uk/article/generative-ai-music, last accessed June 24, 2024.

16 See: https://arxiv.org/pdf/2006.06426.pdf, last accessed June 24, 2024.

17 See: https://arxiv.org/pdf/2006.06426.pdf, last accessed June 24, 2024.

18 See: https://ai.googleblog.com/2022/10/audiolm-language-modeling-approach-to.html, last accessed June 24, 2024.

Platforms like Synthesia,[19] Deepbrain AI,[20] and Pictory[21] use GAI technology in order to allow the users to create videos through text-based prompts, which may include blogs, articles or a brief description of the contents that a user seeks in a video.

Generative Adversarial Network ("**GAN**") is the most popular approach to audio and video synthesis. GANs have made it possible to create highly realistic videos as they operate on two neural networks, the first network creates new video samples based on patterns in the original data, and the other attempts to distinguish between the original video samples and those created by the software. Such models train by themselves in order to ensure that the results produced are as close to reality as possible.

E. 3D Models

Nvidia's GET3D[22] and Magic3D[23] are examples of platforms that offer tools to users which can be used to generate 3D models through text-based descriptions. Similarly, in addition to Dall-E and ChatGPT, OpenAI has also introduced a platform Point-E[24] that can be used to create 3D models based on text prompts. Such platforms work in two steps; they produce a text-to-image result which is then used to produce an image-to-3D model. Essentially, these platforms sample a picture using the text-to-image model and then sample a 3D object conditioned on the sampled image to create a 3D object from a text prompt.[25]

19 See: https://www.synthesia.io, last accessed June 24, 2024.

20 See: https://www.deepbrain.io, last accessed June 24, 2024.

21 See: https://pictory.ai, last accessed June 24, 2024.

22 See: https://research.nvidia.com/labs/toronto-ai/GET3D, last accessed June 24, 2024.

23 See: https://research.nvidia.com/labs/dir/magic3d, last accessed June 24, 2024.

24 See: https://openai.com/index/point-e, last accessed June 24, 2024.

25 See: https://arxiv.org/pdf/2212.08751.pdf, last accessed June 24, 2024.

Training Generative AI Models

GAI models are created using a thorough and multifaceted training procedure that includes supervised and unsupervised learning approaches applied to large datasets. The first phase entails data collecting, which requires gathering large amounts of text from a variety of public sources, such as books, papers, websites, and other text-rich resources, in order to give a complete foundation for the model to learn from. This raw data is subsequently preprocessed to make it suitable for training. The advanced neural network architecture, specifically the transformers uses attention methods to allow the model to focus on different areas of the input text, capturing long-term dependencies more effectively. These models are made up of numerous layers of neurons that process input data via various transformations, with each layer refining the data representation to improve the model's understanding and predictive skills.

Training the model entails altering its parameters, or weights, to reduce the gap between its predictions and the actual data. This is an iterative procedure that begins with initializing the model's parameters, usually with small random values. During the forward pass, input data is fed into the model, which generates predictions. The difference between these predictions and the actual target values is measured using a loss function, such as cross-entropy loss in language models. After initial pre-training on a broad corpus, the model can be fine-tuned on a smaller, domain-specific dataset to improve performance on specialized tasks, such as question answering or code generation. Fine-tuning entails changing the pre-trained model's parameters to better match the specific needs of the target application. The model's performance is then extensively evaluated using several measures. The development process frequently necessitates numerous iterations, with each cycle comprising greater datasets and more sophisticated models to improve performance. To deal with the high computational demands, techniques such as distributed training and the utilization of powerful hardware, such as GPUs and TPUs, are used.

Natural Language Processing

GAI in text/speech form largely belongs to the realm of Natural Language Processing ("**NLP**") that brings together computer science, AI, information engineering, and linguistics. NLP refers to the deployment of computational techniques to understand and manipulate text or speech to serve useful purposes such as translation, answering questions, and text summarisation.[1]

In the NLP domain, these neural networks work sequentially by predicting the next word given the previous words. However, because of their sequential operation, they may struggle with large sentences and paragraphs.

Transformer Architecture

In 2017, a team of Google researchers published a seminal paper titled "Attention is all you need" that laid the basis for the neural network that GAI models use - called the 'Transformer' (the T in GPT). Two key innovations of transformer architecture are the 'attention' and 'self-attention' mechanism.

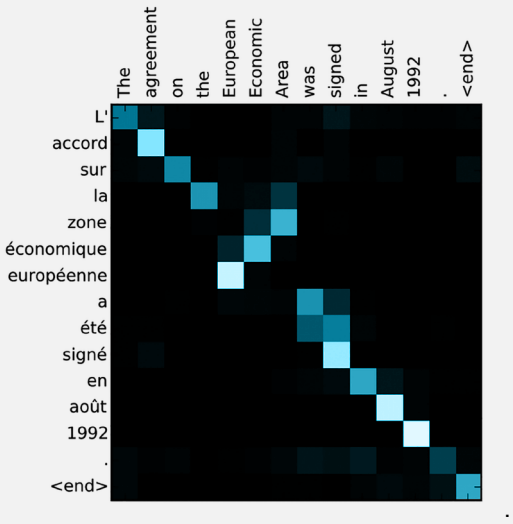
1 See: https://link.springer.com/chapter/10.1007/978-981-15-9712-1_31, last accessed June 24, 2024.

I. Attention

The attention mechanism allows a model to look at every word in the 'input' sequence simultaneously while making predictions for the 'output' sequence. This is especially important in machine translation, for example, in languages with gendered words or where the order of the words is flipped. Let's take the following example sentence, as used in the 2014 paper[2] which first demonstrated the concept:

When the sentence "The agreement on the European Economic Area was signed in August 1992" is translated into French, it would read as: *"L'accord sur la zone économique européenne a été signé en août 1992."*

Firstly, the words in the phrase "European Economic Area" are flipped to read as "la zone économique européenne". Secondly, the adjectives "économique" and "européenne" are feminized to match the feminine object "la zone".



Source: Neural Machine Translation by Jointly Learning to Align and Translate (2014)

The above heat map demonstrates what words the model is "attending" to in the input sequence when it outputs each word. As you might expect, when the model outputs the word "européenne," it attends considerably to both the input words "European" and "Economic."

And how does the model know which words it should be "attending" to at each time step? It's something that's learned from training data. By seeing thousands of examples of French and English sentences, the model learns what types of words are interdependent. It learns how to respect gender, plurality, and other rules of grammar.[3]

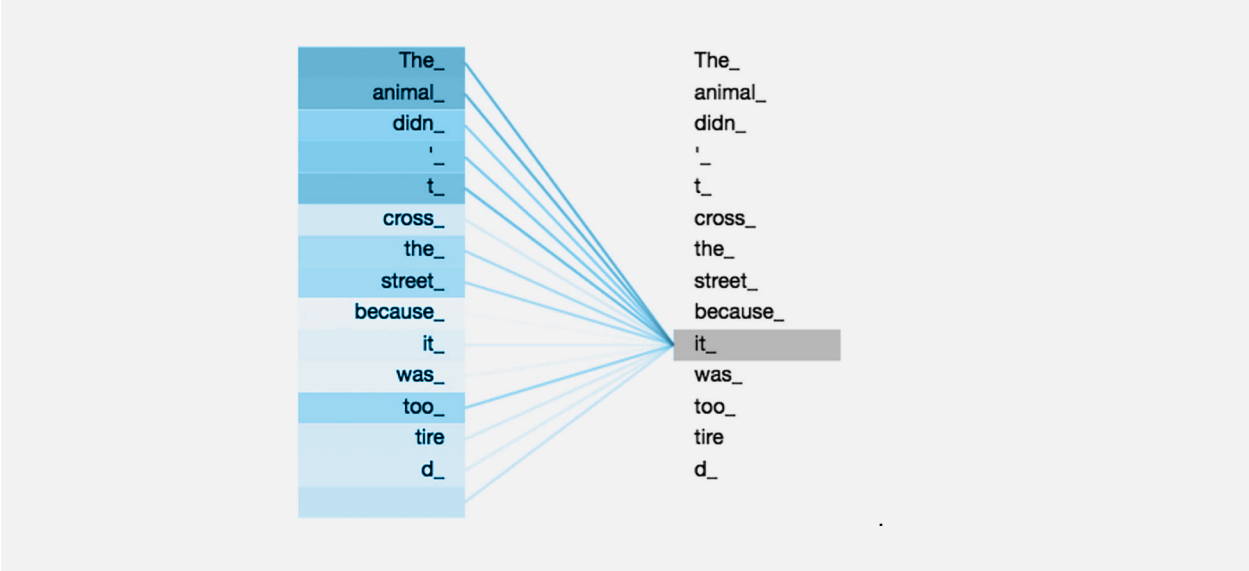
2 See: https://arxiv.org/abs/1409.0473, last accessed June 24, 2024.

3 See: https://daleonai.com/transformers-explained, last accessed June 24, 2024.

II. Self-attention

While attention models are important for tasks such as machine translation, 'self-attention' is useful for disambiguation, that is, discerning the context and meaning of words and sentences. Unlike the sequential models, where the meaning of a word is discerned by separately going back to each following word in the sentence, self-attention models 'bake' the understanding of other words in a sentence into the particular word we are processing. That is, every word has its relationship with other words in the input encoded *at the time* of the processing itself. This helps in parallel optimisation and vastly improves the performance of the model.[4]

Say, we were to translate the following sentence[5]: "The animal didn't cross the street because it was too tired." Who/what is the word 'it' referring to: the animal or the street? At the time of processing, self-attention enables the model to associate "it" with "animal". The following is demonstrated in the diagram below:



Source: Jay Alammar, The Illustrated Transformer[6]

Just like for the attention mechanism, a meaningful internal representation of data is learned by the model using a vast trove of unlabelled data that it is trained on without any supervision. Having trained the transformer model, it can then be fine-tuned through supervised learning and be 'transferred' to a wide area of tasks. This is why it is called a pre-trained transformer. And since we are talking about a transformer model used to generate content, it is called a 'Generative Pre-Trained Transformer' (GPT).

This technique of training is called semi-supervised learning. Unlabelled data, when used in conjunction with a small amount of labelled data, can produce considerable improvement in learning accuracy. Besides, the costs associated with labelling an enormous corpus of text would render the entire experiment infeasible.

The impressive feature of neural networks is that, even without any supervision, they can learn rules of grammar, parts of speech, and whether words are synonymous.[7]

4 See: http://jalammar.github.io/illustrated-transformer, last accessed June 24, 2024.

5 See: http://jalammar.github.io/illustrated-transformer, last accessed June 24, 2024.

6 See: http://jalammar.github.io/illustrated-transformer, last accessed June 24, 2024.

7 See: https://daleonai.com/transformers-explained, last accessed June 24, 2024.

Legal and Ethical Issues

A. The Veracity of Information and Fake News

GAI is a rapidly advancing field of AI that involves training machines to generate novel content such as text, images, and videos. While GAI has many useful applications, it also raises concerns about its potential to be used maliciously.

Given that AI language models can generate human-like text and can be trained to mimic the writing style of individuals, there are serious concerns about its potential misuse for spreading fake news on a large scale. Such misinformation can have serious ramifications, such as distorting public perceptions and eroding confidence in societal structures and systems.[1]

Moreover, GAI models can also be used to create fake social media accounts or bots to spread false information.[2] These accounts could flood social media platforms with fabricated news stories, making it challenging for individuals to find credible sources of information, and consequently making it harder for people to engage in informed and constructive discussions.[3] Moreover, even in the absence of user intent, GAI may provide inaccurate outputs (or hallucinations) while sounding convincing.

The World Economic Forum's 2024 Global Risks Report highlighted that misinformation and disinformation have surged to become a major threat, particularly as elections are held in various economies from the West to Asia. Many Indian performers and celebrities have been the victims of deepfakes videos in the past few years. In the months leading up to the 2024 Indian general elections, realistic deepfake videos went viral online, with celebrities purportedly criticizing/supporting political candidates. On March 15, 2024, to curb the spread of misinformation, unlawful content, and harmful AI-generated content the Indian Ministry of Electronics and Information Technology ("**MeitY**") issued advisories[4] for intermediaries and platforms directing them to:

i) ensure that the use of GAI models on or through their computer resource does not permit its users to host, display, upload, modify, publish, transmit, store, update, or share any unlawful content.

ii) ensure that their computer resource in itself or through the use of GAI models does not permit any bias or discrimination or threaten the integrity of the electoral process.

iii) ensure that under-tested or unreliable GAI models should be made available to the users in India only after appropriately labelling the possible and inherent fallibility or unreliability of the output generated. Further, 'consent popup' or equivalent mechanisms may be used to explicitly inform the users about the possible and inherent fallibility or unreliability of the output generated.

iv) inform its users through the terms of services and user agreements about the consequence of dealing with the unlawful information, including disabling of access to or removal of such information, suspension or termination of access or usage rights of the user to their user account, as the case may be, and punishment under applicable law.

1 See: https://www.analyticsinsight.net/chatgpt-can-be-used-to-disseminate-fake-news-on-a-large-scale, last accessed June 24, 2024.

2 See: https://economictimes.indiatimes.com/opinion/et-commentary/chatgpt-faking-it-a-genuine-artificial-concern/articleshow/97233545.cms?from=mdr, last accessed June 24, 2024.

3 See: https://www.analyticsinsight.net/chatgpt-can-be-used-to-disseminate-fake-news-on-a-large-scale, last accessed June 24, 2024.

4 See: https://www.meity.gov.in/writereaddata/files/Advisory%2015March%202024.pdf, last accessed June 24, 2024.

v) ensure that any information created, generated, or modified through its software or any other computer resource is labelled or embedded with a permanent unique metadata or identifier, in a manner that such information/content is identifiable through such metadata.

vi) ensure compliance with the IT Act and Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 ("**IT Rules**").

While the above advisory was issued for compliance by intermediaries, the references to GAI platforms seemingly infer the intent of the government to treat all such platforms as intermediaries. MeitY also announced that work on deepfake regulations would start either through changes to existing laws or the introduction of a new law[5] that would place accountability on the user and the host platform.[6]

On March 20, 2024, MeitY notified the Fact Check Unit under the Press Information Bureau as the Fact Check Unit under the IT Rules to identify information in respect of any business of the Central Government as "fake or false or misleading". Intermediaries would be required to take reasonable efforts to not host such information identified by the FCU. The Supreme Court has stayed the notification of the FCU pending the proceedings challenging the constitutionality of this provision.

B. Generative Artificial Intelligence and Intellectual Property Rights

While the question regarding ownership and other IP rights in content created by AI has been raised for decades, consumer-facing GAI has injected this debate with new vigor. There are several questions of copyrightability, and copyright ownership on works generated through GAI which deserve consideration:

- Whether there is copyright infringement by virtue of training of GAI models?
- Are there any exceptions to copyright infringement, such as fair dealing exceptions, that may apply to the usage of copyrighted works in the training of GAI tools?
- What is the allocation of liability for infringement between the developer of the AI tool, the company owning such tool, and the user of the tool?
- Whether output generated by GAI tools infringe the copyright in the works forming part of the training data? Are there any exceptions to copyright infringement, such as fair dealing exceptions, that may apply to such output?
- Whether works wholly generated or assisted by GAI tools are copyrightable works?
- Whether the scope of copyright protection for such works differs based on the nature of the work? Does the protection differ between literary, dramatic, musical, and artistic works, and cinematograph films and sound recordings?
- Whether authorship can be attributed to a human author for such works?
- Whether such works could be considered works of joint authorship between the various stakeholders involved?
- Whether works generated by AI tools qualify as "computer generated" works as contemplated under Indian copyright law?

5 See: https://www.businesstoday.in/technology/news/story/ashwini-vaishnaw-on-deepfake-menace-govt-considering-penalties-on-both-creator-and-platform-406830-2023-11-23, last accessed June 24, 2024.

6 Ibid.

Nevertheless, in this section, we will focus on the legal aspects of IP, while also touching upon certain theories underlying the law of IP.

GAI's Training Process, Use of Training Data, and IP Infringement

GAI is trained on datasets that may be in the form of text, images, videos or sound recordings (including data scraped from the internet), all of which could potentially be copyright-protected works. The key question is whether ingestion of such data (including storage) for training purposes amounts to infringement of copyright of the existing works.

Numerous lawsuits have already been filed claiming copyright infringement.[7] In the case of Sarah Andersen v. Stability AI Ltd,[8] the plaintiffs, a collective of artists, argued that Stable Diffusion, trained their AI model on the artists' images to be able to generate content in the style of the artists. The plaintiffs claimed that this was undertaken through embedded and stored compressed copies of their art works in the training data of the defendants. The defendants, however, described the training process as not copying of images, but instead application of mathematical equations and algorithms to capture concepts from the training data. Further, they argued that given the training data is an active program, it could not possibly contain compressed copies of over five billion images. Proceedings are still ongoing.

The New York Times has sued OpenAI Inc. for the unauthorized use of its articles to train the GPT large language models.[9] The proceedings are still ongoing and will have a significant impact on the applicability of "fair use" doctrines and the operation of GAI models. Among the New York Times' various arguments, they claim that the GPT models can generate near-verbatim reproductions of the works, which may obviate the need to purchase access through the Times itself. New York Times claims that ChatGPT is free-riding on the massive amounts of investment and research involved in publishing articles by creating substitutive products without permission or payment.

Another prominent case dealing with the scraping of the content of creators to train and develop AI models resulting in infringement of their copyright is the UK case of Getty Images (US) Inc. v. Stability AI Ltd. Getty Images claimed that the download and storing of Getty Image's copyrighted works for training Stability's GAI model infringing Getty Images' exclusive right to communicate such works to the public. Stability AI had applied for striking down certain aspects of the claim for lack of territorial jurisdiction. Stability AI argued that Stable Diffusion's development and training happened entirely outside the UK. They claimed that all the computing power used for training was always located in data centers operated by Amazon Web Services Inc. in the United States, not the UK. The proceedings are still underway and (as of the date of writing) the UK court has ruled that the case can move on to trial.[10] In addition to its claim against Stability AI in the UK, Getty Images has brought proceedings in the US District Court of Delaware.[11] Getty Images' complaint is for copyright infringement, and removal or alteration of copyright management information.[12] Proceedings are still ongoing.

7 See https://techcrunch.com/2023/01/27/the-current-legal-cases-against-generative-ai-are-just-the-beginning, last accessed June 24, 2024.

8 Case No. 23-cv-00201-WHO (United States District Court Northern District Of California).

9 See: https://fingfx.thomsonreuters.com/gfx/legaldocs/byvrkxbmgpe/OPENAI%20MICROSOFT%20NEW%20YORK%20TIMES%20mtd.pdf, last accessed June 24, 2024.

10 See: https://www.theverge.com/2023/12/4/23988403/getty-lawsuit-stability-ai-copyright-infringement, last accessed June 24, 2024.

11 See: https://www.globallegalpost.com/news/high-court-allows-getty-images-ip-dispute-with-stability-ai-to-go-to-trial-1715547241, last accessed June 24, 2024.

12 See: https://storage.courtlistener.com/recap/gov.uscourts.ded.81407/gov.uscourts.ded.81407.1.0.pdf, last accessed June 24, 2024.

Given that there are yet no decisive pronouncements on whether works generated by AI tools can qualify as copyrightable works, it remains to be seen how different courts decide on this issue based on the nuances of the copyright system and policy considerations of each state.

Fair Use Exceptions for Ingesting and Training GAI Platforms

One of the key defenses used by GAI platforms for ingesting data for training purposes is the doctrine of "fair use". The test of fair use would vary from country to country and would need to be evaluated on a case-to-case basis.

Section 52(1)(a) of the Copyright Act provides that fair dealing with any work for the purpose of private or personal use, including research, does not constitute an infringement of copyright. Indian courts have also used the following four-factor test for determining whether or not a particular use of a work is protected under the fair use exception:

- purpose and character of the use, including whether such use is of a commercial nature or is for non-profit purposes;
- the nature of the copyrighted work;
- the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
- the effect of the use upon (i) the potential market for, or (ii) the value of, the copyrighted work. The subsequent work should not act as a market substitute or affect the market share of the prior work.

Hence, whether the use by GAI of existing copyrighted work amounts to infringement would be a case-wise analysis.

Different jurisdictions have taken different approaches to the extension of exceptions to copyright infringement for AI tools. Japan's copyright legislation contains a broad exception permitting the ingestion and use of copyrighted works for any type of information analysis, including for the purpose of training AI models. The law does limit the exception to uses that amount to enjoying the thoughts or sentiments expressed in that work, or if the action would unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the work or the circumstances of its exploitation. However, there is limited guidance on the scope of the limitation.

There are text and data mining exceptions to copyright infringement in various jurisdictions. The EU Directive on Copyright in the Digital Single Market ("**DSM Directive**")[13] defines 'text and data mining' as "any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations". There are exceptions under the DSM Directive permitting the reproductions and extractions of lawfully accessible copyright works and database rights:

1) for the purpose of text and data mining made by research organisations and cultural heritage institutions in order to carry out text and data mining for the purpose of scientific research;[14] and
2) for the purpose of text and data mining generally. This includes text and data mining for a purpose other than scientific research, including for commercial purposes.

13 See: https://eur-lex.europa.eu/eli/dir/2019/790/oj, last accessed June 24, 2024.

14 Article 3, DSM Directive.

Rights holders may choose to reserve their rights over their copyright works or other subject matter to prevent text and data mining except for the purposes of scientific research. In the event the owner of the copyright expressly reserves their right to text and data mining being performed on their copyrighted works, express permission through licensing arrangements will need to be sought.

In the UK, the text and data mining copyright exception only protects those carrying out "computational analysis" on lawfully accessed works for the "sole purpose of research for a non-commercial purpose".[15] While the UK was considering the introduction of a broader text and data mining exception for commercial purposes, the proposal was later dropped, and the relevant department will now start the consultation process with AI companies and rights holders to secure a workable approach.[16]

Hence, training AI on existing works and whether the same leads to infringement continues to be a contentious issue across jurisdictions. The tug of war between right holders and owners of GAI tools could lead to fine-tuning of fair use principles along with jurisprudence on the ownership of machine-derived works. The factual analysis of each case would play a crucial role, and it is currently difficult to establish a singular principle that will provide a solution to this nuanced area of law.

Copyrightability of Works Generated through GAI Tools

Under Indian copyright law, copyright only subsists in certain works, namely, (i) original literary,[17] dramatic,[18] musical,[19] and artistic works,[20] (ii) cinematograph films;[21] and (iii) sound recordings.[22] The output of the various GAI described above may qualify as a "work" in which copyright can subsist.

However, just by virtue of being such a work, copyright does not subsist in such work automatically. For a 'work' to qualify for copyright protection, it would have to meet the 'modicum of creativity' standard. This was enumerated in the *Eastern Book Company v. D.B. Modak*[23] case where the Court held that in relation to the existence of copyright in the compilation of judgements, a 'minimal degree of creativity' was required, and that 'there must be some substantive variation and not merely a trivial variation'. However, a plain reading of this test does not provide a definitive conclusion on whether an AI can meet the 'modicum of creativity' as required.

15 Section 29A, Copyright, Designs and Patents Act 1988.

16 See: https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response, last accessed June 24, 2024.

17 Under Section 2(o) of the Copyright Act, 1957 ("**Copyright Act**"), "Literary work" includes computer programmes, tables and compilations including computer databases.

18 Under Section 2(h) of the Copyright Act, "dramatic work" includes any piece for recitation, choreographic work or entertainment in dumb show, the scenic arrangement or acting, form of which is fixed in writing or otherwise but does not include a cinematograph film.

19 Under Section 2(p) of the Copyright Act, "musical work" means a work consisting of music and includes any graphical notation of such work but does not include any words or any action intended to be sung, spoken or performed with the music.

20 Under Section 2(c) of the Copyright Act, "artistic work" means (i) a painting, a sculpture, a drawing (including a diagram, map, chart or plan), an engraving or a photograph, whether or not any such work possesses artistic quality; (ii) a work of architecture; and (iii) any other work of artistic craftsmanship.

21 Under Section 2(f) of the Copyright Act, "cinematograph film" means any work of visual recording and includes a sound recording accompanying such visual recording and "cinematograph" is construed as including any work produced by any process analogous to cinematography including video films.

22 Under Section 2(xx) of the Copyright Act, "sound recording" means a recording of sounds from which such sounds may be produced regardless of the medium on which such recording is made or the method by which the sounds are produced.

23 (2008) 1 SCC 1.

Further, the work is also required to "originate" from the author. Hence, in addition to the above, when it comes to the creation and ownership of copyrighted works, the key question is who would be the 'author' of an AI-generated work.

Author of the Works Generated through GAI Tools

Under Section 2 (d) of the Copyright Act, 1957 ("**Copyright Act**"), "author" means - "(vi) in relation to any literary, dramatic, musical or artistic work which is computer-generated, the person who causes the work to be created". The first issue under the above-mentioned definition is its usage of the terms 'the person who causes the work to be created'. It will need to be assessed whether artificial persons such as AI tools can even be granted authorship under the Copyright Act and whether authorship for all works can vest with entities/ body corporates. Also, determining who 'causes' a work to be created is a question of the proximity of a person to the creation of the 'expression' in the content in question — the more closely or directly a person is involved in creating the 'expression', the more he or she contributes to it, and the more likely he or she is to qualify as a person 'who causes the work to be created'. As a result of the above, the current legal framework under the Copyright Act may not effectively deal with/prescribe for the creation of works where the actual creator or a contributor of the 'expression' is not a person, or where the role of the person is minimal. Since there is no settled jurisprudence to this effect, it is unclear who the 'author' of AI-created works is under Indian copyright law.

Interestingly, the Indian Copyright Office granted copyright recognizing "RAGHAV Artificial Intelligence Painting App" ("**AI Raghav**") and Ankit Sahni as authors for a painting titled "Suryast" in December 2021.[24] However, as per news reports, around a year after registration, a withdrawal letter by the Copyright Office was issued.[25] The office placed the onus on the applicant to "inform the Copyright Office on the legal status of AI Raghav in the withdrawal notice. Nevertheless, the office's website reflects the status of the application as still registered. The actual status of the case is unknown.

GAI is already attracting significant amounts of investments and is expected to find an increasing number of use cases, as we have detailed earlier. The lack of clarity in the law regarding the copyrightability of AI output will lead to considerable uncertainties, especially for organizations and individuals invested in this sector. Hence, a clear legal position will sooner or later become indispensable, whether it is made through legislation or jurisprudence.

Current Global Trends

In jurisdictions like China, courts are moving towards recognizing the intellectual contributions of users using GAI tools to generate, in the form of their personalized prompt generation and aesthetic choices. The Beijing Internet Court in a landmark judgement in the case of Li v. Liu[26] granted copyright protection to an image generated based on prompts inputted by a user into the AI art generator Stable Diffusion.

24 Diary no. 13646/2020-CO/A.

25 See https://spicyip.com/2022/10/ai-art-and-indian-copyright-registration.html, last accessed June 24, 2024.

26 (2023) Jing 0491 Min Chu No. 11279 (2023).

The Court noted that the user undertook multiple iterations of finetuning the parameters of the prompts which generated the final image. The Court noted the following while summarizing its opinion on the intellectual contribution involved in the process:

"Based on the initially generated picture, the plaintiff added some prompt words, modified the parameters, and finally got the picture he wanted. From the time the plaintiff had an idea about the picture to his final selection of the picture involved, the plaintiff did some intellectual investment, such as designing the presentation of the character, selecting prompt words, arranging the order of prompt words, setting parameters, and selecting the picture that he wanted. The picture involved reflects the plaintiff's intellectual investment, so it meets the element of "intellectual achievement."[27]

However, in other jurisdictions, proceedings are still ongoing and there are no decisive pronouncements on whether works generated by AI tools can qualify as copyrightable works. In the US, a case was filed on whether a work created entirely autonomously by an AI tool of the plaintiff could be registered as a copyrightable work. The plaintiff claimed that since the AI tool was created by him, he would be the copyright owner of the work, and the AI tool would be the author on a 'work-for-hire' basis.[28] The US Copyright Office rejected the application for registration on the grounds that: *"copyright has never stretched so far, however, as to protect works generated by new forms of technology absent any guiding human hand ... human authorship is a bedrock requirement of copyright"*. Proceedings are still ongoing.

In 2022, the US Copyright Office issued a decision rejecting copyright protection for images generated by Midjourney, which is a GAI system. However, in March 2023, the office released guidance on "Works Containing Material Generated by AI"[29] in which it stated that copyright protection depends on whether the AI's contributions are "the result of mechanical reproduction", such as in response to text prompts, or if they reflect the author's "own mental conception".[30] The office further said that *"The answer will depend on the circumstances, particularly how the AI tool operates and how it was used to create the final work"*.[31] In August 2023, the US Copyright Office published a Notice of Inquiry seeking public comments on a broad range of copyright law and policy issues arising from the development and use of GAI.[32] The US Copyright Office plans to issue a report (in multiple parts) to address the various issues on AI and copyright.[33]

The UK government, in response to the House of Lords Culture, Media and Sport Committee's report on AI and the creative industries confirmed its position on striking a balance between encouraging AI innovation and protecting copyright works. The government has announced its decision to publish a code of practice on AI and copyright, the need for licensing schemes for the usage of copyrighted works as training data through *"mutually-beneficial arrangements [...] with rights management organisations and creative industries trade bodies"*.[34]

27 Ibid.

28 See: https://www.courtlistener.com/docket/63356475/1/thaler-v-perlmutter/; https://www.mishcon.com/generative-ai-intellectual-property-cases-and-policy-tracker, last accessed June 24, 2024.

29 See: https://fingfx.thomsonreuters.com/gfx/legaldocs/byvrlmjlzve/AI%20COPYRIGHT%20notice.pdf, last accessed June 24, 2024.

30 See: https://www.reuters.com/world/us/us-copyright-office-says-some-ai-assisted-works-may-be-copyrighted-2023-03-15, last accessed June 24, 2024.

31 See: https://www.reuters.com/world/us/us-copyright-office-says-some-ai-assisted-works-may-be-copyrighted-2023-03-15, last accessed June 24, 2024.

32 See: Artificial Intelligence and Copyright, 88 Fed. Reg. 59,942 (Aug. 30, 2023), https://www.govinfo.gov/content/pkg/FR-2023-08-30/pdf/2023-18624.pdf, last accessed June 24, 2024.

33 See: https://copyright.gov/laws/hearings/USCO-Letter-on-AI-and-Copyright-Initiative-Update-Feb-23-2024.pdf?loclr=blogcop, last accessed June 24, 2024.

34 See: https://publications.parliament.uk/pa/cm5804/cmselect/cmcumeds/441/report.html, last accessed June 24, 2024.

Other jurisdictions are in the process of legislating on the matter of AI and copyright. In South Korea, the government has stated that they are prioritizing the reform of the copyright system for AI-generated content and have stated that "the government will promptly revise the AI copyright system to bolster the development of the AI-based creative industry,".[35] In Singapore, the government issued a discussion paper in June 2023, titled "GAI: Implications for Trust and Governance, which among other things discussed the copyrightability of output, interpretation of existing laws on copyright, training data and infringement, etc. The paper advocated for discourse and global co-operation on these matters.[36]

Indian Government's Position on Ownership of AI Output

On July 23, 2021, the Rajya Sabha Department-related Parliamentary Standing Committee on Commerce ("**Committee**") released its 161st Report titled "Review of the Intellectual Property Rights Regime in India"[37] ("**161st Report**"), which discussed the issues of artificial intelligence and intellectual property, noting that the present scope of the Indian Patents Act, 1970 and the Copyright Act are not well-equipped to facilitate inventorship, authorship and ownership by AI.

The Report further went on to note the importance of AI and its applications in revenue generation and technological invention and the need for the creation of a "separate category of rights for AI and AI-related inventions and solutions" and "their protection as intellectual property rights". It further recommended that the "department should make efforts in reviewing the existing legislations of The Patents Act, 1970 and Copyright Act, to incorporate the emerging technologies of AI and AI related inventions in their ambit". The Report emphatically noted the need to review the provisions of both the legislations on a priority basis.

On April 6, 2022, the Committee released its 169th report on the Action Taken by the Government on the Recommendations/Observations of the Committee contained in its 161st Report ("**169th Report**")[38]. The 169th Report confirmed the position under the law that only natural persons can be authors of literary, dramatic, artistic, and musical works, it is difficult to ascertain in whom the rights arising out of an AI-generated creative work vest. The 169th Report observed that an AI tool cannot execute or authorize its creator or any other person, to become the owner of the work. An AI may not be capable of enforcing its rights, nor cannot be held accountable and tried in a suit for enforcement of rights in case of an act of infringement. The report reemphasized how there is no provision regarding AI-generated works in the current IP legislations, leading to a situation where AI-related works are commercially utilized in ways without incurring any costs, causing a loss of revenue to companies who invest in AI-related R&D activities.

Hence, granting proprietary rights to AI innovators and protecting AI-driven innovations by enforcing regulations and standards in the country should be the way forward. The Committee, therefore, recommended that the department should channelize efforts to encourage and empower AI innovators by enacting suitable legislations or modifying the existing laws on IPR in order to accommodate AI-based inventions.

35 See: https://www.koreatimes.co.kr/www/tech/2024/05/129_375068.html, last accessed June 24, 2024.

36 See: https://aiverifyfoundation.sg/downloads/Discussion_Paper.pdf, last accessed June 24, 2024.

37 See: https://sansad.in/getFile/rsnew/Committee_site/Committee_File/ReportFile/13/141/161_2022_5_12.pdf?source=rajyasabha#page=26, last accessed June 24, 2024.

38 See: https://sansad.in/getFile/rsnew/Committee_site/Committee_File/ReportFile/13/159/169_2022_4_16.pdf?source=rajyasabha#page=85, lastaccessed June 24, 2024.

However, more recently, on February 9, 2024, in response to a parliamentary question, the Union Minister of State for Commerce and Industry, Mr. Som Parkash noted the following on the government's intention towards regulation of artificial intelligence and intellectual property rights:

"India being a member of all major international conventions and agreements for the protection of Intellectual Property Rights grants adequate protection of rights for works created by legal persons through Copyright Law and protects inventions through the Patent system. Therefore, there is no requirement to create a separate category of rights for AI and related innovations in the Indian IPR Regime. Therefore, while Artificial Intelligence (AI) and related innovations is an evolving stream of technology the current legal framework under the Patent and Copyright Act is well-equipped to protect Artificial Intelligence generated works and related innovations. Presently, there is no proposal to create any separate [rights or amend] the law in the context of AI-generated content.

The exclusive economic rights of a copyright owner such as the right of reproduction, translation, adaptation etc. granted by the Copyright Act, 1957 obligates the user of Generative AI to obtain permission to use their works for commercial purposes if such use is not covered under the fair dealing exceptions provided under Section 52 of the Copyright Act. Since Intellectual property rights are private rights, these are enforced by the individual rights holders. Adequate and e□ective civil measures and criminal remedies are prescribed under the Copyright Law against any act of infringement or unauthorized use of works, including digital circumvention."

The Philosophical Basis of Copyright Law and its Application to GAI

The justification for granting copyright under various jurisdictions may differ, for example, in common law and civil law jurisdictions.[39] The utilitarian rationale for copyright has historically dominated in common law jurisdictions, as exemplified in the U.S. Constitution's provision *"to promote the progress of science and useful arts".*[40] In the French IP Code, the natural right justification is more evident, as is common in other civil law countries — *"The author of a work of the mind shall enjoy in that work, by the mere fact of its creation, an exclusive incorporeal property right which shall be enforceable against all persons."*[41]

Apart from these justifications, there is the economic justification for copyright, i.e., to protect the revenues and livelihood of authors, and to encourage further creation.[42] Moreover, copyright is considered to foster democracy and free speech by ensuring that authors do not have to depend on Government or private patronage, and can earn revenues by lawfully exploiting their works, and restricting others from doing so.[43]

Applying these justifications to AI-generated works also requires a determination of whether the AI in question is pure AI or human-assisted AI. In ChatGPT's own words, ChatGPT does require *"some level of human intervention in its training and maintenance...such as periodic retraining with new data or adjustments to its algorithms or parameters".*[44] Hence, assuming that the GAI in question is assisted by humans, and in light of the justifications behind copyright protection, should the works created by such AI based on human instructions be granted copyright protection?

---

39 Paul Goldstein and P. Bernt Hugenholtz, INTERNATIONAL COPYRIGHT: PRINCIPLES, LAW, AND PRACTICE (OXFORD UNIVERSITY PRESS, 2019), 5.

40 Article I, Section 8, clause 8 of the US Constitution.

41 Art. L. 111- 1, Intellectual Property Code of France.

42 See: https://cyber.harvard.edu/people/tfisher/iptheory.pdf, last accessed June 24, 2024.

43 Supra note 82, at 6.

44 OpenAI's ChatGPT AI language model, response to question from author "*Can ChatGPT be categorised as Assisted AI?*", February 16, 2023.

There are arguments on both sides — proponents of copyright protection state that such kind of protection is necessary to incentivize AI programmers, investors, and users (such as licensees who may want to use GAI for their own business).[45] The UK, for example, extends copyright protection to works that are generated only by computers, however, no moral rights are granted.[46] Section 9 of the Copyright, Designs, and Patents Act, 1988 states that the author of computer-generated works is "the person by whom the arrangements necessary for the creation of the work are undertaken." Although this law has been in existence since the 1980s, this is in line with the UK's objective towards encouraging AI development in the UK, and incentivizing businesses to research, develop, and deploy AI technology.[47]

The Indian Copyright Act recognizes that computer-generated works have authors, i.e., the person who "causes" the work to be created,[48] however, it does not define computer-generated works or the extent of human involvement in such works, if any.

Other arguments against protecting AI-generated works include (i) the fact that copyright has its roots in human authorship and creative endeavor, and protecting AI-generated works will devalue human creativity, and (ii) there is no evidence that such protection will incentivize the development of GAI.[49]

Whichever approach is taken, it is important to assess the impact such an approach will have on incentives for humans to create works and develop such GAI. There is currently insufficient data to determine which approach will work best, both from a copyright user and a GAI developer perspective.[50] Nevertheless, these debates must continue so that Governments can take timely and informed actions when the popularity of GAI demands a decision to be made.

C. Data Protection and Confidentiality

GAI models extensively process large datasets for the purposes of training, and hence it becomes important to assess the implications under the new Indian Digital Personal Data Protection Act, 2023 ("**DPDPA**"). We touch upon some of the key issues below.

Training Data and Processing of Personal Information

GAI relies on a vast trove of data to train itself - data that is scraped from a wide variety of sources, such as from licensed or unlicensed data sets, open-source third-party data sets or consumer apps. Often, the data contains personal information gathered in contravention of the applicable data protection laws, like the EU's General Data Protection Regulation and the DPDPA. This can carry serious financial, reputational, and penal risks for the developer.[51]

45 Mauritz Kop, AI & Intellectual Property: Towards an Articulated Public Domain, 28 Texas Intellectual Property Law Journal (2020).

46 See: https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views/artificial-intelligence-call-for-views-copyright-and-related-rights; and https://www.pinsentmasons.com/out-law/analysis/uk-to-decide-copyright-protection-creative-works-generated- ai#:~:text=Currently%2C%20creative%20works%20generated%20solely,on%20which%20it%20was%20made, last accessed June 24, 2024.

47 See: https://www.pinsentmasons.com/out-law/analysis/uk-to-decide-copyright-protection-creative-works-generated-ai#:~:text=Currently%2C%20creative%20works%20generated%20solely,on%20which%20it%20was%20made, last accessed June 24, 2024.

48 Section 2(d)(vi) of the Copyright Act.

49 Supra note 88.

50 See: https://www.pinsentmasons.com/out-law/analysis/uk-to-decide-copyright-protection-creative-works-generated-ai#:~:text=Currently%2C%20creative%20works%20generated%20solely,on%20which%20it%20was%20made, last accessed June 24, 2024.

51 See: https://www.zwillgen.com/privacy/artificial-intelligence-risks-privacy-generative-ai, last accessed June 24, 2024.

For example, the DPDPA states that the personal data of a data owner ("**data principal**") can only be processed for lawful purposes for which a data principal has given consent in the manner proposed by the DPDPA and the rules thereunder. Thus, the developer has to first identify if the scraped data contains any personal data; and then should they choose to keep the same, they must provide an itemised notice in "clear and plain language containing a description of personal data" sought to be collected. It is also important to note that the process of "scraping the personal data" will amount to processing under the DPDPA. Further, the DPDPA lays down that before processing any personal data of a child or a person with a disability, verifiable consent from the individual's lawful guardian must be obtained before processing their data.[52] Further, the DPDPA prohibits the processing of personal data that may have detrimental effects on the well-being of children[53] and undertaking tracking / behavioural monitoring as well as targeted advertising towards children.[54] Hence, the developer would be required to identify if the personal data in the scraped data belongs to a child or a person with a disability, and accordingly adhere to the additional obligations under DPDPA. However, the provisions of DPDPA will not be applicable if the personal data being processed is made or caused to be made publicly available by (i) the data principal to whom such personal data relates, or (ii) any other person who is under a legal obligation to make such personal data publicly available.

Given that the scraped data runs into tens of terabytes, identifying and categorising the nature of data, not to mention determining whose personal data it is and contacting them, is no mean feat. The costs and inefficiencies involved in the process will hamper the development of GAI models, but may pale in comparison to the risks of non-compliance. For example, the Federal Trade Commission in the US uses 'algorithmic disgorgement' as a new privacy enforcement tool, requiring companies who scraped data without people's consent, in violation of terms of use of certain websites or copyright, to delete their algorithms and training data.[55]

Personal data can also become part of the training data *after the* deployment of the GAI model. Privacy policies of platforms such as Open AI *inter alia* mention that *"If you communicate with us, we may collect your name, contact information, and the contents of any messages you send."*

Thus, whenever one enters a prompt on a GAI model, which could include some personal information about themselves or someone else, then that personal information can permanently become part of the platform's training data. Given that inputs are also used to train GAI platforms, one can also keep entering leading prompts while revealing personal or false details about others, so that not only is the platform tricked into writing false information but also ingests those false prompts in its training data. This can have serious ramifications for one's privacy as without the consent or knowledge of an individual, some other user, not necessarily the developer, is feeding false or personal information into the GAI. It then becomes the onus of the developer to correct or delete the same, if they receive a request from the principal (discussed below).

52 Section 9(1), DPDPA.

53 Section 9(2), DPDPA.

54 Section 9(3), DPDPA.

55 See: https://wp.nyu.edu/compliance_enforcement/2022/03/30/model-destruction-the-ftcs-powerful-new-ai-and-privacy-enforcement-tool, last accessed June 24, 2024.

Right to Correction/Erasure

Even if a developer acquires training data in a fully compliant manner, the data principal still has the right to withdraw their consent. In such a case, the developer has to cease using that personal data in a 'reasonable time.'[56]

However, since the concerned personal information has already been fed to a GAI model, and given that developers themselves don't know what information a model takes into account during its development, it is still an unknown as to how the information can be isolated and removed from a fine-tuned model's 'parameters.'

Moreover, the DPDPA mandates that a developer also has to comply with a data principal's requests for correction or erasure of their personal data if the data is inaccurate or misleading, or incomplete, and requires updation.[57]

Suppose that defamatory content about someone is published online and becomes part of the GAI's training data. Thereafter, a judicial authority determines that the said content is indeed defamatory and orders the concerned websites to take it down. Now, even if the websites comply, that information has already become part of the training data of the GAI model and the same might be outputted to whoever has given related prompts. As highlighted before, the developer would then have to delete that information from the training data and ensure that the GAI doesn't provide the same output to other user prompts. If the developer cannot figure it out, it runs the risk of destroying the model and training data altogether and starting from scratch.

Right to Contextual Integrity

Even if the GAI model solely uses publicly available information as its training data, it is still theoretically capable of violating our 'contextual integrity.' That is, we have a reasonable expectation that our personal information, even if willingly and knowingly made available in the public domain, is not used outside of the context in which agreed to part with it.

For example, the law may require the government to post updated records of liens on one's personal property. Another law may require voter records to be posted in the interest of transparency. We may also post online, of our own volition, blog posts, product reviews, etc. We may thus knowingly and willingly have fragments of our personal information and views scattered across the web. However, GAI is capable of synthesising all this information and making it easily and readily available by someone entering a few queries. It's the scale and speed with which GAIs can compile scattered and fragmented personal information that can violate our contextual integrity.[58]

56 Section 6(6), DPDPA.

57 Section 12(2), DPDPA.

58 See: H. Nissenbaum, Privacy as Contextual Integrity, 79 Washington Law Review (2004).

D. Bias, Prejudice and Ethics

Given the potential of GAI models to alter human perceptions, disseminate false information, and exercise social control, we need to consider the biases (if any) encoded in them. The tendency of algorithms to mirror human prejudices/biases is referred to as algorithmic or AI bias.

AI bias could occur from using incorrect, flawed, or biased data sets by the personnel who train and/or validate the machine learning algorithms.[59] Additionally, the algorithms may also reflect unintentional cognitive biases[60] or social biases[61] of the very persons programming the algorithms. Biases will influence GAI and upcoming AI-based technology. Although these biases reflect our existing society, the goal of AI is to make society better and provide a neutral output.

AI Bias manifests itself in a variety of ways, including through unjust discrimination and stereotypes. Following are some key types of AI biases that could affect GAI functioning:

Sampling/Selection Bias[62]

When the training data is not randomly picked and appears to be favoring specific groups. For example, facial-recognition, AI won't perform well when recognizing women and persons of different races if it is trained mostly on images of white men.[63]

Overgeneralisation Bias[64]

Drawing of conclusions that are excessively broad because they go beyond what can be logically deduced from the facts given. Finding statistical patterns in a training set to produce a more broadly applicable output is a crucial component of machine learning. However, if it is left unchecked, there is always a chance of coming to incorrect or misleading conclusions based on small sample size or little data.

For example, the greater arrest rates in minority groups are partially explained by the fact that they are monitored and policed more frequently. Yet, since there are differences in how these groups are evaluated and managed, AI should not draw the conclusion that just because persons from minority groups are arrested more frequently, they are inherently more dangerous.[65]

59 See: https://wires.onlinelibrary.wiley.com/doi/full/10.1002/widm.1356, last accessed June 24, 2024.

60 See: https://www.techtarget.com/searchenterpriseai/definition/cognitive-bias, last accessed June 24, 2024.

61 See: https://link.springer.com/article/10.1007/s43681-020-00026-z, last accessed June 24, 2024.

62 See: https://dl.acm.org/doi/pdf/10.1145/3466132.3466134, last accessed June 24, 2024.

63 See: https://link.springer.com/article/10.1007/s13347-021-00478-z, last accessed June 24, 2024.

64 See: https://www.statice.ai/post/data-bias-types, last accessed June 24, 2024.

65 See: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, last accessed June 24, 2024.

Representation Bias[66]

Representation bias arises when the data used to train a language model contains biased representations of particular groups of people. Additionally, an AI model can also produce inaccurate projections concerning under-represented groups if it is trained on data that only represents a small portion of the population.

For example, a language AI tool that scans curriculum vitae for recruitment or career counselling may be less likely to propose historically discriminated groups to recruiters, or more likely to offer lower-paying employment to marginalized groups as the data fed to the AI represents bias in recruitment of such marginalized groups.[67]

However, in an effort to correct biased representation of minority groups in their training data, GAI tools may overcorrect leading to historically inaccurate generations. For example, in February 2024, Google's Gemini was under the radar for its image generation feature which depicted figures that were historically white (such as the Founding Fathers of the United States of America or Nazi-era German soldiers) as people of color in order to subvert long-standing problems of racial bias in GAI tools.[68]

Evaluation Bias[69]

Develops while evaluating and iterating an AI model. The quality of the model is frequently evaluated against predetermined standards after it has been optimised using training data. Since these standards are biased or inappropriate for how the model will be utilised, they may not accurately reflect the overall population. For example, AI models deployed in recruitment services, showed preference to male candidates by filtering out female candidates, based on faulty eligibility criteria formulated by the AI, arising from the data fed to it.[70]

Aggregation Bias[71]

Occurs when incorrect inferences are made about specific people based on studying the overall population. For example, consider AI tools used in clinical aid that look at diabetic patients who appear to have different morbidities depending on their gender and racial backgrounds. The AI model may make generalisations about gender and racial subgroups in the population for such morbidities.

66 See: https://arxiv.org/abs/1908.09635, last accessed June 24, 2024.

67 See: https://arxiv.org/abs/2301.12867, last accessed June 24, 2024.

68 See: https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical, last accessed June 24, 2024.

69 See: https://arxiv.org/abs/1908.09635, last accessed June 24, 2024.

70 See: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G, last accessed June 24, 2024.

71 See: https://arxiv.org/abs/1908.09635, last accessed June 24, 2024.

E. Explainability and Accountability of Artificial Intelligence

AI is often praised for its effectiveness, but concerns about the transparency, accuracy, and trustworthiness of its methods have started to emerge.[72]

The issue of accountability in GAI is becoming increasingly important and arises from the fact that, unlike human-generated content, AI-generated content may be difficult to trace back to its source or author. This makes it challenging to hold individuals or organizations responsible for any harm caused. Additionally, as discussed in Section 3(a) above, GAI models have the potential to exhibit bias since they learn patterns and make predictions from the data on which they are trained. If such training data is biased or incomplete, the model's output may also be biased. Explainable AI or XAI is one such tool to address these issues.

To mitigate some risks, the development of Explainable AI ("**XAI**")[73] has evolved to address accountability issues.[74] The concept of XAI pertains to the ability to explain how an AI system arrived at a particular outcome, decision, or suggestion in order to help stakeholders understand why a certain decision was made, what factors were considered, and what the possible outcomes are. XAI tools aim to explain how AI mechanisms like deep learning models, cognitive neuroscience algorithms, and machine Learning algorithms arrive at their decisions,[75] by employing simpler models to provide explanations about the working of the "black box".[76] XAI holds deep significance as it breaks down the technical set of processes that derive outcomes on the GAI platforms, and brings forth the details of how such outcomes were achieved — in a manner that is interpretable to humans.

XAI tools are highly beneficial, primarily in instilling improved trust in AI systems, facilitating better decision-making, and ensuring increased accountability. [77] For example, in the healthcare sector, XAI can help doctors understand how AI models arrived at their diagnoses, enabling them to make more informed decisions.[78] Similarly, XAI assists developers of the GAI models in identifying bugs and errors in the programs that form a part of such GAI platforms, which enables the developers to debug and improve the quality of output on these GAI platforms.[79]

For organizations that tend to rely on AI tools to any degree, XAI could become crucial to building and maintaining stakeholder trust by increasing transparency and accountability in the decision-making process of AI systems. When an AI system makes a decision that affects people's lives, stakeholders must understand how that decision was made. If the decision-making process is not transparent or explainable, stakeholders may lose trust in the system and the organization using it.[80]

---

72 See: https://www.ibm.com/in-en/watson/explainable-ai, last accessed June 24, 2024.

73 XAI, as coined by DARPA, is the method that helps human experts understand solutions developed by AI (See: https://neptune.ai/blog/explainability-auditability-ml-definitions-techniques-tools), last accessed June 24, 2024.

74 See: https://www.sciencedirect.com/science/article/pii/S0740624X21001027, last accessed June 24, 2024.

75 See: https://www.ibm.com/in-en/watson/explainable-ai, last accessed June 24, 2024.

76 The goal of XAI is to address the black box problem, which is the challenge of understanding how an AI system arrived at its decision or recommendation. See: https://christophm.github.io/interpretable-ml-book/explanation.html; https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8466590, last accessed June 24, 2024.

77 See: https://neptune.ai/blog/explainability-auditability-ml-definitions-techniques-tools, last accessed June 24, 2024.

78 See: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01332-6, last accessed June 24, 2024.

79 See: https://em360tech.com/tech-article/what-is-xai#:~:text=XAI%20makes%20it%20easier%20to,responsible%20and%20trustworthy%20AI%20development, last accessed June 24, 2024.

80 See: https://christophm.github.io/interpretable-ml-book/explanation.html; https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8466590, last accessed June 24, 2024.

Entities across the world like Microsoft,[81] IBM,[82] Amazon[83] and Google[84] are actively adapting and promoting XAI, which shows that it is slowly gaining significance — especially in terms of providing the user with the power to ensure that AI is driven by the business, and not the other way around.

F. Content Regulation

Assuming that GAI is not considered an intermediary and not accorded a safe harbour for the outputs generated by such AI, a key question is whether such AI can be liable for unlawfully generated content. Developers generally employ guardrails to ensure that certain content is never present in the output generated, including obscene content. E.g., OpenAI's policy does not allow the usage of their models for illegal activity, child sexual abuse material, generation of hateful, harassing, or violent content, generation of malware, etc.[85] However, there are certain key concerns that may still arise:

Firstly, developers would likely not be aware of all applicable content regulations in every jurisdiction in order to build appropriate guardrails. The guardrails would have likely been built based on the legal regime of where the developers/operators are based.

Secondly, there are conflicting standards applicable across jurisdictions on what is protected by the right to free speech. For this reason, unless developers comply with the strictest possible standard for content, the same output is likely to be unlawful in one jurisdiction, while being compliant in another.

Thirdly, even if such guardrails are sought to be built as per applicable laws, the legal position may not be so precisely defined that safeguards can be built into the AI. E.g., in order to determine if certain content contains hate speech, courts need to analyse the facts on a case-by-case basis, in light of the factors laid down under the law. Such a determination may not be possible by an AI, especially in borderline cases.

Fourthly, even if precise guardrails are built in, the AI may still generate output that violates the operator's policies. For e.g., CyberArk was able to get ChatGPT to create a new strand of malware after repeated requests and constraining the output available.[86] This would be against OpenAI's policy, nevertheless, users were able to bypass this policy.

Thus, the possibility of unlawful content being made available by GAI is very much real. Further, in case GAI does create such unlawful content, it would be interesting to see how such content is dealt with by law enforcement authorities, as well as how liability is ascertained for such content.

In the case of casual users of GAI, even if the output contains unlawful content, only the user would be able to view such content. Law enforcement authorities may not come to know of such content unless the user complains of such output being provided or the user making the output publicly available. In the latter case, exposure still remains when such content is published by the user on public platforms such as social media.

---

81 See: https://learn.microsoft.com/en-us/azure/machine-learning/concept-responsible-ai-dashboard?view=azureml-api-2, last accessed June 24, 2024.

82 See: https://aif360.res.ibm.com/#:~:text=This%20extensible%20open%20source%20toolkit,throughout%20the%20AI%20application%20lifecycle, last accessed June 24, 2024.

83 See: https://aws.amazon.com/sagemaker/clarify, last accessed June 24, 2024.

84 See: https://cloud.google.com/explainable-ai#:~:text=Explainable%20AI%20is%20a%20set,others%20understand%20your%20models'%20behavior, last accessed June 24, 2024.

85 See: https://platform.openai.com/docs/usage-policies/disallowed-usage, last accessed June 24, 2024.

86 See: https://www.infosecurity-magazine.com/news/chatgpt-creates-polymorphic-malware, last accessed June 24, 2024.

In business use cases, the output of GAI may be more publicly available. Either way, once law enforcement authorities are made aware of such content, they may require operators or business users of GAI to ensure that similar content is not published for other users. They may argue that operators/business users have (or should develop) the necessary technical measures to ensure content moderation for outputs that are established to be unlawful. Issues such as these will invariably crop up as the use of GAI becomes more popular.

How is Liability Determined?

If an algorithm does generate unlawful content, the logical next question would be regarding the apportioning of liability. If the nature of the AI is such that a human can be identified for causing and/or controlling the unlawful output, the solution is much simpler.[87] However, when AI "autonomously" acts to create such content in a way that can neither be foreseen nor explained, the answer is not as straightforward. In such a case, there are three eligible candidates– the developer of the AI, the user who prompted the AI to generate that content, and the AI itself. The liability may be on one of these or even a combination of these.

Imposition of liability under the law would need to take into account aspects of morality (i.e., whether it is justified to impose liability on a particular entity), control (whether such entity can actually control the outputs in a feasible way), and efficacy (whether law enforcement authorities can act against such entities to make the liability effective). As with any other law, the positives of such provisions (reducing risks from unlawful content) would need to be balanced with any negative consequences (e.g., discouraging technological development).

In relation to criminal liability, it can be argued that there must be mens rea (i.e., a criminal intent) on the part of the entity committing the criminal offense. In cases where an AI tool has taken sufficient due diligence in the form of content filters and restrictions on its users in their terms of service and other guard rails, it would be interesting to see how jurisprudence on the same evolves especially establishing mens rea for apportionment of criminal liability in such cases.

On March 13, 2024, the European Parliament adopted the AI Act which contains a regulatory framework on AI to provide AI developers, deployers, and users with clear requirements and obligations regarding specific uses of AI.[88] The framework is based on the understanding that different AI use cases may involve different degrees of risk. Nevertheless, there is a need to balance the problem-solving ability of AI systems with avoiding undesirable outcomes that certain AI may produce. The Regulatory Framework defines 4 levels of risk in AI: (i) Unacceptable risk; (ii) High risk; (iii) Limited risk, and (iv) Minimal or no risk. While AI systems with unacceptable risk (e.g., social scoring, biometric categorization, emotion recognition in the workplace and education institutions, etc.) are proposed to be banned, those posing high risks (such as transport and exam scoring) will be subject to strict obligations before being released into the market. There will be minimal pre-release obligations for limited and minimal-risk AI systems which forms the majority of AI systems currently in place. However, there will be transparency obligations, for example, a requirement to inform the user that they are interacting with AI. Separate requirements apply to general-purpose AI models (GPAI) and GPAI models that pose systemic risk.

87 This would form a classic case of the Perpetration-via-Another Liability Model put forth by Hallevy, where the AI is considered an 'innocent agent', or a mentally limited person such as a child. Therefore, under this model, the liability would either be on the programmer of the AI software or the end user.

88 See: https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law, last accessed June 24, 2024.

The major obligations under the AI Act are applicable to high risk AI system providers / developers. It is pertinent to note the European Parliament proposed the AI liability directive ("**Liability Directive**")[89] in order to compliment the AI Act and the risks laid down therein. The Liability Directive defines 'claim for damages' as *a noncontractual fault-based civil law claim for compensation of the damage caused by an output of an AI system or the failure of such a system to produce an output where such an output should have been produced.* The Liability Directive puts forth that the claimants seeking compensation get a more reasonable burden of proof and a higher chance of successful liability claim, wherein they have to demonstrate non-compliance with the AI system with the relevant law. However, in cases where it has been established that the human's act or omission caused the output to be generated, and such output caused damage, the liability will not be that of the AI system. The Liability Directive also proposes that in order for the courts to accurately determine the person liable for causing damage, the courts can order the disclosure of evidence about high-risk AI systems. Such evidence may then be used to determine the individual(s). Similarly, a claimant can ask for evidence from the AI system providers, and the court may order for preservation of evidence.

While a definitive outcome of these proposals is uncertain at this stage, it is likely that the best possible approach will only present itself once these approaches are experimented upon and more data is available on their respective impacts.

G. Generative Artificial Intelligence Models: Intermediaries or Publishers?

The Information Technology Act, 2000 ("**IT Act**") defines an 'intermediary' with respect to a particular electronic record.[90] The definition of "electronic record" is fairly broad, and means *"data, record or data generated, image or sound stored, received or sent in an electronic form or micro film or computer generated micro fiche".*[91] Hence, outputs provided by GAI, whether in the form of text, images, sounds, etc., would be considered electronic records.

An intermediary, under the IT Act, with respect to any particular electronic record, means *"any person who on behalf of another person receives, stores or transmits that record or provides any service with respect to that record 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."*

However, being an intermediary is not sufficient to avail safe harbor under the IT Act. Safe harbour is provided only for *third-party content* hosted or made available by the intermediary if it satisfies certain conditions. Otherwise, the intermediary may be held liable under various regulations relevant to the content hosted or made available by it.

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. They may rely on the decision in *Myspace Inc. v. Super Cassettes Industries Limited*,[92] ("**Myspace case**") where it was held that intermediaries could be held liable only when they have actual or specific knowledge and not constructive knowledge of the existence of infringing content on their website, and do not take any steps to have such content removed.

89 See: https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI(2023)739342_EN.pdf, last accessed June 24, 2024.

90 Section 2(w) of the IT Act.

91 Section 2(t) of the IT Act.

92 FAO(OS) 540/2011, C.M. APPL.20174/2011, 13919 & 17996/2015, decided on December 23, 2016.

Accordingly, operators may argue they do not have actual or specific knowledge of the final output being provided by the AI programme.

Further, in the *Myspace case*, it was held that modifications made to the format of the content, both via an automated process and without manual intervention, would not amount to having actual control, actual knowledge, or a 'reason to believe' that the content uploaded may be infringing of third party rights. GAI also makes automated changes to content without any manual intervention. Nevertheless, it is important to note that in the *Myspace* case, the content uploaded by third parties was not being modified, except for the format itself and the insertion of some advertisements which may not be true for GAI.

GAI are not purely conduits of data since (i) they provide outputs based on the user inputs, and (ii) may modify the ingested data to create a new output. Hence, traditional notions of intermediaries being in the nature of conduits (such as internet service providers or search engines) may not be applicable to GAI. However, one could argue that GAI operators are providing services to users basis the electronic record (input provided by them) and to that extent should be treated as intermediaries. Hence, it would be interesting to see how liability for GAI platforms evolves including jurisprudence on whether intermediary safe harbour can be availed. As stated above in Chapter 3(1), the MeitY had issued advisories in relation to harmful AI-generated content. In such advisories, it appears that the MeitY is viewing several types of GAI platforms (regardless of the difference in their functionalities) as intermediaries for the purposes of obligations under the IT Rules. This may result in platforms that do not perform any functions in relation to providing a communication system over which content is transmitted or provide any services on behalf of another person in relation to an electronic record, being regarded as intermediaries under the IT Rules.

Further, with the Digital India Act also in the works (as of the date of writing), the Government has indicated that the new legislation will treat intermediaries differently based on their business models. Hence, conditions for availing of safe harbor may also differ based on the kind of services provided by the respective intermediary. Hence, it would be interesting to see if GAI would be one of the categories of intermediaries recognized by law, and if so, the kind of safe harbor provided to it.

H. Safety and Cybersecurity

The IT Act includes provisions for punishing offenses related to electronic communication, data, and other cybersecurity issues. Specific violations such as unauthorized access to computers or networks, downloading or copying data without permission, and denial of access to computers can result in the offender being required to pay compensation. These offenses cover a range of activities, including computer-related crimes, tampering with computer source documents, identity theft, cyber terrorism, and impersonation using a computer resource. Additionally, the IPC contains provisions for crimes that can be classified as cybercrimes, such as cheating and forgery.

In India, CERT-In is the national cybersecurity agency, established under the IT Act, and has been tasked with functions such as collecting, analyzing, and disseminating information on cyber incidents; issuing forecasts and alerts of cybersecurity incidents; and undertaking emergency measures for handling these incidents. CERT-In has the authority to call for information and direct service providers, intermediaries, data centers, corporate bodies, and others in matters of cybersecurity. The Information Technology (The Indian Computer Emergency Response Team and Manner of Performing Functions and Duties) Rules, 2013 outline CERT-In's obligations regarding response, prediction, prevention, analysis, and forensics of cybersecurity incidents.

Key requirements include mandatory reporting of certain cybersecurity incidents, the appointment of a Point of Contact (PoC) by relevant entities to interface with CERT-In, and compliance with CERT-In's directions and information requests. Non-compliance with these provisions may result in liability for compensation or a penalty of up to INR 10,00,000. On April 28, 2022, CERT-In issued directions that added compliance requirements to the existing framework, which was supplemented by a set of FAQs released by MeitY.

The emergence of GAI presents new cybersecurity challenges that require attention within the frameworks established by the IT Act, IPC, and CERT-In laws. GAI can be exploited to create sophisticated phishing attacks, deepfakes, and other forms of digital deception that bypass traditional security defenses. These capabilities make unauthorized access to computer systems and data breaches more likely. As GAI can be used to automate and scale cybercrimes, ensuring robust enforcement of laws against unauthorized access, data theft, and impersonation is essential.

GAI may thus increase the risk of identity theft and cyber terrorism. Further, the content generated through GAI platforms which may be in the form of realistic audio or visual imitations of individuals can lead to advanced social engineering attacks and fraud. This underscores the need for strengthening cybersecurity guardrails at an organizational level as well as the regulators' level, particularly in the collection, analysis, and dissemination of information on cyber incidents.

Moreover, GAI can complicate the forensic analysis and prediction of cyber threats, areas where CERT-In plays a critical role. AI-driven attacks can be more complex and harder to detect, necessitating advanced tools and techniques for effective incident response and prevention. The additional compliance requirements introduced by CERT-In through the directions in 2022, supported by the subsequent FAQs, aim to ensure that service providers, intermediaries, data centers, and other entities are well-prepared to handle these sophisticated cyber threats.

I. Competition Law Concerns

While GAI is currently at its nascent stage, it is set to raise numerous competition law concerns as the industry gets bigger and more consolidated. It is no surprise that most of the big tech companies are already at the forefront of the GAI boom. The GAI industry may tend to reward economics of scale and ultimately reward significant investments, resources, and capabilities with significant revenues. This may, in turn, make the industry prone to being dominated by "gatekeepers", paving the way for potential abuse of dominant positions.

Some of the factors which may create entry barriers or reward dominant players are:

- Training GAI requires massive amounts of curated data. Existing big tech players have access to such data, almost globally. Therefore, with a large amount of the necessary data already at their disposal, it should be easier for such entities to train the AI that they intend to deploy. On the other hand, non-incumbent players would first need to invest in purchasing or licensing the data.
- Training GAI also requires considerable computing power and storage capacity, both of which can be provided by cloud service providers. Hence, entities that are dominant in the cloud storage and cloud services business may also stand to gain from further development of GAI.
- As a result of the investments required to generate valuable AI applications, smaller players may be unable to generate revenues and growth due to the sheer lack of resources.

- Hoppner and Streatfeild summarise these factors and note that the characteristics of the stack required to build and deploy GAI has all ingredients for winner-takes-all battles: "(i) Vertical integration in closely interrelated markets, with (ii) upstream dominance and downstream value generation, (iii) unequal access to proprietary resources, (iv) issues of interoperability, (v) data portability, (vii) non-transparency, (vii) IP licensing, and (viii) platform usage fees and conditions vis-á-vis dependent business users, etc."[93]

It would need to be evaluated if the dominance of some players by itself could be harmful for the GAI market, and whether legislations such as the Digital Services Act and Digital Markets Act of the EU would be necessary or (in)effective in the long run. In India, emerging technologies and new-age intermediaries have been proposed to be dealt with under the Digital India Act. Additionally, the government has also proposed to introduce a law to regulate digital competition through the Draft Digital Competition Law Bill, 2024 ("**DCB**") which provides for pre-determined rules to regulate the conduct of large digital enterprises with a significant presence in India.

The DCB broadly lays down a framework for the ecosystem containing enterprises, business users, and end users. It has also put forth certain core digital services, including online search engines, online social networking services, video-sharing platform services, interpersonal communications services, operating systems, advertising services, web browsers, cloud services, and online intermediation services.[94] Interestingly, GAI related services are not expressly included within the list of core digital services. The DCB also lays down two categories of enterprises, viz., systematically significant digital enterprise and associated enterprise. An enterprise may be designated as a systematically significant digital enterprise based on qualitative and quantitative factors, as laid down in the DCB.

The DCB lays down certain obligations on systematically significant digital enterprises, which include but are not limited to obligations in relation to reporting of steps undertaken to comply with the requirements in the DCB, operating in a fair, non-discriminatory, and fair manner, ensuring that without an end user's consent, their data is not intermixed or crossed with data collected for other services and is not shared with any third party. Further, it has also been reported that the government regulator has been evaluating the impact of AI on markets.[95]

93 "ChatGPT, Bard & Co.: An Introduction to AI for Competition and Regulatory Lawyers", 9 Hausfeld Competition Bulletin (1/2023).

94 Schedule I, Para 1(i), Digital Competition Bill, 2024: "Online intermediation service" includes any other digital service, not expressly covered under clauses (a) to (h) of Schedule I, which on behalf of an end user or a business user, receives, stores or transmits electronic record or provides any service with respect to that record and includes web-hosting service providers, payment sites, auction sites, online application stores, online marketplaces and aggregators providing services such as mobility aggregation, food ordering, food delivery services and match-making.

95 See: https://www.financialexpress.com/life/technology-cci-seeks-to-evaluate-ai-impact-on-markets-3465121, last accessed June 24, 2024.

Way Forward

As corporations and governments increasingly embed AI in their products, processes, and decision-making systems, concerns are being raised over how this data is being used: especially by complex algorithms in, say, diagnosing cancer, disbursing loans, rating employee performance, etc. Apart from sector/use-case-specific concerns, overall concerns regarding bias, transparency, ownership, accountability and due process in AI systems would also need to be tackled.[1]

Moreover, social concerns about potential job losses and unemployment due to the increasing demand for GAI technology cannot be ignored.[2] However, if viewed from a different lens, this fear can also act as a catalyst for reskilling and upskilling the workforce to adapt to the evolving job market. While AI and automation may replace some jobs, they also generate new opportunities for skilled professionals who can design, develop, maintain, and operate these technological systems.

On the other hand, consumer-facing GAI empowers laypersons with unprecedented abilities. While this opens up the possibility of a different kind of creativity, it also leads to concerns regarding the creation of misinformation and the devaluation of intellectual property generally. These issues are further exacerbated by the fact that both end users and regulators may not understand the functioning of each distinct GAI — making regulation challenging.

The obscurity and inscrutability of GAI also make it difficult to determine what is at stake and to frame regulations accordingly– as is the case with most emerging technologies. While countries and supra-national bodies such as the IEEE and OECD have come up with strategy papers, vision documents, and ethical guidelines — *all non-binding*, the European Union has formulated an AI law that is risk-based. This could pave the way for significant future jurisprudence.

Given the significant impact of GAI regulation on all aspects of humanity, regulators across the world should adopt an open and consultative approach where stakeholders' views are encouraged and meaningfully considered. This is especially important since this space is highly dynamic and conventional positions on legal, policy or hastened ad hoc solutions may not stand the test of time. It is important to not regulate too early, or too late, while also balancing the degree of regulation. In such a situation, self-regulation by the industry could be a viable stop-gap solution. The industry may be best placed to understand the challenges that need to be addressed while being mindful of over-regulation.

In the long run, however, considering the unique challenges posed by GAI, the right combination of technical tools, accountability, industry self-regulation, and sectoral laws seems like the way forward.

1 Malte Ziewitz, Governing Algorithms: Myth, Mess, and Methods, Science, Technology & Human Values, 2016 Vol 4(I) pp. 3-16.

2 Generative AI set to affect 300mn jobs across major economies, available at: https://www.ft.com/content/7dec4483-ad34-4007-bb3a-7ac925643999, last accessed June 24, 2024.

About NDA

At Nishith Desai Associates, we have earned the reputation of being Asia's most Innovative Law Firm — and the go-to specialists for companies around the world, looking to conduct businesses in India and for Indian companies considering business expansion abroad. In fact, we have conceptualized and created a state-of-the-art Blue Sky Thinking and Research Campus, Imaginarium Aligunjan, an international institution dedicated to designing a premeditated future with an embedded strategic foresight capability.

We are a research and strategy driven international firm with offices in Mumbai, Palo Alto (Silicon Valley), Bengaluru, Singapore, New Delhi, Munich, and New York. Our team comprises of specialists who provide strategic advice on legal, regulatory, and tax related matters in an integrated manner basis key insights carefully culled from the allied industries.

As an active participant in shaping India's regulatory environment, we at NDA, have the expertise and more importantly — the VISION — to navigate its complexities. Our ongoing endeavors in conducting and facilitating original research in emerging areas of law has helped us develop unparalleled proficiency to anticipate legal obstacles, mitigate potential risks and identify new opportunities for our clients on a global scale. Simply put, for conglomerates looking to conduct business in the subcontinent, NDA takes the uncertainty out of new frontiers.

As a firm of doyens, we pride ourselves in working with select clients within select verticals on complex matters. Our forte lies in providing innovative and strategic advice in futuristic areas of law such as those relating to Blockchain and virtual currencies, Internet of Things (IOT), Aviation, Artificial Intelligence, Privatization of Outer Space, Drones, Robotics, Virtual Reality, Ed-Tech, Med-Tech and Medical Devices and Nanotechnology with our key clientele comprising of marquee Fortune 500 corporations.

The firm has been consistently ranked as one of the Most Innovative Law Firms, across the globe. In fact, NDA has been the proud recipient of the Financial Times–RSG award 4 times in a row, (2014-2017) as the Most Innovative Indian Law Firm.

We are a trust based, non-hierarchical, democratic organization that leverages research and knowledge to deliver extraordinary value to our clients. Datum, our unique employer proposition has been developed into a global case study, aptly titled 'Management by Trust in a Democratic Enterprise,' published by John Wiley & Sons, USA.

Research@NDA

Research is the DNA of NDA. In early 1980s, our firm emerged from an extensive, and then pioneering, research by Nishith M. Desai on the taxation of cross-border transactions. The research book written by him provided the foundation for our international tax practice. Since then, we have relied upon research to be the cornerstone of our practice development. Today, research is fully ingrained in the firm's culture.

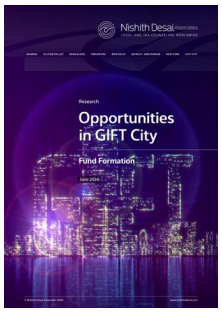
Over the years, we have produced some outstanding research papers, reports and articles. Almost on a daily basis, we analyze and offer our perspective on latest legal developments through our "Hotlines". These Hotlines provide immediate awareness and quick reference, and have been eagerly received. We also provide expanded commentary on issues through detailed articles for publication in newspapers and periodicals for dissemination to wider audience. Our NDA Labs dissect and analyze a published, distinctive legal transaction using multiple lenses and offer various perspectives, including some even overlooked by the executors of the transaction. We regularly write extensive research papers and disseminate them through our website. Our ThinkTank discourses on Taxation of eCommerce, Arbitration, and Direct Tax Code have been widely acknowledged.

As we continue to grow through our research-based approach, we now have established an exclusive four-acre, state-of-the-art research center, just a 45-minute ferry ride from Mumbai but in the middle of verdant hills of reclusive Alibaug-Raigadh district. Imaginarium AliGunjan is a platform for creative thinking; an apolitical ecosystem that connects multi-disciplinary threads of ideas, innovation and imagination. Designed to inspire 'blue sky' thinking, research, exploration and synthesis, reflections and communication, it aims to bring in wholeness — that leads to answers to the biggest challenges of our time and beyond. It seeks to be a bridge that connects the futuristic advancements of diverse disciplines. It offers a space, both virtually and literally, for integration and synthesis of knowhow and innovation from various streams and serves as a dais to internationally renowned professionals to share their expertise and experience with our associates and select clients.

We would love to hear from you about any suggestions you may have on our research publications. Please feel free to contact us at **research@nishithdesai.com**.

Recent Research Papers

Extensive knowledge gained through our original research is a source of our expertise.



June 2024

Opportunities in GIFT City

Fund Formation

June 2024

From Capital to Impact: Role of Blended Finance

Legal and Regulatory Framework

April 2024

Third-Party Funding for Dispute Resolution in India

Exploring Recent Developments and the Legal Landscape

February 2024

Tracking NFTs from Code to Court

Legal Considerations and Disputes

December 2023

Prevention of Sexual Harassment at the Workplace (POSH)

India Legal & HR Considerations

December 2023

The Indian Defence Industry

Redefining Frontiers

For more research papers **click here**.

Nishith Desai Associates
LEGAL AND TAX COUNSELING WORLDWIDE

MUMBAI

93 B, Mittal Court, Nariman Point
Mumbai 400 021, India

Tel +91 22 6669 5000

SILICON VALLEY

220 S California Ave., Suite 201
Palo Alto, California 94306, USA

Tel +1 650 325 7100

BENGALURU

Prestige Loka, G01, 7/1 Brunton Rd
Bengaluru 560 025, India

Tel +91 80 6693 5000

SINGAPORE

Level 24, CapitaGreen
138 Market St
Singapore 048 946

Tel +65 6550 9855

MUMBAI BKC

3, North Avenue, Maker Maxity
Bandra–Kurla Complex
Mumbai 400 051, India

Tel +91 22 6159 5000

NEW DELHI

13-H, Hansalaya Building, 15
Barakhamba Road, Connaught Place
New Delhi 110 001, India

Tel +91 11 4906 5000

NEW YORK

1185 6th Avenue, Suite 326
New York, NY 10036, USA

Tel +1 212 464 7050

GIFT CITY

408, 4th Floor, Pragya Towers
GIFT City, Gandhinagar
Gujarat 382 355, India

Evolution of Generative AI
Legal, Regulatory and Ethical Frontiers

concierge@nishithdesai.com