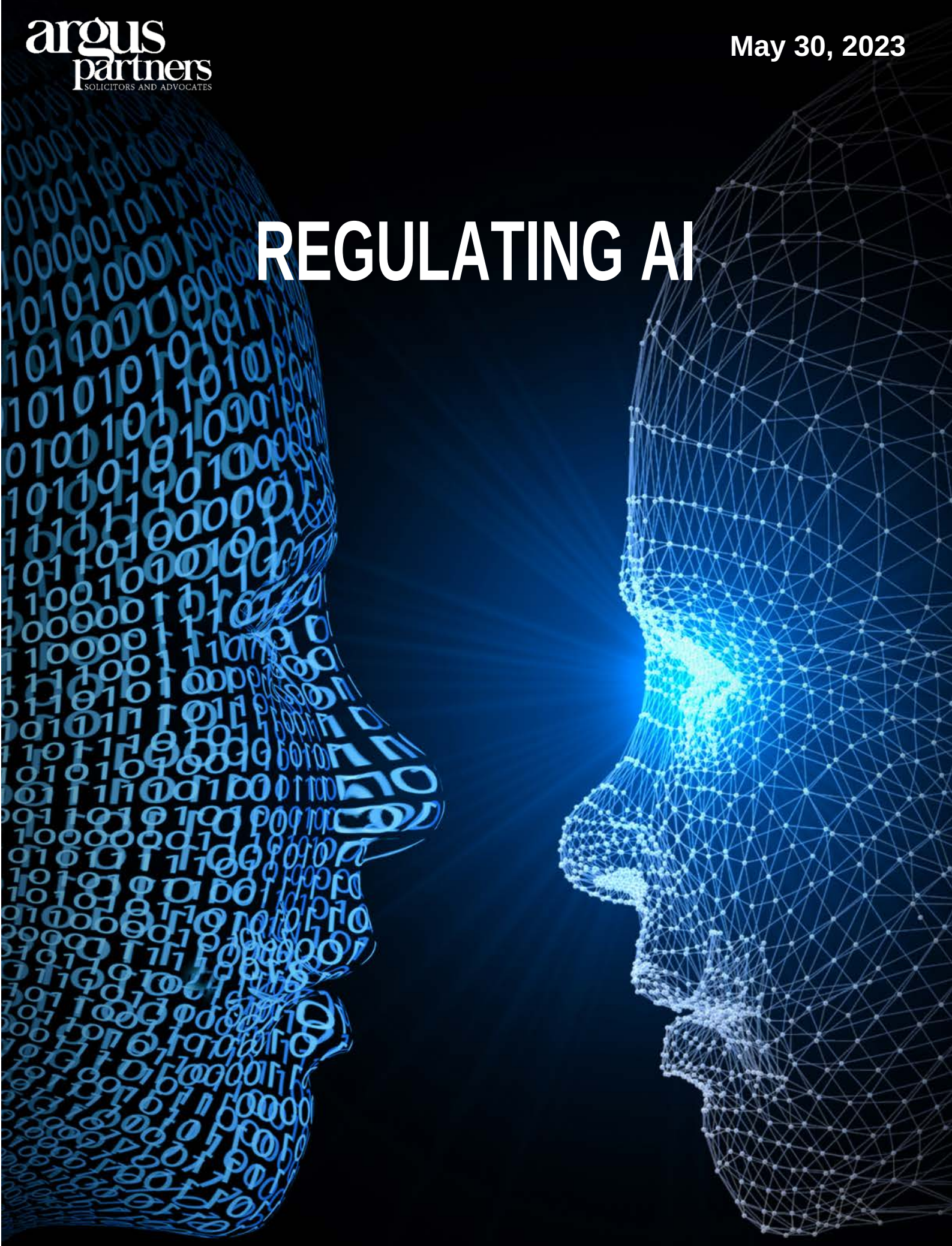
argus partners
SOLICITORS AND ADVOCATES

May 30, 2023

REGULATING AI



MUMBAI I DELHI I BENGALURU I KOLKATA

Table of Contents

I. Introduction

It is undeniable that over the last 2 (two) decades or so, technology has taken a huge leap forward. This is even more true in the case of personal technology, or gadgets, that we use in our daily life for everything, from work to social interactions, consuming news, shopping, hailing rides, and everything in between.

For a long time, the technology industry has been largely self-regulated, with minimal scrutiny from regulators or oversight by government bodies. Legislators have also been slow in enacting laws, and it is often the case that, technological advancements are racing ahead while regulation or legislation is playing catch up.

In the last few years though, with the advent of personal technology and emergence of the gig economy permeating every facet of our lives, there has been an increased focus on regulating the tech industry. Governments around the world have sought to introduce legislation to rein in 'big tech', concerned largely about user privacy, misuse of personal data and competition.

While the European Union (EU) enacted the General Data Protection Regulation (GDPR) (as it is popularly known) back in 2018 and even China introduced the Personal Information Protection Law last year, India is yet to introduce a comprehensive legislation governing the use and processing of data of Indian citizens by large tech companies. The United States, too, lacks a comprehensive federal enactment governing the field.

Even as data protection and privacy related laws are still a work in progress, new technological advancements, and the sudden rise of and interest in artificial intelligence (AI), poses fresh challenges for regulators globally. As disruptive technology like ChatGPT wows the world, and similar offerings from tech giants like Microsoft and Google are now available for use by individuals, is it perhaps time to think of regulating large language models to ensure responsible use of AI?

This article examines the potential risks associated with AI based large language models and makes recommendations for regulatory intervention.

II. What are Language Models (LMs)?

LMs are advanced computer programs that are designed to understand natural human language and generate responses which are human-like. LMs are trained to learn patterns and relationships between words and phrases in natural human language. In order to do this, LMs study and analyse a given set of written work, which could include anything from books to online articles. These written works form the training data set for the LMs and determine their capabilities and limits.

Data is crucial to training LMs. More the data, the better the LM. LMs like ChatGPT and Bard are called 'large language models' because they are trained on vast amounts of data which can include billions of examples of written text. This is typically combined with human supervision during training, where inaccuracies or inefficient responses are corrected manually and the LM is fine-tuned to give better responses. This process also allows the developers to prevent the LM from giving problematic or harmful responses and to enhance accuracy and effectiveness.

The data set that the LM is trained on, is decided by the developers, based on the objective sought to be achieved by the LM. Let's assume an LM is being trained to be used in hospitals. Necessarily, the training data set would include books, articles and written text which concern medical science. Similarly, LMs designed for assisting travellers at an airport, would be trained on written data regarding airport rules and regulations, flight information, etc.

Large LMs like ChatGPT and Bard, are trained on a diverse range of texts from different sources, including billions of web pages and documents concerning various topics. This data set is constantly updated and expanded, allowing the models to keep up with the latest developments in

various fields and gain an improved understanding of human language. Further, interactions with human users also contribute to the LMs' learning and development process.

III. Risks posed by large LMs

Reportedly, over 100-million people today use ChatGPT, for a variety of purposes. While in the beginning, most users may have had some light-hearted fun using the computer programme, increasingly, users of large LMs are finding innovative ways to deploy such computer programmes to assist them in daily tasks like writing emails or letters, creating itineraries, research and even preparing customized resumes and writing computer code.

The sheer range of tasks that large LMs can assist with and the accuracy and effectiveness they already demonstrate, makes them extremely appealing to (at the least) be utilized to complement human effort, if not substitute human involvement entirely. However, use of this technology is not without its risks and may have potential downsides which are discussed below:

A. AI hallucinations and misinformation

A critical issue with any LM is that, unlike humans, it cannot comprehend the meaning of the words and phrases input by a user, or its own response. Rather, it recognises the pattern and sequence of words and phrases used in the input and relies on its training/ data set to generate a response, matching the sequence recognised in the input. This is similar to predictive text/ auto-correct features in smartphone keyboard software that learn the sequence of words in your text messages and use this information to make suggestions and corrections the next time you are typing a similar message.

This lack of comprehension in LMs creates a phenomenon known as 'AI hallucinations'. Since the data set used to train large LMs, albeit vast, is limited, it cannot possibly have responses to every imaginable query. It is limited by its own knowledge. However, since it cannot comprehend its own responses, the LM is unable to determine its limits in responding to certain queries. Instead, analysing the user input, it generates the best possible response from the insufficient data set available to it. This puts the user at risk of receiving false or misleading information. Additionally, even the limited data set available to the LM may consist of false, misleading and biased information, further aggravating the problem. Regardless, large LMs typically present this information in a self-assured manner, as though the response is accurate

B. Potential for bias

As discussed above, responses generated by LMs are limited by the data set on which it is trained. Hence, if the data set includes books and articles containing biased information, naturally, the responses generated will reflect the same bias. While this is not intentional, it is inevitable, since no data set could possibly be free from all forms of bias. The LM cannot critically examine the data set it is trained on, since it lacks comprehension and is hence incapable of eliminating the bias without some form of human intervention.

For example, let`s say an LM is trained on 100 books on the history of India: 70 of them authored by western authors and 30 by Indian authors. It may be likely that the LM's responses will lean toward the western narrative on the history of India. If we expand this to a large LM, trained potentially on billions of books, articles, web pages, etc., it would be difficult to identify the parameters responsible for the bias, unlike in the simple example above. This issue (combined with the sheer complexity of the internal functioning of AI algorithms) has led to the 'black box problem', associated with large LMs. The black box problem refers to our lack of understanding of the actual decision-making process applied by AI systems to arrive at the conclusions reflected in their responses. Simply put, we have no idea how an AI makes its decisions. Consequently, if there is a bias detected in the LMs'

response, we are unable to determine *how* the bias came to be, what is the data being relied on, and which data has been ignored, if at all.

The remedy then, is *post facto* correction of the bias in the LMs' responses. Nevertheless, some biases will inevitably remain, because they may not be easily identifiable or obvious in any particular response *per se*. While explicit bias can be controlled through human intervention, more subtle or normalised bias may pass undetected.

Biased responses may lead to several negative consequences for the users interacting with LMs. Regular usage and eventual day-to-day dependence of users on LMs presenting deep seated but subtle biases, can reinforce stereotypes and even influence personal opinions based on inaccurate information. While the LMs' biases are in fact only a reflection of the prevalent or past societal biases, the ease of usage and mass access to large LMs also escalates the ease of reinforcement of these biases.

C. Potential to aid illegal activity

Like the internet, large LMs can be abused to gain access to content which can aid in commission of illegal acts. But the ease and speed with which LMs could make such harmful content available, in a simple to understand manner, poses an aggravated risk. A commonly highlighted issue is the use of ChatGPT's software coding abilities to create software tools for phishing and scamming. While large LM developers are taking steps to prevent such problematic responses, it is a difficult task. On one hand, while it is difficult to identify which content could be problematic or may have the potential for misuse, on the other, owing to the black box problem, it may be impossible for developers to foresee how the LM would respond to varied users' inputs.

D. Copyright infringement and plagiarism

A significant portion of the data set used in training large LMs like ChatGPT and Bard, may be copyrighted written works. The authors' permission is not sought for this usage, but the collective content of all their works is key to the responses generated by LMs, thus raising concerns around infringement of copyright.

As discussed, LMs cannot comprehend the meaning conveyed by language, but generate responses based on learnt patterns and relationships between words and phrases. Large LMs' responses do not simply copy-paste material from any individual written work. Rather, the sequence of words and phrases identified in written works by multiple authors, on any given topic, are reorganised or re-sequenced by large LMs, to predict a response.

Thus, while an LM's responses may not technically infringe individuals' copyright, and may satisfy the test of fair usage under existing copyright laws, nevertheless, the responses so generated are a nuanced reproduction of the collective works of several copyright holders and hence, a form of infringement of intellectual property rights (IPR). The responses also lack any effort, either to verify or to expand on the information and language already present in the LMs' data set. To compound the problem, large LMs like ChatGPT and Bard do not attribute any credit to the authors they borrow from and may hence be perceived as unethical or guilty of plagiarism.

Adding to all this, these large LMs are necessarily intended to be monetized. For instance, ChatGPT now has a paid subscription model to access its latest features. This means that the developers of ChatGPT are moving toward monetization of their service, while the commercial value of the original works, forming the underlying data set, may be reduced over time, as more and more people turn to the simplified, summarised presentations of these works.

E. Data Privacy

Data, especially big data or large amounts of personal data pertaining to individuals, has always raised privacy concerns. It is possible, or maybe even likely, that in case of large LMs, the vast data set on which such LM is trained, may include personal data of individuals, thus, raising privacy concerns. Some countries like Italy had initially banned ChatGPT owing to its lack of transparency with regard to the data being used for training the LM. This lack of transparency in the training data set, combined with the opaque functioning of large LMs, heightens privacy concerns.

Even where the training data set does not include any direct access to private information, users themselves may divulge personal and sensitive information about themselves or others, during conversations with LMs. Studies show that large LMs have the potential to be manipulated to unintentionally divulge personal information. This apart, large LM algorithms are susceptible to data breach and hacking, putting even users' personal conversations with large LMs at risk.

F. Risk to human jobs and livelihood

Popular media is rife with this one — AI replacing human jobs. AI technology, including large LMs, are already being used in content creation, software coding, customer service, and even translation services. Almost everyone agrees that the current pace of advancements in AI technology will eventually reduce the role of human staff in several, if not most, jobs. How immediate a threat does AI pose to human jobs, or which sectors are most vulnerable, is unpredictable. With time, large LMs could be used in several sectors dealing with complex issues, including medical science, engineering, etc.

If history and the present is any indication, the law will probably be left playing catch-up, unless policy makers and governments get serious about placing well thought out regulations to ensure a smooth transition to an AI-driven economy and human life.

IV. Recommendations for Regulations

During a recent US congressional hearing, Sam Altman (CEO of OpenAI - developer of ChatGPT), highlighted several risks posed by AI and urged the US Congress to regulate large LMs and AI without delay. Although, earlier this year, India's IT minister made a statement in parliament that the government does not plan to regulate large LMs or other AI technology, recent reports suggest that the much-awaited Digital India Act will also seek to regulate emerging technologies, including AI.

On the other hand, the EU and China have already made significant strides toward regulating AI. The EU has proposed the enactment of the Artificial Intelligence Act (AIA), which seeks to categorise and regulate AI systems based on the level of risks they pose. China has also recently published its draft regulations, proposing strong measures around accountability of developers and the training data, aimed at accuracy, objectiveness and protection of IPR, while Russia has resorted to a knee-jerk measure of prohibiting the use of large LMs altogether.

While in most other countries, any form of regulation remains lacking, the broad consensus is that large LMs pose significant risks which need immediate mitigation. Following are some recommendations to be considered for framing a comprehensive regulation for large LMs in India:

A. Independent Regulatory Authority

To comprehensively understand and regulate this nascent technology and mitigate the risks posed, it is imperative to set up an independent statutory regulatory authority to oversee the development and deployment of AI. Such authority may be responsible for framing

regulations and safety standards/ tests to govern AI and also be vested with quasi-judicial powers to determine non-compliances, ordering remedial measures and even impose penalties. The powers and functions of this authority may be similar to that of the Competition Commission of India or the Insolvency and Bankruptcy Board of India. Further, considering the complex and fast-evolving technological space this authority will regulate, it must necessarily comprise domain-level experts in the field of AI, along-with legal experts who will be instrumental in the adjudicatory and rule-making functions of the authority.

B. Research and Development

AI has incredible potential to revolutionize human lives. Thus, regulation must also be aimed at facilitating the development of AI, rather than imposing unnecessary constraints on such development. Responsible development of AI must ensure a secured platform for testing of and researching on large LMs and advanced AI systems, in the form of an AI sandbox. In other words, while in R&D phase, advanced AI systems must be prohibited from being made available to the public at large. In addition to this, regulation should also require developers of advanced forms of AI systems to make appropriate disclosures to regulators at regular intervals, regarding the design and purpose of the AI systems and progress made toward their development.

Such an approach to R&D may be mandated by way of regulation, so as to allow regulators to keep a live tab on the latest developments in the field of large LMs, enabling them to amend and finetune the regulations in advance, rather than playing catch-up.

C. Compliance Reporting and Licensing

Considering the risks associated with large LMs, it is important that public access to these LMs is strictly regulated. Post the R&D phase of any large LM and before public deployment, the developer must be mandated to make a declaration of compliance with extant regulation governing the AI model. This may be achieved by mandating self-filing of a compliance report by the developer, in prescribed format, demonstrating steps taken to ensure compliance with regulations. For instance, if regulation requires the training data to be free of misinformation, the developer in the compliance report would be required to make an affirmative statement that its LM has been trained on a verified data set, which is free from false or misleading information and must also disclose the steps taken during the R&D process to ensure this. Such a compliance report may also be instrumental in determining any alleged future infractions in respect of the given LM.

In addition to this, regulation may also consider a licensing requirement before deployment of large LMs, based on certain identified and objective criteria. However, any proposal for a licensing requirement must be backed with a robust and effective mechanism to ensure that technological advancements don't suffer on account of delays and entry barriers.

D. Mandatory Personnel Requirement

Similar to recently introduced requirements for social media intermediaries under the amended intermediary guidelines, regulation should require AI developers to appoint specific officers to discharge certain roles. This may include a (i) Chief Compliance Officer, responsible for overall compliance with extant law; (ii) Nodal contact person, responsible for coordination with law enforcement agencies and regulatory authority; and (iii) Resident Grievance Officer, responsible for receiving and addressing complaints/ concerns from users.

In addition to this, regulation may also consider mandating formation of an internal committee in the nature of an ethics board, for collective decision making on issues concerning ethical

challenges. This has already been voluntarily and widely adopted in the tech industry, even in the absence of regulation.

E. Regulating Training Data

Training data, being key to responses generated by large LMs, requires it to be regulated and meet certain thresholds set by law, to minimise some of the risks highlighted above.

1. **Exclusion of personal data** - In line with the definition of 'personal data' under the proposed Digital Personal Data Protection Bill, 2022, inclusion of personal data of individuals should be prohibited to be used in the training data set.

2. **Verification of the training data** - Even though, admittedly, a subjective concept, verification of training data may perhaps be critical to address many of the risks highlighted above. It is important to have in place broad principles governing the underlying training data, such as requirements to ensure that the training data is:

    i. free of false and misleading information;
    ii. objective, and free from overt bias or prejudice; and
    iii. not harmful or capable of being utilized to cause harm or assist in illegal activity.

In addition to being a subjective exercise, verification of training data will also necessarily be a laborious task and considered onerous by AI developers. Nevertheless, such verification requirements are critical to ensure that responses generated by large LMs are within regulatory guardrails and meet certain minimum standards of objectivity, accuracy and truthfulness.

At the same time, regulation must consider the very subjective nature of the broad principles coined above and ensure that any proposed consequences for non-compliance are attracted contextually, keeping in mind reasonableness and on a case-to-case basis.

F. Regulating User Interface

The user interface, or UI, of large LMs is the window through which users interact with the LMs. In other words, the webpage displaying the chatbox and the LMs' responses, links to other features available to the users, all together forms the UI. This also includes the manner in which the generated responses are presented to the user.

Regulating aspects concerning the UI of large LMs will perhaps go a long way in educating and informing users, and aid in the overall responsible use of large LMs.

1. **Disclaimer/ disclosure requirement** - Regulation should mandate that large LMs, as part of their UI, must conspicuously display at all times a statement informing and reminding users of certain critical aspects to bear in mind while using an LM. Such statement must disclose that:

    i. The user is interacting with a chatbot/ LM, and not a human being;
    ii. The responses generated are limited by the training data set, and/ or if the LM has access to the internet;
    iii. The responses generated may at times be incorrect, harmful or biased, and advising users to verify the accuracy of the responses;
    iv. The users may not rely on the information solely; and whether the developers take responsibility for the accuracy of the content generated.

2. **Citation requirement** - Large LMs deployed for use by the general public must be mandated to provide citations and reference to the original work(s) based on which a

response is generated. For example, Microsoft's LM — Bing, unlike OpenAI's ChatGPT, presents its responses along with citations or links to the web articles relied on by it to generate the response. This would ensure that due credit is given to human authors who may have penned the original work, thus addressing copyright and plagiarism concerns, while also providing the user a handy option to verify the LM's response.

In addition to this, regulation may also consider mandating large LMs to provide links to original works, where users can purchase such works. Such a requirement may aid in the promotion and commercial sale of such original works.

G. Barring use of AI in certain tasks

Given the current state of large LMs, which lack comprehension, have potential for bias and the black box issue, it may be advisable, for now, to restrict or prohibit the use of such advanced AI systems in tasks requiring human discretion.

Regulators must endeavour to exhaustively identify tasks which may be categorised as high risk and therefore, cannot be delegated to AI systems. Any human decision making that determines or impacts an individual's rights, may not be allowed to be performed by AI systems. For instance, tasks such as judicial decision making, deciding on potential job applications or increments or promotions, approval/ denial of grant of loans or other forms of financial assistance, to name a few.

H. Accountability

Regulation must enable a mechanism to hold developers/ their officers accountable, in cases of real-world harm. Non-compliance with extant regulations must also attract appropriate penalties. While accountability and remedial action are vital to effective regulation, to facilitate and promote the growth of the AI industry, a regulatory regime, perhaps along the lines of the 'safe harbour protection' afforded to social media intermediaries under the Information Technology Act, 2000, may be envisaged for developers of advanced AI systems as well.

V. Conclusion

Artificial Intelligence is here. The pace at which the technology is advancing and being adopted is unprecedented. Learning from past mistakes, regulators must now stand up to the challenge of ensuring responsible use of AI and harnessing the technology for the good. Timely, appropriate regulatory intervention is key to addressing the concerns surrounding AI technology. While developers of large LMs like ChatGPT and Bard are already taking voluntary measures to mitigate these concerns, primarily owing to public scrutiny and criticism, governmental regulations will play a key role in ensuring safe usage of AI technologies like large LMs.

AI is bound to impact almost every facet of life as we know it today. Advanced AI technology, augmenting the human users having access to it, has the potential to create AI haves and have-nots, creating a further, artificial divide amongst populations. As such, lawmakers must also separately address the issue of accessibility and reach, along with the other emerging concerns with AI technology.

One can also not ignore the multiple statements issued by several leading experts in the field of AI, publicly warning of existential threats that advanced AI systems pose to humanity. Several tech industry leaders have gone so far as to call for a moratorium on further research and development of AI, till such time as guardrails are put in place. Does it not then beg the question, if it's time to regulate AI *now*?

Contributed by:

Udit Mendiratta, Partner

Tejas Jha, Associate

DISCLAIMER

You can send us your comments at:

knowledgecentre@argus-p.com

Mumbai I Delhi I Bengaluru I Kolkata

www.argus-p.com

Recent Papers/ Articles

May 2023

Navigating Crossroads Of IBC And RERA

Corporate Restructuring & Insolvency

May 2023

Is Online Rummy Chancier Than An Offline Round?

Technology and Data Privacy

April 2023

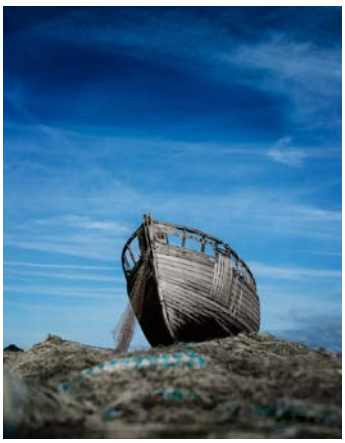
The Digital Personal Data Protection Bill, 2022 – An Analysis

Technology and Data Privacy

April 2023

Rera Regime - The Exemption Conundrum

Real Estate



April 2023

Changes to the Merger Control Regime in India

Corporate and M&A

April 2023

Fintech Primer - II

Technology and Data Privacy

For more Papers/ Articles click here.

MUMBAI
11, Free Press House
215, Nariman Point
Mumbai 400021
T: +91 22 6736 2222

DELHI
Express Building
9-10, Bahadurshah Zafar Marg
New Delhi 110002
T: +91 11 2370 1284/5/7

DELHI
155, ESC House, 2nd floor,
Okhla Industrial Estate, Phase 3,
New Delhi – 110020
T: +91 11 45062522

KOLKATA
Binoy Bhavan
3rd Floor, 27B Camac Street
Kolkata 700016
T: +91 33 40650155/56

BENGALURU
68 Nandidurga Road
Jayamahal Extension
Bengaluru 560046
T: +91 80 46462300

www.argus-p.com I communications@argus-p.com