Data Protection and Generative AI – Policy, Regulation, and the Way Forward

Dr. Stanley Lai, Afzal Ali and Kan Jie Marcus Ho
Allen & Gledhill LLP Singapore

The proliferation of Artificial Intelligence ("AI") has led to paradigm shifts in the context of innovation. With rapid advancement in technology in the past twenty to thirty years, large swathes of data were being generated, collected, and used. It was quickly recognised that this affected all facets of society, and that rules and regulations were urgently required to prevent the unfettered flow and (mis)use of data. Examples of such regulations included the groundbreaking General Data Protection Regulation ("GDPR"), and Singapore's Personal Data Protection Act ("PDPA"). However, just over a decade after the enactment of such rules and regulations, another paradigm shift is on the horizon. Artificial intelligence and generative intelligence are radically transforming how data can be interpreted, used, and presented. It has validly been pointed out that such generative artificial intelligence could bring forth a new epoch of data synthesis and augmentation. This paper discusses how policy and regulations can work to address issues surrounding the use of input data, which is critical to generative AI. Specifically, it will examine whether input data should be considered "personal data" and thus caught by the GDPR or Singapore's PDPA; whether there is a recourse for emotional harm caused by content generated using such data. It will also discuss some of the current limitations and gaps that exist in the current regulatory framework. It is hoped that this discourse will further the continuing dialogue on the intersection between data protection and artificial intelligence, particularly in the domain of Generative AI and Data Protection.

Keywords: Data Protection, General Data Protection Regulation, Comparative Law, Technology Law, Singapore Law

Introduction

The proliferation of Artificial Intelligence ("**AI**") has arguably led to paradigm shifts in the context of innovation, with technologies such as generative AI, machine learning, and cloud computing becoming increasingly pivotal for businesses and organisations. Indeed, Klaus Schwab, Executive Chairman of the World Economic Forum, has persuasively argued that we are now in the fourth industrial revolution, where *"fusion of technologies"* have blurred the lines between the real, digital, and living worlds (Schwab, 2016). With the first industrial revolution having been mainly powered by water and steam power,[10] the second with electricity (Schwab, 2016), and the third with computers and gadgets,[11] it has now emerged, as famously predicted by mathematician Clive Humbly, that "data" – this broad catch-all description for all types of information capable of being stored – is the oil driving the fourth industrial revolution (Charles, 2013).

Historically, the common law has not regarded information as property (*Phipps v Boardman*, 1967). While this position may with time shift, it is fair to say that data is now seen and accepted as having tremendous value. With rapid advancement in technology over the last twenty to thirty years, large swathes of data were being generated, collected, and used. It was quickly recognised that this affected all facets of society, and that rules and regulations were urgently required to prevent the unfettered flow and (mis)use of data. One key legislation which emerged was the General Data Protection Regulation ("**GDPR**"), an overarching data legislation governing the European Union (EU) which sought to provide a comprehensive reform of the existing rules, which was adopted at a time when the internet was still in its infancy (EDPS, *n.d.*). According to the European Data Protection Supervisor, this legislation was needed given that over the last 25 years, technology has transformed our lives in ways nobody could have imagined, and hence a

[10] *Ibid.*

[11] Sakhapov & Absalymova, *Fourth Industrial Revolution and the Paradigm Change in Engineering Education*, MATEC Web of Conferences, 245, 12003 (2018).

review of the rules was needed (*Phipps v Boardman*, 1967). Simply put, the primary purpose of the GDPR was to grant individuals substantive rights in relation to and over their personal data. This was critical at a time when corporations were increasingly unlocking the value of personal data with little or no regulation. Hence, as Hoofnagle, Sloot & Borgesius rightly note, the GDPR *"attempts to put privacy on par with the laws that companies take seriously"*. Indeed, prior to this regulation, it was highlighted that large data companies faced low fines, with there being almost no deterrent effect for the unfettered use of personal data, thereby leading to an imbalance in power (Hoofnagle, Van der Sloot, and Borgesius, 2019). This finally changed following the GDPR's enactment.

Singapore was no different. It recognised that personal data about an individual stood on a different footing from other types of data. The Personal Data Protection Act 2012 ("**PDPA**") was thus enacted to provide a baseline standard of protection for personal data in Singapore. This is the central legislation in Singapore that governs the collection, use, and disclosure of individuals' personal data by organisations (Personal Data Protection Commission, *n.d.*). Chik rightly highlights that this legislation is timely, as *"the digital era poses increasingly greater challenges to the integrity of informational privacy for many reasons"* (Chik, 2013). Following the enactment of this legislation, non-complying organisations risk facing regulatory sanction as well as private civil action should they not handle personal data properly, with the due care that is required as set out in the PDPA.

Just over 10 years after the PDPA's enactment, another paradigm shift is on the horizon. Artificial intelligence and generative intelligence are radically transforming how data can be interpreted, used, and presented. As has been pointed out elsewhere, even within the field of artificial intelligence, the shift has been explicitly evident, with the function of AI having moved from simply analysing expansive datasets to actively generating innovative content (Du *et al.*, 2023). Indeed, consider AlphaGo's victory over the Go world champion in 2016 (Vincent, 2019). to ChatGPT's advanced conversational capabilities,[12] to the creative limits of Midjourney, whose artwork "Théâtre D'opéra Spatial" won the Colorado State Fair (Metz, 2022). These trends will only continue as the full limits of AI are explored. It has validly been pointed out that such generative artificial intelligence could bring forth a new epoch of data synthesis and augmentation, predictive analysis and management, and personalised user interaction (Metz, 2022), which brings in new unique opportunities and challenges for the world ahead.

Generally speaking, generative AI works by using neural networks to identify the patterns and structures within existing data to generate new and original content (NVIDIA, *n.d.*). Through learning the patterns and the structure of their input training data, generative AI tools are able to generate "new data" with similar characteristics AI (Verify, 2024). It is not the purpose of this article to explore the nuances in such training models or the future of generative AI. Instead, the purpose is a more modest one in examining whether existing data protection law is fit for purpose.

It is immediately evident that the use of data to train AI may engender potentially thorny legal issues. Take, for example, a situation where input training data is used to generate defamatory content, which then causes emotional distress. What duties do such AI companies owe to data subjects, if any? When does input data cease to become "personal data" (and consequently fall outside the remit of the PDPA?) Ye, Yan, Li, & Jiang (2024) opine that the rapid development of generative AI has arguably increased personal data risks, particularly in the context of AI pre-training. This is because generative AI consumes vast amounts of personal data while operating in a "black box." Yet, personal data is needed to complete the deep learning procedures that are required for the AI to gain its full potential. The use of such personal data, therefore, attracts scrutiny under the GDPR and PDPA, which were not specifically enacted with Generative AI in mind.

[12] ChatGPT, OpenAI, GPT-4 is OpenAI's most advanced system, producing safer and more useful responses.

In this regard, this article seeks to address some of these questions by examining existing GDPR regulations as well as provisions in the Singapore Personal Data Protection Act in an attempt to identify the possible lacunas in the evolving world of data protection law. The authors hope that this will further the continuing dialogue on the intersection between data protection and artificial intelligence, particularly Generative AI.

1. AI in the context of the GDPR and Singapore's PDPA

Ye *et al.* highlight that OpenAI uses three primary classes of data to train ChatGPT: data that is publicly available on the internet, data that it licenses from third parties, and data from its users or its human trainers. Although conversations with generative AI may not "overtly include direct identifiers like real name or phone numbers," they touch upon user's life experiences, work status, as well as recent thoughts, which can potentially reveal one's identity (Ye, *et al.,* 2024). These issues are best illustrated with the following hypothetical. Take John, an individual who enters his personal data as input into the ChatGPT system, and information relating to his own personal life, such as his hobbies, this being the fact that he likes to play the trumpet, and he then uploads a photo of himself onto the AI input system. John thinks that only he can access the data about himself. But what he does not know is that his data enters the open pool of training data. Thereafter, another anonymous user prompts ChatGPT to generate a funny drawing of a man playing the trumpet – leading to ChatGPT generating a relatively playful take on John's own image and going to the extent of naming this image John when prompted by another user. This is then published online, leading the real John to suffer emotional distress. Who, in this case, should be liable, if at all? What exactly is the personal data that is involved? Is generated data that is "inferred" by the AI considered personal data as well?

2. Is Inferred Personal Data Personal Data?

To answer this question, it would be appropriate to begin with the definition of personal data in the GDPR before examining specific provisions in the PDPA. Article 4(1) forms the definition of personal data, which reads that – (European Parliament and Council, 2016).

"Personal data" means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier to one or more factors specific to the physical, psychological genetic, mental, economic, cultural, or social identity of that natural persons".

From the definition, a key plank of personal data involves the concept of *identifiability*. Stated briefly, identifiability is about the conditions in which a set of data – even if ***not*** linked to a person – is still considered as personal data because it is possible to identify a person from existing data. In this regard, Recital (26) provides further guidance, highlighting that the objective factors which one should consider would be the costs and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments (European Parliament and Council, 2016).

Similarly, in the context of the PDPA, *s* 2 defines personal data as data, whether true or not, about an individual who can be identified – (a) from that data or (b) from that data and other information which the organisation has or is likely to have access to (Government of Singapore, 2020).

This concept of identifiability may be difficult to apply in the context of Generative AI. Given that generative AI systems are often trained by large data sets, including the input data that can be personal data, the issue that arises is whether inferred data (i.e. output data) **is** personal data which is then governed by the GDPR or PDPA. Consider this situation: Assume that a person's physical or mental health information can be inferred from input entered by a person with regard to his daily routine or his food consumption information by the AI. The question then turns towards whether this physical or mental health information is personal data. Indeed, the "inference" by the AI might ultimately not be valid, for example, if it is simply a probabilistic guess by the algorithm. The answer to this question can have far-reaching consequences, particularly as deeming such information as personal data triggers all the data protection obligations, be it under the GDPR or PDPA.

To answer this question, inspiration might possibly be drawn from how past cases in the European Union have been decided.

The European Court of Justice ("**ECJ**") was presented with two requests in two sets of proceedings. In this joint case, individuals sought to obtain a copy of various administrative documents that was drafted with regard to their residence permits. The officials, at first instance, refused these requests (CJEU, 2014). The officials argued in this case that although it was true that information provided could constitute personal data, information which required an abstract legal interpretation cannot be deemed to be personal data (CJEU, 2014). The ECJ held that the input data (such as the applicant's name, date of birth, and the like), as well as the holding by the Minister (that the residence permit was to be denied), were personal data (CJEU, 2014). What was not personal data, however, was the **legal analysis** by the Minister in reaching his decision. This is because the legal analysis was simply information "about the assessment and application by the competent authority of that law to the applicant's situation, that situation being established *inter alia* by means of the personal data relating to him which that authority has available to it" (CJEU, 2014).

However, the decisions do not all speak with one voice. In a different case heard by the ECJ, involving a Data Protection Commissioner's refusal to give an individual access to the corrected script of his examination, somewhat surprisingly, the ECJ held that the examiner's comments, which included the examiner's reasoning, were regarded as personal data. The ECJ held that the content of an examinee's answers was personal data – in addition to information as it related to his handwriting (*Peter Nowak v Data Protection Commissioner,* 2017). The ECJ went even further, holding that the information in the comments of an examiner with respect to the candidate's answers is information relating to the candidate (*Peter Nowak v Data Protection Commissioner,* 2017). Perhaps recognising the far-reaching consequences of its decision and the potential absurd results that might arise if taken too far, the ECJ then held that although such comments constituted personal data, the right to rectification (one such right as provided to data subjects under DPR) did not extend to the correction of an examinee's answers or the examiner's comments (*Peter Nowak v Data Protection Commissioner,* 2017). It reasoned that the assessment of whether personal data is accurate or complete must be made in light of the purpose for which that data was collected. That purpose exists, as far as the answers submitted by an examination candidate are concerned, in being able to evaluate the level of knowledge and the competence of that candidate at the time of the examination. Such errors in any answers do not represent inaccuracy, the existence of which would give rise to a right of rectification. Indeed, such a holding applied to the examiner's comments as well (*Peter Nowak v Data Protection Commissioner,* 2017). Notwithstanding this, the right of access continues to subsist, given that it was personal data about the candidate (*Peter Nowak v Data Protection Commissioner,* 2017).

How, then, does one deal with inferred data about someone created by generative AI? In answering this question, the Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679 states that "profiling can create special category data by inference from data which is not special category data in its own right but becomes so when combined with other data. For example, it may be possible to infer someone's state of health from the records of their food shopping combined with the data on the quality and energy contents of their food" (European Commission, 2018). In this context, profiling has been defined by the Guidelines to be "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular, to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements" (European Commission, 2018). Relying on this guidance, it is argued that in the context of profiling, such inferences (or instances of inferred data) ought reasonably to be considered as personal data. However, going further, should all inferred data created by AI from personal data about an individual be itself regarded as personal data? This article argues that there is good reason to consider such inferred data as personal data. Indeed, this conclusion is supported by the conclusion of the working party, which stated that in the case of automated profiling, a data subject ought to have the right to access both the input data and the conclusions which could be inferred from such data (Article 29 of the Data Protection Working Party, 2016). Such a conclusion would have the effect of requiring AI developers to set clear boundaries and policies in the context of generative AI as to what output can come out of the AI system.

How then may these principles be extended in the context of Singapore's PDPA? According to the PDPA, personal data is data, whether true or not, about an individual who can be identified – (a) from that data or (b) from that data and other information which the organisation has or is likely to have access to (Government of Singapore, 2020). In particular, the Advisory Guidelines to the PDPA states that there are two principal considerations to determining whether something constitutes personal data. The first consideration would be the **purpose** of the information, and the second would be whether a subject would be **identifiable** from that data (Personal Data Protection Act Advisory Guidelines, 2022). The PDPA advisory guidelines do not go further to address inferred data – making this issue a novel one for Singapore. We might find further guidance in the Personal Data Protection Commission's Guide to Basic Data Anonymisation Techniques, which states that it is possible for "certain information" to be inferred from de-identified data but admits the "problem of inference is not limited to a single attribute, but may also apply across attributes, even if all have had anonymisation techniques applied" (Personal Data Protection Commission, 2024). Although a useful starting point, it does not entirely address the questions surrounding inferred data. In this regard, it is suggested that Singapore ought to follow in the EU's footsteps and deem those inferences created by generative AI in the context of profiling to be considered as personal data.

Drawing from this, we consider that conclusions or inferences generated by generative AI could properly be considered as personal data, hence covering the vexed issue of profiling as well.

With this, we turn now to examine the issue of profiling in more detail and its struggles with existing Data Protection law, particularly what rights data subjects should have over data in which they are profiled.

3. The Quandary of Profiling in Data Protection Law – What rights might a Data Subject have?

As Du and others posit, the exponential growth of generative adversarial networks ("**GAN**") has been a foundational technique of generative AI (Du *et al.*, 2023). Brownlee explains GAN in simple terms, highlighting that it is a way of using "generative modelling", which "is an unsupervised learning task in machine learning that involves automatically discovering and learning the regularities or patterns in input data in such a way that the model can be used to generate or output new examples that plausibly could be drawn from the original dataset". According to Brownlee, this is a clever way of training generative AI. The way GAN works involves two key components – the first being the generator model, and the second the discriminator model. The generator model creates new examples using the data set that is provided, whilst the discriminator model performs the function of discriminating the real from the fake. This process is repeated many times, at least until the discriminator can be tricked only half the time, following which the generative AI would then be at an adequate level to generate inferences for the user (Brownlee, 2019).

The problem that data protection law has with this development is as follows. There are huge swatches of input data, some of which are personal data, which could at any one time be sent into the generative AI to be processed by the GAN. Inferences can then be drawn from such data – hence the term "profiling".[13]

We return now to the example of John and the trumpet. Assume further that a generative AI model has been developed to identify the correlation between educational qualifications and number of instruments played. For every individual in such a case, there would be a whole host of personal data, such as educational level, music preferences, as well as instruments played. When ran through the GAN network, the algorithm will generate examples using the input data, whilst the discriminator will then draw conclusions until it reaches a reasonable level of accuracy. How the GAN would handle these data would be through the drawing of correlations, for example, how higher education might be linked to an increased number of instruments played. This correlation and the algorithm developed is likely not personal data. Instead, this is group data and does not fall within the definition of personal, data whether presented in the GDPR or PDPA.

In this case, assume that John then inputs his data into the generative AI model, which then makes a prediction as to the number of instruments that he plays. Here, one might properly argue that the data provided by John is personal data, whilst the inference as to the number of instruments he played is inferred data, which belongs to him as well.

The implications of such a conclusion can indeed be far-reaching. If it is John's personal data, should it then be within John's rights to require the generative AI company to accord him obligations *vis-à-vis* his inferred personal data? As Ye *et al*. (2024) correctly point out, this does not appear to accord with the common understanding of Generative AI .[14] Quach has posited that the output of GPT-2 included at least 0.1% of personal information, including names, addresses, and the like (Quach, 2021). Indeed, the CEO of OpenAI himself admitted that some ChatGPT users could access other's conversation histories as a result of problems with the GPT open-source database (Haughey, 2023).

Wachter and Mittelstadt have argued towards a right over inferred data, which, according to them, would be an *ex-ante* justification to be given by a data controller (*i.e.*, the AI company) as to whether an inference is reasonable, albeit such rights should only apply to "high-risk inferences" drawn through big-data analytics which are privacy-invasive or damaging or have low verifiability. The reasons why such a right is required, according to them, is because "such data draw on highly diverse and feature-rich data of unpredictable value, and create new opportunities for discriminatory, biased, and invasive decision-making" (Wachter and Mittelstadt, 2019). These scholars argue that presently under the GDPR, such individuals are granted little to no oversight of how their personal data has been used to draw inferences about them, in effect according "economy class" status to such data. This is particularly the case as it relates to the data subject's right to know (Articles 13-15), rectify (Article 16), delete (Article 17), object to (Article 21), or portability (Article 20) (Wachter and Mittelstadt, 2019).

The scholars go further to suggest that for an inference to be deemed reasonable, the inference should fulfil the three criteria of (a) acceptability, (b) relevance, and (c) reliability. Limb (a) requires the input data to be normatively acceptable (*i.e.*, race or sexual orientation should be excluded); limb (b) requires the inferred data to be relevant for the chosen processing purpose or type of automated decision (*i.e.*, this requires the data to have a link to the processing purpose); and limb (c) requires the data used must be accurate and reliable (and not from dubious sources) (Wachter and Mittelstadt, 2019).

While, in one view novel, this can be seen as a way forward for Data Protection Law. As examined in Joint Cases C-141 and 372/12, as well as Case C-436/16 above, it is reasonably clear that inferred personal data does constitute personal data when construed in the broad sense and that, therefore, the rights to know, delete, object to, or port are available, albeit the right to rectify has been limited to a certain

[13] This has been defined above.

[14] Ye, Yan, Li, Jiang, *Privacy and Personal Data Risk Governance for Generative Artificial Intelligence: A Chinese Perspective*.

extent, as suggested by Case C-436/16. It is not that far of a stretch to then provide for a right to a reasonable inference. This would avoid any ambiguity in these concepts and provide clarity for generative AI companies when developing their ethical and operational policies to operate within certain pre-defined limits. The authors do not consider that requiring responsible use would hamper innovation. On the contrary, it would foster innovation with the right ideals and boundaries.

This suggestion follows closely to the Singapore Personal Data Protection Commission's ("**PDPC**") Model AI Governance Framework, which is largely undergirded by the principles that the decisions made by AI ought to be explainable, transparent, and fair, as well as the fact that AI systems should be human-centric (Personal Data Protection Commission Singapore, 2020). In this regard, a possible step forward for Singapore's PDPA would be for this right over inferred data to be reasonable to also apply in the context of inferred personal data in Singapore, which would be particularly appropriate in light of the Model AI Governance Framework.

4. The Question of Damages in the Context of Generative AI

We turn now to address one final and perhaps the most practical question in this context – damages. In the hypothetical given earlier, an image created by generative AI has led to John suffering emotional distress. The nub of the issue, therefore, concerns whether John, as a private party, has any cause of action against the AI company for a remedy from emotional distress. This article will, therefore, walk the reader through a two-part analysis. First, it will be considered whether a claim under the relevant data protection statute even arises. Second, it will be considered what exact obligation is typically breached, in the context of generative AI.

For an AI company to *owe* an obligation towards John, it must first owe rights to John as a data controller or data processor. We shall first examine this framework under the GDPR before moving on to our analysis of the PDPA. Pursuant to the GDPR, Article 82(1) entitles "any person who has suffered material or non-material damage…shall have a right to receive compensation from the controller or processor for the damage suffered".[15] A right to sue may, therefore, be created upon breach of the "Rights of the data subject" undergirded by Chapter 3 of the GDPR, such as the right of access, the right to erasure, the right to restriction of processing, or even an omission to provide information where data is collected from the data subject.

In examining this thorny issue, a case decided by the Court of Justice of the European Union ("**CJEU**") might once again prove instructive. In *UI v Österreichische Post AG*, the Court of Justice of the European Union ("**CJEU**") held that Article 82 of the GDPR does not provide for compensation to be payable for the mere infringement of a data subject's rights. In this regard, the CJEU held that the mere infringement of the provisions is not sufficient to confer a right to compensation (CJEU, 2022). By relying on the plain statutory language of the provision, the CJEU held it is clear from the wording of Article 82 that the existence of "**damage**" which has been "**suffered**" constitutes one of the conditions for the right of compensation, as does the existence of the infringement and of a causal link between that damage and that infringement, with the three conditions being cumulative (CJEU, 2022).

Hence, a mere breach of a GDPR obligation is insufficient. What this means would be that in John's hypothetical, he would have to show that the use of the generative AI company had indeed breached one of his rights, and therefore, he would have to prove damage.

In this regard, the Singapore Court of Appeal judgment of *Reed, Michael v Bellingham (Attorney-General, intervener)* is helpful (*Reed, Michael v Bellingham*, 2022). In that case, the Court of Appeal held that emotional distress was sufficient to constitute the "loss or damage" limb under *s* 32(1) of the PDPA. Applying the principles of statutory construction to *s* 32(1), the Court adopted a wide interpretation of the section, noting that there was nothing found in the plain language of the PDPA which expressly

[15] Article 82, General Data Protection Regulations

excluded emotional distress as a type of damage that was covered by *s* 32(1) (*Reed, Michael v Bellingham*, 2022). In doing so, the Court looked towards the statutory rationale of the PDPA, considering the "vast and ever-increasing volume of personal data being collected and processed increases the risk of misuse of personal data", and that *s* 32 "must have been intended to be effective in guarding the right of individuals to protect their personal data" (*Reed, Michael v Bellingham*, 2022). As such, adopting a wide interpretation would serve to further the statutory purpose of the PDPA, allowing the PDPA to provide "robust protection for individual's personal data" (*Reed, Michael v Bellingham*, 2022). As such, the Singapore Courts held that emotional distress was actionable under the PDPA. This is, of course, still subject to a "strict causal link" *vis-à-vis* a breach of the PDPA and the loss or damage suffered, and no legal recourse will be permitted for minimal loss (*Reed, Michael v Bellingham*, 2022).

Given that emotional distress is claimable under the relevant data protection statutes, it would appear critical to identify the obligation that might be breached in the context of generative AI. In other words, the key is to point to what data subject right would be breached in most cases?

The European Commission has highlighted that a data controller is defined as any company or organisation which determines the purpose for which and how personal data is processed. Deloitte provides a good example of the nuances involved in the context of how a generative AI company operating an app like ChatGPT might function. According to Deloitte, a Generative AI system provider (such as OpenAI), would likely operate as a ***data controller*** as it relates to the first layers of training and input data. At the same time, the provider will also likely act as an independent data controller for all data as well. In this regard, it may also play a dual role of being a ***data processor*** – particularly in the case where the AI company simply licenses this "AI engine" to enterprise customers without any embedded data. Hence, a generative AI system provider can clearly be brought under the governance of the relevant data protection statutes.

The above discussion is likely to be critical as AI further develops. In 2024, Italy's data protection authority had informed OpenAI that ChatGPT clearly violated data protection rules. In this regard, the Italian Data Protection Authority had stated that they suspected ChatGPT to have breached Articles 5 (principles relating to the processing of personal data), Article 6 (lawfulness of processing), Article 8 (conditions applicable to child's consent in relation to information society services), Article 13 (information to be provided where personal data are collected from the data subject), and Article 25 (data protection by design and by default) (Lomas, 2024).

Looking at the list above, it would seem the main obligation that OpenAI has failed to comply with would be the obligation to provide certain information where personal data is collected from the data subject. As Lomas explains, ChatGPT was developed using "masses of data scrapped off the public internet", this being information which "includes the personal data of individuals". Amongst the six legal bases to use such information, Lomas highlights that only two possibilities remain – these being that of consent or legitimate interest (given that OpenAI was told by the Italian Data Protection Authority to remove references to "performance of a contract" as a legal basis).

It is unlikely consent can apply as a legal basis, given that consent (other than the privacy policy) is difficult to obtain from millions of users absent a mandated information notice. As far as OpenAI's privacy policy is concerned, it states that data subjects can "*withdraw their consent – where [OpenAI] rely on consent as the legal basis for processing*" (Open AI, 2024). It is, therefore, likely that should OpenAI not provide such a notice to users upon a user operating ChatGPT, it would likely be in breach of Article 13 of the GDPR.[16]

All the same, it is likely that absent consent, the only other legal basis that remains would be legitimate interests – which requires that the

[16] Article 13(1)(d) GDPR.

processing is necessary for the purposes of the legitimate interest pursued by the controller or a third party, except where such interests are overridden by the interests of fundamental rights and freedoms of the data subject which requires protection of personal data, in particular where the data subject is a child.[17] Whether the collection of such personal data to advance generative AI is a legitimate interest has not been decided yet by the CJEU, and this does remain an open question.

A similar position applies in Singapore as well – *s* 13 requires an organisation not to collect, use, or disclose personal data about an individual unless (a) the individual gives his consent, or (b) the collection, use, or disclosure without the individual's consent is required or authorised under the PDPA (Government of Singapore, 2020). Legitimate interests do exist as a valid legal basis to collect, use or disclose personal data in the PDPA as well – though it remains a question as to whether the legitimate interests of the generative AI organisation do outweigh any adverse effects on data subjects. This question will remain an open question to be decided by the Singapore courts.[18]

Returning to the hypothetical of John and the trumpet, it is, therefore, likely that a data controller, such as OpenAI, would be liable for damages for emotional distress. In any event, generative AI companies should ensure that the decisions made by their proprietary AI are explainable, transparent, and fair. This can be done through privacy by design principles, ensuring an appropriate degree of human involvement occasionally, as well as ensuring that the black box of decision making does not become too opaque at times. Such principles are undergirded by the Model AI Governance Framework by the PDPC and would likely serve as a useful roadmap for generative AI organisations to follow.

Conclusion

A few decades ago, none of us would have imagined the capabilities of AI to develop to such an extent, and it is likely that AI will be – and possibly already is – the mantra of the fourth industrial revolution. This article has explored three key issues, (a) whether inferred personal data by generative Artificial Intelligence can be considered as personal data, (b) the rights which data subjects have over such data, and (c) remedies that can be claimed because of a mishap by a generative AI company. This article has then suggested that the Singapore Model AI Governance Framework is a right step forward, particularly as jurisdictions around the world begin to frame their privacy legislation to handle the new epoch of AI generated data. The future is exciting, and the potential risks of generative AI should not be hidden by its immense potential – with strong privacy laws and adequate guidance, it is likely AI can chart an explainable, transparent, and fair path ahead as we move into the future of tomorrow.

References

AI Verify. (2024). P*roposed Model AI Governance Framework for Generative AI – Fostering a Trusted Ecosystem.* Retrieved from https://aiverifyfoundation.sg/downloads/Proposed_MGF_Gen_AI_2024.pdf

Brownlee, J. (2019). A Gentle Introduction to Generative Adversarial Networks (GANs). Machine Learning Mastery. Retrieved from https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/

Charles, A. (2013). Tech Giants may be huge, but nothing matches big data. *The Guardian.* Retrieved from https://www.theguardian.com/technology/2013/aug/23/tech-giants-data

Chik, W.B. (2013). The Singapore Personal Data Protection Act and an assessment of future trends in data privacy reform. *Computer Law & Security Review.* 29 (5), pp. 554–575, doi: 10.1016/j.clsr.2013.07.010.

Court of Justice of the European Union (CJEU). (2014). *Joined Cases C-141/12 and C-372/12, YS v Minister voor Immigratie, Integratie en Asiel and Minister voor Immigratie, Integratie en Asiel v M and S.* Retrieved from https://fra.europa.eu/en/caselaw-reference/cjeu-joined-cases-c-14112-and-c-37212c-judgment Metz, R. (2022). AI won an art

[17] Article 6(1) GDPR.

[18] First Schedule, Personal Data Protection Act.

contest, and artists are furious. *CNN*. Retrieved from https://edition.cnn.com/2022/09/03/tech/ai-art-fair-winner-controversy/index.html

Court of Justice of the European Union (CJEU), (2022). Opinion of Advocate General in Case C-300/21, Bundesrepublik Deutschland v XT. Retrieved from https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:62021CC0300

Du, H., Niyato, D., Kang, J., Xiong, Z., Zhang, P., Cui, S., Shen, X., *et al.* (2023). The Age of Generative AI and AI-Generated Everything. *ArXiv.* doi: 10.48550/arXiv.2311.00947.

European Commission. (2018). Article 29 Working Party - Newsroom. Retrieved from https://ec.europa.eu/newsroom/article29/items/612053

European Data Protection Supervisor (EDPS). (n.d.). *The History of the General Data Protection Regulation.* Retrieved from https://www.edps.europa.eu/data-protection/data-protection/legislation/history-general-data-protection-regulation_en

European Parliament and Council. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*. Official Journal of the European Union, L119, pp. 1–88. Retrieved from https://eur-lex.europa.eu/eli/reg/2016/679/oj

Government of Singapore. (2020). *Personal Data Protection Act 2012 (2020 Rev Ed)*, s 2. Retrieved from https://sso.agc.gov.sg/Act/PDPA2012

Haughey, L. (2023) Are your conversations safe?, *Daily Mail.* Retrieved from https://www.dailymail.co.uk/sciencetech/article-11893689/ChatGPT-creator-confirms-bug-allowed-users-snoop-chat-histories.html

Hoofnagle, C. J., van der Sloot, B. and Borgesius, F. Z. (2019), The European Union general data protection regulation: what it is and what it means, *Information & Communications Technology Law,* 28(1), pp. 65–98. doi: 10.1080/13600834.2019.1573501.

Lomas, N. (2024). ChatGPT is violating Europe's Data Privacy Laws, Italian DPA tells OpenAI. *Tech Crunch.* Retrieved from https://techcrunch.com/2024/01/29/chatgpt-italy-gdpr-notification/

NVIDIA. (n.d.) *What is Generative AI.* Retrieved from https://www.nvidia.com/en-us/glossary/generative-ai/

Open AI. (2024). *Open AI Privacy Policy.* Retrieved from https://openai.com/policies/privacy-policy

Personal Data Protection Act Advisory Guidelines. (2022). Retrieved from https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/advisory-guidelines/ag-on-key-concepts/advisory-guidelines-on-key-concepts-in-the-pdpa-17-may-2022.pdf

Personal Data Protection Commission. (n.d.) PDPA Overview. Retrieved from https://www.pdpc.gov.sg/overview-of-pdpa/the-legislation/personal-data-protection-act#:~:text=What%20is%20the%20PDPA%3F,Banking%20Act%20and%20Insurance%20Act.

Personal Data Protection Commission. (2024). Guide to Basic Anonymization Techniques. Paragraph 4.1(c). Available at: https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/advisory-guidelines/guide-to-basic-anonymisation-%28updated-24-july-2024%29.pdf

Personal Data Protection Commission Singapore. (2020). *Singapore's Approach to AI Governance.* Retrieved from https://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework,

Peter Nowak v Data Protection Commissioner. (2017). CJEU Case C-434/16. Retrieved from https://curia.europa.eu

Phipps v Boardman. (1967). 2 AC 46.

Quach, K. (2021). How to curb GPT-3's tongue. *The Register.* Retrieved from https://www.theregister.com/2021/03/18/openai_gpt3_data/

Reed, Michael v Bellingham (Attorney-General, intervener). (2022). SGCA 60. Retrieved from https://www.elitigation.sg/gd/s/2022_SGCA_60

Schwab, K. (2016) The Fourth Industrial Revolution: What it means and how to respond. *World Economic Forum.* Retrieved from https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/

Vincent, J. (2019). Former Go Champion beaten by DeepMind retires after declaring AI invincible. *The Verge.* Retrieved from https://www.theverge.com/2019/11/27/20985260/ai-go-alphago-lee-se-dol-retired-deepmind-defeat

Wachter, S. and Mittelstadt, B. (2019). A Right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI. *Columbia Business Law Review.* Issue 2.

Ye, X., Yan, Y., Li, J. and Jiang, B. (2024). Privacy and personal data risk governance for generative artificial intelligence: A Chinese perspective. *Telecommunications Policy.* 48 (10), p. 102851, doi: 10.1016/j.telpol.2024.102851.