

DATA IS THE NEW GOLD: A SINGAPORE PERSPECTIVE ON THE DUTY OF CARE CONCERNING A DATASET'S ROLE IN CONTRIBUTING TO BIAS AND AI HALLUCINATIONS

Daniel Seah¹

Amidst the trends and advances in Artificial Intelligence (AI) techniques, the dataset's key role in creating AI harms is a constant. Yet the literature's attention is typically focussed on the machine-learning's role, from which the dataset's role is examined usually in terms of its potential for bias. Datasets also contribute to AI hallucinations. Again, the literature focuses on the composite roles of datasets and AI models in hallucinations, in isolation from a dataset's contributory role in creating bias. Therefore, the overlapping and distinctive roles of datasets, as a dual contributor, are understudied in the literature. Through the legal concept of a duty of care, in the context of AI applications in financial services, this paper surveys the overlapping and distinctive roles of datasets. This paper's focus matters for three reasons. First, it is important to identify and determine which parties, now part of a growing ecosystem of stakeholders, are legally liable for selecting and creating the dataset. Second, the tort of negligence is medium-agnostic and can adapt to governing AI harms, which will influence ongoing legislative efforts that are medium-based. And third, this paper provides a frame of reference to start a broader conversation, beyond Singapore, about how datasets will become more prominent in an AI zeitgeist, and that the common law can play a stabilising effect on the policy makers' impetus to regulate through medium-based legislations.

Key words: AI hallucinations, Dataset bias, Dataset, Negligence, Duty of Care.

I. DATASET'S ROLE IN CONTRIBUTING TO ARTIFICIAL INTELLIGENCE (AI) HARMS

Regarding harms that are created by AI applications, the literature tends to focus on the potentially pernicious role of algorithms or machine-learning.¹ It is within this examination of AI models that the role of datasets,² as a contributor to the AI harm, typically arises. Within this context, dataset bias has received significant interest, especially in the legal literature.³ A common example involves the over or under representation of groups on grounds of ethnicity or educational qualifications. Dataset bias, therefore, mainly concerns the lack of fairness and hence its discriminatory implications against, for instance, classes of individuals.

However, the dataset is also a contributor to AI hallucinations. Although the term "AI hallucination" is relatively new, the cognate phenomenon of inconsistency in open-domain (i.e., conversational AI that uses Natural Language Understanding) chatbots is established.⁴ It is the widespread interest in Generative AI (GAI)⁵ that throws the matter of AI hallucinations into sharp relief. Unlike previous AI applications, which are mainly useful for predictions, GAI's application is compelling. This is because GAI is trained with a large corpus of data to generate new content. In particular, the creation of foundational models such as ChatGPT4, i.e., a specific type of GAI, which is excellent with Natural Language Processing (NLP) with practical downstream applications such as textual prompts (e.g., virtual assistants) and image generation (Midjourney or Dalle-2). Given its

¹ Daniel Seah is Assistant Professor of Law (Education) at the Singapore Management University. He received his PhD in law from UCL (University College London), and is an Advocate & Solicitor of the Supreme Court of Singapore. (danielseah@smu.edu.sg) The author thanks Keisha Chui and Neo Yu Fan for their research assistance, and the anonymous reviewers for their helpful comments.

Received, June 28, 2023; revised, July 12, 2023; accepted, July 30, 2023; published, August 8, 2023

widespread applications, the prospect of AI hallucinations has increased. The literature in AI hallucinations is expanding, but it concentrates on datasets and AI models as composite contributors of hallucinations, in isolation from the problems of bias and unfairness, as well as discrimination, which are also caused by datasets.⁶

To this extent, datasets are dual contributors of legal harms: on grounds of its lack of representativeness, and as one contributor to causing AI hallucinations, respectively. Accordingly, the overlapping and distinctive roles of datasets as a dual contributor are unstudied in the literature. This survey of the overlap and distinction is important for identifying and in determining which parties, now part of a growing ecosystem of stakeholders, are legally liable for selecting and creating the dataset.

It is against this context that this piece focuses on examining the concept of legal proximity, the relationships, between a tortfeasor (the defendant) and victim (claimant), i.e., the duty of care, in the context of AI applications in financial services. This focus on legal relationships matters for three reasons. First, at the international level, there are indications of a shift away from the “autonomy” of machine-learning, especially at the European Union level.⁷ In other words, legal liability must be attributable either to a natural or legal person, although the AI application is autonomous in the sense that it required no human involvement and had caused an AI harm. There is an evolving consensus that stakeholders in an ecosystem of building the AI application must be identified on grounds of transparency.⁸

Second, given the ecosystem of stakeholders who build the AI models, select and create datasets, as well as AI governance frameworks,⁹ it is increasingly likely that the AI harms which ensue are not intentional. This does not mean that there is no legal fault or liability. Therefore, the tort of negligence, a medium-agnostic law, will continue to be relevant through adaptation in AI harms: its relevance lies in the duty of care’s focus on examining legal relationships, which can influence the ongoing legislative efforts, on the international plane, to govern AI harms by setting minimum standards of acceptable behaviour in AI applications.

And third, this paper argues that a focus on the legal relationship between creators of the dataset, and the harm on a potential claimant, provides a frame of reference for future research on the overlapping and distinctive roles of datasets from AI models in creating harm for a potential claimant. However, this paper does not suggest that datasets are single or decisive contributors to AI harms to the exclusion of AI models. Rather, the focus on a dataset’s role as dual contributors of AI harms serves a practical purpose in a *legal* sense. As explained in Part V, this focus provides a frame of reference to promote strategic clarity during litigation in terms of discovery and in determining suitable expert witnesses, i.e., whether the AI harm in question is caused by datasets, AI models, or both.

This paper is structured as follows. Part II explains the manifestations of bias that occur from datasets. Part III explains AI hallucinations and how datasets are one of various contributors to AI hallucinations with consequences for downstream AI applications. Part IV explains the relevance and resilience of the common law,¹⁰ i.e., the duty of care under the *Spandeck* test in Singapore’s law of negligence, in regulating acceptable standards in relation to datasets. Part V evaluates the potential application of the *Spandeck* test to harms brought about by dataset bias and AI hallucinations, and the implications for potential claimants and defendants. Part VI concludes.

II. MANIFESTATIONS OF DATASET BIAS

A. Types of datasets in Discriminatory and Generative AI models

This part explains the meaning of bias in dataset and how legal harms that are not intentional can arise from dataset bias. In terms of machine learning, an AI subset, these are mainly “discriminative” models. In other words, the AI models aid decision-making with recommendations, filtering, or by making predictions.¹¹ These decisions are arrived at through learning the boundaries between the datasets. Examples include supervised learning such as advanced regressions and categorization of data

to improve predictions, and unsupervised learning, i.e., processing input data to create automated customer segments.¹²

GAI, on the other hand, uses generative models which learn the underlying distribution of the data to generate new content from this learned distribution.¹³ Within GAI, foundation models are special because they are trained on a broad corpus of data, and act as a “foundation” for more task specific, downstream applications.¹⁴ An example of a downstream application would include virtual assistants that use natural sounding language for retail banking customers who need help for complex queries. Another example would be robo-advisors who can give specific guidance based on a wide range of retail banking scenarios, which are based on creating training data on simulated client needs and market events to train the AI models.¹⁵

B. How Dataset Bias Arises

Bias can be defined technically to mean an effect that deprives a statistical result of representativeness by systematically distorting it.¹⁶ It can also be defined, in a general sense, to mean an inclination of prejudice towards or against a person, object, or position.¹⁷ This general definition of bias, i.e., the prejudicial implications, is implied in the legal concept of discrimination, which contains direct and indirect discrimination. For instance, with respect to direct discrimination, the EU’s Racial Equality Directive (“Directive”) states:

“Direct discrimination shall be taken to occur where one person is treated less favourably than another is, has been or would be treated in a comparable situation on grounds of racial or ethnic origin”.¹⁸

The same Directive defines indirect discrimination as follows:

“Indirect discrimination shall be taken to occur where (i) an apparently neutral provision, criterion or practice would (ii) put persons of a racial or ethnic origin at a particular disadvantage compared with other persons, (iii) unless that provision, criterion or practice is objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary”.¹⁹

These definitions – technical, general, and legal – reflect two implications. First, it is accepted that bias can – and does arise – without an intention to create bias in the dataset. This is legally relevant because the lack of intention can still cause legal harms from which liability can be determined through, as one example, the law of negligence in the common law.²⁰ Second, the classes of individuals, position, or objects who are subject to discrimination, through bias, are not closed. Although the Directive had focussed on proscribing discrimination on grounds of race or ethnicity, more recent legislations such as the General Data Protection Regulations (GDPR) recognise other classes such as sexual orientation, religious beliefs, and genetic data are protected too.²¹

Against this context, bias in the three senses arises when the dataset is under or representative of certain groups. The bias can be manifested as selection bias (the source for the dataset and who determines this source);²² exclusion bias (who decides what and how a source does not fall into the dataset); reporting bias (observations of a certain type are reported and identified as a source for the dataset, which leads to selection bias); and detection bias (a phenomenon that is given undue observation and forms part of the source as data).²³ The bias can be mitigated through sampling sizes in the dataset, or through weights and mathematical calculations in the AI model’s design.²⁴ Of course, the software engineers can also mitigate these forms of bias through training. However, these measures mitigate and cannot eliminate bias. It is in this sense that this paper uses the expression “dataset bias”. Accordingly, if harm ensues which is attributable to the dataset bias, this is a fault-based liability although there is no legal intention to cause harm.

Another form of bias occurs during the encoding process. This means that the source of the dataset must be presented in a format, i.e., encoding, which is “machine-readable” by the AI model(s) in question. In other words, the data is encoded as values to become machine-readable.²⁵ Yet these values are the result of identifying human characteristics in the dataset. For instance, the different tones

of anxiety, doubt, and happiness for a virtual assistant to produce natural sounding language for a retail banking customer. Therefore, the gender or non-binary nature of a banking customer is a characteristic, an attribute, of the dataset. The value of the attribute could be a number (e.g.: 20) based on the pitch pattern of voices to register happiness and sadness with a median value.²⁶

A final – and crucial – facet of bias potentially arises in the involvement by humans in labelling, i.e., annotating,²⁷ the dataset for training so that downstream applications, such as virtual assistants, are natural sounding in a conversational style for human taste. This is a time consuming and tedious task of sorting data, i.e., by tagging and labelling, for training the AI model.

For instance, engineers of a virtual assistant will design the virtual assistant to behave in a personalised and conversational manner with a retail banking customer. This might include conceiving a range of questions already being raised by existing customers and training the virtual assistant to give the correct answer, especially to complex queries. The quality of this forward facing service depends on the extent and scale of human involvement to rate the virtual assistant's answers, such as the accuracy, helpfulness, cultural sensitivity, and even personal preferences based on different educational and financial profiles.

An example of the granularity in this labelling, to suit human taste, might even include labelling the reflection of a shirt as being distinct from a selfie of a shirt, possibly for an online shopping experience. The humans who perform this labelling are called annotators and are paid hourly rates for their labour.²⁸ As the AI models are trained by datasets for proprietary products, there is scant public information about the oversight of these annotators. It is likely that a vast, and growing, ecosystem of unseen individuals who have signed non-disclosure agreements have been recruited to cope with the increased demand for high-quality GAI models.

The purpose of this account is to underscore that ratings and labelling, by human annotators, can also suffer from errors, inaccuracies, or similar bias such as selection or reporting bias. This phenomenon is complicated by the outsourcing of these labelling tasks to data vendors. Therefore, the key point is that dataset bias occurs at different layers of the dataset being processed, as a source and training data. This implicates the type of legal relationships that potentially connect a claimant's harm (from AI) to more than one defendant.²⁹

However, the dataset bias brought about through annotations might not contribute to a legal harm in terms of direct and indirect discrimination. This is because bias, which implicates discrimination and hence unfairness, is permissible if exceptions apply, such as substantial public interest or statistical purposes.³⁰ Furthermore, although the range of discriminated classes are not closed, there are already established classes such as age, ethnicity, and sexual orientation. In other words, it is possible that claimants who allege bias and do not fall within the established classes will face difficulties in proving the legal harm.

However, the annotations in the dataset can still contribute to AI hallucinations, a conceptually distinct form of legal harm in the context of negligence. Accordingly, there is potentially a large, possibly anonymous, class of individuals who have separately contributed to the final form of the dataset which can cause AI hallucinations. This prospect raises adverse implications for the success of a claimant who sues for negligence.

Put differently, given the disparity in knowledge about an AI application's production between a defendant company and claimant, it is arguable that a defendant technology company would have some advantage – and the deep pockets in a trial – to resist a claimant's argument that there is a legal relationship between the defendant which had caused the AI harm. It is this matter concerning datasets, created by a large class of individuals, as contributors of AI hallucinations, to which we turn in Part III.

III. DATASETS AS CONTRIBUTORS OF AI HALLUCINATIONS

Unlike dataset bias, which implicates issues of discrimination and unfairness, the legal harms of AI hallucinations are conceptually distinct in terms of determining liability in negligence. Generally, AI hallucinations refer to an outcome which is trained by the AI model, i.e., the target, being factually false or is unfaithful based on the dataset (i.e., the source).³¹ In other words, even if the dataset is representative and unbiased, hallucinations can create harm with factual errors, for example, in the form of negligent misstatements.³² Whether the dataset is biased or unbiased, as discussed in Part II, AI hallucinations can also occur: in this situation, the AI harm will need to be carefully identified to establish the legal proximity owed by the defendant (or defendants, if there is a large class of natural and legal persons who have created the dataset) to a claimant.

A. Intrinsic and Extrinsic Hallucinations

Intrinsic hallucinations refer to generated outcome that contains a mismatch between the source and outcome. This is common in large-scale datasets in which the real sentences or tables in the datasets are heuristically selected to be paired as the source and outcome.³³ To re-use the example of a virtual assistant to help retail banking customers: the source might *only* contain the word “pusing” (i.e., a migraine) in Bahasa Indonesia. Yet the generated outcome might state “pusing” to mean a “stroll”, which is correct in Bahasa Malaysia. This can be misleading for a customer with limited knowledge of Bahasa Indonesia and Malaysia, and takes the answer at face value. In fact, in this example, the client needs an accurate answer from the bank because its highly personalised virtual assistant is connecting the client’s request for a specific telemedicine service. In this sense, the output is not faithful due to a mismatch between the source and output because the word “pusing” in Bahasa Indonesia and Malaysia has been heuristically paired, i.e., an approximation, as sources to produce an outcome. This is an example of an intrinsic AI hallucination.

Extrinsic hallucinations, in contrast, refer to a generated outcome which cannot be verified from the source (i.e., an outcome that is neither supported nor contradicted by the source).³⁴ An open-domain dialogue system can generate persona-consistent and informative responses to further engage with a user during a conversation. External resources, such as data from social media posts or web scraping,³⁵ which contain explicit persona information or world knowledge is introduced into the AI system to assist the model generation process. The key point here is that the external resources are part of the source, from which the dataset is derived, to generate an outcome.³⁶

For example, a virtual assistant might state that Ukraine and Russia had just agreed to a ceasefire today. This development, if true, will likely raise implications for a decision to buy or sell certain stocks in a customer’s portfolio. It is noteworthy that extrinsic hallucinations draw from external resources, such as Wikipedia, might not be factually false.³⁷ It might in fact improve the quality of the service that is provided by the AI application. It is this dynamic quality that accords a virtual assistant the natural sounding responsiveness, which is attuned to human taste. Nonetheless, this is an illustration of an extrinsic hallucination because of its unverifiability from the source.

It is possible to mitigate these hallucinations by using annotators to write clean and faithful outputs from scratch since the source might be unverifiable.³⁸ Another way is to pay annotators to rewrite real sentences on the web or the outputs in the dataset.³⁹ Both strategies are task specific and will lack the generalisation that suits human taste in natural sounding conversation. It raises implications of requiring users to accept the risks of hallucinations, in exchange for high quality GAI in the downstream application. Again, these realities raise consequences for a victim to establish legal proximity against a defendant, which are examined in Parts IV and V.

Although this piece is focussed on datasets, AI models (and not just datasets) also contribute to hallucinations. The hallucinations can result from the modelling and training of neural models, i.e., through imperfect representation learning, erroneous decoding, exposure bias, or parametric knowledge bias.⁴⁰ The rationale for a conceptual differentiation of the datasets, from the AI models, in terms of ascertaining legal harms is explained together, in Parts IV and V, with an account of how the duty of care at common law can govern AI harms.

III. NEGLIGENCE AND THE DUTY OF CARE AT COMMON LAW: MEDIUM AGNOSTIC AND ADAPTABLE RULES TO GOVERN AI HARMES

The regulatory landscape on the international plane is fluid.⁴¹ Generally, the regulatory approach suggests a combination of criminal offences and administrative fines. As a regulatory approach, the rationale is not novel since regulations, set by politicians with some form of democratic mandate, seek to shape communal behaviour by setting minimum standards. However, it is the medium-based nature of the regulatory proposals which directly target AI and even GAI despite their brisk pace of innovations, especially at the EU, which raise doubts about the law's ability to adapt since it is an innately reactive discipline.

It is against this context that the tort of negligence at common law acquires its salience. Although negligence is mainly a civil wrong,⁴² its provenance as a branch of law contains both private action (i.e., financial compensation between natural and legal persons), and also a normative basis to set communal standards of acceptable behaviour for the overall good of society.⁴³ Negligence is an established but evolving tort that will influence the ongoing regulatory approaches and debates, at national and international levels, in determining liability for AI harms. This part examines this argument by concentrating on how the duty of care in negligence, from a Singaporean perspective, can bridge the common law with ongoing regulations to govern AI harms.

A. *Spandeck* Test in Singapore: A single, universal test for the duty of care

Negligence is a fault-based tort. This covers harms that are perpetrated without intention by a tortfeasor. It is in this sense that, first, even without intention to harm the victim, the tort requires proof of a victim being owed a duty of care by the tortfeasor. Second, the tortfeasor has breached this standard (concerning the duty) of care. Third, and importantly, there must be legal causation of the victim's harm for negligence to apply. And finally, the harm (damage) suffered by the claimant must not be remote.⁴⁴

A claimant can fail to establish negligence if any of these four heads are not established. Put another way, a defendant who is alleged to have been responsible for creating the dataset which is connected to the AI harm has four attempts to resist a claimant's arguments on negligence. The difficulty for a claimant is partly the success of negligence as a tort – the types of harms which can be addressed by negligence, a medium-agnostic law at common law, grows incrementally over time under careful supervision by the courts. However, as negligence is also concerned with setting communal standards of acceptable behaviour, the rights of defendants against vexatious negligence suits are also embedded in the claimant's need to prove the four heads. Accordingly, the richness of the common law of negligence should give some pause to politicians who are rushing medium-based legislations to govern AI harms when the structure and basic principles of governing (all types of) harms are entrenched in the law of negligence.

In Singapore, the tort of negligence has settled on an incremental approach, in the landmark case of *Spandeck*, of recognising harms for which a defendant owes a duty of care.⁴⁵ This is a single, universal test in the sense that this approach does not *per se* preclude pure economic losses or any potentially new forms of harms. Put differently, AI harms resulting from dataset bias or data which contribute to AI hallucinations will fall under the *Spandeck* test for consideration.

Under *Spandeck*, a threshold question must be satisfied: was the harm suffered by the claimant *factually* foreseeable? For instance, it is factually foreseeable that an elderly banking customer who is not a digital native, and uses a retail banking app, to invest in the bank's products might suffer some form of negligent banking service. However, distress and anguish are not actionable under negligence and will likely fail even the factual foreseeability threshold.⁴⁶ This threshold question casts the factual enquiry net rather wide as to whether a claimant's interests would be endangered. Its purpose is to filter out vexatious claims by claimants at an early stage. At this stage of answering the threshold question, it is not a legal question as to the reasonableness of the foreseeable harm, which will be addressed below.

C. Determining Legal Liability: *Spandeck*'s Two-Stage Test

The two-stage test in *Spandeck* is designed to determine the *legal relationship* between the claimant and defendant. It is an artificial tool to interpose liability on a defendant (despite the lack of intention), but it also serves as a barrier against a claimant's vexatious claim by requiring a claimant to establish legal proximity with the defendant. This is the meaning of a legal relationship and it is in this sense that *Spandeck*'s two-stage test is a legal question.

To illustrate this point, let us assume that Beta is a job candidate who is interviewing for a role at Bank Gamma. However, candidate Beta mistakenly turned up at Bank Alpha, instead of Bank Gamma, because both banks are physically located alongside each other on the street. While inside the physical confines of Bank Alpha, candidate Beta was badly hit by a fan which had inexplicably detached off the ceiling. Although candidate Beta should not even be inside Bank Alpha, the non-legal response would be that Bank Alpha should be responsible for candidate Beta in some ways. Accordingly, to embrace the legal language of negligence, Bank Alpha owes candidate Beta a duty of care to provide for a structurally safe environment.⁴⁷ Under *Spandeck*, this is a form of physical proximity, in terms of space and time,⁴⁸ which arises between Bank Alpha and candidate Beta, although the latter had no reason to turn up at Bank Alpha without which candidate Beta would have escaped the crashing fan.

To illustrate the expansive concept of legal proximity, we can build on this scenario. We assume that Bank Alpha is wholly owned by Firm Delta (who has appointed Bank Alpha as a local manager in Singapore to run the bank) who is based in Dubai, a duty of care can arguably arise, which is decided by the courts on case by case basis.⁴⁹ This is because other proximity factors, as identified in *Spandeck*, such as causal or circumstantial proximity can apply.⁵⁰

In the context of professional services, such as banking services, the defendant also has special knowledge concerning the claimant and potentially owes a voluntary assumption of duty (of care), on which the claimant relies.⁵¹ These factors of proximity are not exhaustive and they can be incrementally interpreted to allow for novel types of harms, which will involve AI harms. Ultimately, it is the nature or quality of the event giving rise to the loss and injury, as well as the type of harm caused, which obtains in the final analysis. If one or more of the proximity factors as discussed are established, a *prima facie*⁵² duty of care arises in the claimant's favour at the first stage of the *Spandeck* test.

The second limb of the two-stage is to determine if policy reasons exists to negate the *prima facie* duty of care.⁵³ This requirement serves to consider the communitarian impact if a legal duty exists. Again, this requirement reveals the open-textured nature of harms which can fall under negligence, and also the considerable difficulty for a claimant who must prove that no policy exists after establishing a *prima facie* duty of care. Furthermore, the types of policy are not foreclosed, but the typical policy arguments which can be used to negate a *prima facie* duty of care have included: indeterminate liability (i.e., a potentially large and unverified number of claimants); the availability of other causes of action (contract or defamation, for instance); distributive or corrective justice (insurance or the need for social good to be spread out or to set communal standards of behaviour).⁵⁴

V. POTENTIAL ROLE OF SPANDECK AND THE EU AI ACT IN AI HARMS

At this writing, the EU is trying to approve significant AI laws which will affect not just EU Member States, but also non-EU States which enter into commercial relations or maintain interactions with EU Member States. Two developments bear mention. First, the EU AI Act ("Act") is undergoing a protracted process of approvals within the EU.⁵⁵ In terms of legal proximity between a claimant and the defendant, the Act has adopted a risk-based approach by classified AI such as unacceptable risk, high risk, limited risk, and minimal risk.⁵⁶

For our purposes, biometric classification systems and chatbots are treated as posing limited risk on humans.⁵⁷ Accordingly, a requirement of transparency is imposed on the provider of AI applications: this is a general standard that will be elaborated on by the national laws of the respective EU member

states. However, transparency will typically mean that a user – the potential claimant – will be given adequate information to make informed choices as to whether the AI application should be used. Some form of informed consent is likely to become a feature of transparency in future. Significantly, Generative Pre-trained Transformer (GPT) models which rely on foundation models are subject to stringent transparency requirements although the details as to what counts as “stringent” are still being developed.⁵⁸

The second development is the European Commission’s proposed AI Directive that contains a rebuttable presumption of causality in the claimant’s favour.⁵⁹ As this expression implies, the directive presumes causality because it is hard for the claimant to prove it. Since the directive will be in interpreted together with other EU laws (including the Act in future), the claimant under this proposed directive might arguably need to show, for instance, a breach between the requirement for transparency in non-high risk AI (i.e., limited risk), and the harm which a claimant had suffered from.⁶⁰

This account of the proposed EU laws matters because there is already an in principle, international level recognition that the concept of legal proximity obtains even in proposed, medium-based legislations on regulating AI harms. It is in this sense that the incremental – and at times circumspect – nature of negligence, as common law, can complement the evolving regulations through legislations by politicians.

The in principle recognition might conceivably alleviate some difficulties for claimants against the defendants’ lawyers who argue that, in a test case, legal proximity does not even arise from dataset bias or hallucinations as a *prima facie* duty of care under *Spandeck*. This proposition is based on the reality that there is a large - and possibly anonymous - class of natural and legal persons who are involved in creating the dataset. In other words, causal and circumstantial proximity, which require some form of legal directness between the professional financial services provider such as a bank (defendant) and client (claimant) cannot be established.

In any event, the odds are likely to be stacked against the claimant. Under the EU AI Act, transparency requirements apply to chatbots, a limited risk for AI applications, which are likely to be heavily built up as downstream AI applications by financial institutions. Transparency, though, is not the same in terms of legal blameworthiness when compared with the requirement for human oversight or data tracing requirements (for high risk AI).⁶¹ The rest of this part illustrates these difficulties with examples as to how *Spandeck* might apply to two downstream AI applications.

A. Dataset Bias

Intelligent banking is not new.⁶² These are AI-powered “nudges”, i.e., banking applications (i.e., apps), which use predictive analytics, machine-learning that converts customer data into personalised and intuitive insights to guide customers in performing banking and investment transactions.⁶³ With GAI, the changes lie in the extent of personalisation, through developing virtual assistants from Large Language Models (LLMs).⁶⁴ For instance, it is already possible to service banking customers with parallel services such as guidance on making savings through energy efficient homes. The personalisation could go further to helping the client to understand how this client’s floor space and to make decisions as to whether carpets or wood floors were most energy efficient under the most attractive payment schemes in the market.⁶⁵ One more example involves a high level of personalisation that even helps a client to decide on whether to buy a dog or not by helping the client to understand the financial and health benefits of owning a pet.⁶⁶

Dataset bias in this form of personalised banking services can arise, but the harm on the claimant must be caused by the defendant’s actions (or omission), i.e., the intelligent banking app. Assuming that causation in this sense has been proved, the dataset bias can stem from training data which is unrepresentative because, for example, of the management culture of a bank. As a recent report by the Royal Commission on Robodebt Scheme in Australia shows, the design thinking behind the automated decision making (which involves machine learning) started from the premise that the automation through AI will create budget savings by identifying overpayment for welfare recipients. The key issue

is that, at the outset, the Minister responsible for supervising the Robodebt Scheme had treated the welfare recipients as “cheats”.⁶⁷

Similarly, dataset bias in the context of personalised intelligent banking can emanate from, for instance, a bank’s preferences for high net worth customers compared with vulnerable groups, such as users with underlying medical conditions, or users who have experienced emotional shock or simply lacks literacy.⁶⁸ Culture, values and beliefs (or ideology), as well as the psyche of belonging to a “tribe” shape our thinking and can contribute to dataset bias in this sense.⁶⁹ A claimant will need to identify with precision whether it is the dataset, or the AI model--or both--which has resulted in the AI harm. The permutations of harms are various but, as explained in Part II(B), bias in this sense generally involves discrimination and unfairness.

For instance, in a rapidly ageing population relative to the general population,⁷⁰ if a bank’s dataset is not representative of the ageing population, it is arguable that the bank owes a *prima facie* duty of care for professional services to a 70 year old customer, through its intelligent banking app which is not adequately inclusive of non-digital natives. In other words, the dataset itself (without the AI model) might be the cause of the AI harm on grounds of the causal, circumstantial proximity, as well as the reliance by non-digital native banking customers on the bank to voluntarily use a dataset which is inclusive and representative to deliver professional banking services and advice.

There are two key implications for dataset bias: the nature of the AI harm, and the claimant’s profile. If the extent of the harm suffered by the elderly claimant, for instance, includes both financial loss and psychiatric harm, it might improve the chances of establishing legal proximity at the first limb of the *Spandeck* test. Even if a *prima facie* duty of care can be established based on this example, a claimant will face a formidable hurdle during the second limb of the *Spandeck* test, i.e., whether a policy exists to negate this *prima facie* duty of care.

One policy with potential to negate a *prima facie* duty of care lies in the overall benefits of the dataset for all banking customers, which can also include both local and foreign customers with a range of age profiles and digital savviness. In other words, this policy is a form of distributive justice, i.e., a loss-spreading device, which benefits the broadest segment of customers and not just non-digital natives. The facts of an actual case will determine the outcome, but the courts through *Spandeck* will be able to adapt the law to govern AI harms in this sense.

B. AI Hallucinations

As explained in Part III, intrinsic and/or extrinsic hallucinations can arise from the dataset and not just the AI models. The relevance of datasets to AI harms are likely to increase as the downstream applications of foundation models grow. For example, HSBC has announced its use of natural language processing (NLP) to enhance institutional interaction with the markets. This includes the NLP’s ability to generate bespoke analytics and gain access to HSBC’s cross-asset datasets. The bank said that its global footprint paired with its NLP offering will allow the bank to deliver an advanced pricing and execution interface for institutional investors.⁷¹

Morgan Stanley is reported to be testing a chatbot with Chat-GPT4.⁷² The aim is to help human financial advisors improve the quality of services to its clients by allowing the chatbot to answer queries from its human financial advisors, by reviewing in seconds the investment bank’s extensive research on capital and analyst commentaries, and data resources. It is unclear if this recourse to OpenAI’s Chat-GPT4 also allows the chatbot access to external resources such as web scraping, but the bank has acknowledged that AI hallucinations is a corollary of this service.

Finally, Deutsche Bank has partnered with Nvidia to develop, among other things, a pilot version of a 3D virtual avatar to support the bank employees’ navigation of its internal systems, such as questions related to human resources. It is reported that the partnership will build on the pilot version by exploring immersive experiences beyond internal use and with banking clients.⁷³

These examples only offer a sample of the nascent but different developments based on a financial services provider's business model and needs. As explained in Part III, large-scale datasets are especially susceptible to intrinsic AI hallucinations when the source and outcome are paired heuristically. Additionally, extrinsic AI hallucinations occur in open-domain dialogue system, i.e., the conversational-styled virtual assistants which banks are actively trying to develop to improve the clients' trust in the quality of banking services. Returning to the first limb of the *Spandeck* test, therefore, it is plausible that since professional services are engaged, circumstantial and causal proximity, as well as the voluntary assumption of responsibility, will be widely used by a claimant in the event of an alleged AI harm.

For these reasons, it seems certain that a bank's terms and conditions will be carefully drafted to ensure that banking customers explicitly consent to the use of datasets, and the potential harms of AI models. This measure might satisfy the reasonable standard of care, the second stage of negligence, in which a defendant owes a claimant in the event of an AI harm.⁷⁴ It is also conceivable that mitigating measures being made by a defendant to the dataset, i.e., to debias or to reduce intrinsic or extrinsic hallucinations, will be favourably considered in terms of the reasonableness in discharging the standard of care owed to a claimant who had suffered harm. Finally, it is an open question whether, given the growing ecosystem of individuals who are involved in annotating and creating datasets, it might lead to a failure of establishing even causal proximity at the first limb of the *Spandeck* test.

C. Implications for Production of Documents and Expert Evidence During Litigation

This account of a dataset's role as a dual contributor of AI harms, and the potential application of *Spandeck*, serves a practical purpose during litigation. The relevance of a dataset can arise during discovery, and in making decisions to choose experts to testify on the impact of harms created by the dataset.⁷⁵ As explained in Part I, the focus on datasets is an effort to promote strategic clarity concerning a dataset's role in contributing to bias and AI hallucinations.

In law, a dataset's contributory role to causing AI harms can be argued to be discrete in legal terms, or together with the AI model. It is important to stress that this paper does not argue that, because of this focus on datasets, it is singularly responsible for the AI harms to the exclusion of the AI models. The purpose of this focus is to draw attention to the role of datasets and its impact on AI harms, apart from the AI model's consequences for AI harms which are extensively examined in the literature.

Two examples of the potential benefits of this focus might illustrate this point. First, under the new Rules of Court ("ROC 2021") in Singapore which govern procedural aspects of a trial, the courts are given more judicial control over the litigating parties as to which documents can be ordered by a court to be requested from each litigating party. Therefore, a request by the claimant from the defendant for documents to prepare for trial (including, for instance, documents from and are adverse to the defendants) might yield sufficient basis for the claimant's lawyer to simply focus on datasets as its case theory to anchor the AI harms, or the need to engage both the dataset and AI models to mount the case theory.⁷⁶

And second, under the ROC 2021, the litigating parties must seek permission from the court to engage an expert to assist the court.⁷⁷ Importantly, the parties must strive to agree on at least one expert witness who must possess specialised technical knowledge.⁷⁸ Given the technical nature of AI harms, it is likely that parties will be inclined to apply for permission to do so. The discovery process will also shape the parties' decisions as to the identification of expert, for which the costs of hiring the expert will have to be shared.

Both examples highlight the necessity for strategic clarity. This is because dataset bias, or AI hallucinations in which datasets play a legal role, dataset bias and AI hallucinations, or just the AI models, can form the case theory. Consequently, the case theory will drive the type of evidence and legal submissions before a court. An expert in testifying for dataset bias might not be a reliable witness to testify for AI hallucinations. An expert who is a recognised expert in both might command higher fees and a claimant, with fewer resources compared with a defendant technology company with deeper

pockets, might not be able to agree on a common expert with the defendant. A case theory that argues AI models and datasets contributed to the AI harm might require more time during discovery and effort to analyse the documents. Again, this potentially increases the legal costs for a client's trial and might not be a practical strategy. In other words, there are strategic reasons in law to characterise the case theory with a focus on one contributor to the AI harms (e.g.: datasets and AI hallucinations) although, from a non-legal perspective, the reasons for the AI hallucinations are more complex and can entail both the AI models and dataset.

Therefore, this paper's focus in framing a dataset as a dual contributor of AI harms with overlapping and distinctive roles might help a legal practitioner to determine the practical decisions in terms of requesting relevant documents for trial, and the strategic choices of experts most appropriate to both parties with different case theories, which must now be mutually agreed.

VI. CONCLUSION

GAI is the latest AI technique to have captured the common imagination. The sentiments range from extreme optimism about GAI's breakthroughs to dystopian visions of human obsolescence. Despite the rapid advances in LLMs, GAI will need time to achieve economies of scale before its full impact is achieved across society.⁷⁹ It seems fair to say that GAI is one of many AI techniques that does not signal an endgame – it is not the “end of history” for AI advancement. It is in this sense that the dataset's key role in AI advances will remain a constant. What is changing is the speed in which data can be extracted to build the dataset growing ecosystem of natural persons, and even GAI itself, to build this dataset.

This paper has surveyed the meaning of a dataset in terms of its dual contributory roles in dataset bias and AI hallucinations. Although it has mainly used Singaporean law to examine this issue, within the context of the EU's AI Act, it has suggested a frame of reference to start a broader conversation, beyond Singapore, about how datasets will assert even greater prominence in this AI zeitgeist, and that the common law can play a stabilising effect on the policy makers' impetus to regulate through medium-based laws.

¹ DJ. Fuchs, “The Dangers of Human-Like Bias in Machine-Learning Algorithms” (2018) (1) *Missouri S&T's Peer to Peer* 2; D. Pedreshi, S. Ruggieri and F. Turini, “Discrimination-aware data mining” (2008) *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 560; B. Friedman and H. Nissenbaum, “Bias in computer systems” (1996) 14 *ACM Transactions on Information Systems* 330.

² “Data” and “datasets” are related but distinct, although both expressions are typically used interchangeably. “Data” refers to the raw and unprocessed information, while a dataset is derived from data through, for instance, organisation and formatting for specific tasks, such as a training dataset for AI models. Given a dataset's directness to the AI models and the harms which arise, the expression “dataset” is used in this paper.

³ J. Adams-Prassl, R. Binns, A. Kelly-Lyth, “Directly Discriminatory Algorithms”, (2023) 86(1) *Modern Law Review* 144; E. Ntoutsis et al, “Bias in data-driven artificial intelligence systems—An introductory survey” (2020) 10 *WIREs Data Mining and Knowledge Discovery* 1.

⁴ See S. Roller et al “Open-domain conversational agents: Current progress, open problems, and future directions” arXiv preprint arXiv:2006.12442 (2020).

⁵ I.e., a category of AI techniques that mimic the patterns and characteristics of a dataset, which generates new and original content.

⁶ See Part III.

⁷ I.e., the EU Artificial Intelligence Act, COM(2021)206.

[https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf).

⁸ See Part V.

⁹ For example, see Singapore's *Model AI Governance Framework* (2nd Edition), 21 January 2020; OECD AI Principles, *Recommendation of the Council on Artificial Intelligence*, 22 May 2019.

¹⁰ I.e., judicial decisions, also known as case law or precedent, which are legally binding and guides the interpretation of future legal cases with similar facts and issues. For our purposes, the common law's design is innately adaptable to technological changes because it is medium-agnostic.

¹¹ See discussion paper by the Singapore Information and Media Development Authority and Aicadium, “Generative AI: Implications for trust and Governance” (2023), 4. https://aiverifyfoundation.sg/downloads/Discussion_Paper.pdf

¹² OECD (2021), “Artificial Intelligence, Machine Learning and Big Data in Finance: Opportunities, Challenges, and Implications for Policy Makers”, 17. <https://www.oecd.org/finance/artificial-intelligence-machine-learning-big-data-in-finance.htm>

¹³ I.e., a probability distribution that draws inferences or makes predictions by approximating the underlying distribution of data (since the real-world data is too complex to model a true distribution of data).

¹⁴ Examples include content generation from text to image (Dalle-2), virtual assistants (see Part V), and data augmentation (Pytorch).

¹⁵ E. Digalaki, Y. Wurmser, “ChatGPT and Generative AI in Banking (Reality, Hype, What’s Next, and How to Prepare)”, *Insider Intelligence*, 30 May 2023. <https://www.businessinsider.com/chatgpt-and-generative-ai-in-banking-how-to-prepare-hype-2023-may>

¹⁶ See the definition of “bias” by the European Commission, Collaboration in Research and Methodology for Official Statistics. <https://cros-legacy.ec.europa.eu/>

¹⁷ *EBA Report on Big Data and Advanced Analytics*, January 2020, EBA/REP/2020/01, 38.

¹⁸ Article 2(2)(a) Racial Equality Directive 2000/43/EC.

¹⁹ Article 2(2)(b) of the Racial Equality Directive 2000/43/EC

²⁰ *Clerk & Lindsell on Torts* (Sweet & Maxwell, 19th Ed, 2006), para 8-04.

²¹ Article 9, GDPR, 2016/679. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

²² A source refers to the location or entity from which the data is extracted. Examples of sources include web scraping, surveys, publicly available data (government database, for example), and social media.

²³ E. Ntoutsis et al, “Bias in data-driven artificial intelligence systems—An introductory survey” (2020) *10 WIREs Data Mining and Knowledge Discovery* 1, 4.

²⁴ Generally, see S. Welleck et al, (2019) “Dialogue natural language inference”, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3731–3741; T. Kamishima et al, “Fairness-aware classifier with prejudice remove regularise”, (2012) *7524 ECML/PKDD (2) Lecture Notes in Computer Science* 35-50; F. Kamiran, T. Calders, (2009) “Classifying without discriminating”, *Computer, Control and Communication, IEEE Computer Society*, 1-6.

²⁵ Values refer to the specific data points or measurements associated with each attribute. An attribute is the characteristic or property of the data.

²⁶ Ł. Stolarski, “Pitch Patterns in Vocal Expression of ‘Happiness’ and ‘Sadness’ in the Reading Aloud of Prose on the Basis of Selected Audiobooks”, (2015) *13 Research in Language* (No 2). <https://czasopisma.uni.lodz.pl/research/article/view/2083>

²⁷ Annotating is a process that adds more information to the attributes so that the dataset’s context and usefulness for downstream applications are improved.

²⁸ J. Dzieza, “AI Is a Lot of Work”, *The Verge*, 20 June 2023. <https://www.theverge.com/features/23764584/ai-artificial-intelligence-data-notation-labor-scale-surge-remotasks-openai-chatbots>

²⁹ See Parts IV and V.

³⁰ For example, see Articles 9(2)(g) and 9(2)(j), GDPR. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>

³¹ Z. Ji et al, “Survey of Hallucination in Natural Language Generation”, (2023) *55 ACM Computing Surveys*, No. 12, 248(3).

³² See Part IV; “KY. Low and S. Heng, “Liability of Maker Towards Subject of Negligent Statement”, *Law Gazette* (Singapore), September 2021. <https://lawgazette.com.sg/feature/liability-of-maker-towards-subject-of-negligent-statement/>

³³ Generally, see S. Wiseman et al, (2017), “Challenges in data-to-document generation”, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*; R. Lebre et al, (2016) “Neural text generation from structured data with application to the biography domain”, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2253–2263.

³⁴ Z. Ji et al, “Survey of Hallucination in Natural Language Generation”, (2023) *55 ACM Computing Surveys*, No. 12, 248(3).

³⁵ Generally, see J. Vincent, “AI is killing the old web, and the new web struggles to be born”, *The Verge*, 26 June 2023. <https://www.theverge.com/2023/6/26/23773914/ai-large-language-models-data-scraping-generation-remaking-web>

³⁶ Z. Ji et al, “Survey of Hallucination in Natural Language Generation”, (2023) *55 ACM Computing Surveys*, No. 12, 248(18).

³⁷ Generally, see C. Zhou et al, (2021), “Detecting hallucinated content in conditional neural sequence generation”, *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. Association for Computational Linguistics, 1393–1404.

³⁸ C. Gardent et al, (2017), “Creating training corpora for NLG micro-planning”, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 179-188.

³⁹ A. Parikh et al, (2020), “ToTTo: A controlled table-to-text generation dataset”, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, 1173–1186.

⁴⁰ A. Roberts et al, (2020) “How much knowledge can you pack into the parameters of a language model?”, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*.

⁴¹ “UN council to hold first meeting on potential threats of artificial intelligence to global peace”, *AP News*, 4 July 2023; “EU AI Act: first regulation on artificial intelligence”, *News European Parliament*, 8 June 2023.

⁴² I.e., a private dispute between parties with financial compensation as the typical remedy that allows the claimant a choice to bring a tort claim against the defendant. This is distinct from a negligent act under criminal law, which must be prosecuted without the need for a victim’s consent: see, for example, Sections 337-338, Penal Code 1871 (Singapore).

⁴³ Generally see K. Simmons, “The Crime/Tort Distinction: Legal Doctrine and Normative Perspectives”, (2008) *17 Widener Law Journal*, No.3, 719-732 at 727-728.

⁴⁴ For a comprehensive treatment, see G. Chan, PW. Lee, *The Law of Torts in Singapore (Second Edition)*, (Singapore: Academy Publishing, 2016), Chapters 5-7.

⁴⁵ *Spandek Engineering (S) Pte Ltd v Defence Science & Technology Agency* [2007] 4 SLR(R) 100, para 73.

⁴⁶ *AYW v AYY* [2016] 1 SLR 1183, paras 78 and 107.

⁴⁷ I.e., occupier's liability is subsumed under the *Spandeck* test in Singapore: See *Toh Siew Kee v Ho Ah Lam Ferrocement (Pte) Ltd* [2013] 3 SLR 284.

⁴⁸ *NTUC Foodfare Co-operative Ltd v SIA Engineering Co Ltd* [2018] 2 SLR 588, paras 46-48, 50; *Spandeck Engineering (S) Pte Ltd v Defence Science & Technology Agency* [2007] 4 SLR(R) 100, para 78.

⁴⁹ This prospect raises the separate issues of service outside jurisdiction and whether Singapore is an appropriate forum to litigate (i.e., *forum non conveniens* doctrine), which are beyond the scope of this paper.

⁵⁰ I.e., the directness of the causal connection or relationship between the particular act or course of conduct and the loss or injury (causal proximity), and an overriding relationship (for the purposes of this paper) of a professional and a client (circumstantial proximity): see *Sutherland Shire Council v Heyman* (1985) 60 ALR 1 at 55-56.

⁵¹ *Ramesh s/o Krishnan v AXA Life Insurance Singapore Pte Ltd* [2015] 4 SLR 1, para 244.

⁵² I.e., at first sight and subject to the second limb of the two-stage test.

⁵³ *Spandeck Engineering (S) Pte Ltd v Defence Science & Technology Agency* [2007] 4 SLR(R) 100, paras 83-85.

⁵⁴ Generally, see D. Tan, YH. Goh, "The Promise of Universality (The *Spandeck* formulation half a decade on)", (2013) 25 *Singapore Academy of Law Journal* 510.

⁵⁵ The European Council (representing all 27 Member States) and European Parliament must agree on a common text, for updates, see <https://artificialintelligenceact.eu/developments/>

⁵⁶ See Articles 5 and 6 of the Act's provisional text, EU Artificial Intelligence Act COM(2021)206.

⁵⁷ T. Madiaga, "Artificial Intelligence Act" Briefing EU Legislation in Progress, June 2023, 5.

[https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf)

⁵⁸ R. Brown, "Europe takes aim at ChatGPT with what might soon be the West's first A.I. law. Here's what it means", CNBC, 15 May 2023. <https://www.cnbc.com/2023/05/15/eu-ai-act-europe-takes-aim-at-chatgpt-with-landmark-regulation.html>

⁵⁹ Artificial Intelligence Liability Directive, COM(2022) 496.

⁶⁰ T. Madiaga, "Artificial Intelligence Liability Directive" Briefing EU Legislation in Progress, February 2023. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI\(2023\)739342_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI(2023)739342_EN.pdf)

⁶¹ Generally, see T. Madiaga, "Artificial Intelligence Act" Briefing EU Legislation in Progress, June 2023, 5. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf)

⁶² D. Blakey, "Mastering the AI advantage: how DBS transformed into a digital leader", *Retail Banker (Analysis)*, 30 January 2023. <https://www.retailbankerinternational.com/analysis/how-dbs-transformed-into-a-world-class-digital-leader/>

⁶³ Generally see "Using synthetic data in banking and financial services", Validata Blog, 22 September 2022. <https://www.validata-software.com/blog/item/460-using-synthetic-data-in-banking-and-financial-services> ; T. Davenport, "The Future Of Work Now: AI-Driven Transaction Surveillance At DBS Bank", *Forbes*, 23 October 2020. <https://www.validata-software.com/blog/item/460-using-synthetic-data-in-banking-and-financial-services>

⁶⁴ A well-known example would be OpenAI's GPT-4: it contains parameters that allow the capture of complex language structures, semantics, and context. Open-AI's GPT is trained on vast amounts of text data from the internet, books, articles, and other sources to learn the statistical patterns and relationships in language.

⁶⁵ D. Mistry, "The 'bank of me'", *Fintech Futures (Analysis)*, 30 June 2023. <https://www.fintechfutures.com/2023/06/the-bank-of-me/>

⁶⁶ D. Mistry, "The 'bank of me'", *Fintech Futures (Analysis)*, 30 June 2023. <https://www.fintechfutures.com/2023/06/the-bank-of-me/>

⁶⁷ *Royal Commission into the Robodebt Scheme (Report. 2023)*, 243.

https://robodebt.royalcommission.gov.au/system/files/2023-07/report_of-the-royal-commission-into-the-robodebt-scheme.pdf

⁶⁸ See the United Kingdom Financial Conduct Authority's examples of consumer vulnerabilities in *Guidance for firms on the fair treatment of vulnerable customers* (February 2021), 13. <https://www.fca.org.uk/publication/finalised-guidance/fg21-1.pdf>

⁶⁹ See D. Kahan et al, "Motivated numeracy and enlightened self-government", (2017) 1 *Behavioural Public Policy*, 54-86.

⁷⁰ By 2030, almost 1 in 4 Singaporeans would be above 65 years old.

⁷¹ A. Smith, "HSBC launches new artificial intelligence global markets service for institutional investors", *Trade News*, 23 May 2023. <https://www.thetradenews.com/hsbc-launches-new-artificial-intelligence-global-markets-service-for-institutional-investors/>

⁷² P. Prakash, "Morgan Stanley is testing OpenAI's chatbot that sometimes 'hallucinates' to see if it can help financial advisors", *Fortune*, 15 March 2023. <https://fortune.com/2023/03/14/morgan-stanley-testing-openai-chatgpt-gpt4-to-help-financial-advisors/>

⁷³ "Deutsche Bank, NVIDIA embed AI into financial services", *Frontier Enterprise*, 18 January 2023, <https://www.frontier-enterprise.com/deutsche-bank-nvidia-embed-ai-into-financial-services/>

⁷⁴ I.e., the second stage of proving negligence (standard of care); see Part IV(A).

⁷⁵ Discovery is a process in which parties to a trial request documents to be produced (including the request of documents that might be adverse to the other party's case) to conduct the trial: see Order 11, Rules of Court ROC 2021 (Singapore).

⁷⁶ I.e., the "storyline" of a lawyer's case to the judge, which includes a structure of the arguments and also the type of evidence which must be adduced.

⁷⁷ Order 12, Rules of Court 2021. Generally, see KP Soh, A Zaid Hamzah, "Expert evidence in civil proceedings after the new Rules of Court 2021", (2022) *SAL Practitioner* 12.

⁷⁸ Order 12 rule 1, Rules of Court 2021.

⁷⁹ Generally, see M. Hiltzik, "Artificial intelligence chatbots are spreading fast, but hype about them is spreading faster", *Los Angeles Times*, 13 July 2023.

References:

-
- D. Blakey, "Mastering the AI advantage: how DBS transformed into a digital leader", *Retail Banker (Analysis)*, 30 January 2023. <https://www.retailbankerinternational.com/analysis/how-dbs-transformed-into-a-world-class-digital-leader/>
- R. Brown, "Europe takes aim at ChatGPT with what might soon be the West's first A.I. law. Here's what it means", *CNBC*, 15 May 2023. <https://www.cnn.com/2023/05/15/eu-ai-act-europe-takes-aim-at-chatgpt-with-landmark-regulation.html>
- G. Chan, P.W. Lee, *The Law of Torts in Singapore* (Second Edition), (Singapore: Academy Publishing, 2016).
- Clerk & Lindsell on Torts* (Sweet & Maxwell, 19th Ed, 2006), para 8-04.
- T. Davenport, "The Future Of Work Now: AI-Driven Transaction Surveillance At DBS Bank", *Forbes*, 23 October 2020. <https://www.validata-software.com/blog/item/460-using-synthetic-data-in-banking-and-financial-services>
- E. Digalaki, Y. Wurmser, "ChatGPT and Generative AI in Banking (Reality, Hype, What's Next, and How to Prepare)", *Insider Intelligence*, 30 May 2023. <https://www.businessinsider.com/chatgpt-and-generative-ai-in-banking-how-to-prepare-hype-2023-may>
- J. Dzieza, "AI Is a Lot of Work", *The Verge*, 20 June 2023. <https://www.theverge.com/features/23764584/ai-artificial-intelligence-data-notation-labor-scale-surge-remotasks-openai-chatbots>
- B. Friedman and H. Nissenbaum, 'Bias in computer systems' (1996) *14 ACM Transactions on Information Systems* 330.
- DJ. Fuchs, "The Dangers of Human-Like Bias in Machine-Learning Algorithms" (2018) (1) *Missouri S&T's Peer to Peer* 2.
- C. Gardent et al, (2017), "Creating training corpora for NLG micro-planning:", *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 179-188.
- M. Hiltzik, "Artificial intelligence chatbots are spreading fast, but hype about them is spreading faster", *Los Angeles Times*, 13 July 2023.
- Z. Ji et al, "Survey of Hallucination in Natural Language Generation", (2023) *55 ACM Computing Surveys*, No. 12, 248(3).
- D. Kahan et al, "Motivated numeracy and enlightened self-government", (2017) *1 Behavioural Public Policy*, 54-86.
- F. Kamiran, T. Calders, (2009) "Classifying without discriminating", *Computer, Control and Communication, IEEE Computer Society*, 1-6.
- T. Kamishima et al, "Fairness-aware classifier with prejudice remove regularise", (2012) *7524 ECML/PKDD (2) Lecture Notes in Computer Science* 35-50.
- R. Lebrete et al, (2016) "Neural text generation from structured data with application to the biography domain", *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 2253-2263.
- KY. Low and S. Heng, "Liability of Maker Towards Subject of Negligent Statement", *Law Gazette* (Singapore), September 2021. <https://lawgazette.com.sg/feature/liability-of-maker-towards-subject-of-negligent-statement/>
- T. Madiega, "Artificial Intelligence Liability Directive" *Briefing EU Legislation in Progress*, February 2023.

[https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI\(2023\)739342_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI(2023)739342_EN.pdf)

T. Madiega, “Artificial Intelligence Act” *Briefing EU Legislation in Progress*, June 2023, 5.

[https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf)

D. Mistry, “The “bank of me””, *Fintech Futures (Analysis)*, 30 June 2023.

<https://www.fintechfutures.com/2023/06/the-bank-of-me/>

E. Ntoutsi et al, “Bias in data-driven artificial intelligence systems—An introductory survey” (2020) *10 WIREs Data Mining and Knowledge Discovery* 1.

A. Parikh et al, (2020), “ToTTo: A controlled table-to-text generation dataset”, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, 1173–1186.

D. Pedreshi, S. Ruggieri and F. Turini, ‘Discrimination-aware data mining’ (2008) *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 560.

P. Prakash, “Morgan Stanley is testing OpenAI’s chatbot that sometimes ‘hallucinates’ to see if it can help financial advisors”, *Fortune*, 15 March 2023. <https://fortune.com/2023/03/14/morgan-stanley-testing-openai-chatgpt-gpt4-to-help-financial-advisors/>

J. Adams-Prassl, R. Binns, A. Kelly-Lyth, “Directly Discriminatory Algorithms”, (2023) *86(1) Modern Law Review* 144

A. Roberts et al, (2020) “How much knowledge can you pack into the parameters of a language model?”, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*.

S. Roller et al “Open-domain conversational agents: Current progress, open problems, and future directions” *arXiv preprint arXiv:2006.12442* (2020).

K. Simmons, “The Crime/Tort Distinction: Legal Doctrine and Normative Perspectives”, (2008) *17 Widener Law Journal*, No.3, 719-732 at 727-728.

A. Smith, “HSBC launches new artificial intelligence global markets service for institutional investors”, *Trade News*, 23 May 2023. <https://www.thetradenews.com/hsbc-launches-new-artificial-intelligence-global-markets-service-for-institutional-investors/>

L. Stolarski, “Pitch Patterns in Vocal Expression of 'Happiness' and 'Sadness' in the Reading Aloud of Prose on the Basis of Selected Audiobooks”, (2015) *13 Research in Language* (No 2). <https://czasopisma.uni.lodz.pl/research/article/view/2083>

D. Tan, YH. Goh, “The Promise of Universality (The Spandek formulation half a decade on)”, (2013) *25 Singapore Academy of Law Journal* 510.

J. Vincent, “AI is killing the old web, and the new web struggles to be born”, *The Verge*, 26 June 2023. <https://www.theverge.com/2023/6/26/23773914/ai-large-language-models-data-scraping-generation-remaking-web>

S. Welleck et al, (2019) “Dialogue natural language inference”, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3731–3741.

S. Wiseman et al, (2017), “Challenges in data-to-document generation”, *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*.

C. Zhou et al, (2021), “Detecting hallucinated content in conditional neural sequence generation”, *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics*, 1393–1404.