Biases in an Automated Decision Making System and its Effect on Individual Rights

Ms. Nischali Varanasi*

Abstract

With big data driving today's prediction based technology, one has to bear in mind that all predictive knowledge and patterns collected from individuals through profiling is nothing but technologies programmed to omit certain data either by choice or due to lack of tools for translation of data. This, over a period of time marks the genesis of a bias, whether conscious or unconscious, being ingrained into the systems which then tends to create an asymmetry of knowledge and concentration of power in select few, thus paving way to a programmed 'Digital Tyranny'. This may potentially result in compromise of interest of the ultimate beneficiaries and jeopardise the core reason behind adoption of these technologies, especially if such technology-assisted automated-decision-making systems (ADMs) are used by State administrative bodies. Further, with lack of legislative sanction the adoption of these ADMs by the State becomes ultra-vires.

The objective of this paper is to broach upon the effect of ADMs on an individual's rights and analyse the point where the Rule of Law begins to be eclipsed by an algorithm. This paper endeavours to funnel down essential requirements that need to be factored while building robust, trust-worthy and sustainable ADMs keeping in mind the legitimate expectation of the subjects of a State in order for an administrative decision being well within the realm of law. The study

* Doctoral Student at National Academy of Legal Studies and Research (NALSAR), Hyderabad, Telangana

aims to explicate that, algorithms which mirror into their architecture and design the principles of natural justice and fairness with a pre-emptive approach would not only help keep the biases under check but also withstand the dynamism of a self-learning system.

Bias in an ADM: Introduction

"Success in creating AI would be the biggest event in human history. Unfortunately, it might also be the last, unless we know how to avoid the risks." - Stephen Hawking.

All technological inventions have both an upside and downside to them. An informed use of technology, by recognizing the pitfalls while circumventing the risks that such technology pose, is the key to using technology in a way that enriches the human civilization. Any technological advancement that has the human at the centre holding the reins can fuel the growth of a society. Artificial Intelligence (AI) as a technology has come a long way amidst debates on whether AI assisted technology is a boon or bane. AI is being harnessed in various domains – from early detection and targeted treatment of cancer to crime prediction through pattern analysis. While this is the silver-lining of these technologies the dark side could be use of these technologies to automate attacks such as automated spear phishing, face-recognizing armed drones or drones smuggling contrabands, or subversion of other AI systems by poisoning data sets[1].

While remaining focused on the silver-lining, one needs to understand that the fuel that drives any AI technology, and more so a predictive AI technology, is data. The underlying data is analyzed to understand how it can be used to make future predictions that are more real-time. This is done through data profiling which is the first step in

[1] *Artificial Intelligence and Robotics for Law Enforcement,* INTERPOL, https://www.interpol.int/en/News-and-Events/News/2020/Artificial-Intelligence-and-law-enforcement-challenges-and-opportunities

preparing the dataset data for predictive analytics, and it is essential for clarifying the structure, content (features), and relationships of the dataset for predictive modelling[2]. Profiling of individuals is the practice of analysing correlated big data in a particular space and categorizing individuals, basis certain predefined parameters to create predictive knowledge or patterns which are thereby used in decision-making. Though the big data used for the exercise of profiling is collected from individuals who are flesh and bone, the analysis over a period of time creates potential digital proxy individuals: individuals who are not flesh and bone but still exist and have a digital presence and identity. In Roosendaal's words "*a digital persona is a digital representation of a real-world individual, which can be connected to this real-world individual and includes a sufficient amount of (relevant) data to serve, within the context and for the purpose of its use, as a proxy for the individual*".

Though there seem to be multifarious advantages being achieved through these predictive profiling technologies, due to the creation of proxy individuals over time, profiling tends to pose a threat to the quality of liberal democracy and equality in a society. One has to, in such a scenario, understand the potential risks that are posed by these technologies and design around them to avoid the risks posed.

Along with creating an innate and subconscious dependency on such technologies it also increases the risks of discrimination through 'Digital Stereotyping' which in turn fuels inequality and stigmatization through decision-making processes which are biased and inaccurate. This tends to create a knowledge asymmetry and centralization of power in a select few resulting in a threatened state of 'Digital Tyranny' which may potentially result in the compromise of the interest of the ultimate beneficiaries and jeopardise the core reason driving the adoption of these technologies.

[2] *Data Profiling*, DATAROBOT, https://www.datarobot.com/wiki/data-profiling/

Though profiling creates a sense of power to humans in their own world while taking decisions using these techniques, we tend to forget that all that one can't translate through machine language, or data that is omitted by choice or due to lack of tools for inclusion or through sheer ignorance of the existence of the data itself, becomes absorbed into a data black-hole which then over a period of time marks the genesis of a bias, whether unconscious or conscious, being ingrained into the systems. It becomes important to address this bias, as machine learning ADM systems are developed and trained using historical data and are more than prone to mirror the existing biases into their future self-learning models. This is very akin to a child absorbing and ingraining the stereotypes and biases in the environment in which he or she develops, which if left unprogrammed or unchecked for, consciously and with mindfulness, may trickle down to the next generation and the generations to come. Having said that, a domino effect of risks and consequences arising therefrom are inevitable, especially if administrative or socio-economic decision making by the State or State entities are driven by such ADM tools and technologies wherein the bias has not been proactively identified and eliminated to withstand the dynamism of a self-learning system.

Effect of Profiling Technologies

I. Profiling Technologies and Human Rights

Profiling technologies, if not adequately channelised, may potentially have a *two-pronged effect*: one at a societal level and another at an individual level (who are none other than the people actually forming the society in the former level). These technologies tend to trespass upon the values that primarily drive a society in achieving democracy – Non-discrimination; Equality before Law and Equal Protection of Laws; Rule of Law and Right to Due Process of Law; Right to Life and Personal Liberty including Right to Privacy and Data Protection, Right against Self-Incrimination, which in-turn affects Autonomy and Self-Determination including Informational Self-Determination. Informational Self-Determination would mean that an individual has a choice to share the personal information sought so that the choice when exercised does not affect the existence of that very choice, and is further insulated against the un-limited collection, storage, use and sharing of personal data and information produced therefrom. Only when such control exists on the information produced the individual is said to have Informational Self-Determination[3]. The control that an individual has over the data and information produced about himself, is a one of the (necessary but insufficient) preconditions for him to live a life said to be 'self-determined'. More so in today's age where personal data (genetic and/or digital) have become proxies for persons, with the intensification of governmental 'identity projects' this aspect has gained significance[4].

These values have been upheld by various human rights charters internationally. Right to Privacy and Data Protection are recognised as inherent and inalienable human rights under Article 12 in the Universal Declaration of Human Rights, 1948 which states that, "*No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks*"[5]. Further, the International Covenant on Civil and Political Rights, 1976 also recognises this under Article 17 by

[3] 1 BvR 209, 269, 362, 420, 440, 484/83 (1983).

[4] ANTOINETTE ROUVROY AND YVES POULLET, THE RIGHT TO INFORMATIONAL SELF-DETERMINATION AND THE VALUE OF SELF-DEVELOPMENT: REASSESSING THE IMPORTANCE OF PRIVACY FOR DEMOCRACY, REINVENTING DATA PROTECTION? 51 (Serge Gutwirth *et al.*, 2009)

[5] Universal Declaration of Human Rights, G.A. Res. 217 (III) A, U.N. Doc. A/RES/217(III) (Dec. 10, 1948), https://www.un.org/en/about-us/universal-declaration-of-human-rights

protecting the individual against ‘unlawful’ attacks on one’s honour and reputation. This right has also been extended to children through Article 16 of the Convention on the Rights of the Child, 1990. India is a signatory to these international instruments. Further, the Indian Constitution *vide* Part III (Fundamental Rights) also lays down safeguards to protect these democratic values. Hence profiling technology designed and adopted also needs to function within the framework of these rights and values in order to be accepted as systems functioning within the realm of law; not trespassing upon human rights; to be trust-worthy and continue to be a sustainable go-to option for the stakeholders.

II. Journey from ‘Rule of Law’ to ‘Rule of Algorithm’

This two-pronged effect is of significance when a state entity or function relies in its decision-making process on ADM systems trained on data sets collected through technologies with innate biases – subjecting the individual to a biased, arbitrary and unfair treatment by the State; and the State consciously or unconsciously ends up *potentially* practising Redlining – “a discriminatory practice that puts services (financial and otherwise) out of reach for residents of certain areas based on race or ethnicity”.[6] “A systematic denial of mortgages, insurance, loans, and other financial services based on location (and that area’s default history) rather than on an individual’s qualifications and creditworthiness felt the most by residents of minority neighbourhoods”.[7]

Further to this, as the dependency on these technologies increases the society would inevitably undergo a journey of transforming itself from a ‘Rule of Law’ into a society being governed by the ’Rule of

[6] Will Kenton, Redlining (Mar. 22, 2021), https://www.investopedia.com/terms/r/redlining.asp.

[7] *Id.*

Algorithm' where, in the guise of being only the facilitating code, the algorithm actually overrides the law. In this journey the point of check, especially within public administration, would be where the following question is answered: "At what point does the automated decision-making in public administration leave the realm of rule of law and turn it into a rule of algorithm?"[8] This is the point where the algorithm would transform into law thereby replacing the legislation enacted by the Parliament. One way to check this transformation is to have a legislative safeguard on the use of the ADM *perse*.

The decision by a State entity to use an automated system, which decision does not have a legislative approval, renders such decision of the State administration to be qualified as arbitrary, questionable and *ultra-vires* and ends up over-riding the premise of good governance. The use of an ADM, in the absence of a legislative approval, weakens the preventive safeguards ex-ante of law endeavoured to be achieved through administrative proceedings as well as reactive safeguards (ex post), that is the legal safeguards or legal guarantees after the administrative decision is made as the legislations and procedural rules have been created on the fundamental premise that a human would be the decision maker[9]. Hence, it becomes a pre-requisite to have an ADM being used by a State authority in its administration implemented through an Act of Parliament that lays down suitable measures to safeguard the individual rights and interests.

Building Sustainable Trustworthy ADMs

While formalizing the use of an ADM by the Sate through a legislative action being the last step one needs to take in the process of

[8] Markku Suksi, *Administrative Due Process When Using Automated Decision-Making in Public Administration: Some Notes from a Finnish Perspective*. SpringerLink (May 22, 2020), https://link.springer.com/article/10.1007/s10506-020-09269-x

[9] *Id.* at 4.

implementing technology into the State functions, building a robust and trustworthy AI would be a prerequisite.

The State, through its established past practices, is obligated to function in a consistent and predictable manner and this creates certain reasonable legitimate expectation in the minds of the individuals. This, popularly known as the Doctrine of Legitimate Expectation, though not a right *perse* is a doctrine based on the principles of natural justice and fairness and endeavours to prevent the State authorities from abusing their power. Hence, the algorithms built keeping in mind the legitimate expectation of the individuals based on the principles of natural justice can pave the path to check the transformation of 'Rule of Algorithm' into law which may otherwise led to the violation of the fundamental right enshrined vide Article 14 of the Indian Constitution - "*The State shall not deny to any person equality before the law or the equal protection of the laws within the territory of India*". One of the ways to the guarantee of non-arbitrariness enshrined under Article 14 is by imbibing the doctrine of substantive legitimate expectation[10] and mirroring it into the software architecture and design.

To achieve this, it is the authors opinion that a regulatory framework for such systems should be designed by adopting a pre-emptive approach, which approach is built into the system design and architecture itself along with the principles of natural justice and transparency, alongside with a real-world performance monitoring being adhered to by the manufacturer of the ADM.

A decision, especially an administrative one, adopted on basis of an ADM should state that the decision so taken was driven by an algorithm and the main characteristics of such algorithmic implementation should be communicated to the individual in question[11]; such an adoption of the ADM should be authorized through an Act

[10] State of Jharkhand v. Brahmputra Metallics Ltd, (2020) S.C.C, 968.

[11] Suksi, *supra* note 9, at 5

enacted by the Parliament; the individual decision must be subject to administrative recourse and if there be a dispute that arises, the state's decision along with the characteristics of the algorithm should be subject to a judicial review; certain specific cases may mandate a human-in-loop approach and restrict decision making on the basis of an algorithm alone; in areas specific to administrative decision-making the use of rule-based ADM may be preferred and mandated over a machine-learning ADM[12].

The High-Level Expert Group on Artificial Intelligence set-up by the European Commission has proposed seven key requirements that AI systems should adhere to for being deemed trustworthy and ethical which are as under,

i. Human agency and oversight: Human-in-the-loop, human-on-the-loop, and human-in-command approaches should be adopted[13];

ii. Technical robustness and safety: A back-up plan has to be devised and kept ready for implementation in case of any eventuality arising due to adoption of an AI system thereby making the adoption of such systems resilient and secure. This makes the adoption of such systems accurate, reliable and reproducible while minimizing and preventing any unintentional harm that may not have been foreseen[14].

iii. Privacy and data governance: Robust data governance mechanisms to be put in place in order to ensure data integrity and quality along with ensuring legitimate access to such data, while respect for data privacy and protection is being harnessed[15].

[12] *Id*

[13] ETHICS GUIDELINES FOR TRUSTWORTHY AI, (Apr. 08, 2019), 16.

[14] *Id.*

[15] *Id.*

iv. Transparency: Transparency is the key to an effective AI system. All the models, system and data should be transparent with no opacity and this should be the endeavoured while building an AI system. Traceability mechanism including for error localization should be incorporated to aide in transparency. The humans should be made aware that they are interacting with an AI system and further all the known shortcomings as well as capabilities of such a system should be made known to the human in question[16].

v. Diversity, non-discrimination and fairness: Though bias cannot be completely eliminated, it may be minimised by making a conscious choice to avoid it. Further, accessibility to the AI system may reduce discrimination while inching towards diversity and endeavour to incorporate fairness; involving all the relevant stakeholders throughout the lifestyle of the system may also be a step closer to diverse, non-discriminate and fair use of AI.[17]

vi. Societal and environmental well-being: A sustainable AI system should be designed by considering environment as well as society and the impact on them should be assessed.[18]

vii. Accountability: Audit enables the stakeholders to assess the lacunae and pitfalls of a system. Likewise, an audit of an AI system aids in the assessment of the underlying algorithms, its design and data analytics being implemented to produce the outcome. This plays a key role, especially in systems that influence state functions such as administrative decisions that engage with human rights and values influencing a democratic society[19].

[16] *Id.*

[17] *Id.*

[18] *Id.*

[19] *Id.*

Robust Data: Backbone of a Trustworthy ADM

While these principles have been suggested for building a trustworthy AI system, the backbone of building any such systems is the data that is being fed into the system and the analytics in the backdrop of this. Analysing the underlying Big Data using various technologies while balancing the data's integrity with right to privacy, right to data protection and the right to informational self-determination among other rights, of the data subject is the key to any system built on profiling. The term "Big Data" signifies extremely large sets of data that are analyzed to infer patterns, trends and correlations. According to the International Telecommunication Union, Big Data are "a paradigm for enabling the collection, storage, management, analysis and visualization, potentially under real-time constraints, of extensive datasets with heterogeneous characteristics"[20]. Hence, the data is the backbone for any ADM and the principles on which this data is collected and processed certainly influences the robustness of the framework on which these technologies are designed and built. Certain of these principles that could be chosen to influence and curate the data could be,

i. *Collection Limitation Principle*: It is essential that the personal data that is being collected should be collated through fair means, with the consent and knowledge of the data-subject and such collection should be within lawful bounds[21].

ii. *Data-Quality Principle*: Data being collected should be related to the purpose for which it is being collected and to such an extent should be accurate, complete and latest. Data's relevance with regard to the context of it's purpose should be kept in mind so that

[20] ITU RECOMMENDATION Y.3600 BIG DATA – CLOUD COMPUTING BASED REQUIREMENTS AND CAPABILITIES (2015), ¶3.2.1

[21] THE OECD PRIVACY FRAMEWORK (2013), ¶14, https://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf?_ga=2.258498697.2042636541.1621335516-719117420.1621335516

only data that is relevant is being collected, thereby ensuring that data being collected is of high quality, thus keeping the noise out[22].

iii. *Purpose Specification Principle*: While it is essential for the data subject to provide consent for data collection in the first instance, the purpose for which such data is being collected is of equal importance and needs to be communicated to the person in question no later than when the data is being provided by him or her. Further any subsequent use of such data should always adhere to the purpose for which such data was provided[23].

iv. *Use Limitation Principle*: The use of the personal data that is being collected, in whatever way it was made available, is to be strictly used for the purposes for which it was disclosed as aforementioned in the Purpose Specification Principle, except otherwise when the data subject consents for the change in the purpose or the change in purpose is authorised by applicable law[24].

v. *Security Safeguards Principle*: Data integrity of personal data collected is the prerogative of the entity collecting such data. Safeguards against risks of loss or unauthorised access, destruction, use, modification or disclosure of data should be prevented through adoption of necessary security measures[25].

vi. *Openness Principle*: All the stakeholders should be provided with the developments, practices and policies, if any, with respect to personal data. There should be policies capturing these aspects and enable the understanding to what extent the Data Controller

[22] *Id.* at 14.

[23] *Id.*

[24] *Id.* at 14.

[25] *Id.* at 15.

has the data stored, the nature of personal data that they hold, the purpose for which such data is being used and the factors identifying the Data Controller[26]. "A Data Controller means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data."[27]

vii. *Individual Participation Principle*: Every individual should be vested with a right to "a) to obtain from a Data Controller, or otherwise, confirmation of whether or not the Data Controller has data relating to them; b) to have communicated to them, data relating to them (i) within a reasonable time, (ii) at a charge, if any, that is not excessive, (iii) in a reasonable manner, and (iv) in a form that is readily intelligible to them; c) to be given reasons if a request made under subparagraphs (a) and (b) is denied, and to be able to challenge such denial; and d) to challenge data relating to them and, if the challenge is successful to have the data erased, rectified, completed or amended"[28].

viii. *Accountability Principle*: The principles have the accountability vested in the Data Controller to ensure compliance with these measures[29].

While these are some of the key principles pertaining data, it is the author's opinion that for any system to be sustainable, building trustworthiness of the stakeholders in the system is essential which is an ongoing process and not a one-time exercise. In addition to what has been discussed so far, this paper brings to the table questions and

[26] *Id.*

[27] REGULATION (EU) 2016/679 (GDPR) § 4(7) (2016).

[28] THE OECD PRIVACY FRAMEWORK, *supra* note 19, at 15

[29] *Id.*

observations that, according to the author, act as a pragmatic guide towards building and sustaining a trustworthy ADM.

1. What is the whole set of data? Does the programmer or user have the visibility of what the entire set of data is and what part of such whole data is being considered and what part of it is not being considered?
2. How is the system handling this data and why (or why not) is it handling the data in a particular way? Does the user have clarity on why a certain set of data is being handled or not being handled in a particular way?
3. Can one build a system completely insulated from bias? An ADM, in all pragmatic sense, may not be unbiased. Building a system completely insulated from bias being a theoretical argument on one side, the system built may in fact, also be oblivious to the biases it carries, thereby pronouncing itself to be completely free of biases. On the contrary accepting that such a '*unbiased*' system would mean aiming to build a system in pure theory, the question one may necessarily need to ask is - Is there a mechanism that can be built into the system that protects the end user from any potential biases acting from within the system, taking a cue from the transparency achieved through questions 1 and 2 above?
4. How can feed back provided by the environment the ADM is operating within, be used to increase trust in the system thereby making it more acceptable and sustainable in-turn? Essentially this would mean answering - How is feedback being received and ingrained into the system? The feedback could be given by the user and stakeholders of the system including the judiciary, the user of the ADM, the data subject or the party who is affected by

use of an ADM or statistics analysed over a particular period of time.

5. What is the confidence score of the ADM and what are the parameters which have prompted to the threshold of the score?

Pre-Emptive Framework for Design Architecture of an ADM

The Council of Europe Directorate General of Human Rights and Rule of Law[30] has provided principles and guidelines to mitigate the risks and tailored best practices on protection of rights of an individual within the application of Big Data. This, to a large extent, provides a pre-emptive framework within which one can endeavour to design and architecture an ADM that embeds the Rule of Law keeping the Rule of Algorithm in check. It further supports in keeping intact the ex-ante and ex-post safeguards of a legislation, thus ensuring that the ADM does not render itself ultra-vires and encompass the Doctrine of Legitimate Expectation thereby upholding the human rights and values of a democratic society touched upon previously in this article. The best practices that one can choose to integrate into an ADM in the context of the discussion in this paper are,

i. *Ethical and Socially Aware Use of Data*: While processing personal data especially for predictive analysis and decision-making processes, the Data Processors will have to consider the impact of processing Big Data and its social and ethical implications in safeguarding human rights and fundamental freedoms. Data Processor means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller[31]. Personal Data analytics should be done

[30] https://www.coe.int/en/web/human-rights-rule-of-law/home?

[31] REGULATION (EU) 2016/679 (GDPR) § 4 (8) (2016).

in consonance with the ethical values acceptable within the social fabric of the community and without prejudicing norms and values of such a community, as identified through ethics committees, while also safeguarding the human rights of the data-subjects[32].

ii. *Preventive Policies and Risk Assessment:* A precautionary approach of adopting preventive policies to mitigate the risks and impact from the use of Big Data at both an individual as well as a societal level is advocated to ensure that data subject are well protected from the identified risks from processing of their personal data. As touched upon in the preceding section of this paper - Profiling technologies and human rights, the Big Data analytics will have a two-pronged effect: at a micro level on the individual's privacy and data protection; as well as at the macro level on the collective dimension of the societal rights. Hence adoption of preventive policies and risk mitigation strategies advocated would be required to take into consideration the legal, social and ethical impact of Big Data analytics and safeguard the right to equal treatment and to non-discrimination[33].

Data Controllers are advised to examine the likely impact of the data processing on the rights and fundamental freedoms of data-subjects: Identify and evaluate risks of processing activities involving Big Data and any undesirable outcome on the individual's rights and fundamental freedom – right to personal data protection and right to non-discrimination and their social and ethical impacts. Develop a 'by-design' and 'by-default' approach to mitigate the risks by introducing appropriate technical and organizational measures taking into account from

[32] GUIDELINES ON THE PROTECTION OF INDIVIDUALS WITH REGARD TO THE PROCESSING OF PERSONAL DATA IN A WORLD OF BIG DATA (2017), ¶5, https://rm.coe.int/t-pd-2017-1-bigdataguidelines-en/16806f06d0

[33] *Id.* at 5.

the earliest design stages through the entire process of data management, by implementing legal principles along with programming products and services with adequate data protection safeguards. The 'by-default' approach means that, "the measures that safeguard the rights to data protection are the default settings and they are built to ensure that only personal information necessary for a given instance of processing is processed."[34]

Feedback from the environment within which a system operates is crucial for analysing the effectiveness of the system developed and the solution provided. This assessment would be effective when the participation of the stakeholders is encouraged to analyse the effect of Big Data analytics being carried out through these systems and also the effect they have in the design of data processing, and the assessment that is being carried out. The legal, social, ethical and technical implications need to be analysed and experts with adequate professional qualifications and knowledge should be chosen to carry out the analysis of the impact of the system. Once the impact on the fundamental freedoms of the data subjects is assessed, these principles urge the Data Controller to seek supervisory guidance as a next step to mitigate risks foreseen in the space of breach of fundamental rights. In addition to this the Data Controller is required to regularly review the results of the assessment of the system by the stakeholder and the experts and document such assessment along with the solutions integrated to reduce the risks which in-turn are to be considered while taking into account the possible administrative sanctions.

iii. *Purpose Limitation and Transparency:* This principle requires the Personal Data to be processed for a specified and limited purpose which has already been made known to the Data Subject. No data

[34] *Id.* at 5.

processing that, in the view of the data-subject, is unexpected, inappropriate or objectionable should be carried out with the data that is provided by such person in question. "In order to comply with the requirement of free, specific, informed and unambiguous consent and the principles of purpose limitation, fairness and transparency, controllers should also identify the potential impact on individuals of the different uses of data and inform data subjects about this impact."[35] This leaves a choice to the data-subject to share the personal information sought and this provides autonomy and informational self-determination to the data subject while also guarding the data subject from self-incrimination. The principle of Purpose Limitation hence guards against unlimited collection, storage, use and sharing of personal data and information produced therefrom.

The aspect of transparency that is broached upon in this principle requires the assessment process, envisaged by the previous principle of Preventive Policies and Risk Assessment, be made known publicly keeping in mind the sensitivities of secrecy and confidentiality of the data-subject. It proposes that all confidential information of the data-subject be provided as a separate annex which shall not be made public. However this may be provided to the supervisory authority while assessing the risks and weighing the solutions.

iv. *By-design approach*: Basis the assessment of risk identified for an ADM the Data Controller is required to imbibe a by-design approach. Data Controllers or Data Processors, as the case may be, are required to analyze the design supporting the data processing being done through their system and shall work with an agenda to reduce, to the maximum extent possible, the "redundant or marginal data, avoid potential hidden data biases

[35] *Id.* at 6.

and the risk of discrimination or negative impact on the rights and fundamental freedoms of data subjects, in both the collection and analysis stages."[36] The principles further suggest that Data Controllers and Data Processors are required to analyze the robustness and functionality of the by-design solution being adopted on a test data set in order to simulate the real time functionality of the ADM and if the results from these simulations are within the acceptable parameters, in which case the use of the by-design solution may be implemented on a large scale. This testing of the by-design solution would enable the assessment of the inherent bias of the ADM when such system is functioning within the identified parameters that are being used to analyze the data. Such analysis of the underlying bias provides a direction on minimizing or altering the use of data being fed into the system which in turn helps in mitigating the risks identified and reduce the occurrence of unfavourable outcomes which may have arisen during the simulation phase. One way the risk identified can be reduced is through adoption of pseudonymisation of data while keeping intact the applicable data protection principles[37]. Though the principles discourage use of non-sensitive information to derive sensitive information, in the event that sensitive information is being derived from non-sensitive information the same safeguards as were to be extended to sensitive information is proposed under the principles.

v. *Consent and anonymization*: Although free, specific, unambiguous and informed consent is professed by the principles, the data subject can reserve the right to withdraw their consent. Further consent is deemed not given freely where there is imbalance of power between data subject and controller unless

[36] *Id.*

[37] *Id.*

the controller demonstrates that there is no imbalance of power. Further anonymization of data is advocated to ensure the effectiveness of de-identification.

vi. *Human intervention in Big Data decision-making:* A human is at the centre of the decision making and this preserves the human autonomy in the process. Where Big Data driven decisions are likely to affect individual rights the person who is the decision-maker shall provide the underlying reasoning for such processing along with the consequences thereof to the data-subject so affected by such a decision. Further, the Big Data considered for such decision making should be interpreted within the circumstances and context within which it was generated[38].

By working within this pre-emptive framework, the endeavour to design a robust, trustworthy and sustainable ADM can be achieved. Further, the inquiries posed by the author are also predominantly answered by working within this framework. However, one aspect that requires to be delved into additionally is the author's question regarding the completeness and wholesomeness of the data; the visibility of its '*completeness*' to the programmer or the user and the reasons for such data being handled or not handled in a particular way. This can be answered in the affirmative by ensuring Context-Based analytics of the Big Data being analyzed. *Context* is nothing but the cumulative history that is derived from various data points about people, places and things and is essential to an analytic decision process[39]; *Cumulative Context* would be the analysis of how these entities relate over time[40]. Big Data's

[38] *Id.* at 7.

[39] Lisa Sokol & Steve Chan, *Context-Based Analytics in a Big Data World: Better Decision*, IBM (Aug. 20, 2013), http://www.redbooks.ibm.com/redpapers/pdfs/redp4962.pdf

[40] *Id.* at 5

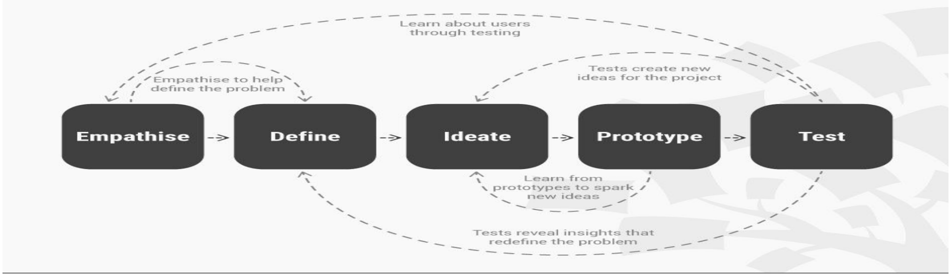
potential in the aide of reasoning urges us to think differently about how one can perform analytics. The Big Data analytics when carried out in a context analytics environment enhances the accuracy of decision-making. The data so generated may be structured or unstructured and may vary in its veracity, volume and velocity. This new environment can provide enhanced capabilities: Increasing the accuracy models and patterns through contextual discovery and visualization analytics; Discover data that is difficult to otherwise mine through such increase in accuracy of these models and pattern assessments; Real time assessment and analysis based on situations and thereby realizing the impact of new data on models and patterns; Discover from such contextual analysis of patterns of life various behavioural predictions. Sometimes vanilla data, devoid of it's context, is important in and of itself[41]. However, most times the relevance of a particular observation being made is better understood and analysed when such data is analysed in its context. Therefore, contextual analytics of historical data results in more accurate predictive data models which when deployed give us a more accurate understanding of the underlying data driving the ADMs.

Conclusion

In the words of statistician George Box "All models are wrong, but some are useful". This paper presumes that no ADM is devoid of bias, however its usefulness is determined by the framework within which it is designed and the realm within which it operates. Every system can be bifurcated into two broad phases – the design phase and the operations phase. This paper endeavours to bring to the discussion table the various threads, from a Rule of Law perspective, that are required to sew through the pieces of designing and operating an ADM for it to be a long-term sustainable solution. The 'design thinking' illustrated below captures the process ideated in this paper while working within the

[41] *Id.* at 2.

boundaries of human rights specifically identified for this discussion, by imbibing principles of natural justice and transparency into the software architecture and design.


Learn about users through testing
Empathise to help define the problem
Tests create new ideas for the project
Empathise
Define
Ideate
Prototype
Test
Learn from prototypes to spark new ideas
Tests reveal insights that redefine the problem

[42]

This leads us to our s*uccess in creating AI which would be the biggest event in human history, as we have paved a path to avoid the risks that we foresee along this journey, in the words of Stephen Hawking.* Further when such *explainable* ADMs are adopted by state run entities to make administrative decisions bolstered by legislative sanctions, these systems aid in faster decision making by the state entities while ensuring substantive legitimate expectation.

[42] Rikke Friis Dam & Teo Yu Siang, *Design Thinking*, INTERACTION-DESIGN.ORG, www.interaction-design.org/literature/topics/design-thinking.