Privacy in the Age of Artificial Intelligence
Table of Contents

Introduction ................................................................................................................................. 3
Defining Artificial Intelligence ..................................................................................................... 4
  Understanding Artificial Intelligence ..................................................................................... 5
  Opening the “Black box” – Explainable AI ............................................................................ 7
Business Models, AI and Personal Data ..................................................................................... 9
What Principles Should Organisations Consider When Dealing with Personal Data?............... 12
How have Organisations Responded? ....................................................................................... 15
Additional Resources ............................................................................................................... 18
Introduction
Before smartphones became a part of our daily lives, personal data online was only used in a small number of transactions such as accessing emails and purchasing items on Amazon or eBay. The advent of smartphones has enabled the proliferation of applications that now fulfil unmet needs in people’s daily lives. In order to access these applications, individuals are required to provide personal information and, in some instances, are asked to provide consent to access other data including personal contact lists. While individuals may eagerly accept such terms and conditions so that they can readily use the application, such terms and conditions (even if they go unread) form a basis for individuals to be wary of how their personal data is used.

As smartphone adoption grew, companies such as Facebook and Google were amassing userbases, which formed large personal datasets at a scale that was previously not achievable before the digital age. These large datasets are then used with algorithms to generate meaningful insights in user activity and preferences, which have been extremely useful to commercial businesses in creating online advertisements targeting specific demographic groups. A prominent example would be using algorithms to automate newsfeeds in Facebook. Here, we see a shift in how data is used – from providing personal preferences to making decisions on our behalf. While individuals were slightly uncomfortable with targeted advertisements, it has been the lack of consent to use personal data for decision-making that has driven individuals to demand more transparency on what personal data is processed and who the data is given to.

In March 2018, it was revealed that Cambridge Analytica had manage to harvest 50 million Facebook user profiles for political purposes. The data collected extended beyond personal information of individuals who agreed to take an online survey that collected personal information – such as including the contact lists of individuals who took the online survey. The profiles harvested were corroborated with data from the US government, for example, on gun ownership, providing a chillingly accurate profile of individuals. With automated newsfeeds, demographic groups with Republican and Democrat slants were fed different information, creating filter bubbles and alternate universes to further shape voter’s opinions. In this example, the implications of automated processes generated through algorithms extend from the digital world into reality.

The use of algorithms is not limited to large Internet companies such as Facebook and Google. Companies from other industries also use unsupervised algorithms to inform decision-making, with reception of the results often varying. In 2015, data of 700,000 individuals were used as training data for a hospital to predict diseases based on patient records, including serious illnesses like liver cancer. The model predicted the onset of psychiatric disorders surprisingly well. Contrastingly, where unsupervised algorithms were also used to rate teachers’ performance in a district in Houston, the teachers filed (and won) a lawsuit citing the use of the AI system as violating their civil rights.

In the examples given, we see an emergence of an entity other than a human being make human-like decisions – giving meaning to the term “artificial intelligence”. These algorithms are now making decisions with ethical implications to the real world. As we continue to embrace the use of artificial

---

3 https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/
4 https://perma.cc/5H25-AQC7
intelligence (AI) for decision making across sensitive and personal spaces, we begin to question who should be accountable for the decision-making process, and how transparent companies need to be in revealing how their algorithms recommend and make decisions. At the same time, individuals are now demanding control over how their data is used and what their data is used for.

Defining Artificial Intelligence

AI has its roots all the way back to 1898, when Nikola Tesla made a demonstration of the world’s first radio-controlled vessel. Fast forward to 1950, Alan Turing published what would come to be known as the “Turing Test”, a test to discern whether a machine’s ability to exhibit intelligence would be indistinguishable from a human’s. Eventually, the term “Artificial Intelligence” was eventually coined by John McCarthy, a Stanford professor and his colleagues in a study proposal for a workshop in 1955.6

The term “Machine Learning” was first introduced by IBM Researcher Arthur Samuel, who introduced a paradigm where programmed machines could be taught to recognise patterns not explicitly coded through the processing of large amounts of sample data. Depending on the use of the algorithms, this dynamic solution could require minimal human intervention for minor changes or absolutely no human intervention at all.

There was a period where AI went through slow speed of progress. However, the 2010s brought about the popularisation of graphical processing units for AI development. Previously only used for image processing, these pieces of hardware were highly adept at high speed mathematics processing, bringing the power of super computers to many research labs. This brought about the rise of Deep Learning. This method of AI (a subset of Machine Learning), relies on large datasets and deep neutral networks to process them. By relying on a simple, repeatable machine learning modules, these highly scalable deep learning networks were able to retain the nuances of complex patterns, leading to significant progress in certain domains of AI, most prominently image recognition, natural language processing and recommendation systems.

---

Some different definitions of AI

**United States Bureau of Census**

Under the Glossary of Selected Abbreviations and Acronyms, AI is defined as “an advanced computer programming language aimed at enabling computers to emulate the human mode of reasoning”. The Organisation for Economic Cooperation and Development (OECD) uses the same definition as the United States Bureau of Census.

**Stanford University**

Nils J. Nilsson, Emeritus Professor of Engineering in the Department of Computer Science at Stanford University, defines AI as “activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment”.

**Singapore’s Infocomm Media Development Authority (IMDA)**

IMDA defines AI as “technologies that aim to simulate human processes or traits, including problem solving”. The IMDA previously referred to AI as “Sentient Technology” in its 2005 5th Infocomm Technology Roadmap, reflecting a shift in the agency’s understanding of AI.

---

Understanding Artificial Intelligence

Figure 1: AI Categories

AI can be broken down into three main categories (Figure 1). The most prevalent form of AI currently is Artificial Narrow Intelligence (ANI), also known as “weak AI”. ANI usually involves a single task and the work delegated is set within a specific or limited context. While the AI may not be able to take on tasks beyond its programmed field, it is able to ferret out correlations and patterns from the data faster than humans. The current applications of machine learning and deep learning mainly use ANI, as most of these algorithms are assigned with task-specific roles. For example, deep learning is used to identify dogs in a huge dataset of photos, but is unable to take on more advanced tasks like altering the photo.

Artificial General Intelligence (AGI) is a tier higher than ANI, or often known as “strong AI”. AGI is when the AI’s understanding and interactions with its environment is on par as that of a human. This means the AI is able to think abstractly, plan, solve problems at a general level, able to innovate, come up with thoughts, and ideas that may not have any precedence. The likelihood of being to teach an AI to invent something that is not there is extremely difficult, which is why AGI will take several more years before it can be implemented as a mainstream technology.

The last tier AI can be classified into is Artificial Super Intelligence (ASI). This is where the AI would demonstrate intelligence beyond human capabilities, often the likes we see in Hollywood movies. ASI would achieve superiority over even the smartest human brains, including social skills.

Source: [https://acceleratingbiz.com/proof-point/future-artificial-intelligence/](https://acceleratingbiz.com/proof-point/future-artificial-intelligence/)

7 https://www.techopedia.com/definition/32874/narrow-artificial-intelligence-narrow-ai
8 https://bdtechtalks.com/2017/05/12/what-is-narrow-general-and-supera-artificial-intelligence/
Researchers do not have an exact timeline as to when AGI and ASI will be reached, but some researchers surveyed in 2015 predicted AI could be better than humans at more less everything in about 45 years.\(^\text{10}\) (These same researchers however, also predicted in 2015 that it would take 12 years for an AI to be better than humans at Go – where in 2017 Google’s DeepMind subsidiary then developed an AI which beat the world’s human Go champions).\(^\text{11}\)

Figure 2: Use of deep neural networks

Machine learning and deep learning form the foundation of most of today’s current AI applications. Machine learning is a subset of AI, a method that gives models the ability to learn without being explicitly programmed to do so;\(^\text{12}\) deep learning is a subset of machine learning that uses neural networks to make computation feasible.\(^\text{13}\)

A network’s ability to reason is embedded in the behaviour of thousands of simulated neurons or neural networks, which have been organised and arranged into hundreds, or even thousands of intricately connected layers; each layer comprising of neurons (Figure 2). Neural networks are able to extract features of the data it has been fed based on algorithms embedded into its programming. The deep learning algorithms can either “classify” or “cluster” the data.\(^\text{14}\)

Algorithms are the pieces of software code, which are a set of unambiguous instructions that a computer can follow to execute. However, the computers will also have to be fed with data and be hosted in an environment in which they can act, before they can be qualified as an intelligent agent. AI cannot function without the data or without the right environment to process this data.\(^\text{15}\)

---


\(^{13}\) [https://blogs.oracle.com/bigdata/difference-ai-machine-learning-deep-learning](https://blogs.oracle.com/bigdata/difference-ai-machine-learning-deep-learning)


In the classification process, neural networks are provided a labelled dataset to train on. This means the pre-determined set, or knowledge, must be taught to the network before it can correctly classify the data according to the labels provided. This is also known as “Supervised Machine Learning”.

In a clustering process, the neural networks discern and group the unlabelled data where similarities are detected. In absence of defining pre-specified attribute, correlations are sought to create a relational database. This is commonly known as “Unsupervised Machine Learning”. Thus, the more data the algorithm is able to train on, the more accurate the final output it likely to be.\(^\text{16}\)

The third type of programming for machine learning is “Reinforcement Learning”, in which an algorithm is provided with a set of rules and constraints. The algorithm is then left to self-learn the best way for it to achieve its goals. Quite often, a reward or point system is involved to ensure the machine learning algorithm will conceive a final output that optimizes outcomes.

**Opening the “Black box” – Explainable AI**

One of the key problems that arise from deep learning is that of the “black box”. Due to the nature of deep neural networking, it is very difficult to ascertain how a decision is made by the AI.\(^\text{17}\) Without this explanation, many other fundamental questions become difficult to answer, such as:\(^\text{18}\)

1. Why did the AI system do that?
2. Why didn’t the AI system do something else?
3. When did the AI system succeed?
4. When did the AI system fail?
5. When does the AI system give enough confidence in the decision that you can trust it
6. How can the AI system correct the error?

---

\(^\text{16}\) IEC Standardisation Management Board

\(^\text{17}\) https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/

Researchers have used a few methods to understand what is going on inside the “black box”. The first way is to control the input that goes into the “black box”. Known as Layerwise relevance propagation, this technique limits the maximum level of correlation a set of variables have on each other. For example, in an algorithm for predicting good leaders, this technique is used to ensure that a current lack of senior women in a data sample will not be used to suggest that gender is an indicator of good leaders. In doing this, that question of “did the AI choose these people because of gender” is made “transparent”.

Figure 3: Techniques for Explainable AI
Figure 4: Model tries to predict whether a patient has flu. “Black box” is opened up by picking up that words like “sneeze” and “headache” contributed to the prediction


The second technique used is by probing the inputs and outputs of the model. By testing variations of the input data, researchers are able to extract what are the most important features of the original input that led the model make a certain prediction. For example, Marco Ribeiro, currently a graduate student at the University of Washington in Seattle, created a programme called Local Interpretable Model-Agnostic Explanations (LIME) to understand which inputs led to the model’s decision. In one of the examples given in his paper,19 by sensitivity testing a model used to predict flu using slight variations of the patient symptom data, LIME picks out “sneeze” and “headache” and “no fatigue” as the reasons why a flu is diagnosed. Doctors are then able to relate this to their own knowledge to make a decision whether the model was coming to conclusions in a trustworthy manner and whether they should put their trust the model’s prediction.

The third technique is the most fascinating as it uses a neural network to explain the “black box” of another network. By using a simple and explainable “encoder” network, different hidden assumptions can be tested in the main model, revealing gaps in the method used by the original AI model to make predictions. These techniques provide the basis for organisations to identify hidden biases, misconceptions and weaknesses in the AI.

We can see through the examples above that although AI models are able give recommendations without human intervention in the model, it is still necessary to have humans involved in the decision-making process. In instances when 1) the cost for machine error is too high, 2) there are class imbalances, the machines cannot answer a question with a high level of confidence, and 3) there is little data available at the present,20 it is recommended that humans regulate the decision-making process, i.e. human-in-the-loop.

Business Models, AI and Personal Data

Corporations have outsourced functions including accounting, HR management systems, and other business functions to other companies, usually smaller ones, to manage costs. The development of models and solutions using AI are done in a similar fashion – outsourced or developed by third-party companies. Another reason for having another company develop solutions using AI is these companies

---

typically have concentrated expertise in building models, software development and system integration (e.g., Google, Amazon, Microsoft etc). For example, Airbnb leverages on Google’s open source library that supports machine learning, TensorFlow, to train datasets to identify room types. Corporations across different sectors are increasingly looking to AI to solve longstanding problems. In March 2018, OCBC established a data science team to develop AI solutions across the bank’s services including insurance and loan financing to offer customers targeted and tailored products and services that are contextually relevant based upon machine learning from customer data.

21 https://medium.com/airbnb-engineering/categorizing-listing-photos-at-airbnb-f9483f3ab7e3
23 https://zencity.io/product/

Figure 5: AI Stakeholders

Source: TRPC

The recent advancements in AI have also contributed to the rise of start-ups using AI to provide solutions for governments. ZenCity, a start-up that recently won at Microsoft’s Innovate.AI competition, is able to aggregate information from users on social media on topics like traffic management, education to provide an understanding to governments on citizens' sentiments. To an extent, developers are able to develop plain vanilla models that can be adapted according to specific use cases.

While we have seen the benefits of AI, for example, through more accurate recommendations and through giving agency to victims, questions still surround whether use of AI would come at the expense of data privacy. For example, Cashé, a start-up in India that offers loans to unbanked young professionals by creating a credit scoring for them, uses personal information from users’ social media (e.g. number of friends on Facebook) to gauge their credit worthiness. While personal data in this instance is used to create an app for a good cause, questions arise on whether data should only be used for the purpose it is obtained for.

The GDPR and Artificial Intelligence

Although the EU’s GDPR does not explicitly mention regulating ‘artificial intelligence’, there are a number of provisions within its framework that looks to regulate how AI may process the data of EU residents.
For example, Article 4(7) differentiates the roles and responsibilities of data controllers and data processors. Where third-party AI developers may use personal data in their models, the roles and responsibilities of the data controller and data processor are immediately relevant. A data controller, according to the European Union (EU)’s General Data Protection Regulation (GDPR) Article 4(7), is “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,” while according to GDPR Article 4(8), a data processor is “a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller”. A data subject is the individual whose personal data is being collected.25 A third-party AI developer in this instance would be categorised as a data processor. As individuals increasingly put pressure on organisations to take data privacy seriously, data processors have the implicit obligation of ensuring the privacy of the data they use. This means that the data should not be available for other stakeholders, and if necessary, anonymised to the extent where it would be difficult for the personal data to be reidentified.

Article 5 which defines the “principles relating to processing of personal data” includes principles on ‘awfulness, fairness and transparency; purpose limitation; data minimization; accuracy; storage limitation; integrity and confidentiality; and accountability.26 While meant to ensure better protection of EU citizens, some of the principles run contrary of AI systems, which tend to first collect as much data as possible, and then only analyse the data after. This makes complying with the purpose limitation and data minimization principles challenging.

Crucially GDPR Article 22 looks to regulate “automated individual decision-making, including profiling” which protects data subjects from decisions based solely on automated processing, including profiling, which produces legal effects or similarly significantly affects the data subject.27 This implies that EU citizens must be offered alternative decision making, with a human being involved, for important decisions such as applications and assessments on mortgages, credit loans, health insurance, job interviews, performance appraisals, school admissions, court rulings, etc. The three exemptions to this however, include (a) where it is necessary for entering into, or performance of, a contract between the data subject and a data controller; (b) is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject’s rights and freedoms and legitimate interests; and (c) is based on the data subject’s explicit consent.28 While this may seem like a great obstacle in using AI systems, organizations can comply by getting explicit consent from data subjects, or working closely with data subjects to ensure transparency on how decisions are made.

Managing Regulation and Consent

Most privacy regulations today around the world were enacted before the use of the Internet of Things (IoT), big data, and AI were a consideration, and where consent provision was defined much simpler. In today’s digital age consumers are creating more data than ever before, yet it is not always clear how and when this data is used, what consent is provided for, nor how consent is provided. IoT use, from consumer, corporation, and government use has also led to the generation and collection of even more data, leading to the use of more and more big data analysis which is now used across sectors. CCTVs

25 https://eugdprcompliant.com/what-is-data-subject/
around the island for example, capture images of individuals among other things, and while they may primarily be used for security or traffic purposes, nevertheless create data sets which can then be used for other purposes. In such instances, are individuals offered the option to provide consent on their data? Likewise, how does one ‘consent’ to Facebook collecting one’s data if one does not even have a Facebook account? Throw in the use of different technologies, automated AI decision making, and cross-sectorial use of data, the issue of regulating data and defining consent becomes all the more complex.

For example, developed for Smart City use, Huawei’s + AI Digital Platform combines the use of different technologies and applications including AI, IoT, big data, a geographic information system, video, cloud, converged communications, and security which can be used across different sectors including smart public safety, environmental protection, transportation, government, education, and agriculture. What is essentially a single service now involves a range of different technologies and use among different industries. The question now for regulators, is how to develop an enabling privacy framework that regulates across different technologies and industries in the face of AI. What current regulations (general and sectorial) remain relevant, and where requires revisions or clarity to be further provided. This likely remains an ongoing and progressive endeavour calling for different stakeholders to work together and work towards.

What Principles Should Organisations Consider When Dealing with Personal Data?

Through the use of digital technologies, there is an ever-increasing amount of data being produced, collected, and analysed today. An adequate regulation or framework for an organisation’s accountability, privacy, and social responsibility around AI will aid in making the procedures for collection of data more secure and more trustworthy. In addition, by incorporating frameworks or policies that are accurate and effective, governments will be able to moderate and manage corporations that seek to abuse data it derives from its users.

And while some of the existing general and sectoral data protection frameworks already regulate how AI systems can process personal data, it is increasingly likely that these frameworks are becoming outdated and do not have the sufficient safeguards necessary to protect against the new privacy threats posed by AI systems.

Around the world, regulators and policymakers are grappling with the challenges in establishing future-proof and technology neutral privacy frameworks to regulate this. For example Article 5 of the EU’s GDPR prescribes principles relating to the processing of personal data which have implications for the development and use of AI systems. In Singapore, the Personal Data Protection Commission (PDPC) has released a discussion paper title “Artificial Intelligence (AI) and Personal Data - Fostering Responsible Development and Adoption of AI” which proposes an accountability based framework on the ethical, governance, and consumer protection issues related to AI. The proposed governance

32 https://www.pdpc.gov.sg/Resources/Discussion-Paper-on-AI-and-Personal-Data
framework put forth by the PDPC seeks to strike a balance between maximizing the benefits of AI without compromising privacy and accountability.

As developers, corporations, and governments continue to embrace AI, building and instilling trust in how AI systems are used will go a long way in shaping the effectiveness of AI-driven applications and systems. For example, people may choose not to use a health app which reminds them when to take their medications if they fear the data collected is not anonymized or are unsure how the information is used.

The following lists some principles organizations can consider when their AI systems interact with personal user data to help build trust and understanding on how data is used.

Transparency: As they participate in the digital economy, consumers are commonly faced with an informational asymmetry as to what data and how much is generated, shared, and processed by their devices, networks, and platforms. There is an anticipated proliferation of more smart and connected devices in homes, offices, and even public spaces, which will further accelerate data generation, sharing and processing. Thus, the importance of educating and creating awareness on when and how AI systems are used, regardless if the decision-making process affects them.

According to the UNI Global Union, a transparent AI system is one in which it is possible to discover how, and why, the system made a decision, and even acted the way it did. Transparency does not refer to simply making the source code available (nevertheless important for audit and inspection purposes), but also that it is made available in a clear and simple manner for users to understand what the system is doing and why. Being transparent to its users is fundamental for organizations in establishing not just trust, but also understanding in how AI systems affect them.

Use of anonymization: When processing or sharing data, organizations should practise the use of anonymization, or the act of deleting or encrypting personally identifiable information to make the data irreversible and where the original data set is not able to be recreated. When used with other data sets, anonymized data should not allow the identification of individuals from the data. Once a personal data set is anonymized, it is no longer considered personal data, and is thus not covered by data protection legislation. However, anonymization should not be seen as a means of reducing regulatory burden, but also as a means of mitigating the risks of data breaches and accidental disclosures.

The effectiveness of anonymization only goes as far as to the extent the data is irreversible. As the Information Commissioner’s Office of the UK (ICO) points out, it may not be possible to establish with absolute certainty a particular dataset is irreversible, especially when taken together with other data that may exist elsewhere. Taking this into account, the focus should not be about eliminating the risk of re-identification completely, but mitigated so the risk of reidentification is extremely remote. Data anonymization techniques include attribute suppression, character masking, pseudonymisation where the PDPC has published a guide on basic data anonymization techniques.

34 http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf
**Data/purpose limitation:** Even when anonymization is not possible, personal data collected should be limited to the purpose for which it was originally collected for, and which consent was provided on. Under the GDPR, the purpose limitation principle specifies that personal data collected must be specified, explicit and lawful, and if the data is further processed for other purposes, it must not be incompatible with the original purpose. The GDPR also makes exceptions for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes.

However, critics point out that AI can be used to potentially uncover unexpected correlations or new purposes for using the data from which it was originally collected for. While data and purpose limitation will help prevent arbitrary reuse of data, healthy governance frameworks can help ensure new purposes and new benefits from the data can still be achieved. Even when a new purpose is discovered, the use of new and innovative ‘notice-and-consent’ models can help ensure consent is obtained for the newly discovered purpose. For example, the use of graduated consent, where users can continuously work with developers to provide consent to different uses of their data throughout the development process.

**Data minimisation:** This principle helps to minimize the amount of data collected and processed by establishing at the onset how the data will be applied and processed to determine what data is relevant and necessary for the purpose. Under the GDPR data minimisation is restricted for data that is adequate, relevant and limited to what is necessary for the purposes to be processed. This helps avoid excessive data collection and the use of continuous assessment of the actual requirements for processing will help ensure that only relevant and necessary data is used.

Likewise, personal data should only be kept for as long as is necessary for the purpose for which it is being processed, where good information governance and appropriate retention schedules developed at the onset can help adhere with data minimisation.

**Privacy by design:** Also called privacy by default under the GDPR, privacy by design considers all the above principles of anonymization, data limitation and data minimization and also includes technical and organizational measures at each stage of the data collection and processing chain. This means considering the privacy and security requirements in each system, operation, and product throughout the full lifecycle in which the data is collected, stored, and processed.

Security measures could include access controls, audit logs and encryption, data segregation so personal data is kept separate from other forms of data, and ‘sticky policies’ that may include attaching conditions and constraints to data that specify how the data should be treated when used for different purposes or by different parties.

---

38 [https://gdpr-info.eu/art-5-gdpr/](https://gdpr-info.eu/art-5-gdpr/)
40 [https://gdpr-info.eu/art-5-gdpr/](https://gdpr-info.eu/art-5-gdpr/)
42 [https://gdpr-info.eu/art-5-gdpr/](https://gdpr-info.eu/art-5-gdpr/)
Conducting privacy impact assessments: Less of a principle and more of a good practise, conducting privacy impact assessments (PIA) can help identify and mitigate privacy risks before the actual processing of personal data. While AI systems may involve new and innovative, complex and sometimes unexpected/unintended uses of personal data, the use of PIAs well help organizations better assess the risks and impacts involved in the use of an AI system processing personal data, and to ensure sufficient and proportional safeguards are in place.

An example of the key tasks involved in a PIA from the PDPC include:

- Identifying the personal data handled by the system or process, as well as the reasons for collecting the personal data
- Identifying how the personal data flows through the system or process
- Identifying data protection risks by analysing the personal data handled and its data flows against PDPA requirements or data protection best practices
- Addressing the identified risks by amending the system or process design, or introducing new organisation policies
- Checking to ensure that identified risks are adequately addressed before the system or process is in effect or implemented

How have Organisations Responded?

Singapore

The Personal Data Protection Commission (PDPC) has released a discussion paper on fostering responsible development and adoption of AI. Designed to create a baseline for discussion, the paper touches on the types of governance frameworks that should be established, as well as the regulatory clarity that should be present. If the governance frameworks are technology-neutral and not drafted prematurely, this would allow the technology to develop unhindered and undistorted. Policies or regulations that adopt a human-centric stance whilst setting transparent, comprehensible, and fair baseline requirements; would create an environment of trust amongst consumers in AI deployments.

Decision making involving AI should be made: i) explainable, ii) transparent, iii) fair, and constantly human-centric. The two key principles put forward in this paper are:

1. Decisions to be made by or with the assistance of AI should be explainable, transparent, and fair to all consumers.
2. AI systems, robots, and decisions made by either should be human-centric.

The Monetary Authority of Singapore (MAS) has released its fairness, ethics, accountability and transparency (FEAT) principles to promote responsible use of artificial AI and data analytics in finance. Under fairness, AI decisions need to be both justifiable and accurate, with regular reviews made to validate accuracy and relevance, and to minimize unintentional bias. Ethics refers to the AI decisions being held to the same ethical standards as human driven standards, and aligned with the firm’s own ethical standards, values, and codes of conduct. Accountability refers to both firms’ internal accountability for AI decisions, and external accountability where data subjects have the means to

enquire and feedback on AI decisions. Lastly transparency ensures data subjects are communicated with and provided explanations on what data is used, how, and the consequences AI decisions made have on them.

The **Infocomm Media Development Authority** (IMDA) has established an “Advisory Council on the Ethical Use of AI and Data” to advise and work on areas for the responsible development and deployment of AI. The Advisory Council, which was appointed by the Minister of Communication and Information, will also seek to develop ethics standards, reference governance frameworks, and publish advisory guidelines, practical guidelines, or codes of practice. The latter subject to voluntary adoption by the industry. The Advisory Council will also assist in the following ways:

1. Engage with relevant stakeholders on ethical and related issues that may arise from private sector use of AI and data, as well as advocate on consumers’ expectations and acceptance of such use.
2. Engage with the private capital community to develop awareness to incorporate ethics considerations during the primary investment decisions into business that may develop or adopt AI technology.
3. Establish a “Legal and Technical Expert Panel” and a “Panel of International Experts” to support the Advisory Council and for global perspectives.

A five-year Research Programme on the Governance of AI and Data Use has been established at the **Singapore Management University** (SMU), to focus on scholarly research on policy, legal, regulatory, governance, ethics, and other issues pertaining to AI and data use. In addition, the programme will also assist the Advisory Council and ideally be a worldwide centre for AI knowledge. The programme was initiated by the IMDA and the National Research Foundation.

**United States**

The New York City Council has enacted a law on algorithmic decision-making transparency to provide the public information on how AI is deployed and used in the city. To do so, a task force is set to be created to provide recommendations on how information on agency automated decision systems may be shared with the public and how agencies may address situations where people may be harmed by agency automated decision systems. 48

**EU**

The High-Level Expert Group on AI in the EU made up of representatives from academia, civil society, as industry is tasked to support the implementation of the European strategy on AI, including recommending on future-related policy development and on ethical, legal and societal issues related to AI, including socio-economic challenges. 49 This includes proposing draft AI ethics guidelines, which cover issues such as fairness, safety, transparency, the future of work, democracy and incorporate the impact on the application of the Charter of Fundamental Rights, including privacy and personal data protection, dignity, consumer protection and non-discrimination.


**New Zealand**

The New Zealand government has released its Government Algorithm Transparency Report providing a review of the government’s use of algorithms through fourteen self-assessed government agencies on their respective use of their algorithms. One of the recommendations made is for published information to better explain how algorithms may inform decisions affecting ordinary people.

**The Partnership on AI**

The Partnership on Artificial Intelligence to Benefit People and Society founded by Amazon, Facebook, Google, DeepMind, Microsoft, IBM, and later Apple, is a technology industry consortium which works together to study and formulate best practices on AI technologies, increase public awareness and understanding of AI, and serve as an open platform for discussion and engagement about AI including its influences on people and society. One of its Work Pillars includes on “Fair, Transparent, and Accountable AI” and developing best practices on the development and fielding of fair, explainable, and accountable AI systems.

---


51 [https://www.partnershiponai.org/about/](https://www.partnershiponai.org/about/)
Additional Resources

1. Personal Data Protection Commission of Singapore (PDPC)
   - Discussion Paper: Artificial Intelligence (AI) and Personal Data - Fostering Responsible Development and Adoption of AI, [https://www.pdpc.gov.sg/Resources/Discussion-Paper-on-AI-and-Personal-Data](https://www.pdpc.gov.sg/Resources/Discussion-Paper-on-AI-and-Personal-Data)

2. Monetary Authority of Singapore

3. KPMG