

*Neither a person nor an apple can be diverse. Diversity is the property of a collection of people—a basket with many kinds of fruit.*

— Scott E. Page

# AI Governance and Ethics Framework for Sustainable AI and Sustainability



Dr Mahendra Samarawickrama (GAICD, MBA, SMIEEE, ACS(CP))

May 18, 2022



<https://bit.ly/AIESG>

# Copyright Notice

Copyright © 2022 Mahendra Samarawickrama

ISBN: 978-0-6454693-0-1

This work is licensed under a Creative Commons “Attribution 4.0 International” license.



This report was submitted to the consultation process of The Australian Department of the Prime Minister and Cabinet for the regulation of artificial intelligence (AI) and automated decision making.

## Third party copyright

Wherever a third party holds copyright in this material, the copyright remains with that party. Their permission may be required to use the material. Please contact them directly.

## Attribution

This publication should be attributed as follows:

M. Samarawickrama, “AI Governance and Ethics Framework for Sustainable AI and Sustainability,” *Submission in response to the Department of the Prime Minister and Cabinet issues paper Positioning Australia as a leader in digital economy regulation - Automated Decision Making and AI Regulation*, Apr. 2022, ISBN: 978-0-6454693-0-1.

# AI Governance and Ethics Framework for Sustainable AI and Sustainability

Dr Mahendra Samarawickrama (GAICD, MBA, SMIEEE, ACS(CP))

## Executive Summary

AI is transforming the existing technology landscape at a rapid phase enabling data-informed decision making and autonomous decision making. Unlike any other technology, because of the decision-making ability of AI, ethics and governance became a key concern. There are many emerging AI risks for humanity, such as autonomous weapons, automation-spurred job loss, socio-economic inequality, bias caused by data and algorithms, privacy violations and *deepfakes*. Social diversity, equity and inclusion are considered key success factors of AI to mitigate risks, create values and drive social justice. Sustainability became a broad and complex topic entangled with AI. Many organizations (government, corporate, not-for-profits, charities and NGOs) have diversified strategies driving AI for business optimization and social-and-environmental justice. Partnerships and collaborations become important more than ever for equity and inclusion of diversified and distributed people, data and capabilities. Therefore, in our journey towards an AI-enabled sustainable future, we need to address AI ethics and governance as a priority. These AI ethics and governance should be underpinned by human ethics.

*Keywords:* AI, Governance, Ethics, Sustainability, ESG (Environmental, Social, and Governance), SDGs (Sustainable Development Goals), DEI (Diversity, Equity, and Inclusion), Social Justice, Framework

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Human Ethics</b>	<b>2</b>
<b>3</b>	<b>AI from the Consequentialism Perspective</b>	<b>4</b>
<b>4</b>	<b>AI from the Utilitarianism Perspective</b>	<b>8</b>
4.1	Bias . . . . .	8
4.2	Diversity . . . . .	11
4.3	Impartiality and Localisation . . . . .	13
4.4	Equity . . . . .	14
4.5	Inclusion . . . . .	15
<b>5</b>	<b>Complexity in AI Governance</b>	<b>16</b>
<b>6</b>	<b>A Framework and a Model for AI Governance</b>	<b>18</b>
6.1	KITE abstraction framework . . . . .	18
6.2	Wind-turbine conceptualised model . . . . .	20
6.3	People, Culture and Mission . . . . .	21
<b>7</b>	<b>Adaptation of the Framework</b>	<b>23</b>
<b>8</b>	<b>Conclusion</b>	<b>26</b>

# List of Figures

2.1	AI is a capability which can transform values of human, data and technologies towards social justice. . . . .	3
2.2	The algorithm of linear regression which fit a straight line cross the data points. Note that human decisions on selection of optimisation problem, data, algorithm, and parameters. How can we ethically govern these decisions? . . . . .	3
3.1	UN Sustainable Development Goals (SDGs) [1]. In 2015, United Nations member states adopted these 17 SDGs as their 2030 agenda for sustainable development. . . . .	5
3.2	Analysis of positive and negative impact of AI on the UN SDGs [2]. Figure courtesy of [2]. . . . .	6
3.3	Results of Australia's SDG assessment [3]. Note the goals in which Australia is <i>off track</i> and needs <i>breakthrough</i> . Figure courtesy of [3]. . . . .	6
3.4	Risk landscape of Australia's SDGs [3]. Australia need to focus these concerns aligning with accelerated economic developments. Figure courtesy of [3]. . . . .	7
4.1	The nature of intuitive decision-making [4]. Figure ©Australian Institute of Company Directors. . . . .	9
4.2	Decision making errors and biases [5]. Figure ©Australian Institute of Company Directors. . . . .	10
4.3	Cultural diversity of Australia and interesting facts. Figure ©Australian Human Rights Commission [6]. . . . .	12

4.4	Categorising continuous variables is important for diversifying the service. Modelling of each category of continuous variable independently, as shown in the figure <b>a</b> can lead to loss of information and poor predictions. On the other hand, modelling the entire data set with a single higher-order polynomial might overfit the model. The figure <b>b</b> shows mathematically complex restricted-cubic-spline regression lines, which can flexibly and accurately model complex and non-linear relationships [7]. . . . .	13
4.5	Relationship of body weight to surrogate measures of fat mass (sum of four SFT) and fat-free mass (height <sup>2</sup> /resistance) in Australians of Aboriginal (filled squares, solid line) and European (open circles, broken line) ancestry [8]. . . .	14
4.6	How K-means clustering getting unsuccessful in non-Gaussian data distribution. The dashed line denotes separating the computed cluster boundaries; filled dots, cluster centres [9]. By bringing reasonable insight, the K-means clustering can be enhanced. . . . .	15
6.1	KITE abstraction framework for AI governance [10]. It aligns with the broader ESG purpose, fundamentally the <i>why</i> aspect of the golden circle. . . . .	19
6.2	Wind-turbine conceptualised model for AI governance [11]. It helps directors address <i>how</i> and <i>what</i> aspects of AI governance. . . . .	20
6.3	The proposed AI governance tools help the corporate board, human resource (HR) and management to orchestrate culture, people and mission towards humanity and sustainability. . . . .	22
7.1	Sustainable AI for Sustainability. Businesses should position their IT, data science, and AI capabilities to address social justice and sustainability strategies. DEI (diversity, equity and inclusion) would be a key success factor of those initiatives. . . . .	24
7.2	Development of sustainable AI as a core competency. AI has been identified as a key enabler for ESG and sustainability. . . . .	25

# 1. Introduction

AI has been identified as the new electricity [12]. Data has been considered the oil for the digital economy. This is also considered the 4<sup>th</sup> industrial revolution. From this perspective, have we thought about the sustainability of the new electricity: AI?

When the steam engine was deployed in the 1<sup>st</sup> industry revolution and electricity was generated in the 2<sup>nd</sup> industrial revolution, sustainability had not been a concern. Humans' rush to economic advantages from the 1<sup>st</sup> and 2<sup>nd</sup> industrial revolutions caused many problems in the long run, such as climate change and related environmental and humanitarian crises [13]. By the time we retrospect and think about the sustainability of power and energy generation, it has caused significant damage to humanity. Therefore, we mustn't be making the same mistake in the 4<sup>th</sup> industrial revolution: AI.

AI governance is a complex process as AI has autonomous decision-making capability. Consequently, AI can create fundamental risks in human dignity, human rights and human autonomy [14], [15], [16]. Hence, AI ethics and governance must be realized from the very beginning when humans initiate artificial intelligence. Therefore AI ethics should be underpinned by human ethics [17].

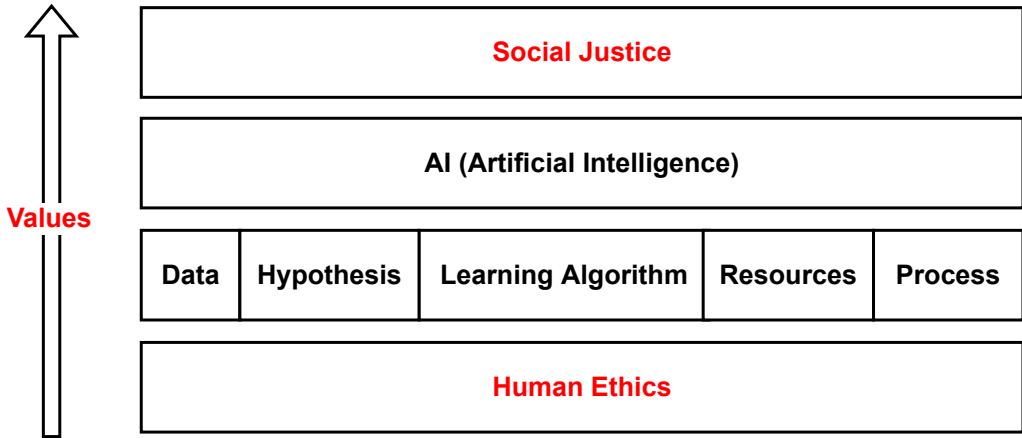
## 2. Human Ethics

*Consequentialism* and *Utilitarianism* can be identified as two broad categories of human ethics. Consequentialism is a theory that says whether something is ethical or not depends on its outcomes or consequences. In this way, the focus is on outcomes rather than the overall benefit or process. In contrast, in Utilitarianism, the ethical nature is decided based on whether the process is optimised to maximise the overall benefit to the society rather than the outcomes. These two different ethical perspectives sometimes create a dilemma, where we may see a decision is ethical in the Consequentialism perspective but not ethical in the Utilitarianism perspective and vice versa. Therefore, the leaders need to understand both perspectives and make sure AI realisation can be justifiable in both perspectives as much as possible.

Human should consider AI as a capability rather than an agent. AI should not take autonomy wherever human dignity is a concern. The fundamental purpose of AI is to transform the values of human, data and technologies towards social justice (see Figure 2.1) by optimising the Consequentialism and Utilitarianism perspectives of human ethics.

In technical perspective, humans are accountable for their decisions on AI implementations:

- bias mitigation,
- problem selection,
- opportunity cost evaluation for social justice,
- data selection and sampling,
- insight (features) incorporation,
- algorithm selection,
- hyperparameter tuning,



**Figure 2.1:** AI is a capability which can transform values of human, data and technologies towards social justice.

- regularisation, etc.

Figure 2.2 shows the basic touch-points of human decision making in a simple form of an AI algorithm, *linear regression*. Note how human decision-making influences a typical AI solution in data, hypothesis, algorithmic, resource and process perspectives. Many tools (e.g., MLOps, ModelOps, AIOps, XOps, DataOps) enable and facilitate deciding and fine-tuning all of those factors and aspects. Our ethics, knowledge and risk appetite determine **why how and what** we do, which is why AI governance and ethics are important.

$$\text{Hypothesis: } h_{\theta}(x) = \theta_0 + \theta_1 x$$

$$\text{Parameters: } \theta_0, \theta_1$$

$$\text{Cost Function: } J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

$$\text{Goal: } \underset{\theta_0, \theta_1}{\text{minimize}} J(\theta_0, \theta_1)$$

**Figure 2.2:** The algorithm of linear regression which fit a straight line cross the data points. Note that human decisions on selection of optimisation problem, data, algorithm, and parameters. How can we ethically govern these decisions?

### 3. AI from the Consequentialism Perspective

AI can support 79% of the United Nations 17 Sustainable Development Goals (SDGs) (see Figure 3.1) [2], which is the foundation of ESG and Social Impact strategies planned to realise by 2030. In 2015, United Nations member states adopted these 17 SDGs as their 2030 agenda for sustainable development [1]. This agenda establishes a shared framework for peace and prosperity for a sustainable future for people and the planet. The framework supports environmental, social and corporate governance (ESG) for sustainability.

In *Consequentialism* perspective of AI ethics, UN SDGs provide a globally acceptable ethical framework for AI governance. However, depending on governance and ethics of AI, there can be pros and cons in AI applications. Figure 3.2 shows how AI impacts positively and negatively on each UN SDGs.

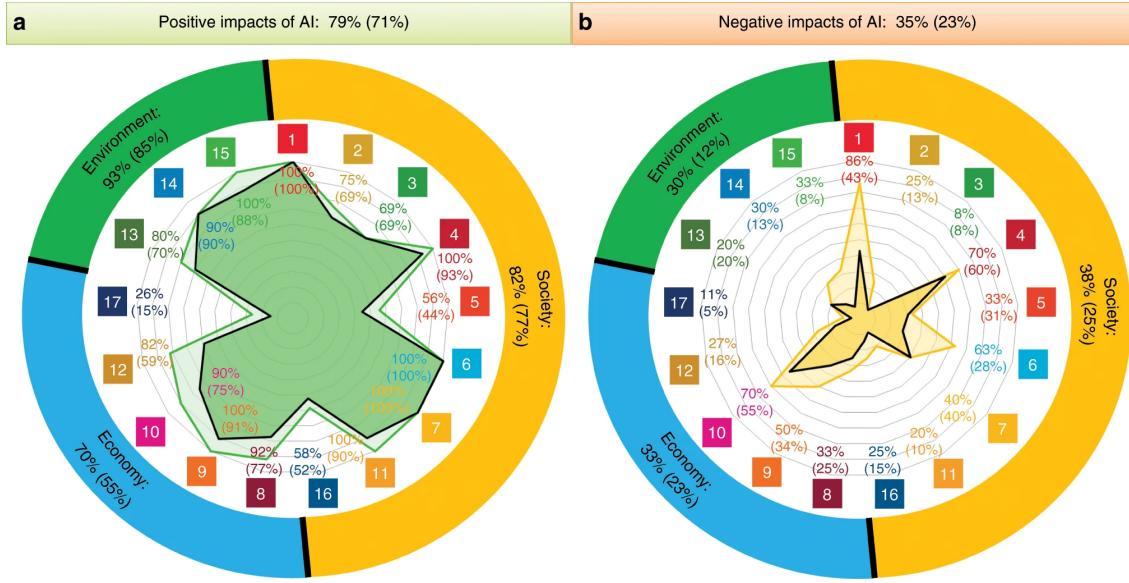
The UN SDGs are an urgent call for action by all countries - developed and developing - in a global partnership; the Australian organisations must address this diligently. Australia still has a long journey ahead in achieving UN SDGs. Figure 3.3 illustrates the results of Australia's SDG assessment [3]. Note the goals in which Australia is *off track* and needs a *breakthrough*. Moreover, Figure 3.4 summarises the Australian concerns related to unsatisfactory progress in each UN SDG analysed in [3]. Therefore, in the economic acceleration effort with AI, the government should focus on achieving UN SDGs effectively, which will promote AI ethics, governance and *AI for sustainability*.



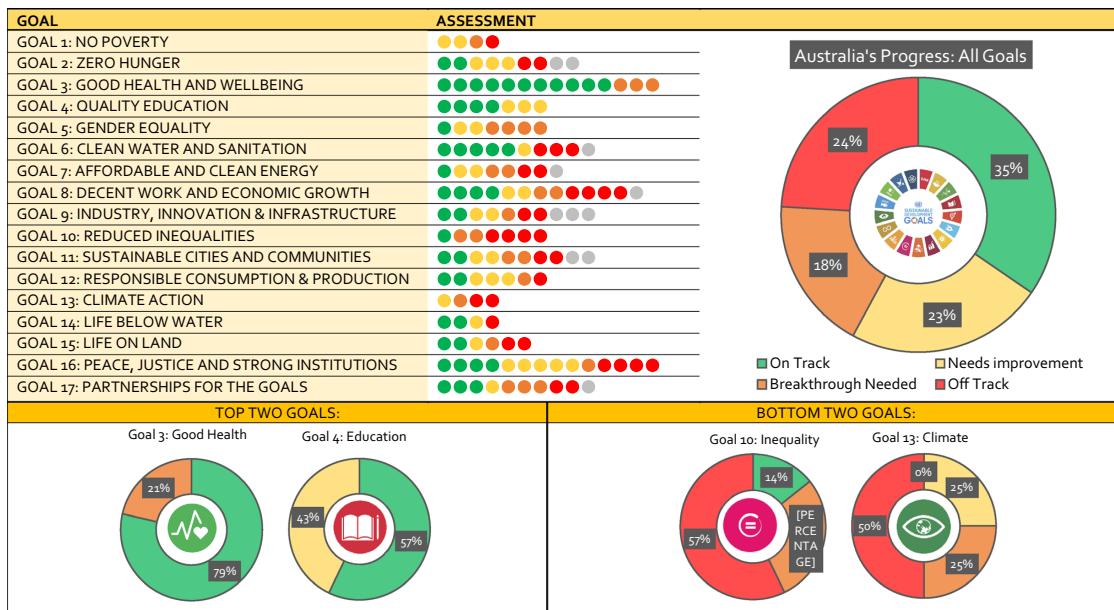
# SUSTAINABLE DEVELOPMENT GOALS



**Figure 3.1:** UN Sustainable Development Goals (SDGs) [1]. In 2015, United Nations member states adopted these 17 SDGs as their 2030 agenda for sustainable development.



**Figure 3.2:** Analysis of positive and negative impact of AI on the UN SDGs [2]. Figure courtesy of [2].



Summary of results from Australia's SDG assessment. Coloured dots represent the assessment outcome for each individual indicator: 'On Track' (green), 'Needs Improvement' (yellow), 'Breakthrough Needed' (orange), 'Off Track' (red), or 'Not Assessed' (grey).

**Figure 3.3:** Results of Australia's SDG assessment [3]. Note the goals in which Australia is *off track* and needs *breakthrough*. Figure courtesy of [3].

GOAL	Where Australia is 'Off Track'	Where a 'Breakthrough is Needed'
GOAL 1: NO POVERTY	Adequacy of welfare payments	Income poverty rate (50% median income)
GOAL 2: ZERO HUNGER	Obesity; Agriculture expenditure	
GOAL 3: GOOD HEALTH AND WELLBEING		Suicide mortality rate; psychological distress; Alcohol consumption
GOAL 4: QUALITY EDUCATION		
GOAL 5: GENDER EQUALITY		Violence against women; Female share of domestic work; Gender superannuation and pay gaps
GOAL 6: CLEAN WATER AND SANITATION	Water affordability; Waterbird abundance; Water-related ODA	
GOAL 7: AFFORDABLE AND CLEAN ENERGY	Energy affordability; Electricity prices	Renewable energy share of final energy consumption and electricity
GOAL 8: DECENT WORK AND ECONOMIC GROWTH	Household debt; Material footprint; Underemployment; Long-term unemployment	Multifactor productivity; Youth employment
GOAL 9: INDUSTRY, INNOVATION & INFRASTRUCTURE	Investment in knowledge-based capital; higher education expenditure on R&D financed by industry	Public infrastructure investment
GOAL 10: REDUCED INEQUALITIES	Distribution of net worth, income and wealth; Labour share of GDP; Non-performing loans	Gini coefficient; Palma ratio (income)
GOAL 11: SUSTAINABLE CITIES AND COMMUNITIES	Housing affordability; Homelessness	Mortgage stress; Urban population density
GOAL 12: RESPONSIBLE CONSUMPTION & PRODUCTION	Hazardous waste	Food waste
GOAL 13: CLIMATE ACTION	Disasters; Greenhouse gas emissions	Per capita GHG emissions
GOAL 14: LIFE BELOW WATER	Great Barrier Reef coral cover	
GOAL 15: LIFE ON LAND	Total forest area; Red List Index (threatened species)	Biodiversity protection (freshwater)
GOAL 16: PEACE, JUSTICE AND STRONG INSTITUTIONS	Sexual assault; Prison population; Civic engagement; Ability to have a say on important issues	Trust in institutions
GOAL 17: PARTNERSHIPS FOR THE GOALS	Government revenue and expenses; Net ODA	Revenue lost due to tax avoidance; Policy coherence; Multi-stakeholder development effectiveness

**Figure 3.4:** Risk landscape of Australia's SDGs [3]. Australia need to focus these concerns aligning with accelerated economic developments. Figure courtesy of [3].

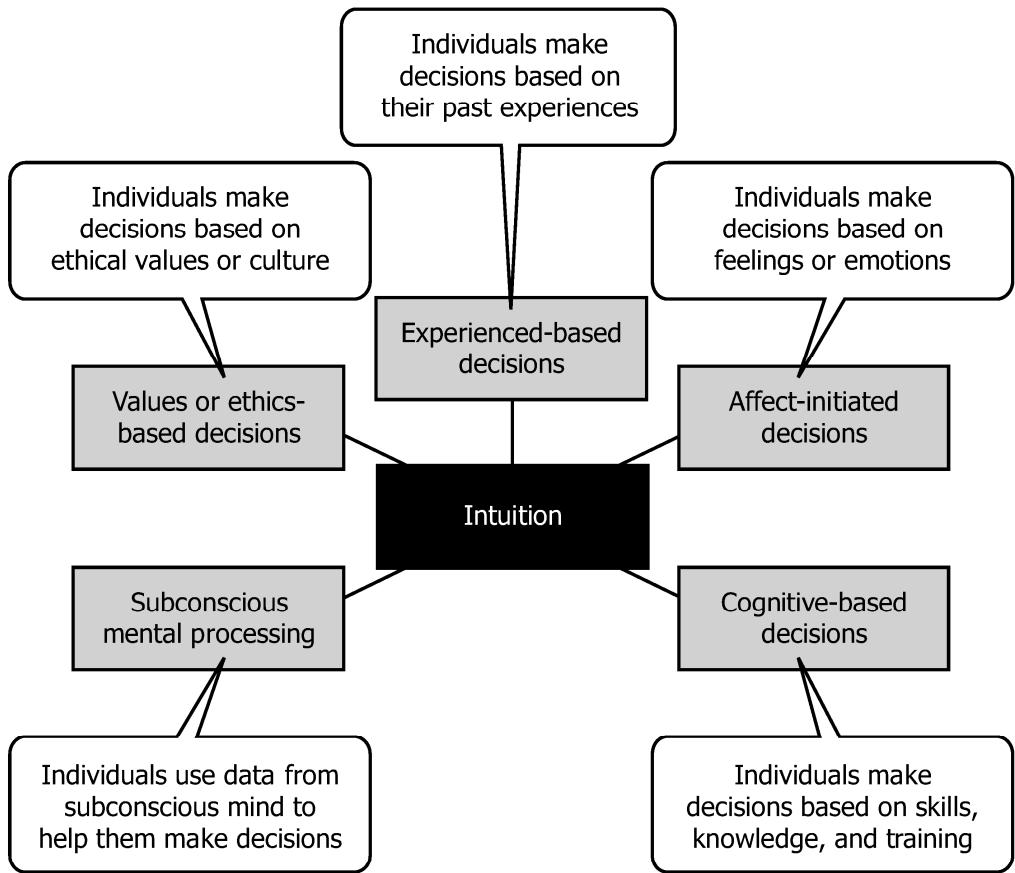
## 4. AI from the Utilitarianism Perspective

In the Utilitarianism perspective of AI ethics and governance, the motivation would be to maximise the overall benefit to the society instead of morality. In this perspective, leaders are encouraged to look into the more granular level and customised design and implementations rather than premeditated norms, moral conventions or solutions (which are more focused on the Consequentialism perspective). The following are important design concerns when focusing on AI ethics and *sustainability of AI* from the Utilitarianism perspective.

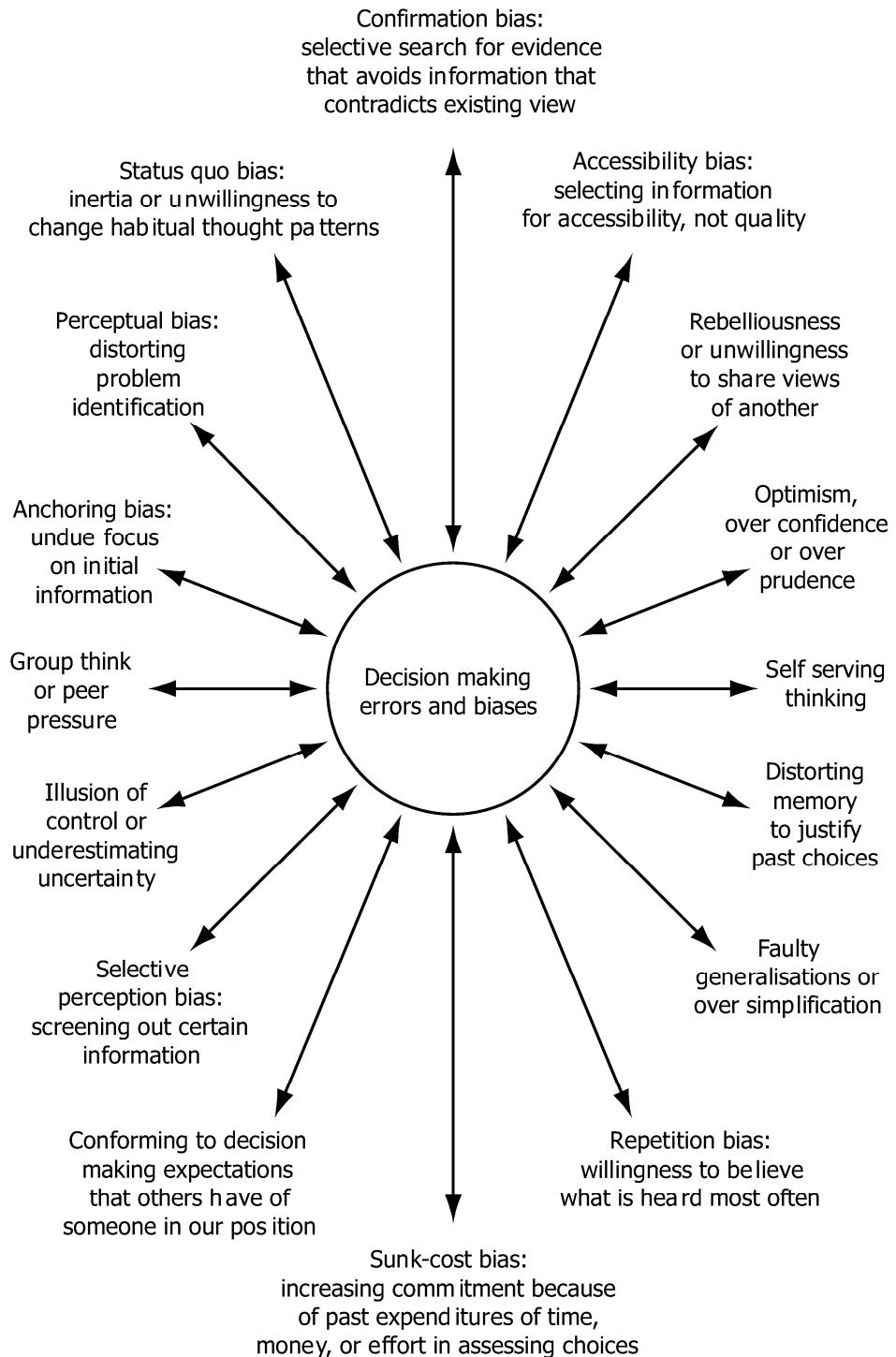
### 4.1 Bias

Bias in data, algorithms and people is the fundamental cause of the failure of AI implementations. Unlike many other applications, AI is introduced to involve autonomous, semi-autonomous or prescriptive decision making. Therefore, it is important to mitigate the biases in AI to maximise social justice. The leaders should be self-aware, conscious, and avoid intuitive decisions on AI implementations, management and governance. Figure 4.1 shows the traits of intuitive decision-making. The collaborations, partnerships and working as a distributed network are recommended by the 17<sup>th</sup> UN SDGs to overcome those traits by promoting diversity, equity and inclusion in people realising AI.

It is understood that each individual has their own biases, traits and ways of thinking. That is why collective decision making with a diverse group is more effective than individual decision making. Figure 4.2 shows various decision-making errors and biases that leaders should be aware of when forming, norming and driving AI strategies and transformation. Diverse perspectives, more information, more alternatives, and different thinking styles are key success factors of Utilitarianism perspectives of AI ethics, which help democratise AI, avoiding disparities and meaningful participation and representation [18].



**Figure 4.1:** The nature of intuitive decision-making [4]. Figure ©Australian Institute of Company Directors.

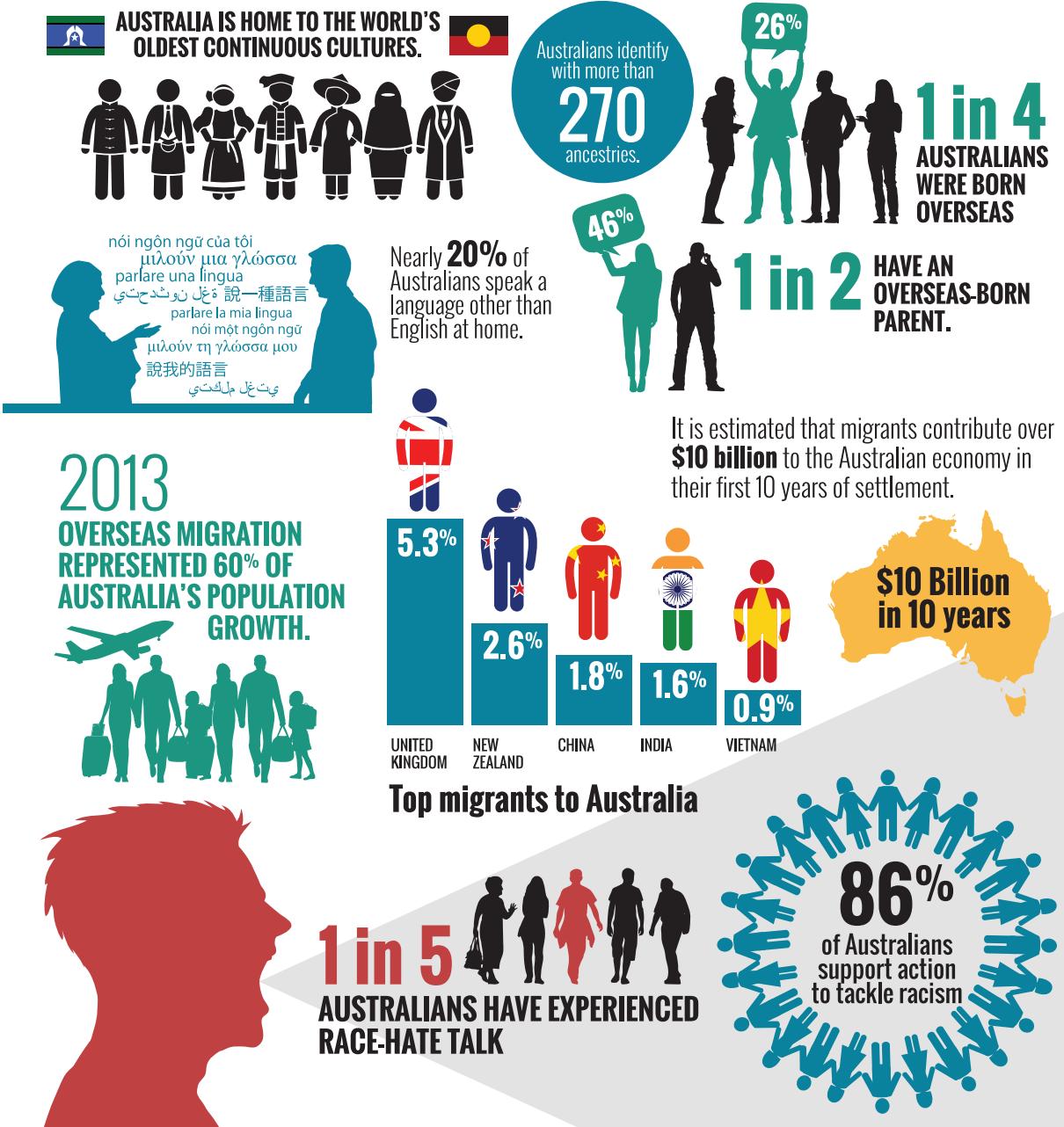


**Figure 4.2:** Decision making errors and biases [5]. Figure ©Australian Institute of Company Directors.

## 4.2 Diversity

Australia has vibrant multicultural community (see Figure 4.3). This is one of the uniqueness of Australia. The Aboriginal and Torres Strait Islander peoples' culture is the world's oldest continuous culture. Australians can be related to more than 270 ancestries. Since 1945, almost 7 million people have migrated to Australia. This rich culture is one of the greatest strengths of its economic prosperity. Therefore, it is important to consider this great diversity when mitigating biases and promoting inclusions in AI initiatives.

Leaders should bring diversity to AI solutions by enabling equity and inclusion. “*Neither a person nor an apple can be diverse. Diversity is the property of a collection of people—a basket with many kinds of fruit*” [19]. Gender equality and reduced inequalities are key focuses in sustainability addressing through 5<sup>th</sup> and 10<sup>th</sup> UN SDGs. On the other hand, the Australian anti-discrimination law was established to eliminate all forms of discrimination which is an integral part of promoting diversity [20].



2014 Face the Facts [www.humanrights.gov.au/face-facts](http://www.humanrights.gov.au/face-facts)

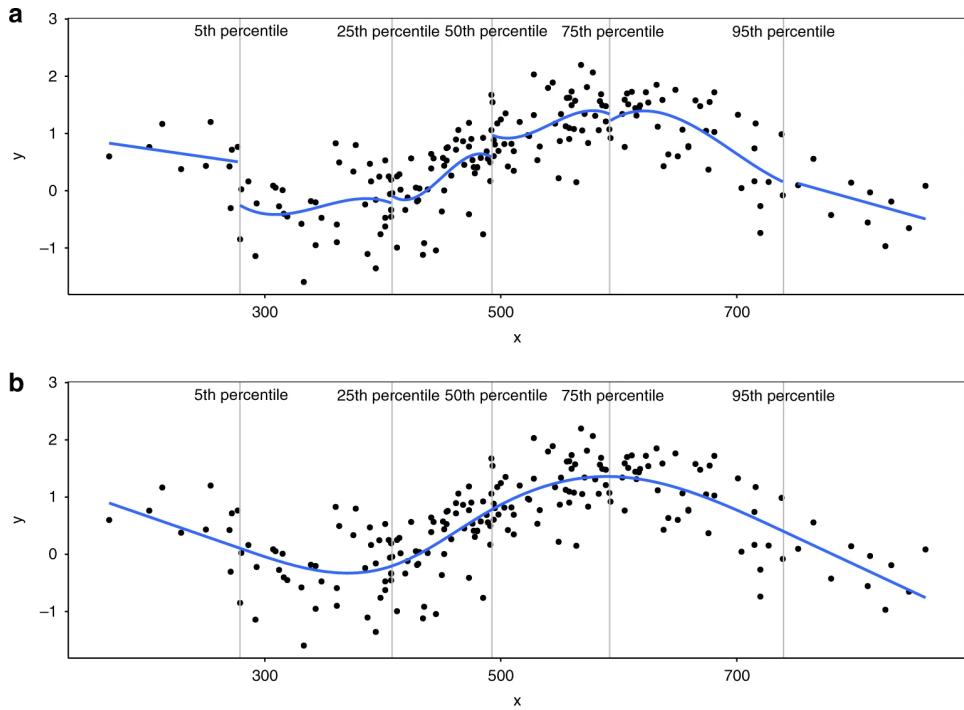


**Figure 4.3:** Cultural diversity of Australia and interesting facts. Figure ©Australian Human Rights Commission [6].

## 4.3 Impartiality and Localisation

Impartiality and localisation are two important objectives in an equitable AI solution. When managing impartiality, retaining fairness to locality is equally important. If the AI model is generalised across the entire population, it may be justified as an impartial solution but might not be fair for minority groups. Even deploying locally optimised multiple models may create injustice to people at the margins of the segments and cause issues from the impartiality perspective.

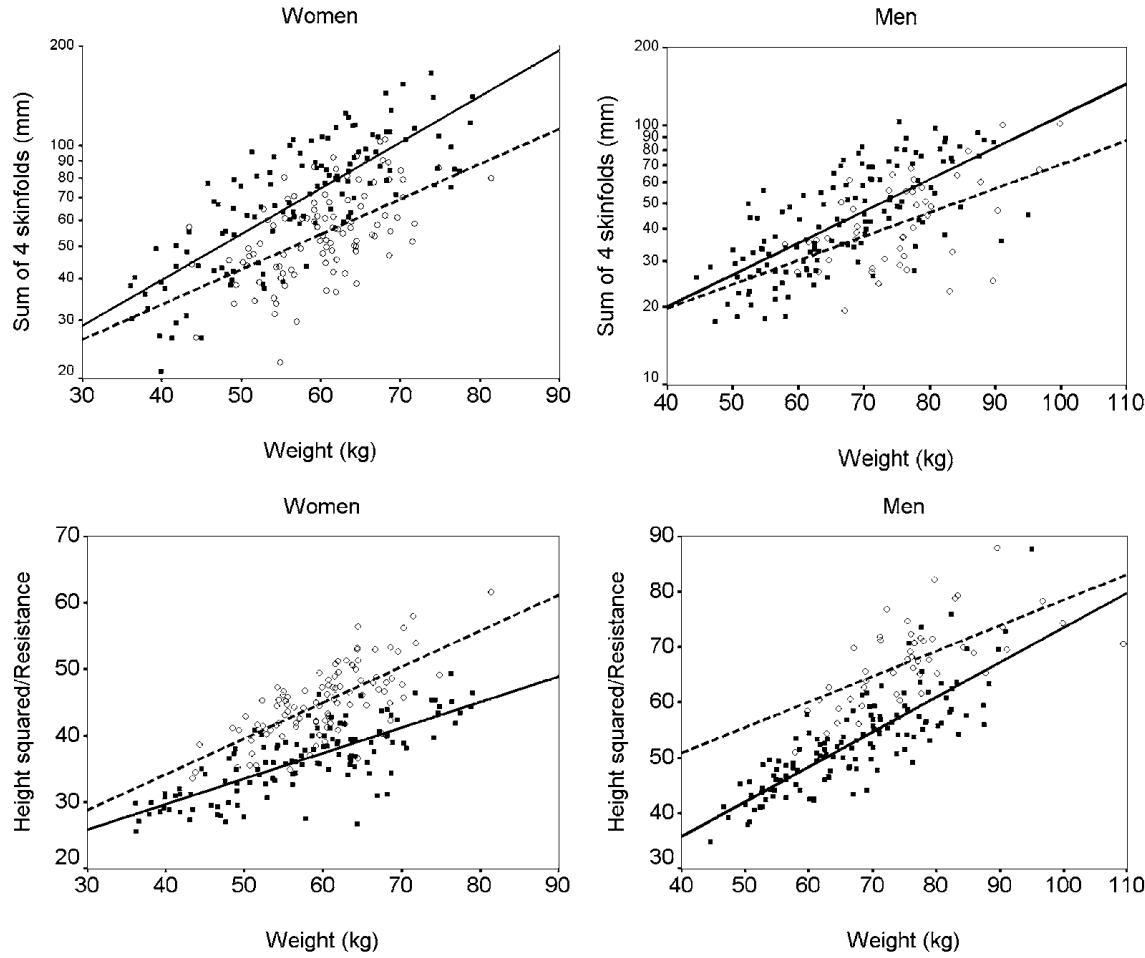
Figure 4.4 shows two modelling strategies on complex and diversified data points. In machine learning, regularisation techniques generalise the model while mitigating overfit. Sometimes, the regularisation may neglect the minority requirements. Therefore, the model complexity on data should be determined by accounting impartiality and localisation of the solution.



**Figure 4.4:** Categorising continuous variables is important for diversifying the service. Modelling of each category of continuous variable independently, as shown in the figure **a** can lead to loss of information and poor predictions. On the other hand, modelling the entire data set with a single higher-order polynomial might overfit the model. The figure **b** shows mathematically complex restricted-cubic-spline regression lines, which can flexibly and accurately model complex and non-linear relationships [7].

## 4.4 Equity

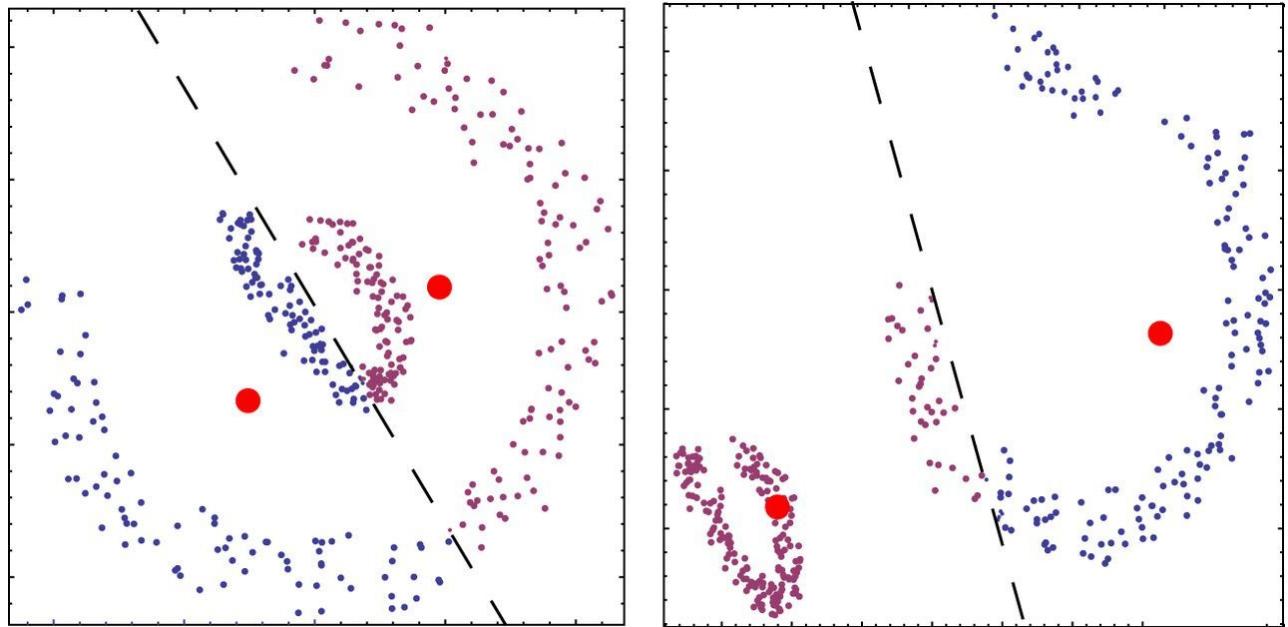
Equity is an important concern in social justice, which is quite relevant to the Australian multicultural society. Bringing AI equity to relevant groups is important when creating values or making decisions from an ethical perspective. For example, Aboriginal and European Australians have a significantly different body fat distribution and fat mass for given body weight or BMI. By research, it has been identified that (see Figure 4.5) BMI ranges valid for the majority of Australians to determine weight status may be inappropriate in Australian Aboriginal people [8].



**Figure 4.5:** Relationship of body weight to surrogate measures of fat mass (sum of four SFT) and fat-free mass (height<sup>2</sup>/resistance) in Australians of Aboriginal (filled squares, solid line) and European (open circles, broken line) ancestry [8].

## 4.5 Inclusion

Reducing *overfit* of an AI algorithm by regularisation and/or dimensionality reduction may disregard important attributes related to minority groups. Therefore, data scientists should bring the right amount of data insights to the design to enhance inclusiveness, which can be considered a controlled bias. For example, most of the time, the initiation of hyperparameters is important at the start of unsupervised learning. This intentional bias can enhance the quality of an AI solution. Poor control of machine learning is difficult to be compensated for and can lead to undesirable outcomes (see Figure 4.6) [9].



**Figure 4.6:** How K-means clustering getting unsuccessful in non-Gaussian data distribution. The dashed line denotes separating the computed cluster boundaries; filled dots, cluster centres [9]. By bringing reasonable insight, the K-means clustering can be enhanced.

## 5. Complexity in AI Governance

The AI spectrum is quite broad [21]. From IoT sensor management to smart city development, different stakeholders should look into different perspectives such as social justice, strategy, technology, sustainability, ethics, policies, regulations, compliance, etc. Moreover, things get even more complex when different perspectives are entangled. As examples,

1. Environmental and Social: AI has been identified as a key enabler on 79% (134 targets) of United Nations (UN) Sustainable Development Goals (SDGs) [2]. However, 35% (59 targets) may experience a negative impact from AI. While the environment gets the highest potential, the society gets the most negative impact from AI and creates social concerns,
2. Environmental and Technology: Cloud computing is promising with the availability and scalability of resources in data centres. With emerging telecommunication technologies (e.g., 5G), the energy consumption when transferring data from IoT/edge devices to the cloud became a concern on carbon footprint and sustainability. This energy concern is a factor that shifts the technology landscape from cloud computing to fog computing [22],
3. Economic and Sustainability: Businesses are driving AI, hoping it can contribute about 15.7 trillion to the world economy by 2030 [23]. On the other hand, the UN SDGs are also planned to achieve by 2030 in the areas critically important for humanity, and the planet [1]. The synergy between AI economic and sustainability strategies will be essential,
4. Economic and Social: Businesses are driving AI, hoping it can contribute about 15.7 trillion to the world economy by 2030. However, the research found that 85% of AI projects will fail due to bias in data, algorithms, or the teams responsible for managing them [24]. Therefore, AI ethics and governance for the sustainability of AI became a key

success factor in economic goals in AI.

5. Economic and Ethical: Still, no government has been able to pass AI law except ethical frameworks or regulatory guidelines [25]. Therefore, there are many emerging AI risks for humanity on our way to economic prosperity, such as autonomous weapons, automation-spurred job loss, socioeconomic inequality, bias caused by data and algorithms, privacy violations, and deepfakes [26].

On the other hand, the complex differences in AI applications don't necessarily mean there are no similarities in other perspectives such as cultural values, community or strategy. For example, similar organizations may work on different sustainability goals for social justice. Such differences in AI strategy should not obstruct the partnership and collaboration opportunities between them.

## 6. A Framework and a Model for AI Governance

When addressing AI governance requirements, the complexity of the AI can be identified as the main challenge [21]. Unlike any other technology, AI governance is complex because of its autonomous decision-making capability and influence on people's decision-making. Hence, AI governance is entangled with human ethics, which must be realised where artificial intelligence is applied or influenced. We introduced a framework and model with the simple golden circle in mind. They help directors find solutions for why, how and what questions when governing AI. First, the innovative KITE conceptualised abstraction framework helps directors drive the purpose of AI initiatives to address key success factors. With the support of the KITE abstraction framework, the innovative Wind-turbine conceptualised model helps to develop a comprehensive AI strategy for organisations. These frameworks and models help drive AI for sustainability in more structured, systematic, transparent, and collaborative ways.

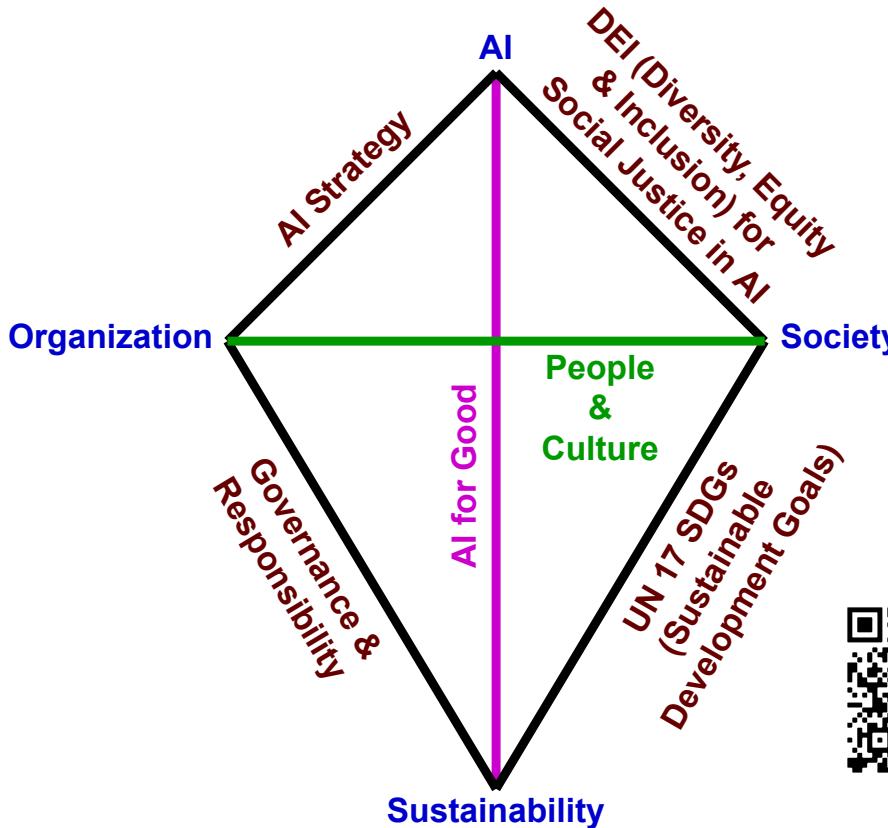
### 6.1 KITE abstraction framework

The KITE abstraction framework (see Figure 6.1) [10] helps directors govern AI aligning with the broader ESG purpose, fundamentally the *why* aspect of the golden circle. Irrespective of the complexity of the AI application, this framework analyses the four key dimensions of

1. AI,
2. Organisation,
3. Society, and
4. Sustainability.

The interdependencies of these dimensions enable addressing of AI strategy, AI for Good and

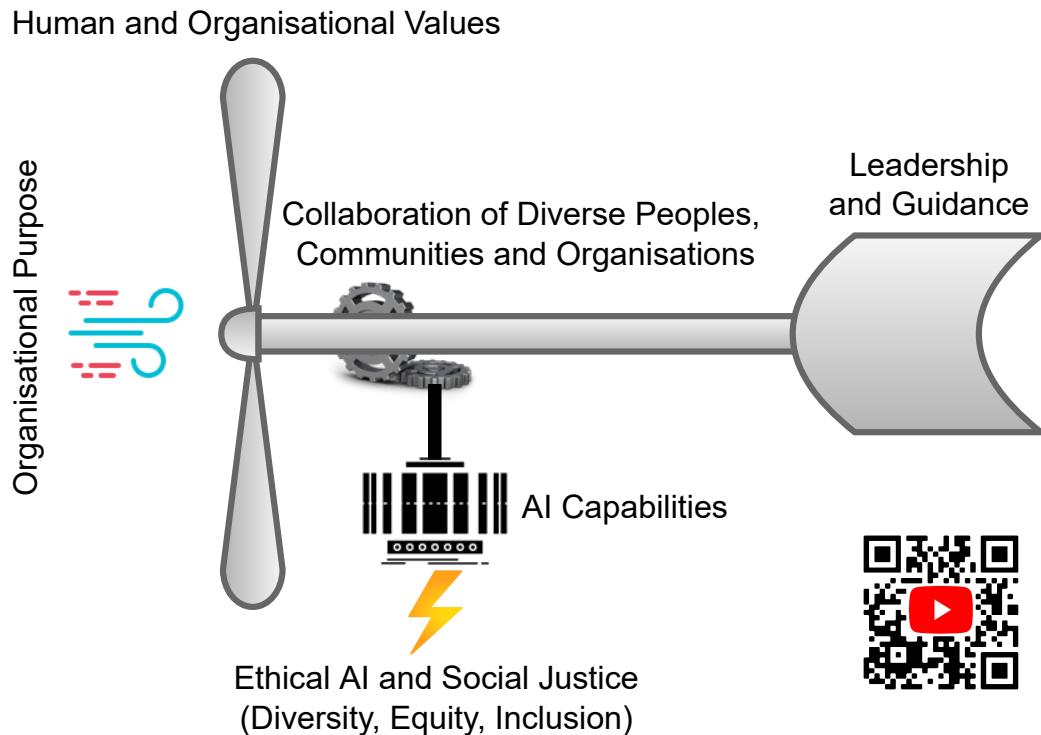
United Nations Sustainable Development Goals. Further, it helps mitigate AI risks due to biases by bringing social diversity, equity and inclusion to AI governance. As illustrated in the diagram, it helps organisational governance and responsibilities by guiding the orchestration of people, culture and AI mission towards sustainability.



**Figure 6.1:** KITE abstraction framework for AI governance [10]. It aligns with the broader ESG purpose, fundamentally the *why* aspect of the golden circle.

## 6.2 Wind-turbine conceptualised model

The wind-turbine conceptualised model (see Figure 6.2) [11] helps directors address *how* and *what* aspects of AI governance. The model helps oversee AI processes supporting social justice with social diversity, equity and inclusion. From the organisational perspective, this model directs the AI initiative towards humanity and sustainable development goals (SDGs) for minimising human suffering. Further, this model helps oversee the operations and management, represented by the tail of the wind turbine. The front-faced multi-blade rotor represents the values and policies (e.g., seven fundamental principles) that ethically and efficiently address humanitarian needs, risks and suffering. The wheels in the gearbox represent the community, partners and volunteers who are continually helping with diversity, equity and inclusion. Finally, the generator represents the Data and AI capabilities that drive the AI innovation and transformation for sustainability. In summary, directors can oversee the full spectrum of the AI processes, stakeholders, and management.



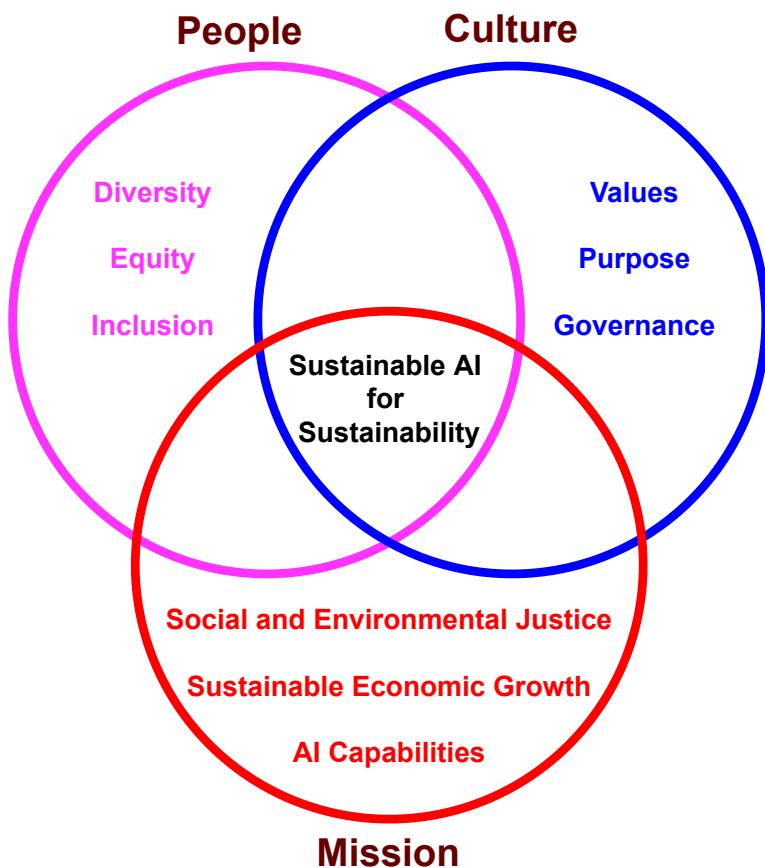
**Figure 6.2:** Wind-turbine conceptualised model for AI governance [11]. It helps directors address *how* and *what* aspects of AI governance.

## 6.3 People, Culture and Mission

To make sure AI for good programs serve the purpose of serving humanity and sustainability, it is important to mitigate the biases in decision making in leadership, management and governance while managing the projects that enhance social justice. These make sure we can realise AI ethics and sustainable development goals.

However, to minimise biases and enhance social justice, it is required to bring social diversity, equity and inclusion to the leadership, management and governance. Only then can we achieve utilitarianism and consequentialism perspectives of human ethics which can underpin the AI ethics for serving humanity and sustainability. Our framework helps all stakeholders including communities, volunteers and partners to collaborate on sustainable development goals and social justice.

From the corporate governance and management perspective, this framework helps the corporate board, human resource (HR) and management to orchestrate culture, people and mission towards humanity and sustainability. Figure 6.2 illustrates how the synergy between corporate culture, people and mission can drive AI ethics towards sustainable AI and goals [11].

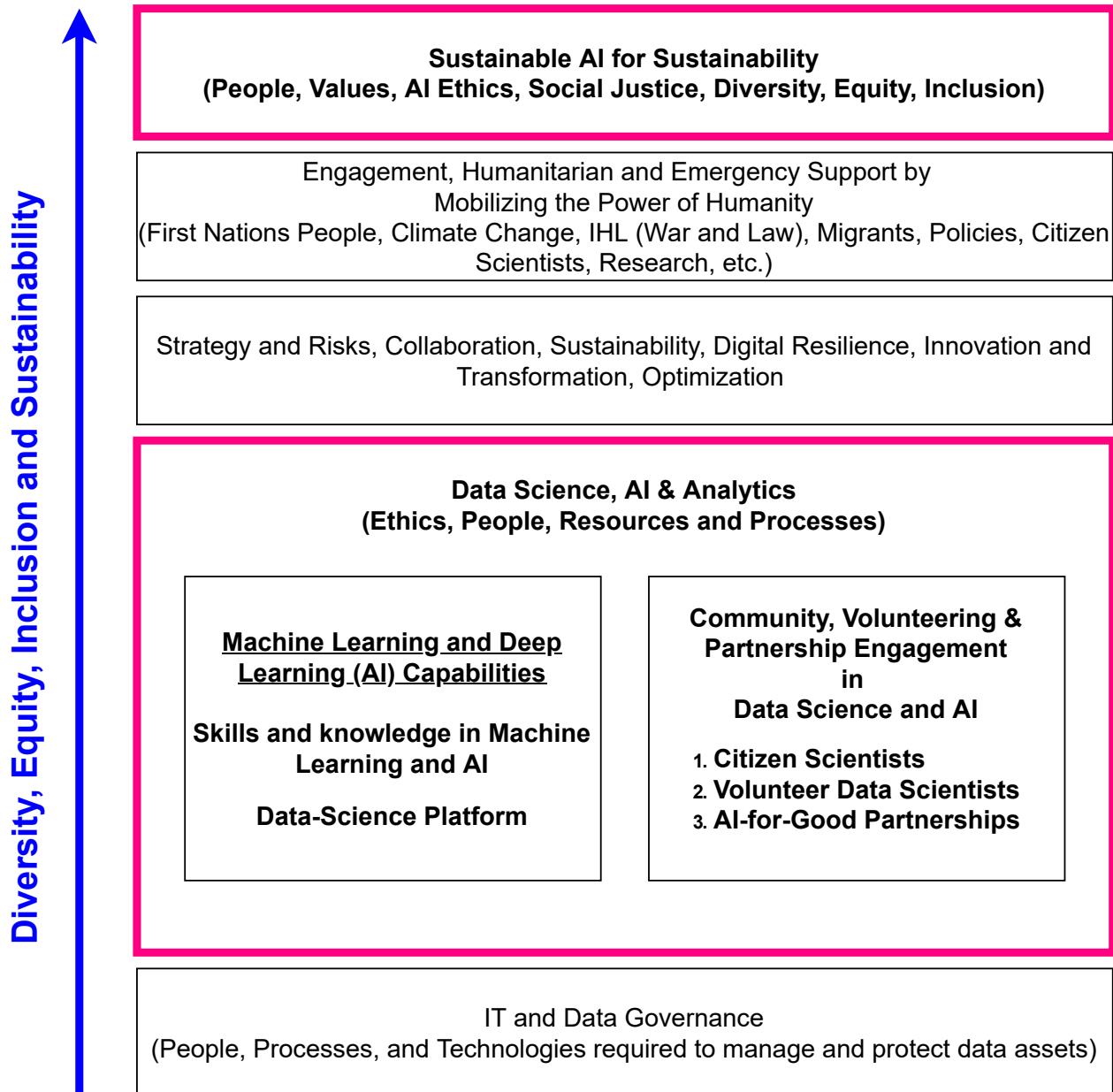


**Figure 6.3:** The proposed AI governance tools help the corporate board, human resource (HR) and management to orchestrate culture, people and mission towards humanity and sustainability.

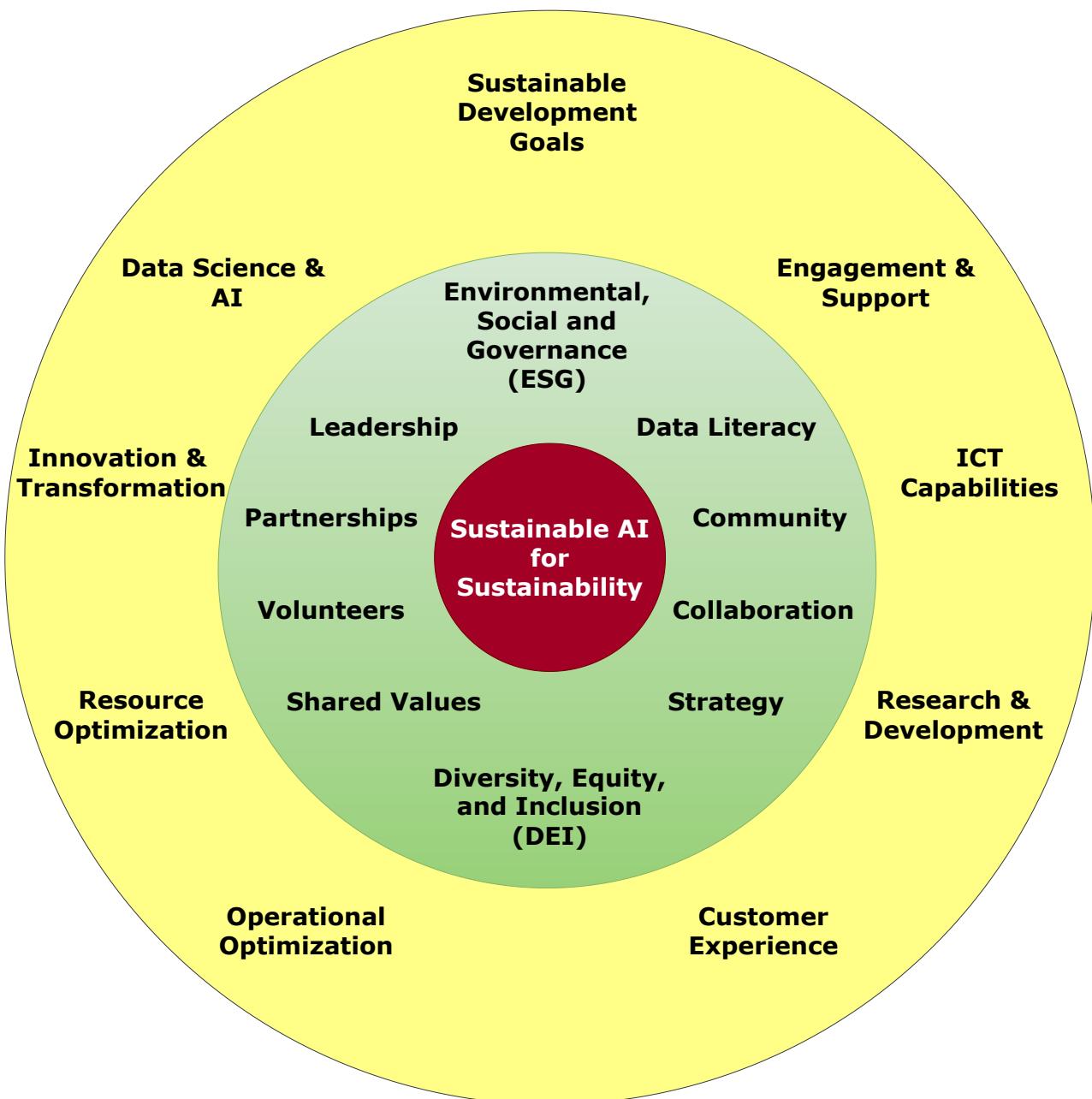
## 7. Adaptation of the Framework

The adaptation of the proposed framework helps creating values based on the data wisdom [27]. It helps AI innovation and transformation towards social justice. As shown in Figure 7.1 [11], the data science and AI as a service layer supports business strategies of ESG by leveraging data and IT assets while enhancing DEI (diversity, equity and inclusion), brand advocacy, customer experience (CX), and return on investment (ROI).

The proposed framework establishes synergy between AI governance and social justice by mobilizing the organizational culture towards AI-driven innovation and transformation. A greater social diversity, equity and inclusion can be expected in AI initiatives which enable ethical inclusion, processes and outcomes in AI. The sustainable AI and sustainable development goals will be a primary focus in AI developments that drive business objectives and corporate social responsibilities. The ideation of this strategy can be illustrated by Figure 7.2 [11].



**Figure 7.1:** Sustainable AI for Sustainability. Businesses should position their IT, data science, and AI capabilities to address social justice and sustainability strategies. DEI (diversity, equity and inclusion) would be a key success factor of those initiatives.



**Figure 7.2:** Development of sustainable AI as a core competency. AI has been identified as a key enabler for ESG and sustainability.

## 8. Conclusion

AI would be a key capability for future prosperity. Good governance of AI is very important to mitigate AI risks and create values. AI frameworks and standards are emerging to govern AI aligning with human ethics and emerging environmental, social, and corporate governance (ESG) principles. In brief, diversity, equity and inclusion (DEI) together with social and cultural values can make AI initiatives vibrant and sustainable. Further, it will mitigate biases related to AI, including biases in data, algorithms, people, and processes. This book's recommendations will help leaders orchestrate people, culture, and mission toward sustainable AI for social justice.

## References

- [1] UN General Assembly (UNGA), “A/RES/70/1 transforming our world: the 2030 agenda for sustainable development,” *Resolut* 25, pp. 1–35, 2015. [Online]. Available: [https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A\\_RES\\_70\\_1\\_E.pdf](https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_70_1_E.pdf)
- [2] R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Felländer, S. D. Langhans, M. Tegmark, and F. F. Nerini, “The role of artificial intelligence in achieving the sustainable development goals,” *Nature Communications*, vol. 11, no. 1, Jan. 2020. [Online]. Available: <https://doi.org/10.1038/s41467-019-14108-y>
- [3] C. Allen, M. Reid, J. Thwaites, R. Glover, and T. Kestin, “Assessing national progress and priorities for the sustainable development goals (SDGs): experience from australia,” *Sustainability Science*, vol. 15, no. 2, pp. 521–538, Jul. 2019. [Online]. Available: <https://doi.org/10.1007/s11625-019-00711-x>
- [4] L. A. Burke and M. K. Miller, “Taking the mystery out of intuitive decision making,” *The Academy of Management Executive*, vol. 13, no. 4, pp. 91–99, 11 1999. [Online]. Available: <http://ezproxy.lib.uts.edu.au/login?url=https://www.proquest.com/scholarly-journals/taking-mystery-out-intuitive-decision-making/docview/210531328/se-2?accountid=17095>
- [5] J. S. Hammond, R. L. Keeney, and H. Raiffa, “The hidden traps in decision making,” *Harvard business review*, vol. 76, no. 5, pp. 47–58, 1998.
- [6] “Face the facts: Cultural diversity | australian human rights commission,” <https://humanrights.gov.au/our-work/education/face-facts-cultural-diversity>, (Accessed on 04/12/2022).

- [7] J. Gauthier, Q. V. Wu, and T. A. Gooley, “Cubic splines to model relationships between continuous variables and outcomes: a guide for clinicians,” *Bone Marrow Transplantation*, vol. 55, no. 4, pp. 675–680, Oct. 2019. [Online]. Available: <https://doi.org/10.1038/s41409-019-0679-x>
- [8] L. Piers, K. Rowley, M. Soares, and K. O’Dea, “Relation of adiposity and body fat distribution to body mass index in australians of aboriginal and european ancestry,” *European journal of clinical nutrition*, vol. 57, no. 8, p. 956—963, August 2003. [Online]. Available: <https://doi.org/10.1038/sj.ejcn.1601630>
- [9] T. Lorimer, J. Held, and R. Stoop, “Clustering: how much bias do we need?” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 375, 2017.
- [10] M. Samarawickrama, “KITE: An Abstraction Framework for Reducing Complexity in AI Governance,” October 2021. [Online]. Available: <https://unstats.un.org/unsd/undataforum/blog/KITE-an-abstraction-framework-for-reducing-complexity-in-ai-governance/>
- [11] M. Samarawickrama, “Social justice and sustainability by leveraging data science and AI: Interview with Dr. Mahendra Samarawickrama,” 2021. [Online]. Available: <https://hyperight.com/social-justice-and-sustainability-by-leveraging-data-science-and-ai-interview-with-dr-mahendra-samarawickrama/>
- [12] S. Lynch, “Andrew Ng: Why AI Is the New Electricity,” <https://www.gsb.stanford.edu/insights/andrew-ng-why-ai-new-electricity>, march 2017.
- [13] N. J. Abram, , H. V. McGregor, J. E. Tierney, M. N. Evans, N. P. McKay, and D. S. Kaufman, “Early onset of industrial-era warming across the oceans and continents,” *Nature*, vol. 536, no. 7617, pp. 411–418, Aug. 2016. [Online]. Available: <https://doi.org/10.1038/nature19082>
- [14] A. Laitinen and O. Sahlgren, “Ai systems and respect for human autonomy,” *Frontiers in Artificial Intelligence*, vol. 4, 2021. [Online]. Available: <https://www.frontiersin.org/article/10.3389/frai.2021.705164>

- [15] L. Zardiashvili and E. Fosch-Villaronga, ““oh, dignity too?” said the robot: Human dignity as the basis for the governance of robotics,” *Minds and Machines*, vol. 30, no. 1, pp. 121–143, Jan. 2020. [Online]. Available: <https://doi.org/10.1007/s11023-019-09514-6>
- [16] M. Boni, “The ethical dimension of human–artificial intelligence collaboration,” *European View*, vol. 20, no. 2, pp. 182–190, Oct. 2021. [Online]. Available: <https://doi.org/10.1177/17816858211059249>
- [17] M. Eggleton, “Award-winner warns of the failures of artificial intelligence,” pp. S4–S5, 2022. [Online]. Available: <https://www.afr.com/technology/award-winner-warns-of-the-failures-of-artificial-intelligence-20220313-p5a4b3>
- [18] B. C. Stahl, “Concepts of ethics and their application to AI,” in *SpringerBriefs in Research and Innovation Governance*. Springer International Publishing, 2021, pp. 19–33. [Online]. Available: [https://doi.org/10.1007/978-3-030-69978-9\\_3](https://doi.org/10.1007/978-3-030-69978-9_3)
- [19] S. E. Page, “Making the Difference: Applying a Logic of Diversity,” *Academy of Management Perspectives*, vol. 21, no. 4, pp. 6–20, 2007. [Online]. Available: <http://www.jstor.org/stable/27747407>
- [20] “Australia’s anti-discrimination law | attorney-general’s department,” <https://www.ag.gov.au/rights-and-protections/human-rights-and-anti-discrimination/australias-anti-discrimination-law>, (Accessed on 04/12/2022).
- [21] E. R. Goffi, A. Momcilovic *et al.*, “Global Trends in AI 2022: Food for thought from GAIEI experts,” 2022. [Online]. Available: <https://globalethics.ai/global-trends-in-ai-2022-food-for-thought-from-gaiei-experts/>
- [22] E. Baccarelli, P. G. V. Naranjo, M. Scarpiniti, M. Shojafar, and J. H. Abawajy, “Fog of everything: Energy-efficient networked computing architectures, research challenges, and a case study,” *IEEE Access*, vol. 5, pp. 9882–9910, 2017.
- [23] PwC, “PwC’s Global Artificial Intelligence Study: Sizing the Prize,” 2017. [Online]. Available: <https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html>
- [24] Gartner, “Gartner Says Nearly Half of CIOs Are Planning to Deploy Artificial Intelligence,” 2018. [Online]. Available: <https://www.gartner.com/en/newsroom/press-releases/>

2018-02-13-gartner-says-nearly-half-of-cios-are-planning-to-deploy-artificial-intelligence

- [25] L. Floridi, “Establishing the rules for building trustworthy AI,” *Nature Machine Intelligence*, vol. 1, no. 6, pp. 261–262, May 2019. [Online]. Available: <https://doi.org/10.1038/s42256-019-0055-y>
- [26] M. Perc, M. Ozer, and J. Hojnik, “Social and juristic challenges of artificial intelligence,” *Palgrave Communications*, vol. 5, no. 1, Jun. 2019. [Online]. Available: <https://doi.org/10.1057/s41599-019-0278-x>
- [27] M. Samarawickrama, “Keeping AI honest,” pp. 52–53, March 2022. [Online]. Available: <https://aicd.companydirectors.com.au/membership/company-director-magazine/2022-back-editions/march/ai-ethics>

## Author's Biography



**Dr Mahendra Samarawickrama** (GAICD, MBA, SMIEEE, ACS(CP)) is the ICT Professional of the Year 2022 in the ACS Digital Disruptors Awards. He is a highly accomplished leader having an impressive track record of driving visions, technology innovations and transformation towards humanity, social justice, and sustainability. He is a founding director of the Centre for Ethical AI and the Centre for Sustainable AI. He supports the formation of organisational Environmental, Social, and Governance (ESG) strategy and drives ESG projects leveraging emerging technologies. He specialises in directing AI, Data Science and Customer Experience (CX)-focussed teams on building state-of-the-art capabilities. He is an author, inventor, mentor, advisor and regularly speaks at various technology forums, conferences and events worldwide. Many of his publications and frameworks related to AI governance and ethics are spotlighted in national and international forums.

As the Manager of the Data Science and Analytics team in the Australian Red Cross, he has developed an AI governance and strategy framework crucial to the business' successful deployment of Data Science and AI capabilities to mobilise the power of humanity. He built the Volunteer Data Science and Analytics team from the ground up, supporting the Australian Red Cross's strategic goals. He is supporting the business for personalised engagement of customers for disaster resilience in these demanding times of pandemic, natural disasters, and global conflicts. He is also a co-author of the IFRC data playbook and contributed to the data science and emerging technology chapter for AI governance, ethics, and literacy. In all these processes, he valued diversity, equity and inclusion. In recognition of this, his team became finalists in 1) the Diversity, Equity and Inclusion in Action Award in the 2021 IoT Awards, 2) the Best Use of Technology to Revolutionise CX Award in the 2021 Ashton Media CX Awards, 3) the Service Transformation for the Digital Consumer for Not-for-Profit/NGO in 2022 ACS Digital Disruptors Awards, and contributed to winning the CX Team of the Year Award in 2021 Ashton Media CX Awards. All of these awards are prestigious national awards.

He is an industry collaborator who actively leads technology innovation-and-transformation initiatives and partnerships toward humanity, social justice and sustainability. In this per-

spective, he is an Advisory Council Member in Harvard Business Review (HBR), an Expert in AI ethics and governance at Global AI Ethics Institute, an industry Mentor in the UNSW business school, a senior member of IEEE (SMIEEE), an honorary visiting scholar at the University of Technology Sydney (UTS), an Advisor for Data Science and Ai Association of Australia (DSAi), and a graduate member of the Australian Institute of Company Directors (GAICD).

He has recently established a YouTube channel and a Twitter channel to share his knowledge with the community. With a PhD in Computer Science and Masters degrees in Business Administration and Project Management, he brings the capacity to steer organisations through the complex, data-driven problems of our time.

