

AI ASIA PACIFIC
INSTITUTE

Fairness in AI: Impact and Opportunities

Table of Contents

About the AI ASIA PACIFIC INSTITUTE	2
Foreword.....	3
1 Introduction	4
2 Definitions	7
3 Fairness in the AI Context	8
4 Responsible Use of AI	12
5 The “S” Factor	14
6 Bias in AI	20
7 Conclusion and Recommendations	24
Appendix A	26
Appendix B	27
Endnotes	29

Copyright

© 2023 AI Asia Pacific Institute. All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, including photocopying and recording, or by any information storage and retrieval system.

Important disclaimer

The reader is advised and needs to be aware that the information contained in this publication may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific and technical advice. To the extent permitted by law, the AI Asia Pacific Institute (including its members and advisors) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.



About the **AI ASIA PACIFIC INSTITUTE**

The AI Asia Pacific Institute addresses the social, legal and ethical risks of artificial intelligence through international cooperation.

Foreword



I am delighted to present this report on the role of fairness in artificial intelligence (AI) and its implications for governments, financial institutions, and investors. As the landscape of AI continues to evolve at an unprecedented pace, it is crucial that we engage in thoughtful discussions about the ethical and societal dimensions of this transformative technology.

The rapid growth of private investment in AI, as highlighted in the Stanford AI Index Report 2023, indicates AI's immense potential and opportunities across various industries. However, with great power comes great responsibility. AI must be developed and deployed in a fair and responsible manner, addressing the inherent biases and risks that can emerge if left unchecked.

Fairness is a fundamental principle that should guide the development and use of AI systems. As this report explores, fairness considerations increasingly influence investment decisions and shape government policies. We must recognise that AI has the potential to perpetuate existing inequalities and biases if not properly addressed. Therefore, incorporating fairness into the design, development, and deployment of AI systems is imperative to ensure equitable outcomes for all.

The report also sheds light on the growing significance of environmental, social, and governance (ESG) considerations in the AI domain. Fairness plays a pivotal role in the social sphere of ESG, encompassing non-discrimination, transparency, and accountability. By embracing ESG principles and integrating fairness into AI strategies, we can not only mitigate risks but also leverage AI's potential for positive social impact.

While this report does not provide prescriptive guidance applicable to every organisation, it offers a comprehensive list of key elements to consider. It is my hope that the insights and recommendations presented here will empower leaders to navigate the complex landscape of AI, making informed decisions that prioritise fairness and uphold societal values.

I extend my deepest appreciation to the AI Asia Pacific Institute team that contributed to this work: their shared commitment to advancing the dialogue on fairness in AI has been instrumental in shaping the content and recommendations found within these pages. I would also like to thank RepRisk, an ESG data science firm combining machine learning and human intelligence to identify ESG risks, for their illumination of the ESG landscape for the purposes of this report.

As we embark on this journey towards a fair and responsible AI future, let us seize the opportunities and address the challenges with unwavering commitment. Together, we can ensure that AI contributes to a more inclusive and equitable world for all.

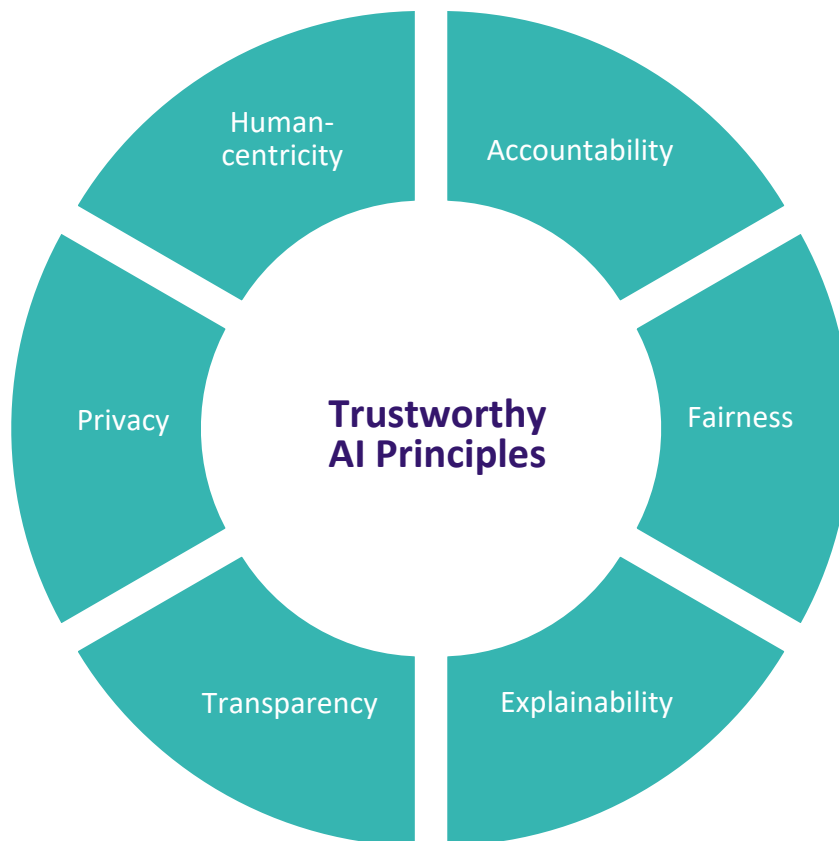
Kelly Forbes
Executive Director
AI Asia Pacific Institute



1 Introduction

Private investment in artificial intelligence has grown rapidly over the past decade, the Stanford AI Index Report 2023 noted that in 2022 there were 3,538 AI-related private investment events, representing a 12% decrease from 2021 but a sixfold increase since 2013.¹ AI-related events that raised over \$1 billion, however, did not decrease and are now over double what they were in 2021. The trends suggest a short-term funding decrease but a greater potential for long-term growth.

The global leader in private investments for AI is the US, with a total of \$47.4 billion invested in 2022. That is 3.5 times the amount invested by China, at \$13.4 billion, and 11 times what was invested in the UK, at \$4.4 billion. This growth in private investment in AI reflects the increasing importance of AI technology across a wide range of industries and applications, as well as the potential for significant returns on investment in the field.

The proliferation of AI technology has sparked numerous debates surrounding the principles that ought to guide its development, implementation, and use. Following extensive industry consultation as part of the 2021 Trustworthy Artificial Intelligence in the Asia-Pacific Region report,² the AI Principles Comparison table was proposed. Six principles to encourage the development of Trustworthy AI³ were consistently found in the region:⁴ human-centricity; fairness; explainability; transparency; privacy; and accountability (Trustworthy AI Principles).⁵





The focus of this report is on fairness as a guiding principle of AI. As AI receives growing attention, the responsible use of AI, which includes fairness considerations, is taking an increasingly active part in investors' decisions and how governments assess long-term public investment and encourage private investment in research and development. The perils and increasing risks associated with AI places a burden on government and investment leaders to correctly assess and monitor the AI industry effectively. The responsible use of AI is also becoming an evolving Environmental, Social and Governance theme among institutional investors, particularly in the 'social' sphere.

ESG data assesses the adverse impacts of ESG risks on the related entities. These considerations can offer a baseline for capturing impact in the arena of AI management and fairness to determine and mitigate potential risks, such as developing better internal policies and practices.

In the social sphere, fairness is evaluated based on a company's treatment of its employees, customers, suppliers, and other stakeholders. This includes, but is not limited, to non-discrimination in hiring; promotion; pay; and other employment practices.

The use of AI systems to automate many of these employee management functions raises questions of fairness, as it creates the risk of, *inter alia*, biased data, algorithmic bias, and a lack of transparency. This has necessitated the implementation of upskilling policies and practices that demonstrate an internal commitment to fairness, while also addressing the impacts of AI-driven automation. The same is true for governments, which increasingly have to evaluate their population's readiness for technological disruption. A parallel strategy should be adopted for AI automation to ensure a balanced societal impact.

At the intersection of governance and AI, implementing fairness in every approach is equally paramount, as AI systems have the potential to perpetuate and intensify existing inequalities and biases. For instance, an AI system trained on data reflecting historical discrimination may inadvertently make biased decisions. To prevent such scenarios, fairness must be incorporated into the design, development, and deployment of AI systems, while ensuring human supervision over AI decision-making.

On the other hand, governments and the investment industry are in a unique position to continue to shape the AI industry. The power dynamics in AI are often tilted towards technology companies, who are superior in technical expertise and resources, data access and ownership, and market influence. But governments and investment leaders can still influence control through regulatory frameworks, funding, incentives and policies.

The promise of AI for government and investment leaders is multifaceted and has the potential to revolutionise various sectors and aspects of governance. But to capture the promise of AI, financial and government leaders need to increasingly understand the role fairness has in this technological disruption and the trade-offs that are associated.

The aim of this report is to increase awareness about the role of fairness in AI and to propose practicable options for the adoption of key mechanisms that governments, financial institutions and investors can consider for AI adoption. As part of this analysis, the report will consider ESG factors and related use cases, expanding on the role of AI in advancing these considerations.

This report does not provide step-by-step guidance, which differs greatly for every organisation depending on local context, but instead offers a comprehensive list of key elements for governments and investors to consider. When combined, these elements will optimally position leaders at the forefront of future opportunities in the sector.

To write this report, the AI Asia Pacific Institute has relied on valuable internal and external expertise and input.

David Hardoon pioneered the regulator and central bank adoption of data science as well as the establishment of the Fairness, Ethics, Accountability, and Transparency (FEAT) principles, first-of-a-kind guidelines for adopting Artificial Intelligence in the financial industry, as well as establishing the MAS-backed Veritas consortium.

Hardeep Arora brings to the table over two decades of expertise in analytics and data science technology, specifically within the financial services sector. Currently residing in Singapore, he serves as the Head of AI Engineering for Temus (A Temasek and UST joint venture). Hardeep's expertise extends to the design and implementation of scalable data systems with embedded intelligence. He has been instrumental in numerous client engagements, including the MAS Veritas initiative, which focuses on enhancing internal governance around the application of AI and the management and utilisation of data.

Janet Wong is an industry veteran in investment stewardship in the asset management industry. She works on engagement with corporate directors and management on investment issues involving corporate governance, social and environmental sustainability. She specifically leads social engagement issues including supply chain and human rights globally.

Joseph (Joe) Negrine is a Tuckwell Scholar at the Australian National University, completing a Bachelor of Law (Hons) and a Bachelor of Arts. His research interests concern emerging technologies, the environment, and how they relate to access to justice. He is a student editor of the ANU Journal of Law and Technology.

Leesa Soulodre is the General Partner of R3i Ventures and R3i Capital, an innovation advisory and AI focused Venture Capital fund operating in the US, Europe and Asia Pacific. Leesa has served over 400+ multinationals and a plethora of start-ups in 19 sectors. She has enabled more than 50 technology driven global reputation projects for the world's largest companies and empowered more than 200 financial institutions with data-driven strategies for responsible investment. She is a Clinical Professor in Entrepreneurship and Complex Problem Solving at SMU Cox Business School and an Adjunct at IE Business School, Luxembourg School of Business, and Singapore Management University.

Philippa Penfold managed many technology implementations and today supports HR functions and companies with their digital transformation. Pip frequently speaks on the Future of Work and the role HR needs to play in finding synergy between humans and artificial intelligence within organisations. She educates HR on how to map their talent challenges and design their technology ecosystem to meet their challenges today and in the future. Pip also provides specialist support and expertise to emerging HR technology companies as an Advisor.

This report would not have been possible without the input of RepRisk who contributed data samples. RepRisk is an ESG data science company combining artificial intelligence, machine learning, and human intelligence to identify and assess business conduct risks.

2 Definitions

Algorithmic bias	Systematic, repeated errors in an AI system that privilege one category over another in an unintended manner.
Counterfactual fairness	A fairness metric that checks whether a classifier produces the same result for one individual as it does for another individual who is identical to the first, except with respect to one or more sensitive attributes.
Demographic parity	A fairness metric that is satisfied if the results of a model's classification are not dependent on a given sensitive attribute.
ESG	Environmental, Social, and Governance. ESG is a framework used to assess a company's business conduct for related risks that could materialize into reputational, compliance, and financial impacts for the company and its stakeholders.
Equalised opportunity	A fairness metric that checks whether, for a preferred label (one that confers an advantage or benefit to a person) and a given attribute, a classifier predicts that preferred label equally well for all values of that attribute.
Fairness trade-off	<p>The inability to satisfy certain fairness criteria simultaneously, such that improving one criterion negatively affects the ability to satisfy another criterion.</p> <p>This is not to be confused with the fairness-utility trade-off, which concerns the trade-off between a model's accuracy and its fairness.</p>
Function creep	The widening of the use of a technology or system beyond the purpose for which it was originally intended.
Generative AI	AI systems that generate new content (text, images, or other media) based on their training data.
Large Language Model (LLM)	An AI model that uses deep learning techniques to recognise, generate, and/or summarise textual data. LLMs are a particular type of generative AI.
Loss function	Also known as a cost function or objective function, is a crucial component used to measure the error or the dissimilarity between the predicted output and the actual target or ground truth.
Threshold optimization	Refers to the process of determining the optimal threshold value that is used to make decisions or classifications in a machine learning model.
Transparency	An umbrella term encompassing concepts such as the explainability and interpretability of AI systems. Similarly to fairness, there are competing conceptions about what constitutes transparency.

3 Fairness in the AI Context

There is not a universally applicable definition of fairness for AI systems. To design AI systems that are fair, it is necessary to define fairness objectives that can be encoded into the system using mathematical language. The application of fairness in AI is a complex and contested concept, and there is no one-size-fits-all approach for every situation. Fairness in AI is highly dependent on context and difficult to quantify.


Ideas regarding what fairness in AI means for organisations were proposed as part of the Veritas initiative in Singapore,⁶ one of the first initiatives aimed at placing philosophical discussions about fairness into the financial and AI context. The Veritas initiative holds ramifications extending beyond Singapore as it serves as a commendable illustration of how collaborative engagement among stakeholders can foster the responsible use of AI. The consortium proposes four guiding fairness principles for AI systems, contending that AI systems:

- (i) Should not unjustifiably, systematically disadvantage individuals or groups of people;
- (ii) Should not unjustifiably use personal attributes as input factors;
- (iii) Should use accurate and relevant data and models, with minimal intentional bias; and
- (iv) Should behave “as designed and intended.”⁷

Fairness has been referred to as “an ‘essentially contested’ concept”, meaning that attempts to improve a system may require “fairness trade-offs.”⁸ Quantitative definitions of fairness seek to formalise different philosophical understandings of fairness. Three such examples are as follows.⁹

First, there is the concept of 'demographic parity'. This principle demands equal outcomes for groups distinguished by a specific characteristic, such as race or gender. For instance, in the context of hiring, demographic parity would mean equal hiring outcomes for all groups. If there are 100 female-identifying applicants and 100 non-female-identifying applicants, and 50 of each are hired, demographic parity is achieved. This principle of demographic parity is rooted in the goal of addressing historical or systemic biases and ensuring equal opportunities for all demographic groups. By achieving demographic parity, organisations aim to eliminate any disparities or underrepresentation that may exist based on certain demographic characteristics.

A second approach to fairness is the 'equalised opportunity' approach. This idea, rooted in Rawlsian justice, equates fairness with promoting the interests of the most vulnerable groups in society. Here, fairness is measured by the actual outcome being equal for all groups. In the hiring example, equalised opportunity requires that the probability of being hired, given that the applicant is qualified, be equal for all genders. Developing the example, assume 50 non-female-identifying applicants are not qualified, and only 40 female-identifying applicants are not qualified.



The demographic parity metric would require an employer to hire 50 members of each group, even though there are more qualified female-identifying applicants. This would result in qualified female-identifying applicants being overlooked in favour of unqualified non-female-identifying applicants. Using the equalised opportunity approach requires hiring *all* qualified applicants, regardless of gender. 60 female-identifying applicants would be hired, and 50 non-female-identifying applicants would be hired. This would result in a workforce that is 60% female-identifying and 40% non-female-identifying.

Last, there is 'counterfactual fairness'. This principle suggests that a decision is fair if it remains the same in a hypothetical scenario where the individual belongs to a different demographic group. This approach considers non-protected features that are not proxies for the protected characteristic. In essence, counterfactual fairness invites us to envision a world where individuals are assessed based solely on their qualifications, merits, and relevant characteristics, irrespective of their demographic background. It prompts us to consider how decisions would unfold if the individuals were part of different demographic groups, effectively removing the influence of factors such as race, gender, or age from the equation.

Counterfactual fairness challenges us to question our assumptions and biases, urging us to envision a world where individuals are evaluated based on their inherent worth and abilities, liberated from the constraints of demographic categorizations. By considering this thought-provoking principle, we can strive for decision-making processes that are not only free from discrimination but also uphold the values of fairness, equal opportunity, and genuine meritocracy.

However, achieving a comprehensive quantitative fairness measure that satisfies all these definitions is challenging and, arguably, unrealistic due to the underlying philosophies which can make outcomes diametrically opposed. Researchers from Cornell and Harvard University have shown that it is almost impossible to meet all conditions simultaneously, except in very specific cases. This highlights the complexity of defining and achieving fairness in any context, including AI applications.¹⁰

As guidance and to simplify the implementation of fairness in AI systems, this present report proceeds on the basis that, for algorithms to be considered fair, they should not systematically disadvantage individuals or a group unless these decisions can be justified."¹¹

The elements that are relevant to this report on the intersection of AI and ESG will be aligned with the fairness standards and recommended practices as part of the Veritas initiative (see Appendix B).

USE CASE (Financial Institution)

Overview

A financial institution based in Singapore applied the Veritas 'FEAT Fair Assessment Methodology'¹² on their credit scoring model, which analyses customers applying for their first credit card or unsecured loan with a bank. Importantly, the referred financial institution observed that the full set of fairness measures should not be applied blindly without much consideration, but rather "carefully select[ed] a relevant subset of fairness measures for each attributed ... on a case by case basis."¹³ In this case, the model predicted that female applicants were slightly more likely to be granted a loan compared to male applicants.

Impact

While the 'demographic parity' measure gave the appearance of unexplained systemic disadvantage, other measures were considered. The 'equalised opportunity' and 'predictive parity' measures revealed a lower discrepancy in approval rates, as they accounted for the fact that female applicants were more likely to repay their loans.

Recommendation

This case demonstrates the importance of using various fairness criteria to enable organisations to understand why any 'unfair' treatment exists. Contrasting different measures improves transparency and our understanding of whether the differential treatment within groups is justified.¹⁴ Such results should be analysed in conjunction with the broader economic and social environment of a particular country. Singapore is a high-income economy, where the female labour force participation rate in 2021 was 64.2 percent. Other countries might differ in results due to potentially different labour rates.

When evaluating how 'fair' a system is, therefore, the first task should be to understand the fairness philosophy and objectives that were chosen. This is contingent on a range of features, including "the system's exact purpose, the consequences of its operation, the people affected, and the values of the people responsible for its design."¹⁵

In practice, it is unlikely that any AI system will be considered universally fair. This is because different measures of fairness conflict with one another, and there will often be disagreement about which measure is the most appropriate. To address this challenge, the goal is to improve the operation of AI systems with respect to the fairness objectives selected by their operators. This requires making difficult trade-offs between different kinds of fairness for different groups of people affected by the AI system.

CASE STUDY (A trade-off example: AI in Hiring)

Overview

A company is using an AI system to assist with recruiting, assessing new job applicants. The company wants to ensure fairness in its hiring process and considers two fairness measures: demographic parity and equalised opportunity. In this case, assume there are two groups: Group A and Group B.

Demographic parity focuses on achieving equal outcomes for different demographic groups. The AI system is designed to hire an equal number of candidates from each group. However, due to variations in qualifications and skills, Group A has a higher proportion of qualified applicants compared to Group B. To achieve demographic parity, the AI system would need to hire a higher percentage of unqualified applicants from Group B to match the number of hires from Group A. This trade-off may result in the company compromising the quality of its hires by favouring demographic parity over individual qualifications.

On the other hand, equal opportunity focuses on ensuring that candidates have an equal chance of being hired if they possess the necessary qualifications. In this case, the AI system would prioritise hiring the most qualified candidates regardless of their demographic group. This approach aims to eliminate biases and provide equal opportunities for all applicants. However, it may result in outcome disparities, as one demographic group may have a higher proportion of qualified candidates than the other. This trade-off means that the company may achieve fairness regarding equal opportunity but may not achieve demographic parity.

Impact

This example demonstrates the trade-offs that companies face when implementing fairness measures. Different fairness criteria may conflict, and achieving one measure of fairness may come at the expense of another. Balancing these trade-offs requires careful consideration of the organisation's goals, values, and context.

Recommendation

Ultimately, the choice of fairness objectives for an AI system will depend on the system's purpose, the people it affects, and the values of those responsible for its design. To create a system that aligns with the values of the society in which it operates, it is necessary to carefully consider the implications of the system's operation and prioritise fairness in a way that is meaningful and appropriate for the context. By diligently navigating these considerations, it becomes possible to foster a system that not only adheres to societal values but also upholds fairness as a paramount principle in its operations.

4 Responsible Use of AI

The responsible use of AI involves using the technology in ways that prioritise ethical considerations, minimise potential harms, and contribute positively to society. It encompasses a range of principles, practices, and guidelines that guide the development, deployment, and management of AI systems. The responsible use of AI can contribute to addressing social implications and ethical concerns, and strengthen governance and risk management practices. By integrating Trustworthy AI Principles into their operations, organisations can align their AI strategies with sustainable and ethical goals, promoting long-term value creation while minimising negative impacts.

Similarly, by embracing Trustworthy AI, governments can harness the benefits of AI technologies while mitigating related risks and ensuring that AI is used in a manner that aligns with the responsibility to uphold societal values and the inclusivity of their citizens. Theoretically, governments could assess ESG risk data related to AI in order to identify and gain a competitive advantage of similar heretofore risks in their own operations. After utilising ESG data to gain insights into the risk associated with AI, governments can then use this information to shape their own regulations and internal policies concerning AI.

While AI has the potential to benefit environmental stewardship, social well-being and inclusion, and accountability, current governance structures, policies, and systems do not fully support those endeavours. Fairness, as one of the Trustworthy AI Principles, plays a large role in building a foundation for overcoming significant hurdles in a number of sectors.



The implementation of AI poses fairness challenges for various industries:

Sector	Challenge	Real-World Fair Case
Healthcare	Biases can occur in medical diagnoses if the AI algorithms are trained on unrepresentative datasets, leading to the risk of incorrect diagnoses and treatments.	To mitigate this, some medical institutions are leveraging interpretable AI models for explainable results. These models not only predict but also provide explanations for their predictions, helping doctors better understand AI-based diagnosis.
Finance	Financial institutions have faced criticism for discriminatory lending practices, often due to biases in historical data.	To counter this, some institutions have started using AI to implement fair credit scoring algorithms that minimise discrimination and improve loan accessibility.
Education	There can be unintended biases in AI systems used for college admissions. This could lead to unfair decisions that disproportionately affect certain groups of students.	To overcome this, some educational institutions have adopted bias-aware admission systems that consciously avoid these biases.
Transportation	With the advent of autonomous vehicles, issues related to safety and ethics have gained prominence.	To ensure fair and safe decision-making, some companies are developing ethical frameworks for AI-driven decision-making in their autonomous driving systems.
Social Services	In the context of social services, the fair distribution of resources can be a significant challenge.	Some organisations have begun to use AI to identify resource disparities and inform more equitable distribution strategies.
Retail	In an era where consumer data is a valuable asset, protecting customer privacy is a top priority.	Some retail companies are using privacy-preserving AI techniques, such as differential privacy, to analyse data without violating individual privacy rights.
Energy	The energy sector faces challenges related to optimising energy consumption and improving sustainability.	AI can be used to predict and manage energy usage patterns, aiding in the transition towards more sustainable practices.

This table showcases different sectors, the specific challenges they face in relation to the responsible use of AI, and real-world cases that seek to address those challenges. It provides a concise sectoral overview and actionable insights for the responsible use of AI.

The above implications impose challenges that must be acknowledged. It compels the implementation of practices that will balance transparency, analysing scenarios on a case-by-case basis and seeking to demonstrate activities that avoid any business conduct risks associated with fairness misalignment. While fairness is critical in also accelerating environmental stewardship, the focus of this report will be on the “S” factor.

5 The “S” Factor

AI is increasingly influencing issues that fall within the social dimension of ESG, such as employee welfare, diversity and inclusion, and workforce transformation, which are now merging with the growing influences of AI. The following paragraphs offer an analysis of AI’s growing role in the social dimension from each of these elements’ perspectives and reflect on how government and organisations leaders can fairly navigate these social factors.

Employee welfare:

AI technologies are increasingly being used to monitor both job-related and non-job-related behaviour and data. There are two particular ways in which this trend has accelerated in the last few years, partially due to the COVID-19 pandemic. Firstly, companies have collected health-related data as part of preventive health schemes for workers, blurring the line between work and personal life. Secondly, with remote work becoming more prevalent, AI systems are reportedly monitoring employees in their homes. These practices raise significant privacy concerns throughout the entire data life cycle, especially when workplace monitoring data is merged with non-job-related data.

Additionally, there is a risk of ‘function creep’, where the collected data is used for purposes other than what was initially communicated to employees. The use of AI systems for people management, such as hiring algorithms, can produce risks of introducing biases. If an algorithm is trained on biased historical data that favours certain demographics, it can perpetuate discrimination against marginalised groups who are equally qualified for job opportunities.

Furthermore, there is often a lack of accountability structures and transparency to protect workers. Employees are frequently left with little or no explanation regarding AI-based monitoring practices, creating a sense of unease and uncertainty. While companies may have legitimate interests in preventing workplace misconduct, the invasive nature of quantifying social interactions and performance goals through AI-based monitoring practices is often disproportionate.

Under Fairness Lenses

Overall, the reportedly extensive and invasive use of AI-based monitoring practices raises significant privacy risks, undermines worker rights, and highlights the need for accountability, transparency, and addressing biases in these systems. While workforce practices are constantly changing due to the implementation of technologies, company’s policies, and internal practices still need to evolve to align well with these changes. For organisations seeking to place fairness at the foundation of these practices, two recommendations will be increasingly needed:

1. The adoption of transparency in policies and internal practices, for example, identifying and providing information in relation to AI-based decisions to employees; and
2. The development of accountability procedures, ensuring that employees remain to have the alternative to dispute AI-based decisions and that any errors or biases can be rectified.

USE CASE (AI-empowered employee surveillance system)

Overview

A British financial institution was reportedly investigated by the UK Information Commissioner's Office in relation to their use of monitoring software. This software was allegedly used to track the time that employees spent at their desk, and the time taken to complete certain tasks. Where employees appeared unproductive, the software would recommend a range of measures to improve their efforts.

Impact

The use of the monitoring software caused significant backlash from the public, the Trades Union Congress, and the institution's staff. In response to these concerns, the financial institution claimed it would change its use of the software to only track anonymised data.

Recommendation

While such tools can potentially improve an organisation's overall productivity and efficiency levels, it is equally necessary for leaders to develop and implement robust internal governance frameworks in preparation for upcoming policies and regulations in this area. Minimum levels of transparency, such as disclosure of the use of such technologies, their purpose, and reach, are likely to be increasingly expected.



Diversity and Inclusion:

As it will be shown in Chapter 4, social discrimination is a significant issue related to AI, which particularly reflects concerns over technology reinforcing or deepening existing inequalities. The impact of these issues has been perceived over the last few years through hundreds of reported negative AI use cases and is often exacerbated by media attention. In reality, the impact of AI on diversity and inclusion can encompass both advantageous and detrimental effects, depending on how it is designed, implemented, and used. There is a counterargument where AI can, in fact, also be used to address diversity and inclusion challenges, highlighting potential accessibility discrepancies.¹⁶ A recent study has revealed the potential of LLMs from an inclusion perspective in policy communication.¹⁷

USE CASE (Harambee Youth Employment Accelerator)¹⁸

Overview

Harambee is a social enterprise working to solve youth unemployment in South Africa with big data and machine learning (ML) solutions.

Impact

Harambee has engaged with more than 1 million young individuals, amassing what they consider the largest dataset on youth employment in South Africa. Harambee collaborates with industry analysts to analyse these extensive datasets and offer comprehensive visualisation capabilities. These partnerships aid in tasks such as determining the proximity of job seekers to potential employers and facilitating matches with accessible locations. Whenever job openings emerge in a specific area, Harambee employs algorithms to extract location coordinates from nearby job candidates and computes their commuting variables to optimise job placement.

Recommendation

Harambee's utilisation of AI/ML technology in its operations serves as an illustration of organisations striving to establish structures and frameworks that promote the responsible use of AI. Harambee demonstrates awareness of the potential negative consequences associated with AI usage in their work, as evidenced by their cautious deployment primarily in test cases. The formation of an ethics council to supervise their AI algorithms played a crucial role in their achievements.

Under Fairness Lenses

It is important to note that, while AI can contribute to advancing diversity and inclusion, it should be designed and implemented with care. AI systems are only as unbiased as the data they are trained on. If the training data contains biases, such as gender or racial biases, the AI system can perpetuate and amplify those biases. This can result in discriminatory outcomes, further exacerbating existing inequalities. Regular monitoring, evaluation, and ongoing training from humans of AI models are necessary to ensure that they continue to operate in an unbiased and fair manner.

By incorporating considerations of diversity and inclusion, AI systems can be designed to be more responsive to the diverse needs of society, uphold fundamental rights, and reflect the current values of contemporary societies. Given the potential effects of AI technologies, especially on vulnerable individuals and groups, continuous scrutiny and adaptability are essential.

Workforce Transformation:

AI technologies have the potential to reshape the workforce by automating repetitive tasks and augmenting human capabilities. This transformation can have both positive and negative social implications. On one hand, AI can unburden employees from mundane tasks, enabling them to focus on more creative and fulfilling work. On the other hand, there may be concerns about job displacement and the need to reskill or upskill workers to adapt to AI-driven changes. The recent breakthroughs in generative AI are certainly adding to these concerns.

As a response, there have been more research attempts to identify and provide guidance on the extent to which different sectors will be affected. Goldman Sachs recently published a study estimating that roughly two-thirds of U.S. occupations are exposed to some degree of automation by AI.¹⁹ They further estimate that, of those occupations that are exposed, roughly a quarter to as much as half of their workload could be replaced.²⁰ Importantly, however, the study claims that only some of that automated work will translate into layoffs.

While historically, jobs displaced by automation have been offset by the creation of new jobs, advances in AI are expected to have far-reaching implications for society. Building human capacity for using AI is crucial to integrate AI technologies into the world seamlessly. A recent study by economist David Autor found that 60% of today's workers are employed in occupations that didn't exist in 1940.²¹ This implies that the technology-driven creation of new positions explains more than 85% of employment growth over the last 80 years.

Under Fairness Lenses

In the social context, the principle of fairness extends to a company's interaction with its employees, customers, suppliers, and other stakeholders. It encompasses maintaining equitable practices in recruitment, promotion, compensation, and other aspects of employment. The advent of technological advancements, particularly AI, has amplified the scope of this dimension. AI plays a critical role in automating certain functions, which necessitates not only fair and unbiased AI systems, but also proactive strategies for managing AI's impacts on employment. These strategies must ensure equitable opportunities for upskilling and reskilling, thus demonstrating a robust internal commitment to fairness. This commitment must also include mitigating potential negative impacts of AI-driven automation on the workforce, such as job displacement, with a focus on protecting vulnerable groups and ensuring equal opportunities for all.

CASE STUDY (Finnish UBI experiment)²²

Overview

In 2017 and 2018, Finland conducted a two-year trial where a random sample of 2,000 unemployed individuals received a monthly Universal Basic Income (UBI) instead of traditional unemployment benefits. The purpose of the experiment was to explore how UBI could address job disruption and provide financial security to individuals in the face of automation and changing labour market dynamics.

Impact

While the Finnish UBI experiment did not lead to a permanent nationwide implementation of UBI, it provided valuable insights into the potential benefits and challenges of such a system. The results indicated that UBI did not significantly increase employment levels during the trial period, but it positively impacted participants' well-being, perceived economic security, and trust in social institutions.

Recommendation

Although the Finnish UBI experiment is a noteworthy example, it's important to note that job disruption and the policies to address it are complex issues that vary across countries and contexts. Different governments employ various strategies such as reskilling programs, job placement services, and social safety nets to address job disruption and support individuals in the face of changing labour market dynamics.

AI's impact on the social aspect of ESG is both transformative and multifaceted. By leveraging AI responsibly and ethically, organisations have the opportunity to enhance employee well-being, foster diversity and inclusion, improve customer satisfaction, and drive positive social change. However, it is crucial for companies to navigate the potential risks and challenges associated with AI deployment, such as privacy concerns, bias, and the equitable distribution of benefits.

Generative AI systems like ChatGPT are rapidly evolving and are already automating a range of tasks that were once done by humans. This is expected to have a significant impact on the job market, with many entry-level jobs being at risk of being replaced by AI. The typical entry-level jobs that will be automated are those that require synthesising large amounts of data and creating content. For numerical data, this may include computer programmers, market analysts and accountants. For other text-based data, this may include paralegals, journalists and graphic designers. The inaccessibility of entry-level positions would limit prospective employees from gaining work experience that would equip them with the knowledge and relationships to attain higher positions. Importantly, even if generative AI systems do not automate *all* positions for a specific job, they may significantly reduce the number of available positions. This is because labour efficiency will improve if humans become more productive when using generative AI technologies. As noted by Oxford University economist, Dr Carl Benedikt Frey, the increased competition for the remaining positions could lead to lower wages.²³

The responsibility in how to fairly make this transition is to be shared among organisations, investors and governments. AI-implementing companies can share a commitment to AI deployments that do not decrease employee job quality and account for the potential displacement of workers. Investors possess a role to account for the downside risks posed by practices harmful to workers and the potential value created by worker-friendly technologies. Equally, governments can adopt domestic policies and incentives towards upskilling and reskilling the workforce for the jobs of the future.



6 Bias in AI

It is important to note that while AI can offer a wide range of opportunities across different fields and industries, it also comes with challenges and risks such as ethical implications, privacy concerns, and potential biases that need to be addressed to ensure the responsible use of the technology. This chapter provides a more comprehensive examination of bias as a significant risk associated with AI and its impact within the context of social considerations.

Bias may arise due to modelling choices or discrimination (whether historic or contemporary) in a system's dataset. In some cases, attempts to reduce some kinds of biases may create or exacerbate others.²⁴ As algorithms become increasingly powerful and are relied on to make decisions with further-reaching consequences, the potential for bias to disadvantage individuals and groups increases.

Often, datasets will contain insufficient data on minority groups such that the AI system cannot be properly trained.²⁵ This leads to biases that can cause negative outcomes to these groups. There are two risks worth highlighting. First, this can undermine a company's effort to create a diverse, inclusive company. Such efforts can be hindered by biases in various stages of a company's operations that use AI systems, including hiring decisions; resource allocation; marketing plans; pricing; and customer engagement.²⁶ Second, the reputational risk of being regarded as using biased AI systems, or failing to mitigate bias, may dampen investor sentiment in a company, posing an acute financial risk. Even if the manufacturers of AI systems *should* be regarded as the culpable parties, it is foreseeable that the companies using the systems may also suffer these financial risks.

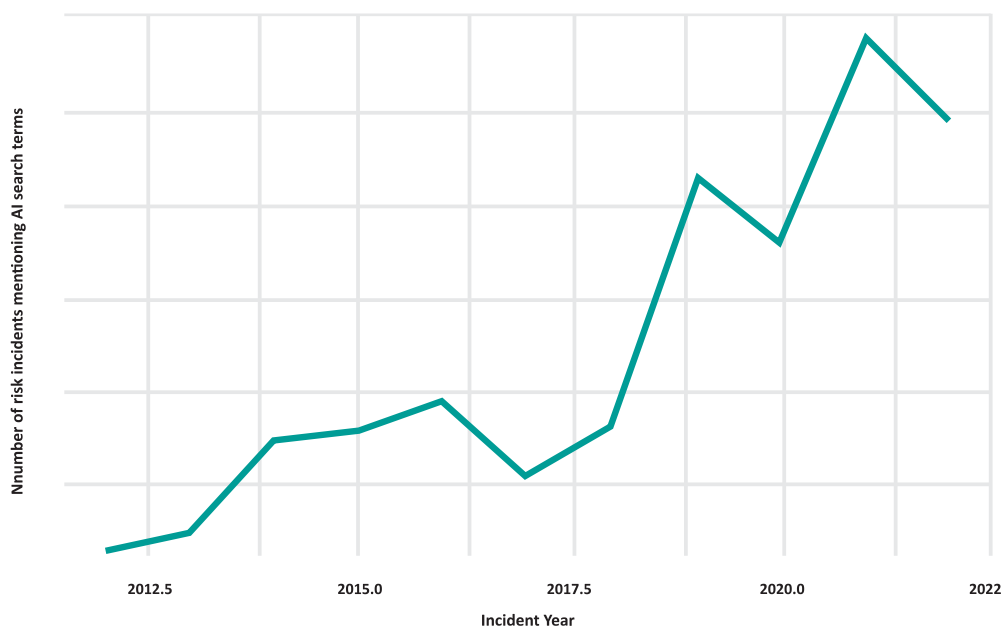
In addition to bias and closely related is *social discrimination*, which arises when AI bias favours the interests of dominant groups in society, "entrench[ing] the status quo."²⁷ A further concern relates to human rights abuses. This includes the fundamental human right to privacy. The collection of personal data in both private and personal settings poses risks to individuals—whether in the form of data breaches containing sensitive information; erroneously drawn inferences (for example, that an individual is a terrorist threat); or through opaque decisions.

The effects of bias and other shortcomings in AI systems can lead to *harmful products*. Credit scoring systems, for example, have demonstrated lower levels of accuracy²⁸ when assessing the creditworthiness of low-income and minority homebuyers. A misallocation of credit is not the only negative consequence of this bias. Individuals who are inadequately rated may miss out on opportunities to demonstrate their trustworthiness, preventing them from accumulating assets and building wealth.

Contrastingly, other products become harmful when used by malicious agents. AI-driven *fraud* is an increasing risk that public and private agents must be aware of. Examples of this fraud include intentionally misleading, generated information; and scam messages. When used at scale, this can affect all industries, as agents can manipulate online readings and write fake reviews, generate fake documents online, and automate mass scam messages.²⁹ In March 2023, cybersecurity vendor Bitdefender cautioned individuals about a fraudulent investment scheme being promoted on a counterfeit version of ChatGPT.³⁰ As new AI systems proliferate, internet users will need to be wary of their authenticity. It is also foreseeable that malicious agents will use AI-generated messages to, for example, represent themselves as selling financial products.

Finally, there are concerns that AI systems can fuel *anti-competitive practices*. When companies use the same pricing algorithms, for example, there is a risk that their pricing strategies become interdependent, keeping prices above the competitive level.³¹

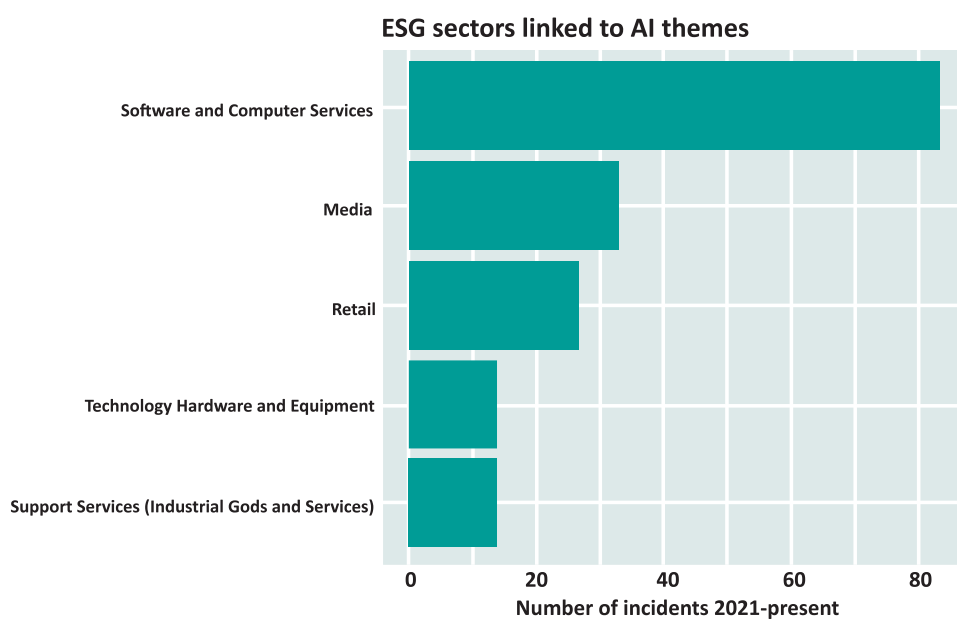
Risk Incidents that include an element related to AI are on the rise.³²



Source: AI Asia Pacific Institute | Chart: Fairness in AI: Impact and Opportunities

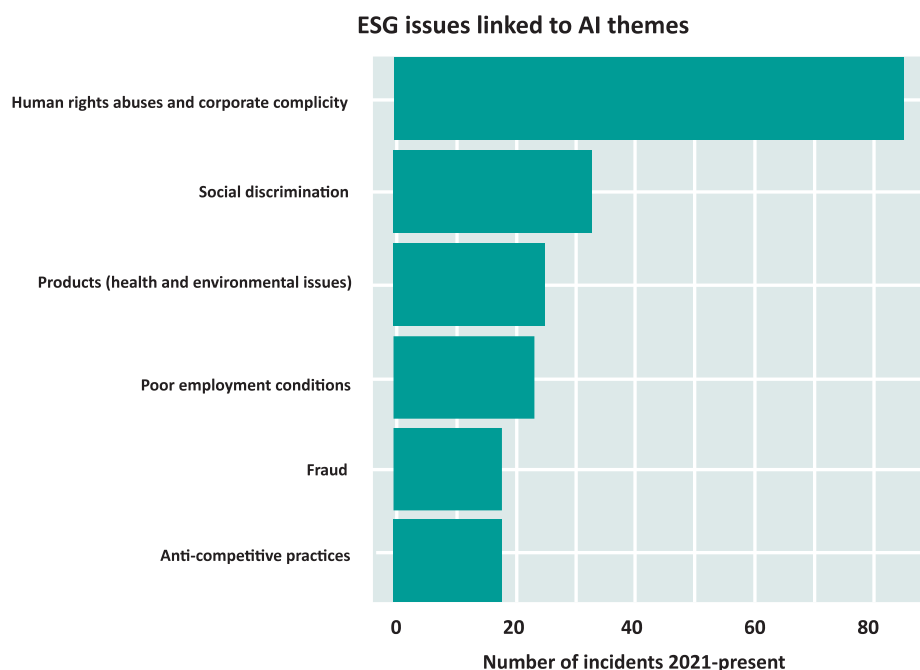
There are approximately 3x as many risk incidents with an element related to AI now than five-ten years ago.

The presented tabular data indicates that a significant portion of AI themes within ESG risk incidents are in the sectors of software and computer services. Nonetheless, notable incidents also extend to the media and retail sectors, wherein social media platforms assume a prominent role.



Source: AI Asia Pacific Institute | Chart: Fairness in AI: Impact and Opportunities

Human rights, social discrimination, and product issues are the most recurring ESG issues relating to AI. Noting that human rights abuses and corporate complicity can often take the form of aforementioned use cases under the “S” factor, involving the detriment of employee welfare, for example. As also covered under the “S” factor, social discrimination is another critical issue for AI, with concerns over technology reinforcing or deepening existing inequalities.



Source: AI Asia Pacific Institute | Chart: Fairness in AI: Impact and Opportunities

Financial Institutions’ Practices to Address AI Risks

Across the banking sector, a proactive approach to mitigating the aforementioned risks, particularly bias, is increasingly being adopted. This typically involves pre-processing, in-processing, and post-processing strategies akin to those proposed by IBM's AI Fairness 360 Toolkit and Microsoft's fairlearn.py package, for example.³³

In the pre-processing phase, many banks are taking steps such as data cleaning and eliminating features directly related to protected classes like race, gender, and age. These measures aim to ensure that the data used to train their AI models is as bias-free as possible.

During in-processing, or the model training phase, banks are leveraging specific techniques to encourage fairness in their algorithms. This can include strategies such as penalising unfairness in the loss function used for model training. Routine bias and fairness audits are also being conducted to evaluate algorithm performance.

In the post-processing phase, banks are taking steps to adjust the outputs of their AI models. Techniques like threshold optimization are commonly used to ensure decisions based on risk scores produced by their algorithms are fair.

In addition to these measures, many financial institutions are now prioritising transparency, making efforts to clarify how their algorithms function and the factors influencing decision-making processes. This is often achieved through comprehensive documentation and sharing some level of information with the users, considering the proprietary nature of these algorithms.

Nonetheless, comprehending the inherent trade-offs entailed in AI implementation and the evaluation of fairness goals is crucial. The advent of language models has introduced fresh challenges and trade-offs when it comes to ensuring fairness. Recent research indicates that although there is a distinct correlation between performance and fairness, fairness objectives and biases can be in conflict. It has been observed that language models exhibiting superior performance on specific fairness benchmarks often exhibit heightened gender bias.³⁴

Fairness can be incorporated into AI systems' design, development, and deployment through various approaches and considerations. These include:

1. **Data Collection and Preparation:** Ensuring that the training data used to build AI models is representative and diverse is crucial. It involves collecting data from a wide range of sources and ensuring that the dataset is balanced and free from biases.
2. **Bias Detection and Mitigation:** Analysing the training data and AI models for potential biases is essential. Techniques such as statistical tests, fairness metrics, and algorithmic audits can help identify and address biases. Mitigation strategies may involve reweighting the data, modifying the training process, or using adversarial techniques.
3. **Inclusive and Diverse Development Teams:** Promoting diversity and inclusion within the teams responsible for developing AI systems can help incorporate different perspectives and mitigate biases. Involving individuals with diverse backgrounds and experiences can contribute to a more comprehensive and fair approach.
4. **Transparent and Explainable AI:** Enhancing transparency and interpretability of AI systems enables stakeholders to understand the decision-making process. This involves developing algorithms and models that can provide explanations for their outputs, ensuring accountability and enabling identification of potential biases or unfair outcomes.
5. **Regular Evaluation and Monitoring:** Continuous evaluation and monitoring of AI systems' performance for fairness are important. Regular audits, impact assessments, and ongoing monitoring can help identify and address any emerging biases or unfair outcomes.
6. **User Feedback and Redress Mechanisms:** Establishing mechanisms for user feedback and redress is crucial. It allows individuals affected by AI systems to report concerns, provide input, and seek remedies if they believe unfair treatment has occurred.
7. **Collaboration and Ethical Guidelines:** Engaging with diverse stakeholders, including policymakers, ethicists, and civil society organisations, can contribute to the development of ethical guidelines and standards for AI fairness. Collaborative efforts help ensure that a broad range of perspectives are considered in the design and deployment of AI systems.
8. **Regulatory and Legal Frameworks:** Governments and regulatory bodies can play a role in promoting fairness by establishing guidelines and regulations specific to AI systems. These frameworks can outline requirements for transparency, accountability, and fairness in AI development and deployment.



By incorporating these approaches, AI systems can be designed, developed, and deployed with fairness as a fundamental principle, reducing biases and promoting equitable outcomes in the financial sector.

7 Conclusion and Recommendations

The promise of AI for government and investment leaders is multifaceted and holds the capacity to transform various sectors and aspects of governance. The AI industry is characterised by prominent power dynamics, wherein government and investment leaders possess a crucial opportunity to shape and influence the advancements within the industry. This position bestows upon them both an opportunity and a responsibility to effectively navigate this technological transition.

In order to harness the potential of AI, leaders must grasp the significance of fairness across different sectors. ESG considerations offer a baseline for capturing impact and developing better internal policies and practices to mitigate AI's potential risks. To foster stronger, trustworthy and equitable growth in AI, a higher degree of collaboration is imperative between public and private entities. The public's confidence in these systems becomes increasingly valuable in the face of potential reputational loss and new regulatory requirements. The ensuing recommendations outline strategies through which governments and investment organisations can endeavour to fully adopt the promise of AI:

1. Under the "S" factor, government leaders carry the crucial responsibility of preparing the labour market for the shift towards an AI-integrated world. However, to implement fairness considerations, this transition must also actively involve the private sector and the investment industry, ensuring an equitable distribution of influence and responsibilities. To ensure a successful and fair transition, investment should focus on two main areas:
 - a) Firstly, a commitment to **collaboration** and open dialogue with all stakeholders should be prioritised. This involves actively encouraging discussions around policy developments related to the use of AI. Key topics such as universal basic income must be approached with fairness in mind, ensuring that such policies consider and address potential disparities and don't disproportionately benefit or disadvantage specific groups.
 - b) Secondly, investment must also aim to empower the entire workforce to effectively use AI technologies. This involves developing and implementing inclusive **upskilling and reskilling** programs that provide equal opportunities for all employees, regardless of their current skills or job roles. Such efforts will ensure a fair transition into an AI-integrated workplace, mitigating the risk of deepening social inequalities due to unequal access to opportunities and resources.
2. Organisations should be **transparent about their chosen fairness measures**, demonstrating an understanding that—in most cases—choosing one measure over another will lead to fairness trade-offs. To this end, they should **justify why the chosen measures are suitable for their AI system's purpose, the groups it will affect, and the values of the organisation**. An understanding of the relevant trade-offs will draw attention to the existing biases and other associated risks of deployed AI systems. Greater transparency should be strived for in underscoring these biases and explaining steps taken to mitigate them.

- 
- 
3. **Preserving a 'human-centered' approach** is a crucial guiding principle in AI, especially when viewed through the lens of fairness. Ensuring human supervision over AI decision-making not only enhances accountability to all stakeholders but also helps prevent inaccurate or undesirable outcomes that could disproportionately affect certain groups. This principle of human oversight helps ensure that AI systems are designed and used in ways that respect human rights, provide equal benefits, and do not lead to unfair discrepancies in outcomes. It maintains the balance of power and decision-making authority, avoiding over-reliance on AI systems and ensuring fairness in their impact on all individuals and communities.
 4. **Build fairness into the system:** Incorporating fairness into AI systems' design, development, and deployment is essential for mitigating biases and promoting equitable outcomes. This involves diverse approaches, including collecting representative and unbiased data, detecting and mitigating biases through statistical tests and fairness metrics, fostering inclusive development teams, ensuring transparency and interpretability of AI models, conducting regular evaluations and monitoring, establishing user feedback and redress mechanisms, collaborating on ethical guidelines, and considering regulatory frameworks.

The potential of AI for government leaders and the financial and investment industry is enormous, but it also carries significant responsibilities. The foremost priority in AI design, development, and deployment should be fairness. We have an opportunity to influence the future of this industry and set an example by ensuring that AI systems are unbiased, inclusive, and provide equitable outcomes.

To achieve this, collaboration between the public and private sectors is crucial. We need to invest in well-informed regulations, empower organisations to address AI bias, and promote transparency and accountability. By embracing these strategies, we can create a future where AI technology serves the greater good, fosters trust, and upholds the values of our nations.

These strategies will guide us through the technological transition and allow us to shape the future of AI for the benefit of society, our financial institutions, and the environment.

Appendix A

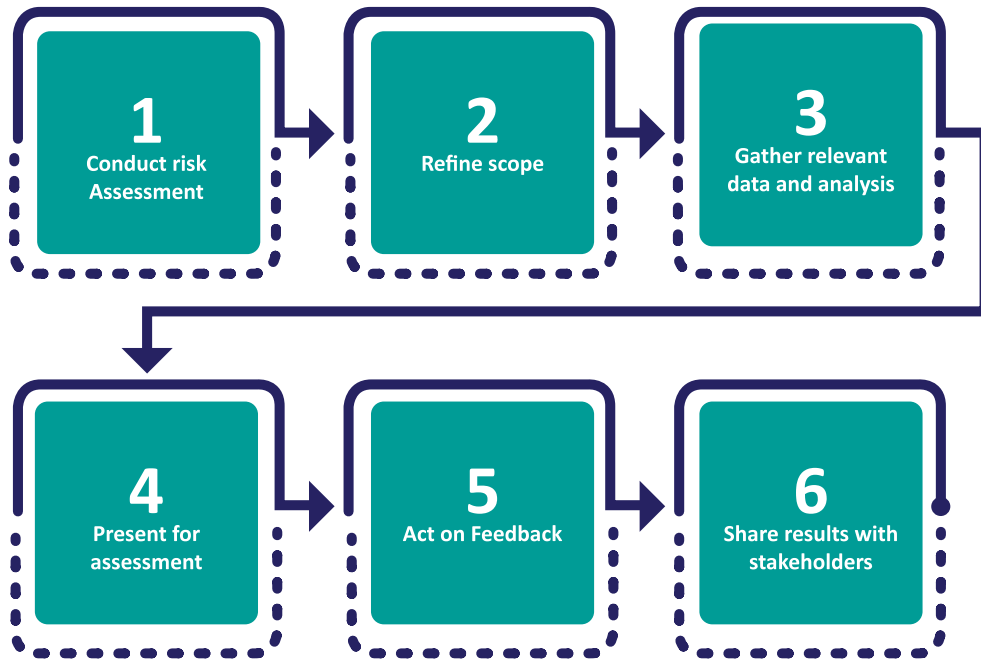
The AI Asia Pacific Institute has published additional reports on the governance of artificial intelligence. Please visit the Research pages on our website for further details.

Key publications

- 2021 Trustworthy Artificial Intelligence in the Asia-Pacific region
- 2022 Trustworthy Artificial Intelligence in the Asia-Pacific region
- 2023 Policy Brief: ChatGPT and Other Generative AI Systems

Appendix B

The Veritas initiative provides a methodology for assessing alignment with their FEAT Fairness Principles. This includes an ordered list of questions, set out in five parts, which are intended to be answered by the owners and developers of AI systems. An example of how the FEAT Fairness Assessment Methodology can be applied is provided in their 'Document 1 Report', which has been reproduced below.³⁵





AI ASIA PACIFIC
INSTITUTE



contact@aiasiapacific.org

Endnotes

- 1 Nestor Maslej et al, *Artificial Intelligence Index Report 2023* (Stanford: Institute for Human Centered AI, 2023), 188, <https://aiindex.stanford.edu/wp-content/uploads/2023/04/HAI_AI-Index-Report_2023.pdf>.
- 2 AI Asia Pacific Institute, *Trustworthy Artificial Intelligence in the Asia-Pacific Region* (2021), <<https://aiasiapacific.org/wp-content/uploads/2021/07/2021-Trustworthy-Artificial-Intelligence-in-the-Asia-Pacific-Region.pdf>>.
- 3 Trustworthy AI is a term used to describe AI that is lawful, ethically adherent, and technically robust. For more information: <https://aiasiapacific.org/wp-content/uploads/2021/07/2021-Trustworthy-Artificial-Intelligence-in-the-Asia-Pacific-Region.pdf>
- 4 The same principles also appear globally as part of international frameworks such as the UNESCO Recommendation on the Ethics of Artificial Intelligence.
- 5 Using the AI Principles Comparison table and extensive industry consultation, we analysed existing developments across four countries: China, Australia, Singapore, and New Zealand. Based on our research findings, we arrived at the following principles to encourage the development of Trustworthy AI in the region: human-centricity, fairness, explainability, transparency, privacy, and accountability, which we refer to as Unified Principles.
- 6 Veritas aims to enable financial institutions to evaluate their Artificial Intelligence and Data Analytics (AIDA)-driven solutions against the principles of fairness, ethics, accountability and transparency (“FEAT”) that the Monetary Authority of Singapore (“MAS”) co-created with the financial industry in late 2018 to strengthen internal governance around the application of AI and the management and use of data.
- 7 Veritas Consortium, *Veritas Document 1: FEAT Fairness Principles Assessment Methodology* (Singapore: Monetary Authority of Singapore, 2020), 34–35, <<https://www.mas.gov.sg/-/media/mas/news/media-releases/2021/veritas-document-1-feat-fairness-principles-assessment-methodology.pdf>>.
- 8 Ibid., 17.
- 9 For a more detailed explanation of these definitions and their shortcomings, see Michelle Seng Ah Lee, “Context-conscious fairness in using machine learning to make decisions,” *AI Matters* 5, no. 2 (2019): 23–29, <<https://sigai.acm.org/static/aimatters/5-2/AIMatters-5-2-07-Lee.pdf>>.

This list is by no means exhaustive. Professor Arvind Narayanan identifies up to 21 definitions of fairness: see Arvind Narayanan, “Tutorial: 21 fairness definitions and their politics,” March 2, 2018, video, <<https://www.youtube.com/watch?v=jIXluYdnyyk>>.
- 10 Jon Kleinberg, Sendhil Mullainathan and Manish Raghavan, “Inherent Trade-Offs in the Fair Determination of Risk Scores,” *arXiv preprint*, 1609.05807 (2016): 4, 17, <<https://arxiv.org/pdf/1609.05807.pdf>>.
- 11 See, eg, Veritas Consortium, *Veritas Document 2*, 11, 56, 68.
- 12 Please refer to Appendix B.

- 13 Veritas Consortium, *Veritas Document 2: FEAT Fairness Principles Assessment Case Studies* (Singapore: Monetary Authority of Singapore, 2020), 112: <<https://www.mas.gov.sg/-/media/mas/news/media-releases/2021/veritas-document-2-feat-fairness-principles-assessment-case-studies.pdf>>.
- 14 Ibid., 113–114.
- 15 Veritas Consortium, *Veritas Document 1*, 17.
- 16 Paul R Daugherty, H James Wilson and Rumman Chowdhury, “Using Artificial Intelligence to Promote Diversity,” MIT Sloan, November 21, 2018, <<https://sloanreview.mit.edu/article/using-artificial-intelligence-to-promote-diversity/>> ; Alexandra Kahn, “Project Us – an AI-powered platform promoting inclusivity in the digital workplace,” MIT Media Lab, April 19, 2023, <<https://www.media.mit.edu/posts/project-us-an-ai-powered-platform-promoting-inclusivity-in-the-digital-workplace/>>.
- 17 Anne Lundgaard Hansen and Sophia Kazinnik, “Can ChatGPT Decipher Fedspeak?,” *SSRN* (March 24, 2023). <<http://dx.doi.org/10.2139/ssrn.4399406>>.
- 18 Araba Sey, Oarabile Mudong, Case Studies on AI Skills Capacity-building and AI in Workforce Development in Africa (Research ICT Africa), <<https://researchictafrica.net/wp/wp-content/uploads/2021/07/AI-Capacity-Case-Studies-Final.pdf>>.
- 19 “Generative AI could raise global GDP by 7%,” Goldman Sachs, April 5, 2023, <<https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html>>.
- 20 Ibid.
- 21 David Autor et al, “New Frontiers: The Origins and Content of New Work, 1940–2018,” Working Paper No 30389, *National Bureau of Economic Research* (August 2022): 12, <<https://www.nber.org/papers/w30389>>.
- 22 Jouko Verho, Kari Hämäläinen and Ohto Kanninen, “Removing Welfare Traps: Employment Responses in the Finnish Basic Income Experiment,” *American Economic Journal: Economic Policy* 14, no. 1 (February 2022): 501–522.
- 23 Jacob Zinkula, “Even if ChatGPT doesn't take your job, it could help another human replace you, says the economist who famously concluded AI could eliminate nearly half of US jobs,” *Business Insider*, February 7, 2023, <<https://www.businessinsider.com/jobs-at-risk-replaced-ai-chatgpt-oxford-economist-2023-2>>.
- 24 Veritas Consortium, *Veritas Document 1*, 41.
- 25 This is an unfortunate yet intuitive reality: less populous groups in society will not be sampled as much as larger groups, leading to less accurate predictions. See, eg, Trishan Panch, Heather Mattie and Rifat Atun, “Artificial intelligence and algorithmic bias: implications for health systems,” *Journal of Global Health* 9, no.2 (2019): 2, <<https://doi.org/10.7189%2Fjogh.09.020318>>.
- 26 Robert Bentley, *The Importance of Good Corporate Governance for ESG and AI* (GuyCarpenter, 2022), 2, <https://www.guycarp.com/content/dam/guycarp-rebrand/pdf/Insights/2022/2022.4-Importance-of-Good-Corp-Governance-ESG-and-AI-v3_final.pdf>.

- 
- 
- 27 Mike Zajko, “Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates,” *Sociology Compass* 16, no. 3 (2022): 8.
 - 28 Edmund L Andrews, “How Flawed Data Aggravates Inequality in Credit,” Stanford University Human-Centred Artificial Intelligence, August 6, 2021, <<https://hai.stanford.edu/news/how-flawed-data-aggravates-inequality-credit>>.
 - 29 Laura Harris, “AI Fraud: The Hidden Dangers of Machine Learning-Based Scams,” ACFE Insights, 6 January, 2023, <<https://www.acfeinsights.com/acfe-insights/2023/1/6/ai-and-fraud>>.
 - 30 Aaron Hurst, “Financial scam utilising fake ChatGPT discovered by researchers,” Information Age, March 6, 2023, <<https://www.information-age.com/financial-scam-utilising-fake-chatgpt-discovered-by-researchers-123501934/>>
 - 31 See OECD, *Algorithms and Collusion: Competition Policy in the Digital Age* (2017), 44, <<https://www.oecd.org/competition/algorithms-collusion-competition-policy-in-the-digital-age.htm>>.
 - 32 RepRisk data.
 - 33 Genevieve Smith, “What does ‘fairness’ mean for machine learning systems?,” Haas Berkeley, 2020, <https://haas.berkeley.edu/wp-content/uploads/What-is-fairness_-EGAL2.pdf>.
 - 34 Maslej et al, *AI Index 2023 Annual Report*, 129, 142.
 - 35 Veritas Consortium, *Veritas Document 1*, 26.