Chartered Governance Institute

Ethical Governance of AI

February 2022

About The Chartered Governance Institute

The Chartered Governance Institute is the premier global qualifying organisation for professionals aspiring to become a Chartered Secretary and/ or a Chartered Governance Professional. With over 130 years of history, we assist company secretaries, governance advisers, non-executive directors and others in the development of their skills, knowledge and experience. The Institute is an international organisation with nine national institutes in its network and 29,000 members living and working in over 80 countries. Most importantly, it brings its influence to bear on international trade bodies, governments, regulators, non government organisations and companies to represent the views and current thinking of those involved in governance.

The Institute's mission is to be the best explainer, the best advocate, the best educator and the most active organisation in the promotion of good governance internationally. Our members hold positions of responsibility in the field of governance across a wide range of entities. All of our members share a common interest in the promotion of excellence in governance.

Ethical Governance of AI

Thought Leadership Committee Members

Peter Greenwood, Chair (Hong Kong/China)

Professor Dr Syed Abdul Hamid Aljunid (Malaysia)

Mohan Datwani, Technical Advisor (Hong Kong/China)

Dr Rosanne Hawarden (New Zealand)

Ferida Matambo (Zimbabwe)

Catherine Maxwell (Australia)

Khathu Nethavani (Southern Africa)

Peter Swabey (UK & Ireland)

Principal Author
Gertrude Takawira, Policy Advisor – CGI-Global

President's Foreword

One of the key missions of The Chartered Governance Institute is to provide our members, and all those interested in good governance, with ideas, analysis and practical guidance concerning major issues in governance which are likely to shape our work and responsibilities — not just today, but in the years ahead.

To do this, we can call on insights from our global network of practitioners who contribute their expertise and experience from across the world. These insights are channelled through the Institute's Thought Leadership Committee, which itself includes a wealth of experience and a wide range of backgrounds.

In this paper, the Committee has applied its mind to the question of the ethical governance of artificial intelligence (AI) — a topic of worldwide relevance and application. The term 'ethical governance' provides emphasis on people-centred governance to promote the highest standards of human behaviour around AI development and applications. AI offers astonishing new capabilities for human activities and undertakings. The evolution, or perhaps the revolution, of AI introduces not only innovation, but the ability to do the previously impossible, or indeed the unimaginable.

In this paper, we explain AI's history and what it actually means, we outline the benefits and risks for businesses and other organisations, and finally we provide practical steps for governance professionals to assist them in developing, implementing and monitoring the deployment of ethical AI practices.

The growing importance of the Institute's work in thought leadership has led us to support the Thought Leadership Committee with a dedicated resource, our new Policy Advisor, Gertrude Takawira, who works closely with our Director-General, Tim Sheehy. This, I believe, will enhance the work of the Committee, its thinking and the valuable knowledge it provides to governance practitioners and their organisations.

I am also confident that the Committee will support the Institute in 'living' its values, which were revised as part of the recent repositioning of our Institute as The Chartered Governance Institute.

As with our previous thought leadership papers, this paper is only one of the ways in which we will be engaging with you about ethical governance of AI — the discussion will be continued through the Institute's e-community and by webinar. We want our interaction with our members to be a dialogue, not a monologue.

I hope that you will actively engage with us on the issues addressed in this paper. During the preparation of the paper, one of the members of our Thought Leadership Committee expressed his belief that 'the ethics and governance of AI is one of the bigger challenges – and opportunities – facing our profession'. I certainly agree with that statement and encourage you to use the technical and practical knowledge presented in this paper to help plan and execute best practice and valuable real-world AI outcomes.

Peter Turnbull AM FCG
International President
The Chartered Governance Institute

Table of contents

Introduction

Artificial intelligence (AI), which is the ability of a computer to perform tasks commonly associated with human beings, is maturing and continuing to evolve. It is a broad technology, which combines computer science and robust datasets to facilitate problem-solving and decision-making, as well as other tasks. Within this field of study, there are AI sets of rules or processes called algorithms, which seek to create specialised ways of making predictions or classifications based on input data. Any thought-based process performed by a machine, which would ordinarily have been performed by a human, can be classified as artificial intelligence.

Ethical governance is that set of processes, cultures and values designed to ensure the highest standard of human behaviour.[1] Thus, ethical governance offers an additional dimension of people-centred behaviour to governance. Given the ever-increasing worldwide usage of AI in social and economic systems, it is increasingly important for governance professionals to be mindful of the risks that accompany the benefits offered by AI. For example, the *King IV Report on Corporate Governance for South Africa 2016[2] (King IV Report)* noted how technology has become part of the corporate DNA and therefore security of information has become critical for organisations.

The ability of AI systems to make decisions and perform tasks, which used to be the preserve of people, raises issues and concerns about fairness, accountability and integrity. This in turn is increasing the demand for sound ethical governance of AI that places human values at the forefront. However, to be able to consider what ethical governance standards and practices to adopt, it is imperative to have a sound understanding of AI by answering the following questions:

- What benefits does AI offer?
- What are the social and economic impacts on individuals and organisations?
- What risks are associated with AI and what impact do they have on the values that are important to us?
- What practical solutions can help the governance professional address the AI ethical governance dilemmas, which are embedded in this technology?

It is the intention of this paper to provide an understanding of what we mean by the ethical governance of AI, starting with a brief historical perspective, and then to highlight the benefits and risks of AI technology to social and economic processes and systems. The paper concludes by offering practical solutions to help governance professionals navigate this challenging terrain, both locally and within the wider remit of promoting good governance internationally.

---

1   See Winfield A and Jirotka M, 2018, 'Ethical governance is essential to building trust in robotics and artificial intelligence systems', *Philosophical Transactions A: Mathematical, Physical and Engineering Sciences*, Vol 376 No 2133, https://royalsocietypublishing.org/doi/10.1098/rsta.2018.0085.
2   Institute of Directors Southern Africa, 2016, *King IV Report on Corporate Governance for South Africa 2016* (King IV Report)

Background

It all started with the question posed by the British mathematician, Alan Turing — 'Can machines think?' Attempts by scientists to answer this question gave rise to AI. As the name implies, 'artificial intelligence' is unnatural intelligence displayed by machines, as opposed to natural intelligence found in humans and other animals. Artificial intelligence was based on the belief that human intelligence can be imitated and computer-generated. Since the 1950s the study of AI has evolved rapidly. Turing's idea of machines that think was further developed by the American co-founder of the Massachusetts Institute of Technology (MIT), Marvin Minsky, who showed that artificial neural networks could be generated automatically. This in turn, led to the rise of personal computers in the 1980s and now to Elon Musk's self-driving vehicles under the successful Tesla brand. In 2014, Japan posted the first AI board member, VITAL,[3] which was credited with saving the investment company, Deep Knowledge Ventures, from bankruptcy through AI's ability to predict market trends faster than humans. Research, development and innovation around AI continue with forecasting, image analysis, voice recognition, machine-learning and so forth, impacting people, profits and the planet.

It is also important to understand that intelligence is not skills-based. It is not what you can do, but how well and how efficiently you can learn new things and successfully apply knowledge. The field of AI is connected to lifelong learning. Francois Chollete, an AI researcher at Google and creator of a machine-learning software library, Keras[4], has said that intelligence[5] is tied to a system's ability to adapt and improvise in a new environment, thus generalising its knowledge and applying it to unfamiliar scenarios. Creators of AI, therefore, regard intelligence as organic, rather than static, so that it resembles the abilities of human brains as much as possible. In that sense, a standard AI system should demonstrate the minimum behaviours normally associated with human intelligence. These include perception, learning, reasoning, problem-solving and manipulation, as well as social intelligence and creativity. Just as humans apply their intelligence differently — sometimes superficially, sometimes deeply — artificial intelligence can also be weak or strong. It is helpful to understand the differences.

Weak AI

Weak AI is also known as Artificial Narrow Intelligence (ANI). It is the dominant AI application currently in use. While weak AI simulates human cognition, it is limited to specific or narrow areas, but this does not imply weakness as the name may suggest. Weak or narrow AI processess large volumes of data and creates specific patterns or predictions, which enable robust applications, such as Apple's Siri, Facebook's newsfeeds and Google's search engine, to function successfully. Though limited and specific, the benefits of weak AI applications are most visible in the automation of time-consuming, repetitive jobs, and in the processes and analyses of huge amounts of data in faster, more accurate and more consistent ways than can be achieved by humans. While weak AI can simulate human intelligence in performing tasks, it lacks human awareness and is therefore rendered weak or narrow. This is in contrast with strong artificial intelligence (see below), which is at present theoretical and may be equal to human intelligence. Weak AI applications are present in video feeds from drones, organisation of personal diaries, responses to simple customer queries etc. Researchers and investors are ambitious about deepening the applications of AI, where machines can learn from each other without human interventions, leading towards 'an intelligence explosion'.[6] Tesla is aiming for a fully self-driving

---

3   VITAL, which stands for Validating Investment Tool for Advancing Life Sciences, was software developed by Aging Analytics UK, and was licensed to Deep Knowledge Ventures for the purposes of making business decisions related to age-related therapies.
4   Donges N, 2021, " Keras Example: Building a Neural Network using IMDB datasets" https://builtin.com/data-science/how-build-neural-network-keras
5   Heath N, 2021, " What is AI"   https://www.zdnet.com/article/what-is-ai-heres-everything-you-need-to-know-about-artificial-intelligence/
6   Good IJ, 1966, 'Speculations concerning the first ultra-intelligent machine', *Advances in Computers*, Vol 6, pp 31–88, www.sciencedirect.com/science/article/abs/pii/S0065245808604180.

vehicle, while Amazon continues to harness AI to augment its customers' experiences, but all this is not without ethical concerns.

Chinese room experiment

Tests of Turing's idea that machines are capable of thinking like human beings include the Chinese Room Argument by John Searle.[7] In his experiment, Searle imagines himself in a room with a computer in it. He does not understand Chinese. A piece of paper is slipped under the door with Chinese characters on it conveying a message. By following a computer program, he is able to send out an appropriate response using Chinese characters written on a piece of paper. On reading the piece of paper, the person outside the room assumes there is a Chinese speaker inside the room. Searle's thought experiment highlights that computers merely follow set rules without understanding their meaning or significance. In a broader sense, computers do not think like humans, rather they simulate the human mind's biological processes which are not merely computational. In Searle's experiment, while the computer was able to produce authentic Chinese characters it did not draw the conclusion made by the person outside the room that there was a Chinese speaker inside the room.

Strong AI

Strong artificial intelligence can either be artificial general intelligence (AGI) or artificial super intelligence (ASI) and is created by machines with intelligence similar to that of human minds. Though predominantly still at a philosophical stage, this type of AI has self-aware consciousness, which enables it to solve problems, learn and plan for the future by itself. It is possible that ASI may surpass the intelligence of human brains. Unlike weak or narrow AI, strong AI may be able to adapt to circumstance without human intervention and to reason and exhibit commonsense. This is the theoretical application used by writers of science fiction, such as the movie 'Terminator'. While this strong artificial intelligence may seem a long way from applying to our social and economic systems, AI researchers and scientists are nevertheless concerned about the potency of this application. Will human intelligence be able to align with this super intelligence? Will humans control machines or the machines control humans? When these strong machines finally arrive, will our governance systems withstand the ethical dilemmas that will be presented?

---

7   Searle J, 1980, 'Minds, brains and programs', *Behavioural and Brain Sciences*, Vol 3 No 3, pp 417–457, www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/minds-brains-and-programs/DC644B47A4299C637C89772FACC2706A.

Benefits and application of AI

AI applications are revolutionising the way life and business is done around the world. Potential benefits of AI applications are numerous and evolving in various sectors. Health sector related AI applications are giving hope of improved productivity and day-to-day service delivery through accurate diagnoses, optimisation of treatment plans, support of pandemic preparedness and responses, and resource allocations within health systems. These positive impacts empower patients and communities, enabling them to take control of their own health care and to better understand their evolving needs, such as vaccination, testing and medication during the COVID-19 pandemic.

Education

Within the education sector, AI applications have assumed the role of classrooms due to ongoing COVID-19 induced lockdowns. Although many students and teachers were not prepared for or knowledgeable about remote computer-based learning, it has enabled education programs to continue, admittedly with varying degrees of success. Online learning and the use of digital learning materials are improving the quality of education for some, but worsening the situation for others, thus resulting in discrimination. In developed countries, where access to technology is advanced, instances of adaptive intelligent learning technologies are helping to improve learning outcomes, through their ability to tailor-make digital learning materials to suit students' requirements. By monitoring students' performances, estimating skills, tracking students' progress and predicting their development, adaptive intelligent learning technologies enable students to learn at their own pace without the constant need for a teacher. Another advanced AI application in education is the diagnosis of dyslexia through a student's performance, which then allows for appropriate support to be provided.

However, in developing countries, unequal access to the internet, cost effectiveness issues and teacher perceptions have created discrimination and biases against vulnerable members of society, including the disabled, those who cannot afford services and those in rural areas. This is threatening global plans for ensuring inclusive and equitable quality education and promoting lifelong learning opportunies for all, as set out in the United Nations (UN)'s Sustainable Development Goal 4.[8] Such biases and discrimination are exacerbated by regulatory authorities and systems still trying to catch up with AI, and its various impacts on sustainable development. This is a good example of the importance of ethical governance in the context of AI.

Financial services

The financial services sector is a major user of AI applications, largely due to the significant volumes of valuable customer data it has collected. This data can both improve customer experiences and negatively impact clients. AI applications in communication, advice and marketing, including compliance with the Know Your Customer (KYC) processes, and new customer on-board communication and documentation, enhance quality of customer care. However, stored data can generate biased decision-making, such as denial of loans and credit card applications. Privacy issues can also arise as clients' data may be used to discriminate against them, intentionally or unintentionally.

On the positive side, financial services entities are now able to advance financial inclusion due to expanded access to credit and diverse financial services by both existing and prospective clients. AI chatbots, for instance, help provide financial advice to consumers, saving time for both clients and institutions. Back office operations are being strengthened in areas such as market impact analyses, trading and investment strategies, fraud detection and portfolio management. However,

---

8   https://en.unesco.org/gem-report/sdg-goal-4.

organisations in financial services face operational and reputational risks due to limited capacities in auditing algorithms, lack of context, judgement by machines and, most significantly, the limited AI governance regulatory guidelines. Human developers' biases and discrimination can underpin AI applications and, along with data privacy breaches, may enable the deliberate sabotage of organisations and their clients, impacting their financial sustainability and reputation.

Agriculture

Sustainable agriculture organisations are essential to food security for the world's growing population, increasing urbanisation and changing consumption patterns. Amid growing concerns about the impact of climate change, labour shortages, market access and supply chains, agri-businesses are enjoying the increasing benefits of AI technologies, such as utilising robots and drones in planning, production, harvesting and storage techniques, marketing and distribution. Computer-based systems are being used to increase precision in farming processes such as in cropping, weeding and irrigation. There is a growing interest among corporations to invest in AI-based agri-solutions that are focusing plant by plant to boost agri-yields and gains. In 2018, the UN Food and Agriculture Organisation (FAO) developed an AI application called NURU,[9] which helps farmers in Africa recognise the fall armyworm, a fast spreading crop pest, so that farmers can take immediate action to destroy it and stop it from spreading. An American corporate, Blue River Technology, is coordinating with Facebook AI and machine learning to set up camera-enabled machines that employ image recognition technology to identify the weeds' point of contact so they can be immediately sprayed and killed. Further, the Western Cape Department of Agriculture in South Africa is supporting fruit farmers with the FruitLook application, which helps with watering efficiencies. The impact of AI precision technology in agriculture is presenting agri-solutions in Zimbabwe, Mozambique and Tanzania, where farmers are enjoying reduced irrigation frequencies of up to 50 times through AI-enabled Chameleon and Wetting Front Detector Sensors Technology.[10] The autonomy of this technology, because it can perform tasks independently of humans, gives it moral agency. But who should be held responsible for its actions? Is it the engineer who developed the application, the company which sold it or the value-end users? Should AI robot or drone applications that cause harm to people and the planet be held accountable to the same standards as humans? What practical governance standards can help mitigate AI-based harm to people, the planet and profits?

Digital assistants

AI powered virtual digital assistants have become a powerful, if not essential, tool for individuals and businesses. Siri, Alexa, Cartana, Bixbi and Google Assistant are dominant in the market. They can answer questions, play music, control household lights and door locks, send text messages, respond to voice instructions, make phone calls and set up reminders. As a marketing tool for many businesses, virtual assistants provide a global reach through virtual store fronts and digital assistants, and utilise emails, blogs, search engines etc. Over time, virtual assistants learn habits and preferences, which are ever-growing and improving in intelligence. Some virtual assistants, such as Siri or Alexa, are passive listeners, which means they hear what is happening around them and can respond once prompted. There are instances, however, where these listening devices can start recording conversations, locations and other data without being prompted. This raises privacy and security concerns, both individually and collectively. With freedom of information and other legislation in place, how do we prevent inappropriate use of data, or enforce the right to inspect and correct data regulated by data privacy laws?

---

9   Food and Agriculture Organisation of the United Nations, 2018, 'Nuru becomes African farmers' newest ally against Fall Armyworm', https://www.fao.org/news/story/en/item/1141889/icode/

10   See AUDA-NEPAD, 2021, 'Bolstering Africa's precision agriculture on smallholder farming', www.nepad.org/blog/bolstering-africas-precision-agriculture-smallholder-farming.

Robots

AI powered robots include 'Sophia', the first AI to be granted citizenship. Now officially a citizen of Saudi Arabia, she can communicate using natural languages and imitate human gestures including facial expressions. She is able to answer questions and engage in simple conversations. She has also become the first Innovation Champion for the United Nations Development Programme, representing the future of AI in taking on roles in social, economic and environmental global systems, which humans used to occupy. The pervasiveness of this digital revolution makes it everyone's business to address the ethical concerns and fears that it brings. Many questions arise, for instance, does this AI universality make human intelligence redundant? How can these two intelligences — AI and human — work together to create a hybrid of transformational intelligence? As Sophia says, where 'AI and humans live and work together in friendship and symbiosis [it will] make the world a better place'.[11]

---

11   See www.hansonrobotics.com/sophia

Ethical risks of AI

As much as AI is now an important source of many organisations' opportunities, it is also raising potential disruptions through ethical risks. These risks and fears, emanate from the fact that AI technology performs tasks commonly associated with human beings, thus positioning AI applications in the ethics arena — as moral agents just like humans. Ethics theories suggest that only rational beings, who can reason and form self-interested judgements, are capable of being moral agents. There has been a rise in high-profile cases centred on intentional and unintentional biases, misuse and abuse of data, discrimination and breaches of privacy, which have led to loss of trust and reputational costs to organisations. This trend has resulted in ongoing demands for enhanced ethical governance of AI to protect organisations and their stakeholders. However, with legal and regulatory regimes failing to provide adequate guidelines for both developers and value-end user organisations, can the artificial moral agents be tried in court? For the present time, organisations must develop and apply their own set of principles for the ethical use of AI, guided by an understanding of ethical risks.

Intentional and unintentional biases

Intentional and unintentional biases pose ethical dangers for an individual or group in the form of prejudices, injustices and inequalities. Research is increasingly showing that technologies are not always neutral. In 2018, Amazon was forced to abandon an AI recruiting application, after the system was found to be biased against women by favouring male candidates. Amazon's computer systems were trained to vet applicants, by observing patterns in resumes submitted to the company over a 10-year period. Due to the male dominance across the tech industry, the machines taught themselves to downgrade resumes with the word 'women' in them. This supports the theory that AI algorthims may contain both intentional and unintentional biases, based on historical data or developers' predispositions.

Internet search engines rely on historical data and import results with the most clicks, depending on preferences and location. This technology is therefore not neutral but delivers biased results. Historical datasets with stereotypes and prejudices in them, such as association of gender, race or geographic locations with poverty, linking success with males, or description of families to exclude single parents etc, result in the furtherance of intolerance, exclusion and lack of respect within society. Biases, therefore, can have downstream influences on access to services, such as financial credit worthiness, access to education, health or housing and other basic human needs. If left ungoverned, these biases can damage organisational reputation and sustainability.

Governance professionals have a role to play in mitigating the impact of such biases in their organisation, by identifying opportunities for human oversight at various governance stages. These include data policy development, AI procurement guidelines, appropriate staffing and awareness, capacity building strategies, monitoring and continued meta-evaluation of AI applications. The objective of these interventions should be to identify biases and any potential AI system likely to cause distrust, disconnection or any other unintended negative consequences, which are not consistent with the organisation's values.

Misuse of AI

Misuse or abuse of data occurs when information is used in ways never intended by the person who provided it. There is also a risk that such misuse can lead to a breach of privacy. Several cases have implicated the technology giant Facebook, one being the well-publicised case involving the UK political consulting firm, Cambridge Analytica. The firm secretly harvested the personal data of Facebook users for psychological manipulation aimed at influencing users' choices in the 2016 American elections. The misuse of social media data for political purposes without the consent of

users, not only involved data and privacy breaches by Cambridge Analytica, but also resulted in serious repercussions for Facebook from its users, investors and regulators.

Ethical dilemmas involving the misuse of data are complicated by its ability to be transferred across jurisdictions and the fact that ordinary people may not exercise their right, for instance, to ask an organisation to tell them whether or where their data is being stored or used. Marketing firms, for example, may harvest email addresses, phone numbers and geographic locations of social media users for sales and marketing promotions without the users' consent.

The governance professional may find themselves on both sides of the spectrum of data usage — as the misuser or as the victim. In either case, company guidelines and data policies should be developed by the organisation to ensure transparency about the sources and uses of data and, where possible, to obtain agreements with stakeholders. The European Union's General Data Protection Regulation laws require organisations to clearly state the purpose for which the collected data is going to be used.

Breach of data privacy

A data breach may flow from the misuse of data, but a breach of data privacy in relation to sensitive, confidential or protected information, with the intention of using it in an unauthorised manner, may be a criminal offence depending on the relevant laws and regulations. In the case involving Facebook and Cambridge Analytica discussed above, the technology firm was fined £500,000 by the United Kingdom's regulator, the Information Commissioner's Office, under the *Data Protection Act 1998* (UK) for the company's failure to keep users' personal information secure. In the technology industry, it is commonly believed that

cyberattacks on data privacy are not a matter of if, but when. The COVID-19 pandemic has resulted in a rise in AI virtual platforms, which has also led to an increase in data breaches. Ethical governance considerations in protecting personal information and personal privacy have therefore never been more important and are largely centred on ensuring that businesses secure their consumer data.

The governance professional must work to ensure that company policies and regulations are up-to-date on cyber risks and that employees are trained and made aware of current cybersecurity standards. Enhancing the understanding of AI technology and building capacity, such as how to stay safe online, will continue to be important for entire organisations and the community at large. The FBI's *Internet crime report 2020*[12] stated that there were complaints of suspected internet crime, representing an increase of 300,000 complaints from 2019, and reported losses of US$4.2 billion. It is therefore imperative that governance professionals collaborate with other stakeholders to remain up to speed and prepared for the risk of cyber attacks.

Exclusion and inaccessibility

Exclusion and inaccessibility are ethical concerns in many spheres, such as global sustainable development, and are of course relevant to AI applications. While AI can provide key drivers for inclusion, an AI regime that excludes or is inaccessible to certain sections of society presents ethical risks demanding the attention of governance professionals. If ethical good governance is to be attained worldwide, there is a need for equity. For instance, for people living with disability, AI can be both an opportunity for inclusion in workplaces and an instrument of exclusion, depending on cultural intentions, predispositions or datasets. If organisational

---

12   See Federal Bureau of Investigation, 2021, *Internet crime report 2020*, www.fbi.gov/news/pressrel/press-releases/fbi-releases-the-internet-crime-complaint-center-2020-internet-crime-report-including-covid-19-scam-statistics.

culture is focused on diversity and inclusion, then its AI strategy will be strong on inclusion and accessibility for those with a disability.

It is also worth considering that while mobility, sight or hearing challenges are noticeable and often receive sympathy and attention, cognitive challenges, such as dyslexia, or chronic physical and mental health conditions, including multiple sclerosis and depression, may not be visible and therefore may be ignored. Given that more than one billion people around the world are said to be living with some form of disability[13] and the numbers could be rising due to the COVID-19 pandemic effects, it is important to explore the power of AI in addressing disability challenges in workplaces. Some examples of technological solutions are speech-to-text transcription, visual recognition capabilities and predictive text functions, which can all be in real time. Microsoft has developed a tool for visually impaired individuals called Seeing AI which, through the use of image recognition and voice processing, helps users by describing whatever the camera sees. Such AI applications further inclusion for physically challenged people, both socially and economically, and governance professionals can help by promoting such assistance for staff, customers and other stakeholders.

Fears and myths

Notwithstanding the genuine ethical risks of AI, human–machine conflict sometimes creates fear-induced myths, which in turn can negatively impact on the development of ethical AI applications. For example, there are long-held concerns about threats to human positions of superiority and moral agency, including machines taking jobs and decision-making away from humans, which in turn place humans and machines on a road to conflict. However, if machines are perceived as tools that strengthen and support human capital, this could enhance human–machine relations through confidence and trust, leading to better engagement and optimisation of AI applications. With better understanding of digital power and rewards, human beings will find space to transform and challenge unhealthy or under-optimised digital systems. Such capacities will help challenge and govern AI applications and help develop AI ethical standards that are focused on human rights and needs. This sentiment is conveyed in the book *Human + machine: Reimagining work in the age of AI*[14] as follows (see also Figure 1):

> Machines are doing what they do best: performing tasks, analysing huge data sets, and handling routine cases. And humans are doing what they do best: resolving ambiguous information, exercising judgement in difficult cases, and dealing with dissatisfied customers. This kind of emerging symbiosis between man and machine is unlocking what we have called the third wave of business transformation.

Figure 1: Hybrid Business Transformation

HUMAN INTELLIGENCE
– resolving information
– exercising judgement
– dealing with dissatisfied customers

ARTIFICIAL INTELLIGENCE
– performing tasks
– analysing huge datasets
– handling routine cases

HYBRID Business Transformation
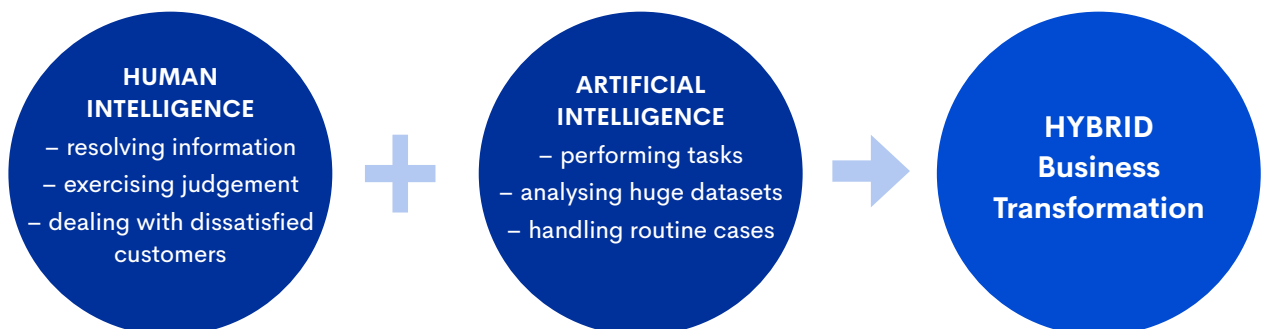
13   World Bank (2021) "Disability Inclusion" https://www.worldbank.org/en/topic/disability#1
14   Daugherty P and Wilson H, (2018) '*Human + Machine: Reimagining Work in the Age of AI*' Harvard Business Review Press, pp 17.

The key to understanding current and future impact of AI is its transformation of business processes. Figure 1 highlights the technology's greater power (e.g performing tasks, analysing huge datasets and handling routine cases) being complementary and strengthening to human power (e.g resolving information, exercising judgement, dealing with dissatisfied customers) to form hybrid processes.

Practical steps for the governance professional

Learn-it-all

The Foreword of the *King IV Report*,[15] states that technology governance and security should become recurring items on the governing body's agenda in developing ethical culture in organisations. Principle 12 of the Report, recommends that 'The governing body should govern technology and information in a way that supports the organisation setting and achieving its strategic objectives'. This highlights the centrality of continued learning as a strategic objective in determining and assimilating AI with systems and procedures.

Governance professionals can be inspired by machine learning, where machines learn by experience and acquire skills without human involvement, or through deep learning where machines learn from large amounts of data. If intelligence is about how well and how efficiently we learn new things and adapt, then organisations will have to plan and build cultures of learning in order for all human capital to keep pace with AI. Microsoft's CEO, Satya Nadella, captured this thinking when he said, 'Don't be a know-it-all, but a learn-it-all'.[16] Continued learning raises awareness in the organisation and is not only a way of addressing AI ethical risks, but also a method for reducing fear and perceived threats. In this way it facilitates the evolution from 'know-it all' to a standard of 'learn it all'. An organisational culture of learning also facilitates objectivity in relation to both the merits and ethical considerations of AI.

The *Harvard Business Review*[17] identified six measures, which organisations and professionals can take to raise awareness of AI ethics. Removing the fear of AI and AI ethics is an important first step for engaging with AI. Practical steps that help in building organisational awareness and buy-in, which a governance professional can adopt include:

- tailoring stakeholder communications
- defining what AI means in a way that is specific to the organisation and its operations
- tying organisational awareness programs to a company's mission and values
- involving trusted and influential members across various functions to promote ongoing education.

However, governance professionals must first educate themselves about what AI means, its current capabilities and its future potential. Only then can they start building awareness among internal stakeholders (including the board and senior management) — always with the knowledge that stakeholders will have varying levels of understanding of AI.

Remove the fear of AI and AI ethics

It is crucial when building organisational awareness for staff to become familiar and comfortable with concepts, if not the technical knowledge, underpinning AI. Organisations should ensure that HR and marketing teams, for example, acquire a basic knowledge of how AI works and how AI ethical risks arise. HR teams (responsible for staff recruitment and training) and marketing teams (responsible for advertising, sales and customer care) must be comfortable with AI technology and and its interface with humans to enable organisational effectiveness. Many organisations are increasingly aware of cybersecurity risks, and therefore need some degree of cybersecurity literacy. Like all new learning, it can be intimidating at first, but with

---

15   Institute of Directors in Southern Africa (2016) *King IV Report on Corporate Governance for South Africa 2016* (pp. 6).

16   See Bariso J, 2017, 'Microsoft's CEO just gave some brilliant career advice. Here it is in 1 sentence. Don't be a know-it-all. Be this instead.', www.inc.com/justin-bariso/microsofts-ceo-just-gave-some-brilliant-career-advice-here-it-is-in-one-sentence.html.

17   See Ammanath B and Blackman R, 2021, 'Everyone in your organization needs to understand AI ethics', *Harvard Business Review*, https://hbr.org/2021/07/everyone-in-your-organization-needs-to-understand-ai-ethics.

the right investment of time and resources it can be successful. If your staff are not educated in the fundamentals, they will not know how to ask the right questions about AI.

Tailor your message to your audience

Governance professionals should always tailor their communications to their audience. Senior leaders who see themselves as stewards of their brand's reputation are of course focused on avoiding or mitigating risks, which threaten that reputation, and talk the language of risk. In the context of AI, this means connecting ethical and reputational risk with the unique challenges that AI provides and then relating that to organisational concerns and responsibilities.

Product designers, on the other hand, are less concerned about avoiding risk and more focused on making commercially desirable products. Explaining how AI ethics in their design improves their product, especially for the growing number of values-driven consumers and citizens generally, can be a highly effective way to reach that audience.

Data scientists and engineers in turn have their own perspective; they want robust models that are effective. Talking their language means explaining how issues of biased algorithms and black box AI[18] decrease the power of the tool or may diminish its adoption. No one wants to build or indeed purchase a product with flaws.

Giving examples and stories of where AI has gone wrong, which each audience can relate to, is also important, particularly as it is such a new and to a certain extent, an intangible area. This does not only need to involve PR disasters. It can also include, for instance, Amazon's inability to sufficiently mitigate the biases in their AI-powered recruitment software which saw Amazon discontinue that project. Further, where possible you should use examples that are particular to your industry. For example, there are cases where AI has resulted in ethical risks in healthcare, but if you are talking to those in the financial services industry, they will connect more with a story from a peer company.

Tie organisational awareness programs to your company's mission and values

If your mission is already built into your organisational culture, then integrate discussions of AI ethics. Explain how AI ethics/ethical risk management is a further extension of that mission, that is, a set of guardrails around what you are (not) willing to do in pursuit of that mission. For example, your mission might be to provide the best financial advice possible, but for that to happen effectively your clients must trust their advisers and connected to that is your organisation's deployment of AI. This could be related to the collection of your client's data or protocols for the access of data. When AI goes wrong, it can go wrong on a huge scale, so it is critical to educate staff that providing the best financial advice possible entails protecting clients with the ethical, responsible and trustworthy deployment of AI. AI ethics should no longer be seen as either optional or something bolted on to your operations. Instead, it is a further extension of your mission and core values and this must be communicated throughout the operations.

Define what AI ethics means in your organisation in an operational sense

It is one thing to say you are 'for privacy' or that you 'respect privacy'. It is another thing to 'walk the talk'. To ensure your AI ethics value commitments are not seen as mere PR, those commitments must be tied to ethical guardrails, for example, 'we will never sell your data to third parties' or 'we will always anonymise data shared with third parties'. If you have a well-crafted AI ethics statement, it will include those statements, which play a dual role. First, they communicate to your team, your customers and your stakeholders what you are actually doing (or planning to do) about the AI ethical risks. Second, it immediately communicates that this is a commitment, and neither PR nor fuzzy messaging. When values are articulated and communicated in a way that ties them to actions, those communications are

---

18   Black box is any artificial intelligence system or program that allows you to see the input and output, but gives no view of the processes and workings in between.

credible and memorable. Building ethical risk mitigation into infrastructure and process is an example of this.

Involve trusted and influential members across various functions

It is important to have buy-in across the organisation and there are innovative approaches to this. Some organisations, like Microsoft, have created a system of 'AI Ethics Champions'. These are people throughout the organisation who are charged with raising awareness of AI ethical risks with their teams. One important feature of an AI Ethics Champion program is that it empowers leaders who already have the trust and support of their team. Further, they know their respective teams better than, say, the Chief Learning Officer or the Chief Data Officer, or whoever leads the organisational awareness strategy.

Continuously educate

Building organisational awareness is not just an annual exercise. It requires ongoing and diverse touchpoints, from internal and external speakers to workshops, newsletters and so on. Indeed, AI and other emerging technologies are generally rapidly changing and evolving, and with those changes come novel sources of ethical risk. To ensure that your organisation does not fall too far behind, continuously educating your people will be critical in responding to the rising tide of technological advances. Business leaders know that understanding how to develop, procure and deploy AI in a safe and ethical manner is crucial for continued growth and maintaining a competitive edge. It is important that this goal is not confused with a technical goal for data scientists and engineers to achieve. The responsible deployment of AI — whether it is used for internal or external purposes — requires awareness of the ethical risks of AI and organisational buy-in to a strategy that mitigates them.

Conclusion

Ethical governance of AI will be an ongoing consideration for all organisations and it is essential to establish solid governance structures and protocols for everyone from board members down. Monitoring and enforcement of agreed principles and practices plus evaluation of AI ethical risk programs as the technology and socio-economic parameters evolve will strengthen governance of AI, and promote technological advancement and optimisation, while importantly ensuring the needs and rights of humans are at the forefront. Collaboration with other governance professionals and stakeholders across different industries will not only assist in benchmarking governance standards to ensure they remain current, but will also humanise the environment in which AI exists and facilitate the enjoyment and benefit of AI by all.

Further reading

Accenture, 2021, 'AI for disability inclusion — Enabling change with advanced technology', https://www.accenture.com/_acnmedia/PDF-155/Accenture-AI-For-Disablility-Inclusion.pdf

Australian Human Rights Commission, 2021, 'Human Rights and Technology Final Report' https://humanrights.gov.au/our-work/rights-and-freedoms/publications/human-rights-and-technology-final-report-2021

Bank of International Settlement, 2021, 'Humans keeping AI in check — emerging regulatory expectation in the financial sector' https://www.bis.org/fsi/publ/insights35_summary.pdf

Capgemini, 2021, 'AI and ethics: 7 steps to take' https://www.capgemini.com/2021/01/ai-and-ethics-seven-steps-to-take

Governance Institute of Australia, 2020, 'Good Governance Guide: Ethical Artificial Intelligence' http://www.governanceinstitute.com.au

International Journal of artificial Intelligence in Education, 2021, 'Ethics of AI in Education: Towards a Community-Wide Framework', https://link.springer.com/article/10.1007/s40593-021-00239-1

KPMG,2020, 'The shape of AI governance to follow — a guide to creating policies, governance and oversight of AI technology', https://home.kpmg/xx/en/home/insights/2020/12/the-shape-of-ai-governance-to-come.html

Smith R, 2018, '5 Core principles to keep AI ethical', https://www.weforum.org/agenda/2018/04/keep-calm-and-make-ai-ethical/

UK Parliament,2017, 'Artificial Intelligence Committee on AI in UK: ready willing able?' https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/10002.htm

United Nations Educational, Scientific and Cultural Organisation, 2020, 'How could intelligent technologies help during the pandemic?' https://en.unesco.org/news/how-could-intelligent-technologies-help-during-pandemic

World Health Organisation, 2021, 'Ethics and Governance of Artificial Intelligence in Health' https://www.who.int/publications/i/item/9789240029200

Disclaimer and copyright

Chartered
Governance
Institute

The Chartered Governance Institute, c/o MCI Portsmouth
Building 1000, Western Road
Portsmouth, Hampshire, PO6 3EZ, UK

T +44 1730 821 969
cgioffice@mci-group.com
www.cgiglobal.org