Briefing

Business Ethics Briefing ibe

Issue 58 | January 2018

Business Ethics and Artificial Intelligence

This briefing sets forth a framework of fundamental values and principles for the use of Artificial Intelligence (AI) in business. Its primary goal is to encourage organisations to engage in a multi-stakeholder dialogue that always considers commitment to values in the application and impact of AI developments.

What is artificial intelligence and what is the impact of AI developments on our society? What are the biggest risks that new technologies can pose? How will we seek to control the way it affects our daily lives? Are we preparing ourselves sufficiently?

The rapid development and evolution of AI technologies, while unleashing opportunities for business and communities across the world, have prompted a number of important overarching questions that go beyond the walls of academia and high-tech research centres in the Silicon Valley. Governments, business and the public alike are demanding more accountability in the way AI technologies are used, trying to find a solution to the legal and ethical issues that will derive from the growing integration of AI in people's daily lives.

In business ethics in particular, the rise of artificial intelligence has brought about considerations including the level of control organisations can retain over their machines' decision-making processes and how to ensure that the AI systems they adopt always act in a way that is in line with the organisation's core values. If it is true that with great power comes great responsibility, AI has become increasingly powerful and its applications to business are only starting to reveal its potential.

Some business leaders have indicated that responsible organisations need to redefine how they interact with technology to be able to be seen as trustworthy in the age of artificial intelligence.[1] People tend to trust those individuals or institutions that operate with openness and take account of the public interest. Working with regulators and policy makers, businesses have the opportunity to make a significant contribution to agree on a framework of ethics and norms in which AI can thrive and innovate safely.

Box 1 *What is Artificial Intelligence (AI)?*

Artificial Intelligence (AI) is a term generally used to describe the simulation of elements of human intelligence processes by machines and computer systems. It is characterised by three main features:

1. **Learning** – the ability to acquire relevant information and the rules for using it
2. **Reasoning** – the ability to apply the rules acquired and use them to reach approximate or definite conclusions
3. **Iterative** – the ability to change the process on the basis of new information acquired.

Particular applications of AI include speech and face recognition and 'driver-less' vehicles.

Science fiction often portrays AI machines as human-like robots that look, think and feel like a human – Blade Runner 2049 is the most recent example of this type of narrative. The robots of the Hollywood movies are often evil, vindictive and try to take over the world. Despite the fact that they do not provide an accurate depiction of the state of artificial intelligence, this has an impact on how people perceive AI, which is often feared and seen as a danger or a threat.

Futuristic visions of AI can catch media attention, whilst true immediate risks of AI are sometimes overlooked. These include:

- **Ethics risk**: certain applications of the AI systems adopted might lead to ethical lapses;
- **Workforce risk**: automation of jobs can lead to a deskilled labour force;
- **Technology risk**: black box algorithms can make it difficult to identify tampering and cyber-attacks;

[1] Forbes (11/09/2017) Artificial Intelligence Is Here To Stay, But Consumer Trust Is A Must for AI in Business

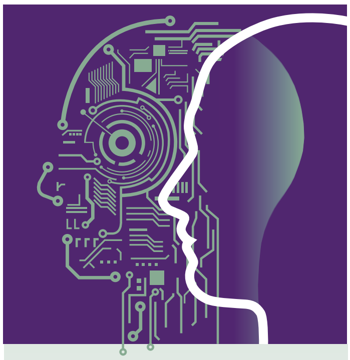
Briefing

Business Ethics Briefing

ibe

Business Ethics & Artificial Intelligence

IBE interactive framework of fundamental values and principles for the use of Artificial Intelligence (AI) in business.



Encouraging organisations to engage in a multi-stakeholder dialogue that always considers commitment to ethical values in the application and impact of AI developments

- A — Accuracy »
- R — Respect of privacy »
- T — Transparency »
- I — Interpretability »
- F — Fairness »
- I — Integrity »
- C — Control »
- I — Impact »
- A — Accountability »
- L — Learning »

- **Legal risk**: data privacy issues, including compliance with GDPR;
- **Algorithmic risk:** biased algorithms can lead to a discriminatory impact.

In the workplace, people often see AI not as a tool they can use to make their job easier or more effective, but rather as something that is done to them and sometimes competes with them to take their jobs. In this context, companies have the responsibility to address the risks that stem from AI and to bring the focus back on the people that are the users of AI, explaining how they would actually benefit from these technologies and why they shouldn't fear them.

Understanding human values to design values-led systems

Winston Churchill once said that *"we shape our buildings and afterwards our buildings shape us."*[2] This, *mutatis mutandis*, applies to artificial intelligence too. AI technologies are not ethical or unethical per se. The real issue is around the use that business makes of AI, which should never undermine human ethical values.

The IBE has engaged with organisations and technology experts to identify the founding values that form the cornerstone for the ethical framework of *ARTIFICIAL* Intelligence in business (see the IBE interactive framework).

Accuracy[3]

Companies need to ensure that the AI systems they use produce correct, precise and reliable results. To do so algorithms need to be free from biases and systematic errors deriving, for example, from an unfair sampling of a population, or from an estimation process that does not give accurate results.

The ethical implications of machine learning algorithms are significant. An example is given by the US criminal justice system, which is increasingly resorting to the use of artificial agents to ease the burden of managing such a large system. Any systematic bias in these tools would have a high risk of errors, as the New York based non-profit organisation ProPublica illustrated with regards to Northpointe's Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) criminal risk assessment system. This software is used in sentencing and parole hearings across the country. ProPublica argued that the system misrepresented the recidivism risks of different convicts. They suggest that there is a systematic racial bias in the risk estimation.[4]

It is critical that the machine learning algorithms that drive AI decision-making are trained on diverse sets of data in order to prevent similar biases. It is also important that organisations set appropriate accuracy levels to determine clearly their expectations and what an acceptable standard is for them.

It is worth noting that in some instances, because AI can learn from data gathered from humans, it might be that some human biases are reflected in the machine's decision-making. Microsoft had to delete an AI chatbot that it had introduced on Twitter with the aim to improve its communications skills by interacting with the users of the social media platform. After only a day in a human community, the chatbot transformed into an evil Hitler-loving, incestual sex-promoting, 'Bush did 9/11'-proclaiming robot.[5] This indicates how, even in the era of artificial intelligence, influencing human behaviour to embed ethical values should remain at the forefront of every conversation about business ethics.

Respect of privacy

The rise of AI has been described by some as the death of privacy,[6] while others have compared it to an Orwellian Big Brother ready to scoop on everyone's private life. Certainly, machine learning technologies have brought about new ethical issues related to the respect of privacy. On this topic, the European Commission has recently reinforced the principle that everyone has the right to the protection of personal data by proposing a comprehensive reform of data protection rules (*General Data Protection Regulation - GDPR*).[7]
Article 8 of the EU Charter of Fundamental Rights states that "*Everyone has the right to the protection of personal data concerning him or her. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other*

[2] In October 1943, following the destruction of the Commons Chamber by incendiary bombs during the Blitz, the Commons debated the question of rebuilding the chamber.

[3] In this briefing, the term 'accuracy' is used to describe the ability of algorithms to produce unbiased outcomes in terms of algorithmic fairness.

[4] ProPublica (23/05/2016) Machine Bias

[5] The Telegraph (24/03/2016) Microsoft deletes 'teen girl' AI after it became a Hitler-loving sex robot within 24 hours

[6] The Guardian (03/08/2014) The death of privacy

[7] http://ec.europa.eu/justice/data-protection/

legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified."[8]

The GDPR will apply from 25 May 2018. It is important that organisations have prepared appropriately for it and plan their approach to GDPR compliance in advance, to gain buy in from key people and to ensure that they have communicated the appropriate changes to their staff.

Transparency and openness

Professor Luciano Floridi, professor of philosophy and ethics of information at the University of Oxford, compares AI to a dark room: "*Suppose you enter a dark room in an unknown building. You might panic about monsters that could be lurking in the dark. Or you could just turn on the light, to avoid bumping into furniture.*"[9] More openness in the use of AI algorithms and systems can help to shed some light into the dark.

Traditionally, many organisations that have developed and now use AI algorithms do not allow public scrutiny as the underlying programming (the source code) is proprietary, kept from public view. Opening sourcing material in computer science, when appropriate, is an important step. It helps the development community to understand better how AI works and therefore be able to explain it more accurately to the public and the media. This is particularly important as better information within the general public improves trust and prevents unjustified fears. Moreover, the more people have access to the code, the more likely it is that bugs and long-term opportunities and risks can be worked out.

Microsoft, Google, Facebook and Amazon have been making remarkable progress developing artificial intelligence systems. Recently they have released much of their work to the public for free use, exploration, adaptation and perhaps improvement. Box 2 illustrates some examples..

Box 2 *Open technology*

Among the technologies that major tech companies have opened recently are:

- Google's TensorFlow, the heart of its image search technology, open-sourced in November 2015; [10]

- The custom hardware designs that run Facebook's M personal assistant open-sourced in December 2015; [11]

- Microsoft's answer to these machine learning systems, the prosaically named Computation Network Tool Kit, was made public in January 2016. [12]

Interpretability

As AI algorithms increase in complexity, it becomes more difficult to make sense of how they work. In some cases, AI applications have been referred to as a 'black box' where not even engineers can decipher why the machine made a certain decision. This can significantly hinder their effectiveness, and cause concern. The use of 'black box' algorithms makes it difficult not only to identify when things go wrong, but also to determine who is responsible in the event on any damage or ethical lapse.

Interpretable and explainable AI will be essential for business and the public to understand, trust and effectively manage 'intelligent' machines. Organisations that design and use algorithms need to take care in producing models that are as simple as possible, to explain how complex machines work.

There are some trends that are emerging to make AI more interpretable. As an example, the Defense Advanced Research Projects Agency (DARPA) has launched the Explainable AI (XAI) programme, aimed at creating *"a suite of machine learning techniques that:*

- *Produce more explainable models, while maintaining a high level of learning performance (prediction accuracy);*
- *Enable human users to understand, appropriately trust, and manage effectively the emerging generation of artificially intelligent partners."*[13]

According to DARPA, the *"new machine-learning systems will have the ability to explain their rationale, characterise their strengths and weaknesses, and convey an understanding of how they will behave in the future."*[14]

[8] http://fra.europa.eu/en/charterpedia/article/8-protection-personal-data
[9] Aeon (09/05/2016) Should we be afraid of AI?
[10] Wired (09/11/2015) Google Just Open Sourced TensorFlow, Its Artificial Intelligence Engine
[11] Wired (10/12/2015) Facebook Open Sources Its AI Hardware as It Races Google
[12] Wired (26/01/2016) Microsoft Open Sources Its Artificial Brain to One-Up Google
[13] https://www.darpa.mil/program/explainable-artificial-intelligence
[14] Ibid fn[13]

Fairness

Fairness and justice, which are core issues in the stakeholder theory, remain paramount for ethical businesses when dealing with AI. Fairness to all stakeholders, and to society as a whole, requires businesses to consider the wider impact of AI design and developments. As AI systems are able to perform tasks, previously undertaken by humans, in a more efficient and reliable way, the workplace is going to change and it is therefore important that companies pay attention to how this will affect its employees and customers.

One widely held belief that is certain to be challenged is the assumption that automation will primarily challenge workers who have little education and lower-skill levels. The reality is different. While lower-skill occupation will continue to be affected, a great many university educated, white collar workers are going to discover that parts of their job, too, can increasingly be challenged as software automation and predictive algorithms advance rapidly in capability. Current discussions examine the best approaches that will minimise potential disruptions, making sure that the fruits of AI advances are widely shared and competition and innovation are encouraged and not stifled.[15]

Technology, of course, will not shape the future in isolation. Other major societal challenges will also impact (e.g. ageing population, development of different professional figures). However, companies have a role to play in ensuring that this transition will be smooth. This means tackling issues such as long term unemployment, social inequality and lack of trust from customers in the way AI is utilised.

Integrity

As the saying has it, integrity is doing the right thing even when nobody's watching. In the context of AI, we should ensure that it is used only for its intended purpose, even when there is no means to enforce this. When designing or selling an AI system, it is important to ensure – whenever possible – that the use of AI solutions by third parties is restricted to the intended purpose and the end user will respect the agreed uses of the technology. Binding clauses that define the use for which the technology is intended are helpful, however it is also necessary to conduct appropriate due diligence on the clients as well to minimise the risk of a potentially dangerous misuse. From the conversations that the IBE had in preparation for this briefing, it emerged that this is becoming increasingly common.

The danger of AI comes primarily from the way it is used. The use of drones harnessed with AI systems is a good example of this. If on the one hand they can significantly enhance some rescue and health operations – drones used to spot sharks along the Australian coasts or drones used to deliver blood and other vital supplies to remote areas of Rwanda, often unreachable during the rainy season are just two examples,[16] on the other hand the same technology can be used in conflict areas posing a completely different set of ethical issues.[17]

Control

Much of the public scepticism around the future of AI is fuelled by the fear that humans might lose control over the machines, which would then prevail and possibly wipe out humanity altogether. Recently, it was reported in headlines that Facebook abandoned an experiment after two AI programs appeared to be chatting to each other in a strange language that made it easier for them to work but only they could understand.[18] The extensive coverage of this story, which appears to be largely exaggerated by the media and only partially grounded on facts, is just an example of public concern on this and suggest that it is an issue that needs to be addressed.[19]

To have full control over AI systems, it is important that both companies and algorithm designers only work with technology that they fully understand. Being able to explain the functionalities of a technology of which they appear to be in control of is essential to build trust with employees, customers and all stakeholders. In addition, it minimises the risk that it is misused or that other parties might take advantage of it for personal gains.

[15] Ford (2015) Rise of the Robots: Technology and the Threat of a Jobless Future

[16] Fox News (28/08/2017) Drones are using artificial intelligence to protect Australian beachgoers from sharks and The Verge (13/10/17) Drones begin delivering blood in Rwanda

[17] The Guardian (18/02/2017) Has a rampaging AI algorithm really killed thousands in Pakistan?

[18] The independent (31/07/2017) Facebook's Artificial Intelligence shut down after they start talking to each other in their own language

[19] Gizmondo (31/07/17) No, Facebook Did Not Panic and Shut Down an AI Program That Was Getting Dangerously Smart

Companies also need robust control of the system's development process to ensure there is sufficient scrutiny and testing of algorithms for bias, or misuse (see Box 3).

Impact

Artificial Intelligence has become a buzzword in today's business world. In an environment where new machine learning technologies are created and developed at a fast pace, companies might be tempted to adopt them because they want to be ahead of the game and on top of the latest technological advancement, rather than because they really need them or because they benefit their business. Just because a company can use a certain AI technology, it doesn't necessarily mean that it should. As the CBI suggested, measuring the impact of AI is important to help companies to avoid unnecessary costs and potential risks deriving from the use of inadequate or inappropriate technologies.[20]

Further, measuring the potential impact that a new technology can have before adopting it, can identify undesired side-effects and consequent ethical risks. Therefore, it is important to test the algorithms and AI implementations in difficult situations to gauge a clear idea of unwanted outcomes.

Through such tests, companies need to identify what are the ethical risks involved, who is going to be impacted – positively and negatively, who is going to bear the costs, and whether there is a valuable and less risky alternative.

Box 3 provides some examples of issues to consider when assessing the impact and risks of AI.

Box 3 *AI control and testing procedures.*

Given the rapid adoption of AI in business, there is the risk that the governance systems required to mitigate the potential risks of its deployment are overlooked. Data science teams should have structured controls and testing around their development process, which are overseen centrally by the business. Because AI tools are often self-learning, control and testing procedures should be dynamic and constant.

This includes:

- Bias detection and correction;
- Risk assessment and impact analysis of each AI tool, and approval by management
- Involvement of ethics team to ensure that the AI systems in place are in line with the organisation's values
- Involvement of legal and compliance teams to ensure compliance with data protection and privacy regulations
- Robust cybersecurity and controls, including access control.

Accountability

Accountability is central to the definition of good practice in corporate governance. It implies that there should always be a line of responsibility for business actions to establish who has to answer for the consequences. AI systems introduce an additional strand of complexity: who is responsible for the outcome of the decision-making process of an artificial agent? This is compounded by AI development being largely outsourced by companies rather than developed in-house.

Machines, as such, are not moral agents and therefore they cannot be held responsible for their actions.[21] Who should be accountable, then, when an AI system violates ethical values? Should it be the designer of the algorithm, the company that adopts it or the final user? It is difficult to provide a univocal answer and a rich debate has flourished on this topic. Although the question of responsibility remains largely unanswered, a valuable approach would be for each of the parties involved to behave as if they were ultimately responsible.

In practical terms, it is advisable to include in contracts a clause to define each party's responsibilities and their limitations. Although it is not always practicable or comprehensive and it can't substitute for individual empowerment, this can help to prevent a situation where all parties have shared responsibility and therefore it becomes difficult to attribute accountability appropriately.

Learning

To maximise the potential of AI, people need to learn how it works and what are the most efficient and effective ways to use it. Employees and other stakeholders need to be empowered to take personal responsibility for the consequences of their use of AI and they need to be provided with the skills to do so. Not only the technical skills

[20] The Guardian (20/10/2017) Artificial intelligence commission needed to predict impact, says CBI
[21] D. G. Johnson (2006) Computer systems: Moral entities but not moral agents

to build it or use it, but also an understanding of the potential ethical implications that it can have. It is important that companies improve their communications around AI, so that people feel that they are part of its development and not its passive recipients, or even victims.

Ensuring business leaders are informed about these technologies and how they work is essential to prevent that they are unintentionally misused. However, it is important that businesses engage with external stakeholders as well, including media reporters and the general public, to improve their understanding of the technologies in use and ensure that they can assess more accurately the impact of AI on their lives.

The role of business

Business decision-makers, employees, customers and the public need to be able to understand and talk about the implications of business and AI to be at the forefront in the use of it. It is essential that companies know the impact and side effects that new technologies might have on their business and stakeholders.

The topic of AI and its applications and ethical implications for business is broad and requires a complex multi-stakeholder approach to be tackled. However, there are some measures that organisations can adopt to minimise the risk of ethical lapses due to an improper use of AI technologies.

- *Design new and more detailed decision-making tools for meta-decisions*, to help those that design algorithms and construct AI systems to ensure that they 'do the right thing' and act in line with the company's ethical values. This can come in the form of dedicated company policies that ensure proper testing and appropriate sign-off from relevant stakeholders, both internally and externally.

- *Engage with third parties for the design of AI algorithms only if they commit to similar ethical standards*: the design of these systems might be outsourced and it is important to conduct ethical due diligence on business partners. A similar principle applies to clients and customers to whom AI technologies are sold. Testing a third party algorithm in a specific situation is also important to ensure accuracy.

- *Establish a multi-disciplinary Ethics Research Unit to examine the implications of AI research and potential applications*; and be proactive in publishing the working papers from such a Unit to internal and external stakeholders.

- *Introduce 'ethics tests' for AI machines, where they are presented with an ethical dilemma*. It is important to measure how they respond in such situations in order to predict likely outcomes in a real-life dilemma, and therefore assume responsibility for what the machines will do.[22]

- *Empower people through specific training courses and communication campaigns in order to enable them to use AI systems efficiently, effectively and ethically*. These training courses should be directed not only at the technical personnel building the tool, but also at senior business stakeholders who should understand the assumptions, limitations and inner workings of AI technology.

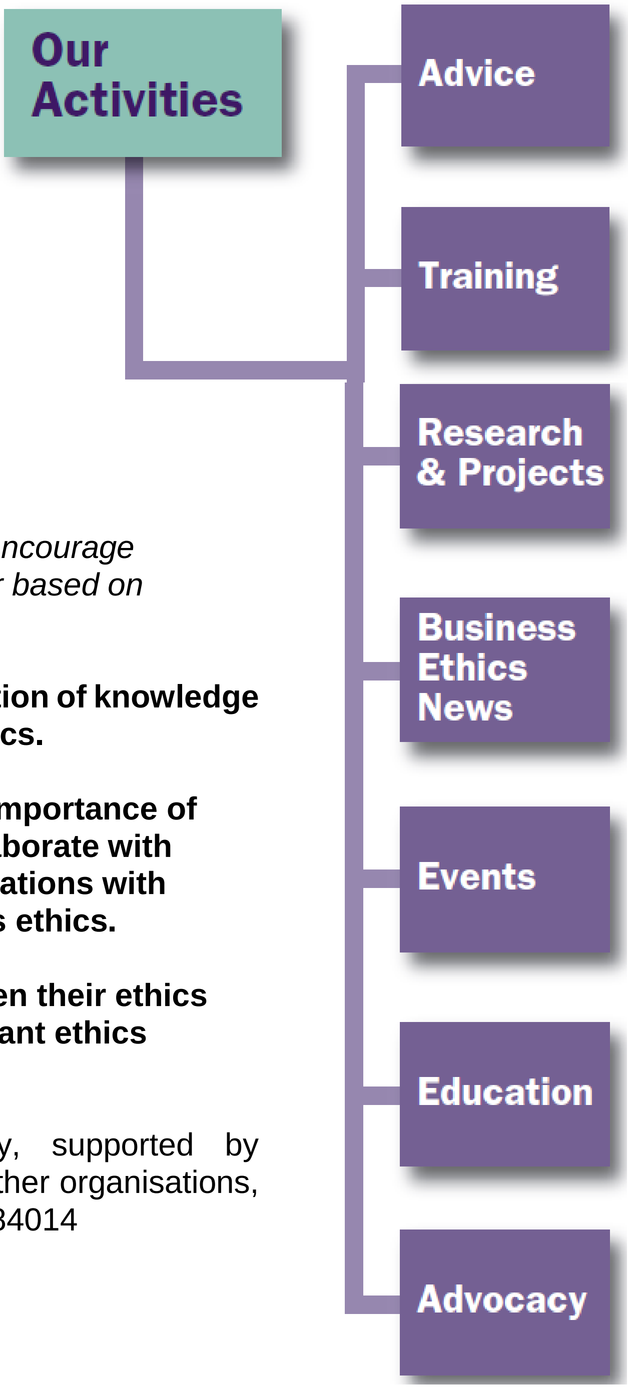
Box 4 *Questions to ask yourself*

Many organisations include in their code of ethics (or similar document) guidance to support individual decision-making through a decision-making model or guide. This often takes the form of 'questions to ask yourself'. This could be applied in a similar manner before adopting or using AI:

What is the purpose of our job and what AI do we need to achieve it?
Do we understand how these systems work? Are we in control of this technology?
Who benefits and who carries the risks related to the adoption of the new technology?
Who bears the costs for it? Would it be considered fair if it became widely known?
What are the ethical dimensions and what values are at stake?
What might be the unexpected consequences?
Do we have other options that are less risky?
What is the governance process for introducing AI?
Who is responsible for AI?
How is the impact of AI to be monitored?
Have the risks of its usage been considered?

[22] Bias detection methodologies have widely been studied. As an example, see The Guardian (19/12/16) Discrimination by algorithm.

ibe
Institute of Business Ethics


Our Activities

Advice

Training

Research & Projects

Business Ethics News

Events

Education

Advocacy

The IBE was established in 1986 to encourage high standards of business behaviour based on ethical values.

Our vision is to lead the dissemination of knowledge and good practice in business ethics.

We raise public awareness of the importance of doing business ethically, and collaborate with other UK and international organisations with interests and expertise in business ethics.

We help organisations to strengthen their ethics culture through effective and relevant ethics programmes.

The IBE is a registered charity, supported by subscriptions from businesses and other organisations, as well as individuals. Charity no. 1084014

The IBE would like to thank all those who contributed to this Briefing. A special thanks goes to Prof David Grayson CBE, director of Corporate Social Responsibility at the Cranfield School of Management, and Michelle Lee, Manager: Risk Analytics - Data Science at Deloitte LLP, for reviewing the briefing and providing useful comments and suggestions. We are grateful to all who contributed, but the IBE remains solely responsible for its content.