ISSN 2732-3781 DOI: 10.52398/gjsd.2022.v2.i2.pp27-35

GiLE Journal of Skills Development

Key Considerations of Ethical Artificial Intelligence That Organisations Need to Consider for Success

Shivaan Munnisunker
Hungarian University of Agriculture and Life Sciences

Abstract

It is argued that while Artificial Intelligence is far from having a consciousness like humans do, its consequences on society are minimal. Thus there is no rush to consider ethical issues. However, Artificial Intelligence applications are being implemented in almost every industry, imposing social unrest and upheavals for businesses. This paper aims to advocate for the importance and urgency of Artificial Intelligence ethics. This paper explores the different areas of ethics and then explains the concept of Artificial Intelligence ethics. A literature review is provided addressing four areas of Artificial Intelligence ethics that leaders must address if they are to win successfully in the industry in which they operate. These areas are biases, data security, explainability, and impact. A case study focusing on the company Strategeion is examined to illustrate the complexities of an Artificial Intelligence system in which a potential candidate for a job was discriminated against because of an error in its learning system.

***Keywords*:** Artificial Intelligence Ethics, Biases, Data Security & Algorithmic Accountability

1. Introduction

Scholars and business leaders agree that Artificial Intelligence (AI) is still in its infancy even though it has furthered its progress with technologies such as self-driving cars, medical diagnoses, and facial recognition. With the arrival of AI products, the world has progressed into a new era where machines are used to make decisions. They not only fulfil consumers' orders, but they use algorithms to make decisions. It has changed the way people react to AI products' decisions and what they expect from these products. In many cases, people unknowingly use AI to generate information based on their preferences and interests. These can come in the form of movie recommendations on Netflix, translations in Google Translate, or sales predictions in Customer Relations Management (CRM) systems (Ouchchy et al., 2020). AI-generated content can be beneficial; however, these recommendations and predictions are sometimes inaccurate. AI algorithms have flaws, especially when they do not have enough data or feedback to learn from (Ransbotham, 2018).

AI-generated content and decisions need to be checked by decision-makers. Discrepancies in terms of decisions will be illustrated in the presented case study. These kinds of inaccuracies might just result in an unpleasant user experience but can become disastrous when AI is used for critical and strategic decisions. Therefore, there is a need for the constant review of AI systems in terms of ethical knowledge, learning, monitoring and engineering. This paper considers the concept of AI ethics by referring to four schools of thought about ethics. These need to be addressed by business leaders to gain the trust of stakeholders when employing an AI system. The case study presented will consider how the trust in a reputable company, Strategeion, was damaged by an AI system they implemented. The unintended errors caused by this system show the importance of constantly monitoring AI systems to ensure that decisions made by these systems are accurate.

2. Literature Review

2.1. Ethics

The terms ‘ethics’ and ‘morality’ are often used interchangeably. Bartneck et al. (2021) define morality as a complex set of rules, values and norms that determine or are supposed to determine people’s actions. In contrast, Dent (2012) defines ethics as the theory of morality. Dent (2012) further explains that it could also be said that ethics is concerned more with principles, general judgements and norms than with subjective or personal judgements and values. There are various schools of thought about ethics, and these are summarised in Table 1.

TABLE 1. CATEGORIES OF ETHICS

School	Interpretation
Descriptive Ethics	This category of ethics is the easiest to understand - it simply describes how people behave and/or what moral standards they claim to follow. Descriptive ethics incorporates research from anthropology, psychology, sociology and history to understand what people do or have believed about moral norms, i.e., different societies have different moral standards (Bartneck et al., 2021).
Normative Ethics	This category of ethics involves creating or evaluating moral standards. Thus, it attempts to determine what people should do or whether their current moral behaviour is reasonable. Traditionally, the field of moral philosophy involved normative ethics - several philosophers tried their hand at explaining what they think people should do and why i.e., “*this action is wrong in this society, but it is right in another*” (Timmons, 2020, p. 5).
Deontological Ethics	Österberg (2019, p. 2) notes that “*deontological ethics is characterised by the fact that it evaluates the ethical correctness of actions on the basis of characteristics that affect the action itself.*” The term deontology or deontological ethics derives from the Greek word ‘deon’, which essentially means duty or obligation. Deontology can thus be translated as duty ethics.
Machine Ethics	According to Guarini (2013), machine ethics attempts to answer the question: what components would it take to build an ethical AI system that could make moral decisions? The main difference between humans making moral decisions and machines is that machines do not have ‘phenomenology’ or ‘feelings’ the way humans do. Machines, however, can process the data that represents feelings. Currently, no AI system or computer can feel and be conscious like a person (Dehaene et al., 2017). Life-like robots have been developed, but these robots do not possess phenomenal consciousness or actual feelings (Bartneck et al., 2021).

Source: Author’s own compilation.

2.2 Areas of AI Ethics for Businesses.

Increasingly as the adoption of AI by businesses continues to grow, four main ethical questions need to be addressed. The four areas of ethical focus are bias, security, explainability and impact. This paper examines these areas because they are all relevant for improving a business's transparency and trust. A business capable of addressing these areas is perceived as credible in the market and among its stakeholders (Appen, 2021).

2.2.1 Bias

Bias is defined as a tendency (known or unknown) to have a preference for one thing over another, which lacks objectivity and influences an outcome (Sun et al., 2020). An example of this in the business world could be deciding to purchase raw material from a supplier simply because the supplier is a relative of the decision maker, rather than using another supplier who could also offer the same quality raw materials; an objective fact that the decision-maker simply chooses to ignore (Bird et al., 2020).

Defining, detecting, measuring, and mitigating bias in AI systems is not an easy task and is an active area of research. Several efforts are being undertaken across governments, non-profit organisations, and industries to enforce regulations to address bias-related issues (PriceWaterhouseCoopers, 2022). AI biases should not discriminate against people based on sensitive data including, but not limited to:

> personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs; trade union membership; genetic data, biometric data, data processed solely to identify a human being; health-related data; data concerning a person's sex life or sexual orientation.
>
> (Aysolmaz et al., 2020, Online).

Concerns of bias can become evident when the AI system accepts or rejects a person for a loan or a job, or it affects suggestions of the type of markets to participate in based on the person's personal data (Aysolmaz et al., 2020).

Table 2 highlights four ways of addressing AI biases.

TABLE 2: RESEARCH AREAS ON MANAGING AI BIASES

Research Area	**Solution to AI Bias**
Algorithmic Awareness	-Training and awareness programs to educate users about the uses of AI systems. -Raise awareness of the existence and causes of biases in AI systems.
Algorithmic Accountability	-Make decision-makers accountable when using AI systems. -Put into place policies that compensate individuals who have fallen victim to erroneous decision-making. For example, online retailers can make wrong decisions about the customer. Therefore, they would offer discounts on the next purchase to correct this. -Investigate black boxes through algorithmic accountability reporting.
Algorithmic Transparency	-Make algorithmic reports public. -Make AI systems more user-friendly. -Reduce algorithmic biases by detecting errors in the input data, which result in the wrong information.
Algorithmic Audit	-Establish audit methods for third parties to determine what algorithms are doing, as well as include the details of the suppliers of the algorithms.

Source: Aysolmaz et al., 2020.

2.2.2 Data Security

A common pitfall businesses face with AI is the lack of a data strategy or governance plan. Data used correctly can help businesses predict market trends and gain deeper insights into consumer spending habits. Typically, this kind of data is confidential and private. Safeguarding the privacy and confidentiality of large volumes of datasets is essential for decision-makers. This is especially important when the data is built into the AI system. In this scenario, attackers may launch inconspicuous data extraction attacks that risk the entire AI system. Another type of attack can come from smaller sub-symbolic function extraction applications or viruses which require less effort and resources. The consequence of data attacks is the loss of sensitive information, which can lead to significant reputational damage and financial losses and can be detrimental to the long-term stability of an organisation (IBE, 2018). Common types of leaked information range from employee/customer data, and intellectual property, to medical records. Over the years, data leaks have increased due to cyber-attacks and insiders leaking information.

Table 3 below lists some of the most significant data leaks businesses have experienced and displays the costs.

TABLE 3: MASSIVE ENTERPRISE DATA LEAK INCIDENTS

Organisation	**Personal Records ($)**	**Breach date**	**Type**	**Source**
Aadhaar	1.1 billion	March 2018	Identity theft	Malicious outsider
LinkedIn	700 million	June 2021	Identity theft	Malicious outsider
Facebook	533 million	April 2019	Identity theft	Malicious outsider
Twitch	7 million	October 2021	Identity theft	Malicious outsider
Nintendo	300,000	April 2020	Financial	Malicious outsider

Source: Tunggal, 2022

2.2.3 Explainability

Adopting an AI system would be successful if it can be explained, understood and trusted by customers and end users (Pásztor, 2018). Developing AI systems based on customer information is common, and customers will therefore want to be sure that their personal information is collected responsibly, handled, and stored securely. Some stakeholders will even want to understand the basics of how their data is being used. AI has evolved to a stage where humans increasingly interact with AI systems. In the workspace, employees will have to develop skills to work with, and make decisions using, AI systems. The following needs to be considered: the communication between AI systems and employees should be simple and easily understood. AI systems should present new discoveries from vast datasets in a presentable manner that business decision-makers understand. This type of information should help businesses to increase profits and attain a competitive edge in the market. It often takes years to collect and process data before it can be analysed. As such, it is essential that the analysis is carefully planned and executed and that any general feedback about the performance of the AI system and its learning process is not lost between studies. AI models should be able to provide an interpretation of datasets. An AI system that works with an employee instead of replacing the employee is preferred (Pásztor, 2018).

2.2.4 Impact

Before any business can decide to implement an AI system. They should investigate the following ethical questions around impact:

- What is my model intended to do?
- What impact will my model's creation have on my business, the people who build my model, my end users, and society?
- What happens when my model makes the wrong decision?

These questions will drive a business to develop an AI model with a net positive impact on all relevant stakeholders. However, avoiding these questions or responding to them inaccurately could result in unintended consequences. An AI system that performs poorly may make discriminatory decisions—for example, AI-powered recruiting tools that show bias against women or facial recognition software that has trouble recognising darker-skinned faces (Appen, 2021).

3. Case Study: Strategeion

3.1. Background

The non-profit company, Strategeion, was founded by a small group of army veterans after having been honourably discharged during the 2008 recession. Strategeion's business model was to help veterans by providing them with job opportunities and support. The veterans were a group that was hit hard by the recession. Since tech companies were appointing recent young graduates from prestigious universities, it was difficult for the veterans with an Information Technology (IT) background to apply for positions, and therefore they found it difficult to fit into civilian life. The IT veterans built on their background of programming experience supporting various military operations with IT expertise. They created the company to enhance the lives of other veterans by creating an online platform that would enable veterans to stay in touch with each other and share their experiences about civilian life. The co-founders had witnessed how problems such as poverty, joblessness and homelessness affected many American communities during the economic downturn. The founders were searching for a technical solution for these problems and vowed to develop services, platforms and technical solutions for the benefit of all. As the company matured, the platform expanded to include a range of services - from social networking to personal blogging and even a location-based search application that helped individuals to move to new communities and to discover local points of interest. The vision for their company was to 'leave no man behind'. Strategeion believed that the best way to fulfil its pledge was to make its source code freely available to the public. Their logic was that the code would help others to serve their own communities with probably different requirements. They felt that making the code open source would provide a means of transparency and public accountability (Princeton University, 2018).

3.2. The Problem

Strategeion enjoyed low turnover rates from veterans and a satisfied workforce as they adjusted well to civilian life. In recognition, Strategeion was listed in Wealth Magazine as one of 'The Pop 100 Companies to Work For In 2013'. This resulted in a surge in job inquiries and applications from the public. The surge of applications outpaced the number of positions, and the Human Resource (HR) department was increasingly overwhelmed by these applications. In their internal communication board, HR complained about the volume of applications and that it was affecting their workflow as they could not complete other work tasks. The leaders interpreted this as a call for help. After exploring several options, the team decided to implement an AI system that employed the use of natural language processing and machine learning to deal with the influx of applications. The leaders expected that the AI system would ease their problem by implementing clever technical tricks to automatically pre-sort resumes according to a candidate's desirability, optimising especially for a projected 'fit' within the company. The AI system was referred to as PARiS. To train the system, HR worked with engineers by giving several resumes from current and previous employees whom they deemed were exemplary or poor fit according to the work and culture of these employees. PARiS would rate incoming resumes according to their match with the ideal types and cast aside those below a set threshold. Over the coming weeks, HR was relieved of the volume of applications as PARiS automatically selected the best candidates. It was consistent and efficient and seemed to represent the company's values by looking for the best fit.

Hara, a promising candidate with a disability who was devoted to computer science, applied for a position at Strategeion; however, she received an automated rejection from the PARiS system within hours of her application. She was surprised since she felt she was the perfect candidate for the role. She felt her application displayed a strong academic background, civic duties, and work experience with non-profit organisations. She was unsatisfied with the outcome and requested further feedback from the company. The request was received by HR, which upon review, noticed that she was indeed a perfect match for the role. Hara's case was used to review PARiS and to investigate the reason for the system's rejection of Hara's application. One concern was that PARiS may have used Hara's disability status as a reason to reject her application. However, the system's engineers assured HR that they had explicitly designed the algorithm so that it would not discriminate against such categories. Upon further investigation, they found that the system required candidates to participate in some or other sport and that it was probably why PARiS rejected Hara's application. Engineers found that the PARiS system highly evaluated athletics and military service participation. Given the overrepresentation of veterans among Strategeion's employees and their tendency to excel at the company, PARiS had learnt to connect a history of playing sports with a 'good fit'. While it was true that many of Strategeion's ex-military employees no longer participated in sports, their resumes typically reflected a history of having done so. Hara had no history of sports on her resume (Princeton University, 2018).

3.3. Outcome

In keeping with the values of honesty, an HR representative reached out to Hara with the findings. He explained about the implementation of the AI system and how it had learnt patterns to select the most promising candidates in the recruitment process. He admitted there were still some bugs in the system. He apologised on behalf of the company and invited Hara for an interview. He assured her that they would work on the shortcomings of PARiS. This did not pacify Hara, who was appalled that Strategeion had delegated recruitment, an area that could have a profound impact on her life prospects, to an AI system. Furthermore, she felt that system had discriminated against her. She blogged about the company's response on her personal website, where her followers joined in the discussion about the AI ethical concerns surrounding PARiS.

3.4. Discussion of the Case

A job influences a person's income, housing choice, family size, health aid options and other essential life (Hasan et al., 2021). Decisions about who is awarded a job may not have life-threatening consequences but can significantly impact the individual applying for the job. Hara believed that her life prospects would be significantly improved by joining Strategeion and was therefore upset by the idea of an AI system deciding her fate. She argued in her blog that human intervention is necessary for such a decision. When human agents reject worthy applicants, they may feel regret. An AI system, on the other hand, feels none of this. Instead, the system applies cold calculations to data to determine access to a scarce resource (e.g., jobs).

Hara was shocked to learn that an AI system reviewed her personal information without her consent. As current and previous employees of Strategeion learnt how PARiS functioned, they were also upset that their personal information might have been used to train the system. Strategeion's use of its employees' personal information for unexpected and undisclosed purposes left them open to allegations that they had violated privacy norms and standards.

Hara had rejected the call for an interview and instead filed an official complaint with the company. The board received and handed this over to their legal team. They had to ascertain whether they committed any legal wrongdoing by using employees' personal information without their consent and whether PARiS contravened the United States anti-discrimination law. If PARiS was coded to discriminate against candidates with disability status, then Strategeion would have most certainly violated the law. However, the investigation showed this was not intentional and was instead the result of redundant encoding, which allowed it to infer such erroneous results from the data. Therefore, the lawyers believed Strategeion had not violated any laws.

4. Recommendations and Future Research

Strategeion had, throughout its history, promoted the notion of fairness through its positive approach to recruiting employees from a group that other companies did not easily select. Despite the leader's approach to hiring these veterans, they had to acknowledge that PARiS eventually failed to live up to the company's values. They would need to find a way to evaluate all applications fairly. A possible solution would be to integrate PARiS with human intervention. An individual trained to spot biases and screen the applications with PARiS would eliminate or prevent a similar future occurrence. Engineers would need to monitor how PARiS learns by sampling the data and processing its uses regularly.

Not much literature is devoted to AI ethics (Eitel-Porter, 2021). This is an area that is still in its infancy. Future work should be devoted to understanding morality and ethics and how they can be implemented in AI systems. AI systems simply use algorithms and calculations based on the data that it is given to present results. As was shown in this case, humans' trust in AI systems' ability to make decisions is more complex. Humans still prefer humans to make decisions, especially decisions that may significantly impact their future.

This paper analyses AI ethics using a qualitative methodology by reviewing various literature and business articles. Some of this literature can be found in reputable journals such as Natural Machine Intelligence, Frontiers in Robotics and Artificial Intelligence and Foundations and Trends in Machine Learning to name a few. For instance, the article reviewed 70 international ethical guidelines and presented four key areas common in existing guidelines. These are values, big data and algorithms, inadequate understanding of AI ethics and values related to transparency and data security (Franzke, 2022). From a corporate perspective, IBM (2022) presents three key insights in their report. These are: business leaders are champions of AI ethics, which grew from 15% in 2018 to a staggering 80% presently. Secondly, more than half the organisations in the report have taken steps to embed AI ethics in their systems. Finally, diversity and inclusion issues are still not well represented, which is essential for mitigating biases in AI. Therefore, corporations have a limited approach to the ethics of AI. A possibility for future research could include a quantitative study on the opinions of decision-makers that rely on AI systems. A study can also be performed to test the validity and reliability of AI systems.

5. Conclusions

The paper advocates for the need to increase research and understanding of the ethics of AI. Evidence had been provided by various examples such as data leaks, supply decisions and through the case study. The case showed that more knowledge is required about AI to support the decision to implement an AI system, along with the possible pros and cons of using such a system. The potential threats and opportunities must be well thought through before implementing a system. A clear strategy that is drawn from the company's vision is needed before the use of AI should be considered. Constant learning and monitoring of AI are required to overcome the effects of biases and unforeseen circumstances.

The business world is increasingly moving towards adopting AI, and while AI will not immediately replace all jobs, people will increasingly have to interact with AI. In the workplace, learning intervention will be required to upgrade a worker's skills so that AI becomes accessible and user-friendly. The consequences of not doing so can result in loss of employee and customer trust and faith. A clear articulation of the business strategy must be embedded within the AI system and considered within the values and culture the business operates with.

References

Appen. (2021). *Ethical AI Techniques to Minimize Bias Throughout the Model Build Process*. Available from https://appen.com/blog/ai-ethics-the-guide-to-building-responsible-ai/

Aysolmaz, B., Dau, N., & Iren, D. (2020). *Preventing algorithmic bias in the development of algorithmic decision-making systems: A Delphi study*. 53rd Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2020.648

Bartneck, C., Lütge, C., Wagner, A., & Welsh, S. (2021). *An Introduction to Ethics in Robotics and AI*. Cham: Springer.

Bird, E., Fox-Skelly, J., Jenner, N., Larbey, R., Weitkamp, E., & Winfield, A. (2020). *The Ethics of Artificial Intelligence: Issues and Initiatives*. Brussels: European Parliamentary Research Service.

Dehaene, S., Lau, H.,, & Kouider, S. (2017). What is consciousness, and could machines have it? *Science*, *358*(1), 486-492. https://doi.org/10.1126/science.aan8871

Dent, J. C. (2012). Morality, ethics, norms and research misconduct. *Journal of Conservative Dentistry*, *15*(1), 92-93. https://doi.org/10.4103/0972-0707.92617

Eitel-Porter R. (2020). Beyond the promise: Implementing ethical AI. *AI and Ethics*, *1*(1), 73-80. https://doi.org/10.1007/s43681-020-00011-6

Ewing, A. (2013). *The definition of good*. (1st Edition). London: Routledge.

Franzke, A. S. (2022). An exploratory qualitative analysis of AI ethics guidelines. *Journal of Information, Communication and Ethics in Society*. Advanced Online Publication. https://doi.org/10.1108/JICES-12-2020-0125

Guarini, M. (2013). Introduction: machine ethics and the ethics of building intelligent machines. *Springer Science*, *32*(1), 213-215. https://doi.org/10.1007/S11245-013-9183-X

Hasan T., Jawaad, M., & Butt I. (2021). The influence of person–job fit, work–life balance, and work conditions on organizational commitment: Investigating the mediation of job satisfaction in the private sector of the emerging market. *Sustainability*, *13*(12), 6622. https://doi.org/10.3390/su13126622

IBE. (2018). *Business Ethics and Artificial Intelligence*. London: IBE.

IBM. (2022). *AI Ethics in Action: An Enterprise Guide to Progressing Trustworthy AI*. Available from https://www.ibm.com/thought-leadership/institute-business-value/report/ai-ethics-in-action#

Österberg, J. (2019). *Deontological Ethics: Assessment*. Philosophical Studies Series.

Ouchchy, L., Coin, A., & Dubljević, V. (2020). AI in the headlines: The portrayal of the ethical issues of artificial intelligence in the media. *AI & Society*, *35*(1), 927-936. https://doi.org/10.1007/s00146-020-00965-5

Pásztor, D. (2018). *AI UX: 7 Principles of Designing Good AI Products*. Available from: https://uxstudioteam.com/ux-blog/ai-ux/

PriceWaterhouseCoopers. (2022). *Understanding algorithmic bias and how to build trust in AI* (p. 1). Los Angeles: PriceWaterhouseCoopers. Retrieved from https://www.pwc.com/us/en/tech-effect/ai-analytics/algorithmic-bias-and-trust-in-ai.html

Princeton University. (2018). *Hiring by Machine*. Available from https://aiethics.princeton.edu/wp-content/uploads/sites/587/2018/12/Princeton-AI-Ethics-Case-Study-5.pdf

Ransbotham, S. (2022). *AI's Prediction Problem*. Available from https://sloanreview.mit.edu/article/ais-prediction-problem/

Singer, M. G. (2004). The concept of evil. *Philosophy*, *79*(308), 185-214.

Sun, W., Nasraoui, O., & Shafto, P. (2020). Evolution and impact of bias in human and machine learning algorithm interaction. *PLoS One*, *15*(8), e0235502 https://doi.org/10.1371/journal.pone.0235502

Timmons, M. (2020). *Normative Ethics*. International Encyclopedia of Ethics. https://doi.org/10.1002/9781444367072.wbiee907

Tunggal, A. T. (2022). *The 66 Biggest Data Breaches*. Available from https://www.upguard.com/blog/biggest-data-breaches

UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO. Available from https://unesdoc.unesco.org/ark:/48223/pf0000381137

Declaration Statements

Conflict of Interest

The author reports no conflict of interest.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

Ethics Statement

No dataset is associated with this manuscript.

Open Access Agreement