---

**Proceedings Paper:**
Wang, Y. orcid.org/0000-0003-1575-0245, Xiong, M. and Olya, H. orcid.org/0000-0002-0360-0744 (2020) Toward an understanding of responsible artificial intelligence practices. In: Bui, T.X., (ed.) Proceedings of the 53rd Hawaii International Conference on System Sciences. Hawaii International Conference on System Sciences (HICSS 2020), 20202-01-07 - 2020-01-10, Maui, Hawaii, USA. Hawaii International Conference on System Sciences (HICSS) , pp. 4962-4971. ISBN 9780998133133

10.24251/hicss.2020.610

---

# Toward an Understanding of Responsible Artificial Intelligence Practices

Yichuan Wang
University of Sheffield
yichuan.wang@sheffield.ac.uk

Mengran Xiong
University of Sheffield
mxiong5@sheffield.ac.uk

Hossein G. T. Olya
University of Sheffield
h.olya@sheffield.ac.uk

## Abstract

*Artificial Intelligence (AI) is influencing all aspects of human and business activities nowadays. Although potential benefits emerged from AI technologies have been widely discussed in many current literature, there is an urgently need to understand how AI can be designed to operate responsibly and act in a manner meeting stakeholders' expectations and applicable regulations. We seek to fill the gap by exploring the practices of responsible AI and identifying the potential benefits when implementing responsible AI practices. In this study, 10 responsible AI cases were selected from different industries to better understand the use of responsible AI in practices. Four responsible AI practices are identified, including governance, ethically design solutions, risk control and training and education and five strategies for firms who are considering to adopt responsible AI practices are recommended.*

## 1. Introduction

Artificial Intelligence (AI), a set of algorithm-based machine, is programmed to self-learn from data and display predictions and intelligent behaviors through artificial neural networks, automated machine learning, robotic process automation, and text mining [1]. AI is capable of responding to real-world problems and arriving decisions in real-time or near real-time manner on behalf of human being [2], [3], [4]. For instance, Chatbots, an AI-enabled service robot, developed by Bookings.com provides real-time 24/7 customer service with the support of 43 languages to answer travel related queries to its customers. With such highly evolved language processing capabilities, Chatbots can interact with customers and provide them with personalized recommendations. It also enables Booking.com to deliver marketing automation thereby simplify routine works accordingly.

AI, as a major shift in the global economy, is influencing all aspects of human and business activities nowadays. It holds the promise to create efficiency and effectiveness by using data generated from an explosion of digital touchpoints [5], [6]. At the same time, it comes with its own concerns relating to privacy concerns, user distrust, data leakages, information transparency, and ethical concerns. Such ethical dilemma and concerns, if they are not well addressed when developing AI initiatives, would lead to the potential loss of credibility for products and brands and hamper the company reputation in the marketplaces. Ethical and societal concerns aroused from AI systems need to be addressed in priority to ensure effective, ethical, and responsible use of AI [7]. However, relatively little attention has been given to understand responsible approaches to the development, implementation, management, and governance of AI.

Indeed, corporate social responsibility (CSR) has become the main preoccupations of organizations in the global marketplaces [8], used in broad domains including areas of policies, programs and actions while interacting with stakeholders [9][10]. For instance, customer retention rate could be enhanced as consumers prefer to purchase from and engage in, socially responsible companies [8]. Likewise, company reputation could be built along with CSR activities [11]. From the CSR perspective, organizations need to embrace the goal of being socially responsible while bringing AI into the business mainstream. However, according to the Cognizant's report, only about 50% of surveyed 975 executives across industries in U.S. and Europe had policies and procedures in place to address ethical concerns while designing AI applications [12]. Although potential benefits emerged from AI technologies have been widely discussed in many current literatures, the sustainable outcomes from business to the society that AI presents is remained unexplored [6]. Specifically, there is an urgently need to understand how AI solutions can be designed to operate responsibly and act in a manner meeting stakeholder expectations and applicable regulations [7], [13], [14].

We seek to fill the gap by exploring the practices of responsible AI and identifying the potential

URI: https://hdl.handle.net/10125/64352
978-0-9981331-3-3

benefits when implementing responsible AI initiatives. Therefore, this study set out to answer the following research questions.

RQ1: What are the practices of responsible AI?

RQ2: What benefits and challenges have been brought by implementing responsible AI practices?

To answer the above research questions, we hope to provide business practitioners a more current comprehensive understanding of responsible AI and both theoretical and practical reference values for the use of AI in a more socially responsible way. In this paper, we begin by providing the historical context of technology use of CSR, and then move on to understanding ethical challenges in AI and the development of AI in responsible practices. We conducted a multiple case study of responsible AI, which leads to the identification of responsible AI practices and the recommendation of responsible AI strategies.

## 2. Literature Review

### 2.1. Technology Use in CSR

Corporate social responsibility (CSR) can be defined as commitments from organizations to the society in improving societal, environmental and economic well-being through different business practices [8], [15]. The relationship between the company's social responsibilities and its financial performance has been documented extensively in the literature [16], [17]. The study from Bernal-Conesa et al. [18] has indicated that the contribution of CSR-oriented strategies is significant to the overall performance of the organizations. From the empirical perspectives, this principle has been incorporated in marketing communications by many organizations in order to enhance stakeholder perceptions and retentions [19]. Thus, CSR is perceived to have increasing importance for increasing enterprises' competitiveness.

CSR domains within the marketing field are classified into seven categories, including employee relations, human rights, diversity, community issues, corporate governance, environmental issues and product issues [20], [21]. Consumers are evinced to have domain-based pro-company responses to CSR practices due to the influence of moral foundations theory (MFT) either individual-oriented or group-oriented [8]. Their reactions towards companies can be moderated through CSR domains in the case of CSR strengths, therefore, properly CSR activates in different aspects need to be organized and lapses of CSR are required to be solved by companies [8].

As digital has become a megatrend in the global economy, new technology gains great popularity among different industries, offering new possibilities and bringing benefits in many aspects of human lives [22]. For example, labor force may be replaced by the intelligent machines [23]. However, concept of the sustainability has changed as it is confronted with the digital transformation, also known as a technological leap [24], leading to the increase in the restraints, from the national laws and international rules, on companies' responsibilities towards society and environment (Bernal-Conesa et al., 2017). Thus, challenges could be posted to organizations for creating sustainability and responsibility in the long run. Inability to communicate the CSR programmers and integrate them into strategies may lead to the failure from achieving full potentials. Moreover, criticisms of CSR vary between companies and industries [20]. Data, algorithms and bots are main areas to be explored during the process of sustainable digitalization [22]. Specifically, although having access to consumer data helps predict their potential moves and create personalized experiences for them, privacy invasions and algorithmic bias derived from the sophisticated use of consumer data cannot be underestimated [25]. Hence, the performance of technologies is required to be aligned with CSR principle and enhance its implementations [26]. In practice, technology could identify the integration points of CSR initiatives, offering corporate strategy to increase the overall integrated level. In addition, it could reduce human bias through the multi-dimensional measurement on the programme performance. Therefore, it is arguable that technical resources can be integrated with human resources, within or across companies, helping develop capabilities to address sustainable concerns and delivering responsible values to stakeholders to obtain sustained benefits [27].

### 2.2. Ethical Challenges in AI

AI is no doubt beneficial to society as it helps to harness empathy and creativity skills of human and leveraging their emotional intelligence [28], [29]. An example is that Siri, assistant of iPhone, is able to recognize user's requests through voice message, and provide them assistance accordingly. It could lessen the uncertainty, reduce the time spent on administration and improve the efficiency in decision-making process based on the data evidence. In practice, the application of AI varies as it is programmed to use specific data to achieve a certain goal [30]. Marketers with such data can provide additional benefits to target consumers in a more efficient way [25].

In recent years, the pace of using consumer data in the marketing field exceeds the academic scholars' analytics [25]. Consequently, negatively unforeseen issues may come along with initial programs and against its positive goals. In addition, the lack of transparency on algorithms, in reality, has caught public attention, leading to the rise of ethical concerns on the use of AI [2]. Ethical issues are associated with the emergence of machine learning, as it allows intelligence system to get access and learn from numerous datasets, to derive its own rules, enhance its behaviors and produce cognitive competence [31]. The ways in which its performances caused ethical reflections, may result in deviating from sustained values and presenting new challenges [28], [29], [32]. For instance, interruptions of systems are of frequent occurrence due to the self-reflection. Programmers' biases might exist as the abilities of AI are initially dependent on human inputs, therefore, it might be problematic as bias can also be replicated from previous events according to the algorithm [2]. Thus, it is argued that intelligence systems are requiring moral reasoning capabilities while facing certain ethical dilemmas [29].

Studies on ethical AI, both from the data and the information system perspectives, have been conducted recently, leading the mitigation of unfair bias. Reinforcement learning (RL) is prospected to prevent ethical issues in the process of intelligent decision-making [32]. It can learn from interruptions while using data, either from humans or from environments, to avoid repetitive problems. In addition, formulating ethical principles to guide the design of AI system and rational algorithms are argued to be effective to ensure the ethics [33]. Nevertheless, it is not an easy task. Research from Robbins [29] states a lack of assistance from ethical norms or policy guidelines to regulate AI developer to achieve a balance between the effective use of AI and the concerns on ethics in the society. Taddeo and Floridi [33] point out that the formulation of ethical principles depends on cultural contexts and the domain of analysis which they could vary.

## 3. Research Method

Our cases were drawn from materials on current and past responsible AI projects from multiple sources such as practical journals, print publications, case collections, and companies', vendors', consultants' or analysts' reports. The absence of academic discussion in our case collection about the utilization of responsible AI is due to the incipient nature of such in this field.

The following case selection criteria were applied: (1) the case presents an actual implementation of responsible AI; (2) it clearly describes the practices of responsible AI. We were able to collect 10 responsible AI cases in different industry (See Appendix 1). Categorizing by region, 4 cases were collected from Northern America, 6 cases from Europe and UK.

Data analysis followed the constant comparison method. Initially data analysis was performed concomitantly with data collection, and continued with an explicit coding stage and an analytical coding procedure stage [33].

In the explicit coding stage, the analysis started by comparing and coding each statement extracted from the case materials into categories. This allowed categories to emerge to fit in an existing category [33]. Relevant statements were labelled and either created as a new code and given a definition, or assigned to the existing codes with memos indicating their relevance and potential properties. Through this process, the statement was broken down into units of meanings. The concept as a basic unit of analysis labels phenomenon representing a practice of responsible AI [35]. After the explicit coding stage, the data were conceptualized, defined and categorized in terms of their properties, which initiates the analytical coding stage.

During the analytical coding stage, the research team compared the properties and dimensions of the emergent categories. In order to constantly analyze and compare the categories, the concept map was employed to visualize the classification [35]. Four dimensions underlying responsible AI practices were identified. They are described in detail in the following sections and visualized in Figure 1.

## 4. Practices of Responsible AI

Responsible AI is a governance framework that uses to harness, deploy, evaluate, and monitor AI machines to create new opportunities for better service provision. It focuses on designing and implementing ethical, transparent, and accountable AI solutions that help maintain individual trust and minimize privacy invasion. Responsible AI places human (e.g., end-users) at the center and meets stakeholder expectations and applicable regulations and laws. Prior to designing and implementing responsible AI, organizations need to understand the practices that will help them drive ethics and trust of AI use. The four practices of responsible AI include: (1) Data governance; (2) Ethically design solutions; (3) Human-centric surveillance/risk control; and (4) Training and Education. These practices are evident

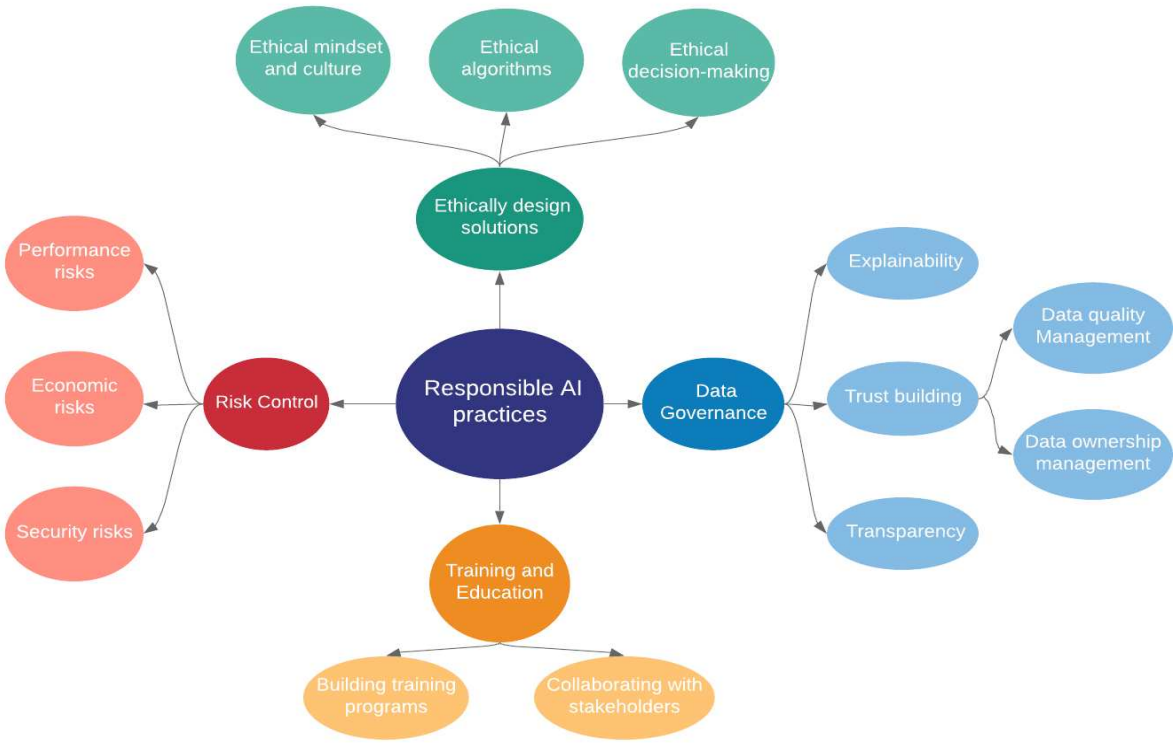


**Figure 1. A concept map of responsible AI practices**

in the real-world cases of responsible AI. These are described in turn below.

### 4.1. Data Governance

Governance of responsible AI focuses on building transparency, trust, and explainability.

**Transparency**. It is important that the organizational use of AI must be transparent to the stakeholders by allowing them fully understand how an AI application processes their data and arrive to specific decisions [36]. According to the Direct Marketing Association (DMA)'s investigation, 80% of surveyed consumers would be very or moderately comfortable with sharing personal data when they know about how digital data is shared and effectively used for marketing purposes [37]. Capital One is making the criteria system of credit card transparent by providing a computational decision with complete explanation to their customers when their credit card applications are accepted or denied [38]. Likewise, Alder Hey Children's Hospital, as one of the largest children's hospitals in Europe, has developed an AI featured digital App called Alder Play. Alder Play has incorporated the cognitive advances in order to present the enjoyable and informative experiences for its young patients. Young patients allow to active their own avatar during their stay, receive awards when completing treatments, and get access to further guidelines and contents accordingly [39]. Alder Play enables healthcare professionals to have access to medical records of patients who are eligible for NHS treatment. Patients and their families would be able to obtain their medical records online. This could largely improve transparency in the clinical processes, thereby enhancing the quality of health services and strengthening the patient engagement.

**Trust building**. Trusted AI is built through high-quality data and consent to use [12]. AI with high-quality data could mitigate biased and inaccurate results generated. To ensure the quality and reliability of data, where the data sources come from, the limitation of data, and data rules to sharpen data error detection should be identified when developing AI algorithms and systems. For example, PwC has employed H2O.ai to build a revolutionary bot named GL.ai, which uses AI algorithms to effective track operational data and transactions and correct errors to maintain accurate purchase histories and interactions for their business customers.

What makes AI workable is its access to personal information [36]. However, widespread access to personal information (e.g., consumer-generated content, online transactional data, and browsing and clicking data) has brought negative impacts to

individual, business, and society [25], [40]. The availability of consumer data gives rise to serious concerns where consumers suffer from privacy invasion, fraud, information leakage, and identity theft, and on the other hand, companies cannot collect consumer data effectively due to the consumers' distrust. These trends have led to a focus on data protection and transparency of data use by the regulators in many countries such as General Data Protection Regulation (GDPR) formulated by the European Union and Act on the Protection of Personal Information (APPI) in Japan. These regulations aim to protect all individuals' rights regarding privacy and personal data and give control to individuals over their personal data. With these regulations came into force, it is crucial for companies to institutionalize the practice of obtaining consent statement or permission from users and reduce ambiguity of data use and make the logic behind automation clear through effective communication with users [12].

**Explainability**. Providing meaningful and personalized explanations about the results generated by AI models could reduce uncertainty and build trust with users [12]. To develop explainable AI, Supplier's Declaration of Conformity (SDoC) proposed by IBM suggests that effective AI systems should be able to interpret algorithm outputs via examples properly and describe the testing methodology [41]. For example, PwC has released its Responsible AI Toolkit to guide companies to accountably harness the power of AI and provide them with personalized advisory services. Likewise, Alder Hey Children's NHS Foundation Trust in Liverpool, UK has driven the intelligent use of digital techniques based on big sets of patient data. Alder Hey's AI systems powered by IBM Watson cognitive analytics enable healthcare professions to interact with young patients and deliver them with personalized health services, thereby improving the quality and experience of care and securing the sound health services [39]. AI-enabled personalized health services have improved patient experiences in terms of familiarization, distraction and reward [42]. Specifically, before patients arrive, 360-degree tours of hospital environments and introductive videos of blood test and x-ray check are available for them to explore the hospital conditions and familiarize with potential treatment experiences. Parents could speak to a virtual assistant called Ask Oli to inquire about the progress of their children's health checks and treatments. Questions are assured to be answered in real time. Additionally, Alder Hey offers young patients with character-based stickers activated by using augmented reality (AR).

## 4.2. Ethically Design solutions

Ethical concerns should be minimized in designing AI solutions in three ways. First, design engineers need to be aware of possible ethical challenges such as artificial stupidity, racist robots, data and cyber security when developing AI systems. To prevent these ethical concerns, AI system allows for human inspection of the functionality of the algorithms and systems [7]. For example, Google has pointed out that concerns on ethical, environmental and societal challenges while applying AI technology need to be addressed across all sectors of society [43]. User-centered AI systems are designed based on Google's concept of general best practices for software systems. As acting a leading role in the development of AI, Google has invested in AI research and announced guidance principles to manage its research fields and product development, thereby influencing its business decisions in a more ethical way [43]. Assessment of responsible AI applications could be made via these objectives, leading to the obligation for Google to form a "responsible innovation team" with experts from a range of disciplines to initially examine its ethical level, and select a council of senior executives to make decisions for more complicated issues [44][45]. In addition, an external advisory group is organized with Google's AI solution developers from a variety of disciplines to avoid unethical AI practices and complement its internal governance [44].

Second, a responsible AI system should themselves be able to make socially significant decisions by a set of ethical algorithms in order to reduce the risk of unethical behaviors [14]. Lessons could be learnt from a ridesharing platform, for instance, the unethical AI algorithm potentially creates unfairness on the distribution of drivers' task assignments and pricing practices. This algorithm exists like a "black box" and helps its drivers evade local transport regulators.

Third, a prerequisite for implementing responsible AI successfully is to develop ethical mindset and culture for organizations and employees. This is critical for reducing any risks when applying AI. H&M Group, for instance, has developed a checklist, along with 30 questions to guide all ongoing and new AI projects to ensure that AI applications are used with fairness, transparency, beneficial results, governance, collaboration, reliability, respecting privacy, focused, and security. Such a practice help H&M to ensure every AI solutions they develop are subject to the comprehensive assessment of risks in its use.

### 4.3. Training and education

Building training programs is another crucial responsible AI practices. Such programs are to equip managers and employees with a deeper understanding of ethical use of AI and data. IEEE's Initiative for Ethical Considerations in Artificial Intelligence Systems[1] is a program designed to promote ethical and responsible AI and ensure AI architects and solutions developers are educated and trained to prioritize ethical considerations of AI [36]. This program suggests that organizations should provide training courses for ethical use of AI in areas such as methods to guide ethical design, and safety and beneficence of artificial general intelligence and artificial superintelligence to those employees who will play a critical support role of responsible AI. Mentoring, cross-functional team-based training and self-study are also beneficial training approaches to help employees develop the ethical AI mindset and culture.

Google has provided a series of advanced technical knowledge online for people to master technical skills. One suggested path is related to Machine Learning (ML) techniques, a subset of AI which could be applied to the datasets generated from the real world. To be specific, Machine Learning Crash Course (MLCC) is designed by Google engineers with the help from university computer science faculties, offering resources with insights of data science and innovative ML approaches for the supplement of study by self-learning. It has featured with lessons including video lectures, actual case studies and practical exercises. For example, a technical module on fairness in 11 language versions has been added to the MLCC by Google, in order to train its staff around the world and help them mitigate bias [45]. Additionally, material rewards from Kaggle Machine Learning Competitions could be given to those who learn new skills with ML challenges. Moreover, training of "Ethics in Technology Practice" project has been developed at the Markkula Center for Applied Ethics at Santa Clara University [45]. It offers assistance for Google users to identify multifaceted ethical issues during their daily work. Besides, Resource Library from Google is available to be accessed to create individual pathway.

Cloud AutoML has been introduced to design the own model by using Google's techniques such as "learning2learn" and "transfer learning" [46]. This could increase the productive level for less-skilled users. The Google Cloud AI Solution provides either prepackaged solutions or personalized model to serve organizations' needs across industries. Moreover, it has shared experiences to improve AI practices, partnered with professionals to apply projects with positive societal effects, and worked with stakeholders to promote thoughtful leadership in this area [43]. Therefore, it could guarantee a long-term development of AI technology as well as its implication.

In addition, PwC has published the articles and white papers to demonstrate their responsible AI experiences [47]. "AI: Sizing the prize" from PwC aims to estimate the percentage of the increase in GDP to be contributed to AI in various regions [48]. From a recent PwC analysis report on the financial services sector, concerns related to augmentation, automation has been addressed, and corresponded advice on the way to adapt AI in the future has been provided. PwC advises exploring AI solutions within explanatory and operational areas, which could help using budget and resources in a more ethical and societal way [48]. In addition, PwC has worked on leveraging AI to fulfil client demands and expectations, thereby sharing its own experiences to help customers to employ the power of AI in the same way [49]. As AI cannot learn without human intervention, consequently, it is vital to train both intelligence machines and staff to acquire appropriate data [50]. Efforts from staff across the whole PwC global network has accelerated the PwC's approach to the AI. It is proved that the advantages of aligning AI innovation with core strategic objectives outweigh operating initiatives in isolation [50].

Another example, reported by Audi AG, is that the "Beyond AI Initiative" is created to address social acceptance barriers of autonomous driving and the future of work by educating development engineers, scientists and other stakeholders.

### 4.4. Human-centric surveillance/risk control

Successful responsible AI requires a series of risk control mechanisms at the design, implementation, and evaluation stages. Several risks should be taken into consideration when developing responsible AI for organizations that includes security risks (cyber intrusion risks, privacy risks, and open source software risk), economic risks (e.g., job displacement risks), and performance risks (e.g., risk of errors and bias and risk of black box, and risk of explainability).

To minimize these AI risks, the first step is to formulate the rules of risk controls, with clearly focused goals, execution procedures, metrics, and

[1] Ethically Aligned Design: A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems, version 1, IEEE Standards Assoc., 2016; standards.ieee.org /develop/indconn/ec/ead_v1.pdf.

performance measures. In other words, a strong data protocol should be defined that provides clear guidelines to proactively identify AI risks that enable organizations to harness data effectively from the time it is acquired, stored, analyzed, and finally used.

Second, organizations should review the data they gather internally and externally and realize their potential risks. AI comes from self-learning through human designed algorithms. It is imperative to ensure the creditability of data so that AI can learn from the right patterns and act according to their input. Once the potential risk of these data has been managed, managers can make better decisions, thereby minimizing cost and complexity.

Finally, a responsible AI system should consider the economic risks such as job displacement, liability, and reputation risks. It is widely acknowledged that future trend of AI will utilize AI approaches to augment and complement human cognitive skills, and focus on human-AI machine interaction and collaboration to bring together the best of each [51].

## 5. Formulating Responsible AI Strategies

Lessons learnt from our selected case studies, we suggest the following five strategies might provide useful guideline for those seeking to develop responsible AI initiative in their organizations.

### 5.1. Emergence of Chief Responsible AI Officers (CRaiO)

Firms increasingly expect that the deployment of AI is aligned with their goals and values of CSR. AI not only enable firms to explore sharper customer insights, but also become a powerful strategic resource to facilitate positive business reputation and brand recognition if it is used in an ethical and responsible manner. However, only 25% of around 250 surveyed companies have considered the ethical implications of AI before investing in it according to the PwC's investigation [52]. This shows that the responsible AI practices in most cases are immature. CRaiO roles should emerge to in response to this need. We define the CRaiO as a role in charge of developing a responsible AI roadmap and policy in conjunction with internal and external stakeholders to make use of trusted AI, integrating the oeuvre of responsible AI to the projects across functional units, and cultivating an inclusive responsible AI culture across organizational and functional boundaries. Creating a CRaiO may require intensively cross-functional collaborations and organizational changes. A careful assessment on organizational resources and capabilities should be taken. Alternatively, as suggested by EY [53], AI ethics multi-disciplinary advisory board can be established to provide advice and guidance to the Board of Directors.

### 5.2. Balancing economic and social sustainability of AI use

AI for sustainability has attracted academic and practical attentions in recent years, particularly discussions on how can AI techniques be applied to find a balance between economic and social sustainable impact for businesses has been excited in diverse disciplines. When applying AI, its societal impact on well-being of humans and environment should be seriously considered. If firms develop AI algorisms with controversial impact on human rights, privacy, and employment, it may lead to the potential loss of credibility for products and brands, and hamper the company's reputation in the marketplaces. Thus, the ultimate goal of responsible AI is to strike a balance between satisfying customer needs with less ethical concerns and dilemmas, and attaining long-term profitability for businesses and services. Ecological modernization theory (EMT) argues the ecological outcomes could be maximized through achieving a balance between economic growth and social sustainability [54]. In this sense, firms should develop their AI solutions by taking the co-creation of economic and social sustainability into consideration. Specifically, firms need to establish policies on ethical governance considering socially preferable approaches, address ethical issues both in the initial design and post-launch stage of AI systems, and place AI ethics as part of the CSR strategy.

### 5.3. Transparent and customer-centric data policy

There is no strategy with AI without a good data quality management. However, with the data protection regulations such as GDPR came into force, firms require to obtain consent statement or permission from consumers if they want to use their information. These regulations have been a double-edged sword for firms, potentially acting as a barrier to behavioural targeting, personalisation of the communications and other promotions plans of marketers. On the other hand, with appropriate data policy, it will improve consumers' confidence in sharing the data with firms for AI use [56].

Furthermore, penalties for the GDPR non-compliance is about ranges from €10-20 million or 2-4% annual global turnover, which is a hefty fine and challenge for small and medium retailers [55]. Although the GDPR is an EU act, but it has a global acts as international marketers that plan to communicate with EU citizens must comply with the

regulations. Thus, persuading customers to share information through transparent and customer-centric data policy may turn these regulations from a threat to an opportunity and may improve their trust towards AI .

### 5.4. Creating socially responsible initiatives with AI

Responsible AI is not just about designing AI to operate ethically and responsibly, what do matter is how AI can be leveraged to advance socially responsible initiatives [57]. For instance, Quantcast, a leading AI company who specializes in AI-driven marketing, optimizes customers' advertising campaigns through using AI-driven real-time insights. Meanwhile, they rely on real-time data and machine learning capability to help their customers ensure brand safety and prevent consumers in the markets from fraud and fake information dissemination. H&M utilizes AI to ensure customer centricity (approaches such as fitting consumers' physical dimensions with their preferred style and incorporating multiple data sources for dynamic analysis), as a result of cutting environmental waste and cost caused by high purchase return rates. These socially responsible initiatives with AI contribute to increased trust and sustainability among consumers.

### 5.5. Carrot and stick mechanism to regulate AI usage

Carrot (reward/incentive) and stick (punishment) mechanism has been widely applied to regulate IT usage [58]. It is important to understand what mechanisms can trigger employees' ethical AI behavior or impede the misuse of AI. Floridi et al. [59] have designed a series of actionable plans to financially incentivize ethical use of AI at the organizational level. First, firms should encourage cross-disciplinary cooperation and debate on technological, social, legal aspects of AI. For example, H&M has created an Ethical AI Debate Club where cross-functional employees and their customers and AI researchers can meet for debates on ethical concerns and dilemmas arise in the fashion industry. Second, developing an inclusive triadic configuration to capture the complex interactions among ethics, innovation, and policy in confluence, it will help firms to ensure AI has ethics as a core consideration and policy is guided facilitating socially positive innovation [59]. Moreover, punishment plays a key role in affecting employees' ethical AI behavior. Firms should develop a monitoring, auditing and punishing mechanism to redress for a wrong caused by AI usage and to moderately punish unethical AI behaviors.

## 6. Conclusion

As being maturing rapidly, AI holds an incredibly power which has created new opportunities for social good. However, the scalability of machine learning might lead to inevitable disruptive impacts, consequently, concerns may be aroused while misusing AI. In practice, only few companies across industries have incorporated AI with a series of practices in a manner consistent with ethical considerations, organizational values, public expectations and societal norms. Attention is urgently needed for research to formulate responsible AI strategies that will enable firms to move forward to leverage AI most efficiently and ethically.

Although our study identifies responsible AI practices which is not only contributing to the disciplinary field of AI and ethics, but also provides practical recommendations for practitioners, it is subject to the limitation of data source but at the same time formulating new directions for future research if primary data can be collected. First, the adoption of responsible AI is still in its infancy. Case materials used in this study mainly came from companies' and consultants' reports. The absence of academic works may result in a potential bias, as companies usually publicize their success stories [60]. Further validation could be undertaken by collecting primary data from consumers, C-level executives, AI software companies, third party organizations and policy makers to fully explore responsible AI practices at the individual, organisational, industrial, and societal levels.

Second, as we found trust plays a vital role in implementing AI, understanding consumers' cognitive appraisals, emotional states, and behavior responses toward irresponsible use of AI enables practitioners to avoid negative consequences. The different scenario of irresponsible use of AI (e.g., ineffective marketing message, identity theft, and invasion of privacy) can be examined through the surveys and field experiments.

## 7. References


[1] Huang, M. H., & Rust, R. T. (2018). Artificial intelligence in service. Journal of Service Research, 21(2), 155-172.

[2] Amershi, B. (2019). Culture, the process of knowledge, perception of the world and emergence of AI. AI & Society, https://doi.org/10.1007/s00146-019-00885-z.

[3] Rai, A., Constantinides, P., & Sarker, S. (2019). Editor's Comments: Next-Generation Digital Platforms: Toward Human–AI Hybrids. Management Information Systems Quarterly, 43(1), iii-ix.

[4] Russell, S. J., & Norvig, P. (2016). Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited.
[5] Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS Quarterly, 36(4), 1165-1188.
[6] Martínez-López, F. J., & Casillas, J. (2013). Artificial intelligence-based systems applied in industrial marketing: An historical overview, current and future insights. Industrial Marketing Management, 42(4), 489-495.
[7] Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. The Cambridge handbook of artificial intelligence, 1, 316-334.
[8] Baskentli, S., Sen, S., Du, S., & Bhattacharya, C. B. (2019). Consumer reactions to corporate social responsibility: The role of CSR domains. Journal of Business Research, 95, 502-513.
[9] Peloza, J., & Shang, J. (2011). How can corporate social responsibility activities create value for stakeholders? A systematic review. Journal of the Academy of Marketing Science, 39(1), 117-135.
[10] Öberseder, M., Schlegelmilch, B. B., & Murphy, P. E. (2013). CSR practices and consumer perceptions. Journal of Business Research, 66(10), 1839-1851.
[11] Cheng, B., Ioannou, I., & Serafeim, G. (2014). Corporate social responsibility and access to finance. Strategic Management Journal, 35(1), 1-23.
[12] Ramaswamy, P., Jeude, J., & Smith, J.A. (2018). Making AI responsible and effective. https://www.cognizant.com/whitepapers/making-ai-responsible-and-effective-codex3974.pdf
[13] Torresen, J. (2018). A review of future and ethical perspectives of robotics and AI. Frontiers in Robotics and AI, 4, 75.
[14] Wallach, W., & Allen, C. (2009). Moral Machines: Teaching Robots Right from Wrong. New York: Oxford University Press.
[15] Sen, S., & Bhattacharya, C. B. (2001). Does doing good always lead to doing better? Consumer reactions to corporate social responsibility. Journal of Marketing Research, 38(2), 225-243.
[16] Lee, M. D. P. (2008). A review of the theories of corporate social responsibility: Its evolutionary path and the road ahead. International Journal of Management Reviews, 10(1), 53-73.
[17] Marín, L., Rubio, A., & de Maya, S. R. (2012). Competitiveness as a strategic outcome of corporate social responsibility. Corporate Social Responsibility and Environmental Management, 19(6), 364-376.
[18] Bernal-Conesa, J. A., de Nieves Nieto, C., & Briones-Peñalver, A. J. (2017). CSR strategy in technology companies: its influence on performance, competitiveness and sustainability. Corporate Social Responsibility and Environmental Management, 24(2), 96-107.
[19] Bhattacharya, C. B., Korschun, D., & Sen, S. (2009). Strengthening stakeholder–company relationships through mutually beneficial corporate social responsibility initiatives. Journal of Business Ethics, 85(2), 257-272.
[20] Servaes, H., & Tamayo, A. (2013). The impact of corporate social responsibility on firm value: The role of customer awareness. Management Science, 59(5), 1045-1061.
[21] Du, S., Yu, K., Bhattacharya, C. B., & Sen, S. (2017). The business case for sustainability reporting: Evidence from stock market reactions. Journal of Public Policy & Marketing, 36(2), 313-330.
[22] Osburg, T., & Lohrmann, C. (2017). Sustainability in a digital world. Springer International.
[23] Pavlou, P. A. (2018). Internet of Things–Will Humans be Replaced or Augmented?. GfK Marketing Intelligence Review, 10(2), 43-48.
[24] Osburg, T., & Schmidpeter, R. (2013). Social innovation. Solutions for a sustainable future. Springer.
[25] Martin, K. D., & Murphy, P. E. (2017). The role of data privacy in marketing. Journal of the Academy of Marketing Science, 45(2), 135-155.
[26] Bocquet, R., Le Bas, C., Mothe, C., & Poussing, N. (2013). Are firms with different CSR profiles equally innovative? Empirical analysis with survey data. European Management Journal, 31(6), 642-654.
[27] Dao, V., Langella, I., & Carbo, J. (2011). From green to sustainability: Information Technology and an integrated sustainability framework. The Journal of Strategic Information Systems, 20(1), 63-79.
[28] Wright, S. A., & Schultz, A. E. (2018). The rising tide of artificial intelligence and business automation: Developing an ethical framework. Business Horizons, 61(6), 823-832.
[29] Robbins, S. (2019). AI and the path to envelopment: knowledge as a first step towards the responsible regulation and use of AI-powered machines. AI & Society, https://doi.org/10.1007/s00146-019-00891-1.
[30] Parkes, D. C., & Wellman, M. P. (2015). Economic reasoning and artificial intelligence. Science, 349(6245), 267-272.
[31] Vallor, S. and Bekey, G. (2017). Artificial Intelligence and the Ethics of Self-Learning Robots, in Robit Ethics 2.0: From Autonomous Cars to Artificial Intelligence. Oxford Scholarship Online.
[32] Arnold, T., & Scheutz, M. (2018). The "big red button" is too late: an alternative model for the ethical evaluation of AI systems. Ethics and Information Technology, 20(1), 59-69.
[33] Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. Science, 361(6404), 751-752.
[34] Glaser, B. G., & Strauss, A. L. (2017). Discovery of grounded theory: Strategies for qualitative research. Routledge.
[35] Strauss, A. L., & Corbin, J. M. (1998). Basics of Qualitative Research: SAGE Publications.
[36] Bryson, J., & Winfield, A. (2017). Standardizing ethical design for artificial intelligence and autonomous systems. Computer, 50(5), 116-119.
[37] Direct Marketing Association (2018). GDPR: A consumer perspective. Available at: https://dma.org.uk/uploads/misc/5af5497c03984-gdpr-consumer-perspective-2018-v1_5af5497c038ea.pdf.

[38] Knight, W. (2017). The financial world wants to open AI's black boxes. MIT Technology Review. https://www.technologyreview.com/s/604122/
[39] Alderheycharity (2017) Download our brilliant new app now, Alder Hey Children's Charity. Available at: https://www.alderheycharity.org/news/latest-news/the-alder-play-app-has-launched/.
[40] Cohen, M. C. (2018). Big data and service operations. Production and Operations Management, 27(9), 1709-1723.
[41] Mojsilovic, A. (2018). Factsheets for AI Services. Available at: https://www.ibm.com/blogs/research/2018/08/factsheets-ai/
[42] Ustwo (2019). Alder Play: Revolutionising patient care for children and their families, Ustwo. Available at: https://www.ustwo.com/work/alder-play
[43] Pichai, S. (2018). AI at Google: our principles, Google. Available at: https://www.blog.google/technology/ai/ai-principles/
[44] Gershgorn, D. (2018) Google created a 'responsible innovation team' to check if its AI is ethical, Quartz. Available at: https://qz.com/1501998/google-created-a-responsible-innovation-team-to-check-if-its-ai-is-ethical/
[45] Walker, K. (2018). Google AI Principles updates, six months in, Google. Available at: https://www.blog.google/technology/ai/google-ai-principles-updates-six-months/
[46] Li, J., & Li, F. F. (2018). Cloud AutoML: Making AI accessible to every business. Internet: https://www. blog. google/topics/google-cloud/cloud-automl-making-ai-accessible-everybusiness.
[47] Faggella, D. (2019). AI in the Accounting Big Four – Comparing Deloitte, PwC, KPMG, and EY, Emerj. Available at: https://emerj.com/ai-sector-overviews/ai-in-the-accounting-big-four-comparing-deloitte-pwc-kpmg-and-ey/
[48] PwC (2019). Sizing the prize PwC's Global Artificial Intelligence Study: Exploiting the AI Revolution, PwC Global. Available at: https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html.
[49] PwC (2019). Artificial Intelligence (AI), familiarity breeds content, PwC UK. Available at: https://www.pwc.co.uk/services/consulting/technology/insights/artificial-intelligence-familiarity-breeds-content.html.
[50] PwC (2019). The responsible AI framework, PwC UK. Available at: https://www.pwc.co.uk/services/audit-assurance/risk-assurance/services/technology-risk/technology-risk-insights/accelerating-innovation-through-responsible-ai/responsible-ai-framework.html.
[51] Pavlou, P. A. (2018). Internet of Things–Will Humans be Replaced or Augmented?. GfK Marketing Intelligence Review, 10(2), 43-48.
[52] PwC (2019). A practical guide to responsible artificial intelligence (AI). Available at: https://www.pwc.com/gx/en/issues/data-and-analytics/artificial-intelligence/what-is-responsible-ai/responsible-ai-practical-guide.pdf
[53] EY (2018). How do you teach AI the value of trust? Available at: https://www.ey.com/Publication/vwLUAssets/ey-how-do-you-teach-ai-the-value-of-trust/$FILE/ey-how-do-you-teach-ai-the-value-of-trust.pdf
[54] Spaargaren, G., & Mol, A. P. (1992). Sociology, environment, and modernity: Ecological modernization as a theory of social change. Society & Natural Resources, 5(4), 323-344.
[55] Wolford, B. (2019). What are the GDPR Fines?. Proton Technologies AG. Retrieved from https://gdpr.eu/fines on 30.05.2019.
[56] Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. PLoS Medicine, 15(11), e1002689.
[57] Jobin, A., Lenca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1, 389-399.
[58] Liang, H., Xue, Y., & Wu, L. (2013). Ensuring employees' IT compliance: Carrot or stick?. Information Systems Research, 24(2), 279-294.
[59] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Schafer, B. (2018). AI4People—An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689-707.
[60] Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technological Forecasting and Social Change, 126, 3-13.

## Appendix 1

The list of responsible AI cases in this study:

- Audi AG (Automobile manufacturing), Germany
- Capital One (Financial and banking), United States
- H&M (Clothing retail), Sweden
- PwC (Professional services), United Kingdom
- Alder Hey Children's Hospital (Health care service), United Kingdom
- Google (Software), United States
- Sage Group (Software), United Kingdom
- IBM (Software), United States
- Quantcast (Software), United States
- Ernst & Young (EY) Global (Professional services), United Kingdom