Responsible AI Systems: Who are the Stakeholders?

Conference or Workshop Item

How to cite:

For guidance on citations see FAQs.

© 2022 The Authors

https://creativecommons.org/licenses/by-nc-nd/4.0/

Version: Version of Record

Link(s) to article on publisher’s website:
http://dx.doi.org/doi:10.1145/3514094.3534187

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online’s data policy on reuse of materials please consult the policies page.

oro.open.ac.uk
Responsible AI Systems: Who are the Stakeholders?

Adwait Deshpande  
School of Computing & Communications  
The Open University  
United Kingdom

Helen Sharp  
School of Computing & Communications  
The Open University  
United Kingdom

ABSTRACT
As of 2021, there were more than 170 guidelines on AI ethics and responsible, trustworthy AI in circulation according to the AI Ethics Guidelines Global Inventory maintained by AlgorithmWatch, an organisation which tracks the effects of increased digitalisation on everyday lives. However, from the perspective of day-to-day work, for those engaged in designing, developing, and maintaining AI systems identifying relevant guidelines and translating them into practice presents a challenge.

The aim of this paper is to help anyone engaged in building a responsible AI system by identifying an indicative long-list of potential stakeholders. This list of impacted stakeholders is intended to enable such AI system builders to decide which guidelines are most suited to their practice. The paper draws on a literature review of articles short-listed based on searches conducted on the ACM Digital Library and Google Scholar. The findings are based on content analysis of the short-listed literature guided by probes which draw on the ISO 26000:2010 Guidance on social responsibility.

The paper identifies three levels of potentially relevant stakeholders when responsible AI systems are considered: individual stakeholders (including users, developers, and researchers), organisational stakeholders, and national / international stakeholders engaged in making laws, rules, and regulations. The main intended audience for this paper is software, requirements, and product engineers engaged in building AI systems. In addition, business executives, policy makers, legal/regulatory experts, AI researchers, public, private, and third sector organisations developing responsible AI systems, and anyone interested in seeing functional responsible AI systems are the other intended audience for this paper.

CCS CONCEPTS
- Computing methodologies—Artificial intelligence  
- Social and professional topics—Professional topics—Computing profession—Codes of ethics  
- Social and professional topics—Professional topics—Management of computing and information systems—Project and people management—Systems planning  
- Social and professional topics—Professional topics—Management of computing and information systems—Project and people management—Systems analysis and design  
- Social and professional topics—Professional topics—Management of computing and information systems—Project and people management—Systems development  
- Social and professional topics—Professional topics—Computing and business—Socio-technical systems  
- Software and its engineering—Software creation and management  
- Information systems—Information systems applications—Computing platforms

KEYWORDS
AI ethics, AI system builders, Corporate social responsibility, ISO 26000:2010 Guidance on social responsibility, Responsible AI systems, Stakeholder identification

ACM Reference format:

1 Introduction
As of 2021, AlgorithmWatch, an organisation which tracks the effects of increased digitalisation of everyday activities had identified more than 170 Artificial Intelligence (AI) guidelines in various stages of development as part of its AI Ethics Guidelines Global Inventory [3] [9]. These guidelines cover a broad range of artefacts [28]: assessments, questionnaires, diagnostics, and checklists for organisations to assess their practices [59]; end-to-end-frameworks for different stages of the AI lifecycle [29] [61]; strategy guides and canvases for simplifying business models and decision-making [2] [69] [39]; design guides for AI systems
In addition, these guidelines are developed by a wide array of stakeholders ranging from technology companies such as Microsoft, Google, and IBM; research institutes such as the Future of Life Institute and the Alan Turing Institute; professional bodies such as the Association of Computing Machinery (ACM), the Institute of Electrical and Electronics Engineers (IEEE); and nation states including United Kingdom (UK), Australia, European Union (EU), Japan, and Canada (to name a few) [64]. Each of these stakeholders has different motivations, approaches, and intended outcomes vis-à-vis responsible AI systems. Because of these numerous guidelines, varying artefacts, and multiple stakeholders, the term ‘ethical AI’ itself has become an umbrella term signifying different aspects of assessing, understanding, and addressing the impacts of AI. Depending on the guideline and the stakeholders, ethical AI covers moral, ethical principles of AI operation, issues of privacy, trust, transparency, responsibility, accountability, and bias in AI systems, criticality of the data collected to train the AI systems, and legal compliance of AI systems [28] [56]. The variety of stakeholders, differences in artefacts, multiple definitions of ethical AI, and different regional/national jurisdictions in which the AI systems operate leads to confusion/complexities/uncertainty. This is particularly true for those involved in designing, developing, and maintaining AI systems (i.e. AI system builders, see section 2). Making decisions that ensure the resultant AI system behaves in an ethical, responsible way is often a task without a clear pathway of execution.

As part of unpacking the complexity of this landscape and helping the day-to-day practice of those engaged in designing, developing, and maintaining responsible AI systems, this paper aims to identify the stakeholders who could be impacted by the responsible AI systems. By identifying the stakeholders who could be impacted at an early stage of the AI lifecycle, the AI system builders can be better positioned to decide which ethical AI guidelines are most relevant for the products and services they are building. By drawing on this list, they can explore the stakeholders’ viewpoints and better understand the consequences of their design and development choices.

The paper presents the findings from the first stage of an ongoing study on responsible AI systems. It relies on content analysis of peer-reviewed and selected pre-print academic literature based on searches conducted in ACM Digital Library (DL) and Google Scholar. For the content analysis, it uses the clause 5.3.2 of the ISO 26000:2010 standard for social responsibility [41] to probe the literature for the following two research questions:

- **RQ01:** Who are the stakeholders likely to be impacted by responsible AI systems?
- **RQ02:** Which stakeholders are expected to address the impacts of responsible AI systems?

Since the stakeholders targeted by an AI system can vary significantly depending on the sector it serves, the aim of the paper is to identify an indicative long-list of potential stakeholders rather than a definitive list. Along with identifying the long-list of stakeholders, the paper reflects on the gaps in current knowledge vis-à-vis responsible AI systems and impacted stakeholders and further areas for investigation. Since the paper is focussed on identifying the stakeholders, it does not focus on identifying specific impacts or distinguishing between the impacts (positive or negative). The paper does not aim to map the impacts to specific stakeholders. Similarly, solutions or remedies to any impacts of responsible AI systems are also not covered as part of the findings discussed. The paper is not intended to critique or improve the existing guidelines. The long-list of stakeholders identified in this paper is aimed at facilitating a different perspective on how responsible AI system development could be approached.

The primary audience for this paper is software, requirement, and product engineers engaged in day-to-day work of building responsible AI systems. Other audience for this paper is business executives, policy makers, legal/regulatory experts, AI researchers, public, private, and third sector organisations developing responsible AI guidelines, and anyone interested in seeing functional responsible AI systems.

The next section (2) discusses definitions of key terms used in the paper. The research approach for the paper is covered in section 3. Section 4 includes the findings with a discussion on the potential stakeholders impacted by responsible AI systems and stakeholders who can address the impacts. The final section (5) discusses the findings and offers some thoughts on further areas for investigation.

## 2 Definitions / terminology

This section provides key definitions and terminology relevant to the discussion in the paper.

- **An AI system** is a deployable data-driven software system, which may use abstractions in the form of one or more machine learning (ML) models and is packaged with a user interface for gathering inputs and presenting outputs. The aim of such an AI system is to assist an end-user to perform a particular task [56] [54]. Apart from ML, AI technologies cover several techniques and applications including natural language processing (NLP), deep learning (DL), and reinforcement learning (RL). For the sake of brevity, the term AI system is used to cover most of the techniques and evolving approaches related to AI.

- **There are multiple definitions of responsible AI in the literature. These identify moral, ethical considerations for the use of AI systems [65] [21]; empirical/experimental approaches to its ethical behaviour [21] in conjunction with societal consequences of AI systems [50] [47]; and human oversight, technical robustness, privacy and data governance, transparency, fairness, wellbeing, explainability, and accountability of the systems [35] [65]. In this paper we focus on the chain of human and organisational control which governs responsible behaviour for the AI system. We use the term...**
3 Research approach

The findings in this paper are based on a literature review of articles short-listed from ACM DL and Google Scholar. The short-listed articles were subject to a conventional content analysis drawing on a set of probes adapted from the ISO 26000:2010 standard clause. In this section we describe the approach to short-list the literature followed by the process for content analysis.

To identify the relevant literature, we conducted two sets of searches: one on the ACM DL and the other on Google Scholar. Initial, exploratory searches on ACM DL and Google Scholar indicated that there is large body of growing research on ‘ethical AI’. For this paper, only peer-reviewed and pre-print academic literature that contained explicit reference to ‘responsible AI’ were identified in the searches.

For the literature search on ACM DL we used the following search string: "responsible AND (AI OR 'artificial int*') AND ((stakeholder AND responsib*) AND ("definit*" OR ‘influence*’ OR “interest*/"))."

The ACM DL searches identified 847 results.

The literature searches on Google Scholar used the following search strings: “Responsible AI”; “Responsible AI + stakeholders”; “Responsible AI + responsibility”; “Responsible AI + definitions”; and “Responsible AI + influence + interest”.

For the Google Scholar search results, the first 200 results were considered for each search string. These results were screened by title and by abstract to identify 51 most relevant results.

For both sets of searches, only results from 2015 onwards were considered to ensure the most relevant and recent findings on responsible AI and stakeholders were identified. The results from ACM DL (847) and Google Scholar (51) were sorted for relevance. Eliminating duplicates, 894 unique results were identified for screening.

The search results identified from ACM DL and Google Scholar searches were then screened for title and abstract. Effectively the results from Google Scholar were screened twice to reduce the risk of human bias in identifying articles manually. Where the title and abstract were insufficiently informative on responsible AI, the main text of the article was screened. The screening was performed over three rounds by one researcher with the second researcher vetting the screening and suggesting any revisions to the inclusion/exclusion criteria. Only articles that contained responsible AI in the title and abstract, defined responsible AI, described stakeholders, participants of responsible AI, and described stakeholder or participant roles in responsible AI were considered for inclusion as part of the screening.

Articles covering other aspects of ethical AI including explainability, trustworthiness, transparency, autonomy, technical implementation details of ethical AI (including responsible AI), data descriptions of ethical AI (including responsible AI), AI responsibility in case of hardware-based AI systems such as robots, drones, autonomous vehicles, and broader discussions about responsible technologies or computing were excluded in the screening. As identified in section 2, these exclusions were intended to ensure a concise discussion on the chain of human and organisational control which governs responsible behaviour for the AI system.

At the conclusion of the third round of screening, 45 articles were identified as being relevant for the discussion presented in this paper. As part of the content analysis, we included only peer-reviewed articles or pre-print versions of articles intended for academic publications. Despite the growing volume of grey literature in the form of policy reports, governmental consultations, and reports by strategy/management consulting sector on AI, such articles were not considered for content analysis.

For the 45 articles identified at the conclusion of the literature screening, we carried out content analysis based on the title, abstract, and the main text of the article. We used a form of conventional content analysis to identify the long-list of stakeholders from the literature [38]. We used the clause 5.3.2 of the ISO 26000:2010 standard on social responsibility [41] to form the probes used for the content analysis. We used ISO 26000:2010 because it is technology-agnostic and therefore could be easily adapted in the context of responsible AI systems. The clause 5.3.2 was used since it also includes questions specifically aimed at identifying stakeholders. The table below shows the original ISO questions and the adapted version of the questions for responsible AI systems. The revised probes are mapped to the research questions (RQ01 and RQ02).
4.1 Who are the stakeholders likely to be impacted by responsible AI systems?

The literature identifies a very broad range of stakeholders when discussing those who are likely to be impacted by responsible AI systems and those who can address the impacts of the decisions and activities of the responsible AI system. Amongst others, these stakeholders include AI researchers, non-AI experts, policymakers, civil society, academics, impacted users, and non-users [60].

Users of the system are identified by [58] as the most relevant stakeholder when the stakeholders to whom responsible AI systems have legal obligations are considered. However, a distinction between users of the system and data subjects whose data is used to train a system is necessary according to [71] when considering stakeholders impacted by responsible AI systems. Non-expert users such as teachers, parents, alumni, and community groups are also described as bringing in the real-world perspective on responsible behaviour on the decisions and activities of AI systems [49]. Although there is consensus in the literature that users are important stakeholders, it is the AI researchers and those engaged in building the AI systems that are more likely to express concerns about an AI system’s decisions and activities [60] [51]. The rise of online activism (via social media and networks) by the various stakeholders is also seen to contribute to increased understanding and acknowledgement of technology companies’ concerns about responsible AI systems [17] [33].

According to [62], the decisions and activities of AI systems are driven by the attributes, preferences, and usage patterns of the dataset used to train the AI models. This could result in users who are underrepresented in the datasets, and in engagement with responsible AI systems, being disadvantaged [62]. For example, women and children are identified as stakeholders that can be disproportionately affected by an AI system depending on their region, age, socioeconomic conditions, and ethnicity as per [62] and the policy guidance on AI for children published by the United Nations Children’s Fund (UNICEF) [22]. Users who lack requisite training and understanding about how AI systems use data and make decisions, may use the system incorrectly or misinterpret the system’s activities [10]. This could worsen the AI system’s ability to behave responsibly [10]. The users of the system and whether the system has an influence on their health, freedom, security, finance, and livelihood are also identified as crucial parameters when identifying the impacted stakeholders [10]. Potential risks of automated and autonomous decision-making by an AI system also heighten the perceived impacts in users and the need to hold the AI system responsible [10].

Apart from the users of the AI systems, those who develop AI systems and those who create policies for their deployment are also key stakeholders likely to be impacted by the responsible AI systems. The next section describes these findings in detail.

4 Findings
system’s decisions or activities [10]. If the task of identifying the harms of AI systems is assigned to a select group of experts, that also poses a risk of introducing bias in its decision-making [49]. To minimise the negative effects on users of the responsible AI system, non-AI experts have a role to reason about the long-term effects of responsible AI systems decisions or activities [49]. In contrast, AI experts, particularly those engaged in day-to-day development and maintenance of AI systems are more likely to consider the effects in terms of the system’s performance and efficiency rather than moral or ethical dimensions of the AI system [25]. AI system builders may be negatively impacted if organisations shift the responsibility of AI systems decisions and activities onto them to minimise scrutiny of business decisions [30]. Similar risk is identified by [32] who argues that driven by economic logic, organisations may lack the structures that protect the AI system builders from negative consequence of AI system’s decisions or activities they [the system builders] have built.

Large and influential multinational organisations in the technology sector are also likely to be subject to the negative impacts of AI systems. To mitigate such impacts, organisations such as Microsoft, Google, IBM, Baidu, Tencent, Intel, Sony, Workday, SAP, and Sage have produced either AI ethics guidelines or a statement to clarify their positions on responsible behaviour of AI systems [64]. Through these artefacts, these organisations appear to have identified their concerns and described their approach to how any intended/unintended consequences of responsible AI systems will be identified and addressed [17]. For example, Microsoft, Nvidia, and IBM have issued software frameworks to identify weaknesses of their responsible AI system offerings [17]. Google [74], Microsoft, and SAP have also identified changes to their business practices to identify and respond to any concerns arising from the decisions and activities of the responsible AI systems they are building [64]. Despite the sizeable number of guidelines issued by technology companies, across the sector technology companies have mainly responded to the potential negative impacts from AI systems through the prism of corporate social responsibility [1]. Similarly, faced with increasing scrutiny and concern around AI systems, universities and research institutes have also been engaged in identifying and rectifying any negative effects to ensure AI systems are more responsible [17].

According to [60], beyond exercising corporate social responsibility and identifying guidelines for ethical AI systems, those engaged in building AI systems in technology companies need to have the ability to veto development of specific behaviours or use of specific data sets. AI system builders have traditionally not been educated in ethical issues and may also lack the right opportunities to raise ethical concerns due to organisational structures [32]. Since companies are primarily driven by economic logic, AI system builders are likely to be excluded from any decisions that do not include technical development and maintenance of the AI systems [32]. Similarly, the responsibility for the decisions and activities of AI systems is often assigned to those engaged in building AI systems, despite them being mostly excluded from business decisions about AI systems as per [30].

Recognising societal impacts of AI systems and the need for the AI systems to be held responsible, several nations including EU nations (especially France, Germany, and Italy), the United States, the United Kingdom, Canada, countries in the Nordic-Baltic region (such as Finland, Sweden, and Denmark), Japan, China, India, Mexico, Australia, and New Zealand have released AI ethics documents identifying their concerns about ensuring that AI systems are responsible [64]. At the intergovernmental level, multiple bodies under the EU have also been active along with the Organisation for Economic Co-operation and Development (OECD), the G7, and the G20 to identify the principles which should govern a responsible AI system [9] [64]. However, such literature is insufficiently informative about the impacts of responsible AI systems for those engaged in day-to-day practice of designing and developing AI system and making it more responsible to the various stakeholders. For example, although guidelines such as European Commission-High-Level Expert Group (EC-HLEG) AI guidelines [35], OECD [54], Montreal declaration [76], and Beijing AI principles [18] are detailed and prescriptive in terms of responsible behaviour of AI systems, these are not legal stipulations. The extent to which these guidelines and declarations have impacted stakeholders engaged in making rules and regulations on responsible behaviour of AI systems is unclear based on the literature we reviewed.

4.2 Which stakeholders are expected to address the impacts of responsible AI systems?

The literature identifies the role of AI system builders, technology companies, and the users of the systems in conjunction with guidelines and frameworks that enable impacts of responsible AI systems to be identified [44] [60] [17] [75] [40] [7]. A wide spectrum of stakeholders including system engineers, project/product/process managers, designers, developers, data scientists, AI/ML experts, regulators, auditors, organisations that fund or own the AI systems, and users of AI technology have a role to play in ensuring that the AI systems are held responsible for specific impacts according to [40] [7].

Although new roles on AI and machine learning responsibility are said to be emerging in technology companies, AI researchers are seen to have primarily responded to the concerns about responsible AI systems through self-organising, unconference events, and creation of formal organisations and conferences [60]. Such formal organisations include the AAAI/ACM Conference on AI, Ethics, and Society (AIES), ACM Conference on Fairness, Accountability, and Transparency (FAccT), research institutes such as the Future of Life Institute, the Alan Turing Institute, The Institute for Ethical AI & Machine Learning, the Institute for Ethics on AI (Oxford university), and the AI Now Institute (to name just a few) [60] [14].

Apart from the developers and engineers who design, develop, and maintain AI systems and enable them to be responsible [40] [7], the literature also identifies a role for: the Human-Computer Interaction (HCI) community [44] [62]; non-
AI researchers [44]; ethics teams specifically designated to analyse and monitor AI systems for responsible behaviour [60]; AI researchers [70]; and public, private, and non-profit organisations that have developed AI ethics frameworks [42]. AI system builders need to not only manage technical performance but also deliver products and services in response to market pressures and business priorities [60]. In this context, the HCI community is also seen to have an important role to play to challenge the implicit assumptions and biases which govern AI systems and their responsible behaviour [62] [48]. The need for the HCI community to draw on organisational behavioural psychology literature to inform the design of AI systems and their decision-making is highlighted by [45] [11] [31]. Although AI researchers and AI system builders may lack the control to steer the commercial imperatives, they are not entirely powerless or irresponsible [27]. They have a crucial role to play in identifying risks to AI system lapses in responsibility and raising awareness about it [27].

Multi-stakeholder sector-specific guidelines [16]; domain-specific guidelines [23]; and technical solutions and empirical knowledge bases [52] also have a role to play in helping non-expert users and non-AI researchers investigate the impacts of responsible AI systems as per [75]. Structural and organisational changes in technology companies are suggested by [75] to ensure that such guidelines and knowledge bases are prepared and available in the public domain. Governance mechanisms such as AI and ethics boards at an organisational-level [13]; regulatory levers in the form of validation and certification agencies for responsible AI systems [16]; and training developers and engineers who design, develop, and maintain AI systems in ethics and human rights [75] are also mentioned in the literature to help address impacts of responsible AI systems.

In the literature, technology companies are primarily seen to focus on the impact of responsible AI systems on users. However, stakeholders such as advocacy groups, policy makers, and community groups are also identified by [43] as being of equal relevance when considering stakeholders who could address the impacts of responsible AI systems through soft-power interventions (e.g. developing manifestos, or online activism to raise user awareness).

To address the impacts of responsible AI systems, the role of internal organisational change at technology companies to ensure that an advisory board is constituted to document and monitor AI-related decision-making is highlighted by [64]. This could also result in changes in workforce hiring practices, composition, and skill sets, with Microsoft, Google, and SAP cited as organisations that have changed their business practices vis-à-vis workplace hiring [64]. To address the impacts of responsible AI systems, the organisations need to invest resources into data collection and curation, system testing, research into the possible social impacts of their system, and, in some cases, technical research as per [5].

As the legal/regulatory obligations on the deployed responsible AI systems develop at national/international level, organisations may have to change existing corporate structures to allay liability risks by ensuring legal compliance and minimising deliberative public scrutiny [60] [43]. The legal obligations of AI systems are likely to be based on use cases that involve blind spots in data, biases in data collection, and how AI systems’ decision-making has been modelled due to such blind spots and biases [62]. As a result, organisational practices on how AI systems harvest user data (including behavioural data) may have to change to address the imbalance of power between the AI system and its users [27]. Organisations need to ensure and publicise the mechanisms through which users are better informed of the legal obligations of responsible AI systems as per [27]. For example, Facebook has a ‘red team’ and technology companies such as Microsoft, Nvidia, IBM, and Google have either released or internally established frameworks or guidelines to address concerns about responsible AI systems and product development [17] [74]. Amongst non-tech companies, the example of H&M Group is noted in the literature for developing a checklist to ensure it uses AI systems in a responsible manner [74]. Professional bodies such as ACM and IEEE; learned societies such as the Japanese Society for Artificial Intelligence and the Royal Society; and research institutes such as the Future of Life Institute and the Alan Turing Institute have also identified potential approaches to address impacts of responsible AI systems. For example, ACM has released a Code of Ethics and Professional Conduct [6]. IEEE has released guidelines for the ethical design of autonomous and intelligent systems [67]. The Japanese Society for Artificial Intelligence has released a set of ethical guidelines to address the concerns about AI systems [68]. The Future of Life Institute has developed the Asilomar AI principles to advocate the use of AI systems for developing tools beneficial for everyday use [26]. The Alan Turing Institute’s work has focussed on AI ethics and safety in the public sector [46]. Such efforts are intended to allay concerns about responsible AI systems (see also [34] [66] [24] [28]).

When the stakeholders responsible for making rules and regulations at national/international level are considered, the question of who has legal obligations arising from holding AI systems responsible has led to uncertainty about the legal status of AI and the prospect of AI systems being accorded legal personhood [47]. However, legal personhood for AI systems raises the challenge of such systems holding their own assets or insurance premiums to compensate in any potential legal situation [72] [73]. In consequence, despite significant debate (see [37] [15]), the consensus appears to be that in the present state of AI system development, the legal obligations will be with humans [24] [66] [20]. However, as AI systems gain increased sentence and become more autonomous, regulations and public policies on legal obligations of AI responsibilities will need to adapt beyond human ethics [20] [27].

To balance the economic gains of AI innovation against societal consequences of responsible AI systems, most of the activities aimed at addressing the impacts of responsible AI systems have been in the form of ongoing investigations, consultations, or declarations of principles to guide the AI system builders. Some examples of existing and ongoing efforts to address the concerns about responsible AI systems are The OECD AI Principles [55] [54]; the Montreal Declaration [76]; the
UK House of Lords Select Committee on Artificial Intelligence report on AI [36]; the EC-HLEG on AI [35]; and the Beijing AI Principals [18]. The EU countries, Australia, UK, United States of America (US), Canada, China, and other G7 and G20 countries have so far focussed on identifying key sets of principles which could guide the development and future laws to address the impacts of responsible AI systems [64] [9]. At this stage of responsible AI system development however, the role of nation states and national/international legal/regulatory agencies has been advisory rather than focussed on legal or regulatory enforcement. Since AI systems cover multiple sectors and industries, there appears to be limited consensus in the literature on the mechanisms, types and extent of governance such stakeholders could implement to address the impacts of responsible AI systems. Technology companies have often argued in favour of corporate self-regulation or collective industry regulation, citing a lack of flexibility or understanding on the part of governments to regulate effectively [64]. The risk of premature regulation stifling AI-based innovation and competitiveness in a rapidly changing geo-political context also appears to have limited the role of nation states in addressing the impacts of responsible AI systems so far [64].

5 Discussion and conclusions

This paper identifies an indicative long-list of potentially relevant stakeholders for responsible AI systems: those who are impacted by responsible AI systems, and those who can address the impacts of responsible AI systems. To identify this indicative long-list of stakeholders, the paper adapted clause 5.3.2 of the ISO 26000:2010 standard on social responsibility to probe peer-reviewed and pre-print academic literature short-listed through searches on ACM DL and Google Scholar. This section considers the usefulness of the ISO 26000:2010 standard for stakeholder identification of AI systems, the long-list of stakeholders identified, and the gaps in current knowledge when considering relevant stakeholders for responsible AI systems.

The paper provides a potential way to use the clause 5.3.2 of ISO 26000:2010 standard for stakeholder identification of responsible AI systems. The paper used the adapted questions to probe the short-listed literature rather than attempt to identify impacted stakeholders for specific responsible AI system features. As a result, the extent to which the ISO 26000:2010 standard and the adapted questions would prove effective in day-to-day practice needs further testing and analysis.

Users of the system are the most relevant stakeholders when considering who is likely to be impacted by responsible AI systems. Users who are underrepresented in the datasets, and in engagement with responsible AI systems are likely to be disproportionately impacted by AI systems. Health, freedom, security, finance, and livelihood are crucial parameters when identifying which stakeholders are impacted. In addition to users, negative performance and efficiency of AI systems and moral or ethical dimensions of the AI system are also likely to impact AI developers and experts engaged in day-to-day development and maintenance of AI systems. Large and influential multinational organisations in the technology sector are the key stakeholders when the negative legal, economic, and regulatory impacts of AI systems are considered.

To address the impacts of responsible AI systems, a wide spectrum of stakeholders including system engineers, project/product/process managers, designers, developers, data scientists, AI/ML experts, regulators, auditors, organisations that fund or own the AI systems, and users of AI technology could play a role. To enable AI system builders to address these impacts, internal organisational changes at technology companies may be needed to improve AI-related decision-making at management/executive level. Individuals and organisations engaged in using, building, and maintaining AI systems, together with legislative and regulatory agencies at national and international level have a crucial role to play in making rules which would address the impacts of the responsible AI systems.

Based on these findings, we can discern three levels of potentially relevant stakeholders when responsible AI systems are considered:

- Individual stakeholders i.e. users, developers, engineers, researchers, AI experts, HCI researchers and experts, non-users of AI systems, non-AI experts;
- Organisational stakeholders i.e. technology companies, professional bodies such as ACM, IEEE, learned societies such as the Japanese Society for Artificial Intelligence, and research institutes such as the Future of Life Institute and the Alan Turing Institute; and
- National/international stakeholders engaged in making laws, rules, and regulations i.e. nation states (EU, US, UK, China, Australia etc.), regional/national legislative/regulatory agencies, OECD, UNICEF etc.

From the perspective of day-to-day practice, stakeholders such as those who design, develop, and maintain the systems, users of the system, and the organisations which fund the development of the systems can be considered to directly impact the development of the AI systems. On the other hand, stakeholders such as the public, private, and non-profit sector organisations engaged in developing the ethical AI guidelines, and stakeholders responsible for developing legal/regulatory rules for AI systems can be considered to play an indirect role in the behaviour of the responsible AI systems. However, as further clarity on the applicability of the ethical AI guidelines emerges, the impact on public, private, and non-profit sector organisations can be expected to become clearer. Similarly, as the national/international stakeholders develop legal/regulatory obligations for AI systems in greater detail, their impact on the day-to-day practice of AI systems can be expected to grow in the future.

The paper identifies a key challenge facing those who build AI systems: even from the perspective of day-to-day practice, the stakeholders beyond the immediate users of the system are likely to be impacted by the AI system (either positively or negatively). Since the focus of the AI system builders is often likely to be the immediate users of the system, the impacts of the system in a financial, legal, or political context is unlikely to be their
priority. The AI system builders may also lack the required knowledge or the power to make decisions should the wider stakeholder ecosystem beyond the end-user of the system be impacted. Despite this, there is an over-emphasis on the role of AI system builders in the current popular discourse. As a result, the role of organisational stakeholders, particularly technology companies, needs increased attention when addressing the impacts of responsible AI systems.

The decision-making dilemma or complexity arising out of the need to consider the stakeholders in an ecosystem context, beyond the immediate users of the AI system is also significantly compounded by the large number of ethical AI guidelines which are currently available. There are multiple definitions and interpretations of responsible AI, ‘what’ it constitutes, ‘who’ is responsible for it, ‘who’ is it responsible to, and ‘how’ those responsibilities should be discharged.

This paper offers an initial insight into the broad range of stakeholders relevant to building responsible AI systems and the nature of their engagement with the system. It highlights that several questions remain unanswered when considering the day-to-day practice of building responsible AI systems. For example, the usefulness of specific types of guidelines i.e. assessments, questionnaires, diagnostics, checklists; or software-based toolkits for day-to-day activities i.e. stakeholder identification, requirement analysis, maintaining software code, or identifying product backlogs, remains unknown and needs further investigation. The legal, regulatory obligations of responsible AI systems appear to be under active development and subject to significant variations at regional, national, and international jurisdictions. The impacts of these obligations at each of the three stakeholder levels need to be identified. The extent to which these obligations would affect the day-to-day practice of those who design, develop, and maintain responsible AI systems is yet to be understood.

This paper also highlights that at each stakeholder level, there are different, potentially conflicting trade-offs to consider. At an individual level, those engaged in designing and developing AI systems are more likely to be focussed on system features to be developed and technical performance. Individual stakeholders may lack training on ethical/moral principles which are grounded in social science disciplines. The users may lack sufficient technical understanding of AI systems to discern how their own choices and preferences shape the behaviour of the AI systems. At the organisational level, financial performance, shareholder value, and competitive market pressures are likely to be highly influential factors which govern the organisation’s priorities in terms of how the AI systems should behave. At the national/international level, existing legal/regulatory rules and obligations, geo-political competition, performance of the local economy, and consensus-building with other legal/regulatory institutions are likely to influence how stakeholders respond to the AI system behaviours and outcomes. As part of ongoing development of responsible AI systems, improved understanding of the interest, impact, and influence of stakeholders at the various levels will be essential for day-to-day work and decision-making for those who design, develop, and maintain AI systems.

ACKNOWLEDGMENTS

The authors would like to thank the reviewers for their incisive and valuable feedback.

REFERENCES


