Human TECHNOLOGY

ISSN: 1795-6889

https://ht.csr-pub.eu Volume 20(3), December 2024, 488–524

A REVIEW OF CAVEATS AND FUTURE CONSIDERATIONS OF RESEARCH ON RESPONSIBLE AI ACROSS DISCIPLINES

Bahar Memarian*
Simon Fraser University
Canada

Tenzin Doleck
Simon Fraser University
Canada

Abstract: *Research on responsible AI has boomed over the past few years. Yet, an understanding of the responsible AI research across primary disciplines is missing in the literature. Through a review of 1224 studies across 7 databases, a total of 177 studies were included for a review of responsible AI across the emerging areas of Business, Engineering, Governance, Health, Science, and Social Science. Using grounded theory and axial and open coding, we extracted and summarized areas, principles, research designs, challenges, findings, and trends in responsible AI research across the emerging areas. The contribution of this work lies in presenting the caveats and future considerations of research on responsible AI holistically and across primary disciplines in higher education.*

Keywords: *artificial intelligence, AI, responsible AI, education, review*

DOI: https://doi.org/10.14254/1795-6889.2024.20-3.4

INTRODUCTION

A healthy society advocates and enforces responsible behaviors by its citizens. Philosophers, historians, and others (e.g., Nietzsche, Sartre, Heidegger, Levinas, and Derrida) have embarked on the notion of responsibility and responsible behavior (Raffoul, 2010). Responsibility is an umbrella and all-encompassing term, but broadly it necessitates taking accountability and ownership of actions, whether they be good or bad (Gold, 2021). It predominantly stems from the philosophy of morality and humanistic growth (Constantinescu et al., 2021). Today, we see conflicts of responsible behavior not only among human citizens but also among artificial intelligence (AI) citizens who have become a part of our world (Yudkowsky, 2008). AI applications and agents, whether sparse or at scale, have come to be used by humans for collaboration and automation in decision-making. This has as a result sparked concerns on who, when, and how should be accounted responsible for actions and their subsequent implications (Giroux et al., 2022; Puaschunder, 2019). Despite the advancements in the technologies and algorithmic designs of artificial intelligence, AI is still missing a complete picture of societal values including responsibility, and may exhibit responsibly indifferent behavior (Cheng et al., 2021). There is a challenge to identify more of the must-have humanistic characteristics of AI (Ozmen Garibay et al., 2023). There is as such a need to understand and envelop responsibility in AI (Robbins, 2020). Neglecting societal needs such as responsibility and solely focusing on mathematical and algorithmic enhancements of AI programs is significantly detrimental. This approach may deter the notions of humanistic traits and have long-term and irreversible impacts on society and humans (Santoni & Mecacci, 2021). Neglecting these issues can be particularly harmful today given the proliferated uses of AI across disciplines and contexts (Sætra & Danaher, 2022).

PRIOR WORK AND GAP IN RESPONSIBLE AI

A summary of some key research on responsible AI can be seen in Table 1. Prior work has, for example, examined the precursors to responsible AI (Mikalef et al., 2022; Schiff et al., 2020), cases for responsible AI (Wang et al., 2020), explainable AI (Arrieta et al., 2020), and professional AI (Alam, 2023). Such research informs and warns of how responsible AI should be conceived and employed but may lack insight into the current state of research around responsible AI, especially across diverse areas. Much of the prior research conducted highlights the importance of responsible AI and holistic ways to address it, as presented in Table 1. Findings from such research often lead to the development of frameworks and best practices to devise AI responsibly. Yet, few prior research, particularly reviews, have examined the current state of research on responsible AI. We are motivated to examine the state of research on responsible AI across diverse disciplines. This survey is needed given that AI applications and techniques may greatly vary across different areas, leading to more distributed rather than universal notions of responsible AI behavior. Through a review of the literature, we are thus interested in examining emergent research trends on responsible AI across diverse areas and also see if similarities and differences exist. In doing so we wish to contribute to our understanding of research on responsible AI at every step of the way and

stage of research and further share challenges remaining with research on responsible AI across different areas.

Table 1. Summary of prior work

Reference	Key foci
(Akbarighatar et al., 2023)	Combined topic modeling with manual content analysis to review the literature on AI maturity and readiness.
(Alam, 2023)	Share the development of a course titled Safety, Fairness, Privacy, and Ethics of Artificial Intelligence" (SFPE-AI).
(Arrieta et al., 2020)	Present a methodology for the large-scale implementation of AI methods in real organizations with fairness, model explainability, and accountability at its center.
(Benjamins et al., 2019)	Share issues and lessons learned in an organization that develops and uses AI, both from a technical and organizational perspective.
(Cheng et al., 2021)	Provide a systematic framework of Socially Responsible AI Algorithms focused on examining the subjects of AI indifference.
(Clarke, 2019)	How organizations may manage responsibly and the need for risk assessment.
(Dignum, 2019)	Share ways in which AI can potentially produce unintended consequences and how to improve our knowledge about responsible AI.
(Mikalef et al., 2022)	Present five explanations for the principles-to-practices gap and the need for impact assessment.
(Schiff et al., 2020)	Present four themes that lead to 15 recommendations for responsible AI.
(Shneiderman, 2021)	Present a responsible AI initiative framework that encompasses five core themes for AI solution developers, healthcare professionals, and policymakers. These themes are summarized in the name SHIFT: Sustainability, Human-centeredness, Inclusiveness, Fairness, and Transparency.
(Siala & Wang, 2022)	Provide a bottom-up mapping of the current state of research at the intersection of Human-Centered AI, Ethical, and Responsible AI (HCER-AI).
(Tahaei et al., 2023)	Share 10 responsible AI cases from different industries to understand the use of responsible AI in practices

GOAL AND RESEARCH QUESTION

Rooted in a grounded theory approach, the goal of this research is to review literature across diverse domains emerging from the studies (i.e., science, engineering, governance, social views, and science) and identify research trends related to responsible artificial intelligence (AI). Grounded theory has been long used in higher education research (Conrad, 1982), enabling a differentiation between interpretivism and an absolute truth from an external reality. In recent years, research has shown that grounded theory has become one of the most commonly used qualitative research approaches (Stough & Lee, 2021) as it allows for generating a theory that is grounded in data (Birks & Mills, 2022; Morse, 2016; Stough &

Lee, 2021; Timmermans & Tavory, 2007). Using grounded theory in the review work hence offers a more structured yet interpretive approach to codification and analysis (Dunne, 2011; Wolfswinkel et al., 2013). We seek to answer the following topic and associated research questions: ***What are emergent trends in research on responsible AI across diverse domains (i.e., science, engineering, governance, social views, and science)***

1. What are the emergent principles of responsible AI?
2. What are the emergent research designs on responsible AI?
3. What are the emergent research challenges on responsible AI?
4. What are the emergent research findings on responsible AI?
5. Putting it together: What are emergent principle-research design-findings-challenges on responsible AI?

We aim to understand research designs, challenges, findings, future work, and implications surrounding responsible AI overall and across the mentioned domains. We first provide background information on responsible AI (RAI). The methods section summarizes our search process and protocol for conducting the systematic review of responsible AI in the literature. The results section summarizes our findings on how the landscape of responsible AI compares across different domains in the literature. The discussion section further synthesizes and presents the implications of research across different domains. The contribution of this work is to offer a picture of the current state and future considerations and directions for research on responsible AI across diverse areas.

METHODS

Using a grounded theory approach, we study and analyze the reviewed studies to obtain key emergent themes of research and similarities and differences across areas (Stough & Lee, 2021; Thornberg, 2012). The conceptualization of the area and its associated attributes such as study gap/motivation, context, design, findings, and so on are informed by reviewing each article and identifying what emerges. In other words, we use open coding to codify emergent topics from the reviewed studies. Wherever possible, we then apply axial coding to thematize and classify the attributes or summarize them in a few constituting words.

SEARCH PROCESS

An overview of our search process is presented in the PRISMA chart in Figure 1. We searched the string: "responsible artificial intelligence" or "responsible AI" in the following databases: PubMed, JSTOR, SCOPUS, IEEE Xplore, Science Direct, Semantic Scholar, Web of Science, and EBSCOhost. We then uploaded the articles to the Covidence software. Covidence is an online tool used for carrying out a systematic search process. It helps with identifying duplicates and organizing the search process. Of the total of 1224 publications uploaded, a total of 460 duplicates were removed from Covidence. Since we were interested in articles that put responsible AI at the foci of their work, we only selected articles from each database having "responsible AI" incorporated into their title and abstract and which were peer-reviewed and disseminated in English.

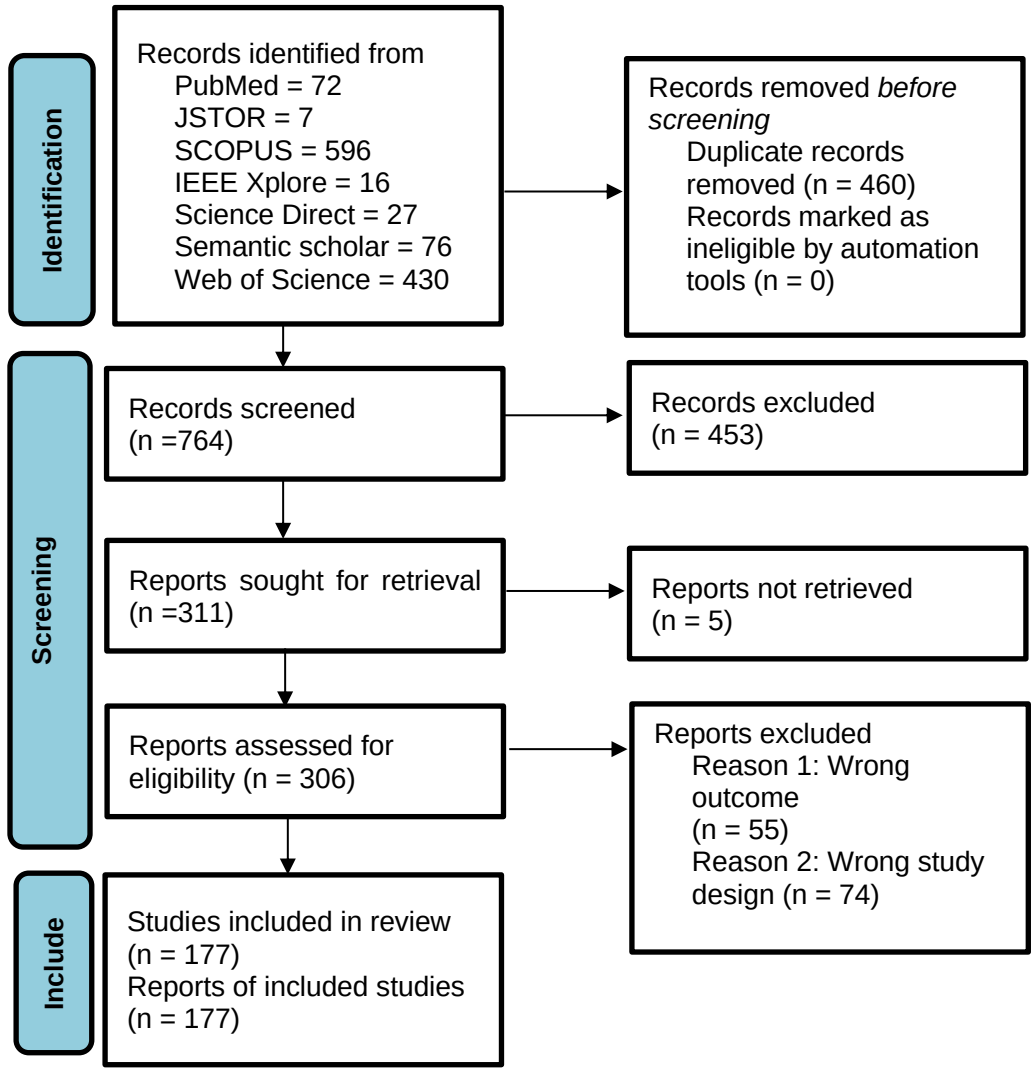

Figure 1. PRISMA Chart

METHODOLOGICAL FRAMEWORK

Our work is rooted in the grounded theory approach, where the coding and thematic analysis of research informs the findings of our research questions. This also meant that we depended on the literature reviewed to identify areas of interest. After undergoing the review process shown in the PRISMA chart in Figure 1, we used the title and keywords of the reviewed studies to attribute their primary area/discipline. A summary of the areas extracted and references can be found in Table 2. After review and thematic analysis of 177 studies included in this review, a total of six areas emerged, namely:

- Business (*N*= 12, 7%): including finance, and fintech.
- Engineering (*N*= 20, 11%): including the application of engineering across disciplines other than health.

- Governance (*N*= 12, 7%): including law, policymaking, and law enforcement.
- Health (*N*= 36, 20%): including medical and biomedical applications.
- Science (*N*= 42, 24%): including data science and pure sciences.
- Social science (*N*= 55, 31%): including perspectives, panel discussions, and practices in teaching and learning.

As can also be seen in Table 2, the highest number of publications were in social science, followed by science, health, and engineering, and the two areas of business and governance the least. The emergent keywords also show to be variable across the areas.

Table 2. Areas and references of reviewed studies

REFERENCE	AREA (% OF TOTAL)
(Asif et al., 2023; Crockett et al., 2023; Fenwick et al., 2018; Ferreira et al., 2023; Hadley, 2022; Haidar, 2024; Maki et al., 2023; Marcoux & Martineau, 2022; Maree et al., 2020; Narayanan, 2023; Papagiannidis et al., 2022; Wach et al., 2023)	Business, *N*= 12 (7)
(Abeywickrama et al., 2019; Alfrink et al., 2023; Bao et al., 2023; De Brito Duarte, 2023; Delecraz et al., 2022; Dominique et al., 2023; Falco, 2019; Faßbender, 2021; Gavrilova, 2023; Hu, 2023; Kamada, 2016; Kuck, 2023; Lee et al., 2024; Lescano et al., 2023; Lewis et al., 2022; Lu et al., 2022; Lu et al., 2023; Onyejegbu et al., 2023; Öz et al., 2022; Ramirez-Amaro et al., 2023)	Engineering, *N*= 20 (11)
(Arora et al., 2024; Azafrani & Gupta, 2023; Boulanin & Lewis, 2023; Brand, 2022; Gianni et al., 2022; Hacker & Passoth, 2022; Krafft et al., 2021; Krarup & Horst, 2023; Li & Yang, 2021; Meerveld et al., 2023; Pak, 2022; Upreti et al., 2023)	Governance, *N*= 12 (7)
(Al-Dhaen et al., 2021; Al-Hwsali et al., 2023; Borda et al., 2022; Cai et al., 2023; Calvo et al., 2023; Charalampakos et al., 2023; Cheng et al., 2021; El-Haddadeh et al., 2021; Fosso Wamba et al., 2021; Gaon & Stedman, 2019; Garg, 2023; Gilbert et al., 2023; Golbin et al., 2020; Gupta et al., 2021; Islam et al., 2023; Iwasawa et al., 2023; Jawad et al., 2021; Johnson et al., 2021; Jörg et al., 2023; Karagianni et al., 2023; Knopp et al., 2023; Kumar et al., 2021; Li et al., 2023; Liao et al., 2023; Liu et al., 2021; Marvin et al., 2022; Müftüoğlu et al., 2022; Prabhudesai et al., 2023; Roemmich & Andalibi, 2021; Rudd & Igbrude, 2023; Sivarajah et al., 2023; Szabo et al., 2022; Trocin et al., 2023.; Wang et al., 2021; Weekes & Eskridge, 2022; Wellnhofer, 2022)	Health, *N*= 36 (20)
(Agarwal & Mishra, 2021; Akbarighatar et al., 2023; Al Hashlamoun et al., 2023; Alibašić, 2023; Arya et al., 2023; Ayoubi et al., 2023; Baeza-Yates & Villoslada., 2022; Bani Ahmad, 2024; Belle, 2019; Bhattacharyya et al., 2023; Blockeel et al., 2023; Christos et al., 2020; Church et al., 2022; De et al., 2023; Deshmukh & Srinivasa, 2022; Feher, 2021; Fritz-Morgenthal et al.,	Science, *N*= 4 (24)

2022; Harfouche et al., 2024; Herrmann, 2023; Heuillet et al., 2022; Jaipuria et al., 2022; Ji et al., 2023; Kapania et al., 2023; Kumar et al., 2023; Liao et al., 2022; Lin, 2024; Liu et al., 2023; Mahoney et al., 2022; Merhi, 2023; Meske et al., 2022; Pisoni & Molnár, 2023; Pushkarna et al., 2022; Rantanen et al., 2021; Rivas et al., 2023; Rizinski et al., 2022; Roy & Salimi, 2023; Satish et al., 2023; Sothilingam, 2023; Souza et al., 2023; Tutun et al., 2022; Walsh et al., 2021; Weiss et al., 2023)	
(Alam, 2023; Aler Tubella et al., 2024; Anagnostou et al., 2022; Ansari et al., 2021; Barletta et al., 2023; Brumen et al., 2023; Buchholz et al., 2022; Buijsman, 2023; Cachat-Rosset & Klarsfeld, 2023; Calvo et al., 2020; Chang & Ke, 2023; Chen et al., 2023; Constantinescu et al., 2021; Contractor et al., 2022; Dennehy et al., 2023; Dholakia et al., 2023; Domínguez Figaredo et al., 2023; Dong et al., 2023; Farina, 2022; Grammatikos et al., 2023; Gwagwa et al., 2021; Hernández & Galanos, 2022; Inuwa-Dutse, 2023; Kapania et al., 2022; Karpagam et al., 2022; Kuennen, 2023; Lo et al., 2023; Lukkien et al., 2023; Mallinger & Baeza-Yates, 2024; Man Li & Crabbe, 2022; Maruyama, 2022; Marzouk et al., 2023; Methnani et al., 2023; Methnani et al., 2021; Mikalef et al., 2022; Minkkinen et al., 2024, 2023; Murray-Rust et al., 2023; Nedunuri, 2023; Ong & Findlay, 2023; Porayska-Pomsta, 2023; Reis et al., 2023; Rodier et al., 2023; Ruttkamp-Bloem, 2023; Sanderson et al., 2022, 2023; Satornino et al., 2024; Saxena et al., 2023; Scepanovic et al., 2023; Sinde et al., 2023; Stahl, 2022; Szczekocka et al., 2022; Tayebi & Garibay, 2023; Teixeira et al., 2023; Yang et al., 2020)	Social science, *N*= 55 (31) Total, *N*= 177 (100)

To help make the data collection process less biased, we employed a coding scheme and charted findings about our research questions. This implies that if each study's description fits the coding, it is charted appropriately. In doing we wish to make our data collection process more systematic and reproducible. A summary of coding for each research question can be found below:

1. What are the emergent principles of responsible AI?
    1. Empirical: mathematical or established theories from the literature as the guiding principle
    2. Social: generic social norms and goals as the guiding principle
    3. Empirical and social: an empirical and social set of principles used by the study
2. What are the emergent research designs on responsible AI?
    1. Quantitative: empirical means of hypothesis testing, conducting studies, and collecting quantitative data with mathematical analysis
    2. Qualitative: thematic means of coding, reviewing, and summarizing qualitative data with the synthesis of information
    3. Mixed: quantitative and qualitative means of data collection and analysis sed by the study
    4. Prescriptive: implies the research findings on responsible AI offer insight, model, or tool that is prescriptive of future actions

3. What are the emergent research challenges on responsible AI?
    1. Empirical: challenges with the mathematical models or theoretical aspects of research
    2. Social: challenges with the social and societal aspects of research
    3. Empirical and social: both empirical and social challenges
4. What are the emergent research findings on responsible AI?
    1. Descriptive: what had been done in the past
    2. Diagnostics: what actions have been done correctly or incorrectly in the past
    3. Predictive: predicts future actions
    4. Prescriptive: offers actionable recommendations for the future
5. Putting it together: What is the Research Landscape on Responsible AI? That is what is emergent principle-research design-findings-challenges on responsible AI?

We analyze data collected for each of our research questions through descriptive summaries of codes and frequency counts of derived text. A complete summary of data collected from our review is summarized in Table 8 and can be found in the Appendix. The descriptive tables present the count of each category against the total counts of that category, leading to a percentage of the area. We use this along with a thematic review of codes to inform emergent trends and highlight strengths and weaknesses in research around responsible AI across the derived areas of Business, Engineering, Governance, Health, Science, and Social Science.

LIMITATIONS

We acknowledge that this study comes with limitations. We broadly search for responsible AI and so our work only provides a high-level picture of research on responsible AI across different areas. We only included studies in English and did not include tertiary publications.

RESULTS

This section shares the findings of the review and coding of each of our research questions.

Principles

A descriptive summary of principle types employed within the areas of reviewed studies can be found in Table 3. Within the areas of Business and Engineering, Social principles were used the most, and the combination of social and empirical principles was used the least. Within the areas of Governance and Social Science, Social principles were used the most, and empirical-only principles were used the least. Within the areas of Health and Science, the combination of social and empirical principles was used the most, and social principles were used the least.

The word ethics is an emergent principle across the six areas and overall. Within the area of Business, principles are more regulatory and in the form of guidelines. Within the area of Engineering, principles are a social examination of technical contexts. Terms that are often associated with responsible technical AI such as transparency, accountability, explainability,

safety, security, accident, and privacy are seen in Engineering. Within the area of Governance, the safety and privacy principles held at an organizational level are an emergent theme of principles. Within the area of Health, principles are a social examination of healthcare designs and systems. Terms that are often associated with responsible medical AI such as transparency, non-maleficence, explanation, justice, and transparency are seen in Health. Within the areas of Social Science, terms such as fairness, transparency, accountability, nondiscrimination, safety, and privacy are emergent themes within the reviewed studies.

Table 3. Descriptive summary of principles across areas

AREA	EMPIRICAL	SOCIAL	EMPIRICAL AND SOCIAL
Business (7)	25.00	58.33	16.67
Engineering (11)	25.00	45.00	30.00
Governance (7)	16.67	50.00	33.33
Health (20)	30.56	27.78	41.67
Science (24)	38.10	19.05	42.86
Social science (31)	14.55	60.00	25.45
Overall	25.42	41.24	33.33

RESEARCH DESIGNS

A descriptive summary of research designs and analyses employed within the areas of reviewed studies can be found in Table 4. Within the areas of Business, Health, and Science, Qualitative studies are reported the most, followed by quantitative, and a combination of qualitative and quantitative studies are reported the least. Within the areas of Governance and Social Science, Qualitative studies are reported the most, followed by a combination of qualitative and quantitative studies, and quantitative studies are reported the least. Within the area of Engineering, quantitative and combination of quantitative and qualitative studies are reported the most and the qualitative challenges are reported the least. Overall qualitative challenges are most reported, followed by quantitative and a combination of qualitative and quantitative challenges.

Examples of quantitative studies include improving classification using hierarchical conditional multi-class, smart contracts, and models that enable comprehension of the implications and consequences of AI use. Examples of qualitative studies include literature reviews, societal frameworks of AI, and qualitative survey questionnaires. Examples of mixed methods studies include a review of the role of decision trees, speculative design and interview on a model, course designs, and mathematical modeling of responsible decision-making.

Table 4. Descriptive summary of research design/analysis across areas

AREA	QUANTITATIVE	QUALITATIVE	MIXED
Business (7)	25.00	58.33	16.67
Engineering (11)	35.00	30.00	35.00
Governance (7)	8.33	58.33	33.33
Health (20)	36.11	41.67	22.22
Science (24)	28.57	52.38	19.05
Social science (31)	10.91	69.09	20.00
Overall	23.73	53.67	22.60

RESEARCH CHALLENGES

A descriptive summary of research challenges within the areas of reviewed studies can be found in Table 5. Except for the area of Science, all other areas report Social or a combination of Social and Empirical challenges the most. For Science, Empirical challenges are reported the most. Examples of empirical challenges are integrating RAI mathematically in the data science lifecycle, inaccurate uses of harmonized systems, and identifying a true set of features for predictive analysis.

Examples of social challenges are controllability and sharing authority between humans and AI agents, addressing unfavorable outcomes, enabling civic participation, and impact of personal experiences. Examples of empirical and social challenges are conflicting goals, democratic shortfalls, opportunity risks, costs and uncertainty, and unparalleled contexts.

Table 5. Descriptive summary of research challenges across areas

AREA	EMPIRICAL	SOCIAL	EMPIRICAL AND SOCIAL
Business (7)	33.33	25.00	41.67
Engineering (11)	30.00	35.00	35.00
Governance (7)	8.33	41.67	50.00
Health (20)	22.22	41.67	36.11
Science (24)	38.10	28.57	33.33
Social science (31)	20.00	50.91	29.09
Overall	25.99	39.55	34.46

RESEARCH FINDINGS

A descriptive summary of research findings with AI within the areas of reviewed studies can be found in Table 6. All areas except for Engineering shared descriptive findings the most, and for Engineering descriptive findings were noted the most. The least shared findings were diagnostic for Business, Prescriptive for Engineering, Predictive for Governance, Health and Social Science, and Diagnostic for Science.

Examples of Descriptive findings include highlighting key factors (e.g., accountability and explainability), needed interdisciplinary discussions, and types of audit planning risk assessments taken by companies. Examples of Diagnostic findings include model checking for human-agent collectives, and proposing models or algorithms that can find deficits in the responsibility of AI. Examples of Predictive findings include the use of the Internet of Things (e.g., medical or IoMT) to continuously learn and help predict patterns in future data or events. Examples of Prescriptive findings include. research roadmap, frameworks, and best practices proposed to enhance the responsibility of AI.

Table 5. Descriptive summary of research findings across areas

AREA	DES=-CRIPTIVE	DIAGNOSTIC	PREDICTIVE	PRES-CRIPTIVE
Business (7)	66.67	0.00	25.00	8.33
Engineering (11)	35.00	45.00	15.00	5.00
Governance (7)	50.00	25.00	8.33	16.67
Health (20)	61.11	22.22	5.56	11.11
Science (24)	78.57	4.76	7.14	9.52
Social science (31)	56.36	23.64	7.27	12.73
Overall	60.45	19.77	9.04	10.73

PUTTING IT TOGETHER

A descriptive summary of research trends can be found in Table 7. As can be seen in Table 7, frequent trends per area are rather low and each study within each area and overall devises different assortments of principle types, research design, and data analyses, and discovers different challenges and findings. Yet, overall social principles, qualitative analysis, social challenges, and descriptive findings tend to be the predominant research trend. A largely qualitative and social type of research on responsible AI may bring both benefits and disadvantages to our understanding of responsible AI. Next, we aim to summarize the findings of our review and present new conceptions and directions for future research.

Table 7. Descriptive summary of frequent research trends across areas

AREA	TOP TREND	MEANING
Business (7)	2331 (*1*)	Social principle, mixed analysis, empirical and social challenges, descriptive findings
Engineering (11)	2221 (*1*)	Social principle, qualitative analysis, social challenge, descriptive findings
Governance (7)	2222 (*1*)	Social principle, qualitative analysis, social challenge, diagnostic findings
Health (20)	1112 (*3*)	Empirical principle, quantitative analysis, empirical challenge, diagnostic findings
Science (24)	3231 (*3*)	Mixed principle, quantitative analysis, empirical challenge, descriptive findings
Social science (31)	2221 (*5*)	Social principle, qualitative analysis, social challenge, descriptive findings
Overall		**Social principle, qualitative analysis, social challenge, descriptive findings**

DISCUSSION

In this work, we sought to gain a big-picture overview of the state of research on responsible AI. Through a review of 1224 studies across 7 databases, a total of 177 studies fit our criteria and explored responsible AI within one of six areas of Business, Engineering, Governance, Health, Science, and Social Science.

SUMMARY OF FINDINGS

We first identified and explored the areas and demographics of publications within the reviewed studies (see Table 1 and Table 2). We found that the highest number of publications were in social science, followed by science, health, and engineering, and the two areas of business and governance the least. This trend may suggest the need to make responsible AI in general and applied sciences streamlined more. What may make the characterization of responsible AI particularly difficult in governance and business can be conflicting financial versus humanistic motives by opposing stakeholders and the more prominent presence of dilemmas about responsible behavior in the two fields.

When examining the key terms, we found emergent associations of management for business, the bias for engineering, decision-making for governance, computing for health, data for science, and education for social science. These terms may signal what is deemed more representative of responsible behavior and outline the differences in perspective that exist surrounding responsible AI across the fields.

The journal followed by conference papers was the emergent type of research dissemination. We also see a surprisingly high number of publications overall in 2023 in comparison to the prior years. These may suggest that responsible AI is a newly found and hot topic that is currently receiving increased attention in the literature with interest in rapid

dissemination in the form of journal and conference articles. The lack of a high number of books on responsible AI across areas also indicates that all areas are struggling to find a standardized and generalizable definition and approach to responsible AI.

A descriptive codification of principles (i.e., either empirical, social, or empirical and social) across the six areas and overall revealed that Social principles are most often employed as building theories and principles of responsible AI. We also found that Business and Engineering, Governance and Social Science, and Health and Science share similar profiles of principles that are largely social. This makes sense given that responsible behavior in nature is a social construct and must have humanistic traits and is thus naturally foreign to empirically guided algorithms and AI applications. The unifying emphasis on social principles around responsible AI suggests that more discussion and enhancement to the social conceptualizations are needed. Further, work is needed to reformulate such social principles appropriately (e.g., accurately, and ethically) into a language that is understandable and readily understood by AI applications.

The principle of responsible AI is often associated with terms such as accountability, transparency and explainability, justice and fairness, non-maleficence, security, and privacy. This may suggest that AI manifests itself the most within the data pipeline and design of systems across different areas, and social construction of and consensus on responsible AI is needed. We also found that within Business the principles are more regulatory responsibility, in Engineering more technical responsibility, in Governance more organizational responsibility, in Health more medical responsibility, in Science more learning and design responsibility, and in social science, more equity and fairness responsibility is needed. These terms again highlight what is deemed more important across different areas and perhaps can hint at where more efforts need to be put into improving principles of responsible AI across different areas.

Research design tends to be qualitative with responsible AI. This makes sense as the fundamentals of responsibility have a subjective, interpretive, and socially constructed nature. We find either quantitative or qualitative studies often create vastly different conceptions of responsible AI. Quantitative studies are mathematical, theoretical, and algorithmic, and often deal with technology, designs, and procedures to make AI responsible in empirical ways. Qualitative studies are interpretive, and subjective, and often discuss and disseminate an understanding of what constitutes AI responsibility. Mixed studies employ a combination of empirical measurements and subjective perceptions and experiences of individuals toward responsible AI. While each research design and data analysis approach offers a unique perspective on responsible AI, more work may be needed to connect and make different research designs interoperable. For example, a challenge for future work may be on how to quantitatively characterize the many notions and needs of responsible AI appropriately.

Except for the area of Science, all other areas report Social or a combination of Social and Empirical challenges the most. For Science, Empirical challenges are reported the most. This finding highlights the dominant role of social gaps and needs that need to be accounted for and addressed with research on responsible AI. For example, even areas such as Engineering that use empirical research design and analysis the most found to be facing social challenges concerning responsible AI the most.

We find an abundance of descriptive findings surrounding responsible AI, and less work on developing diagnostic, predictive, and prescriptive research findings on responsible AI.

Diagnostic models can help inform where AI responsibility went wrong in the past. Predictive models can help hypothesize where AI responsibility may go wrong in the present or future. Prescriptive models can help propose corrective actions or guidelines on how to always safeguard responsible AI use.

Overall, we find research across areas to devise different assortments of principle types, research designs, and data analyses, and discover different challenges and findings. Yet, when looking holistically over the areas combined, we find the emergent research to be using social principles, qualitative analyses reporting on social challenges, and descriptive findings. There may thus need to be more harmonized and continuous research efforts to bridge the gap between empirical theory and practice and qualitative and social notions and needs of responsible AI.

IMPLICATIONS

The review of 177 studies across six areas presents both underlying and different implications. Examples of universal and underlying implications are the ever-expanding notions of ethics, such as accountability, explainability, transparency, justice, non-maleficence, accountability, and privacy. There is such a possibility that RAI's boundaries and terms are infinite and evolving with time. The future progression of AI, along with social-political powers may dictate to what degree humans can intervene and characterize responsible AI appropriately. Below, we present example implications for each of the six areas of Business, Engineering, Governance, Health, Science, and Social Science.

In Business, applications such as smart contracts may safeguard the responsible and ethical growth of AI. Yet, they remain a technique and technology for privacy, and so their expanded and increased use may also be susceptible to hacks and newly found issues. In Engineering, traditional policy development and execution may be evaluated by selected representatives' checks. Identifying who and when may qualify as a representative, however, may become difficult when there are so many unique attributes of each individual, and aggregating them into representative classes or profiles of groups becomes multi-optional. The verification results through counterexamples and simulation results may enable more intuitive and comprehensive model checking for the AI engineer. This presumes that every example fed the model has a clear-cut counterexample with a robust opposing result. If however, the actions and their results are assumed to spread in a fuzzy, rather than a binary spectrum, having a counter-example for every scenario may be overly simplistic or hard to characterize. In governance, stakeholder participation and public confidence may be needed for incorporating AI into governance decision-making processes. A challenge may become the emergence of AI technology itself as a stakeholder, since its intelligence may give it power to have or impose needs of its own. As such public opinion may become understood and explained through the lens of AI technology, which may be narrow and fragmented in ways it understands and expresses the world (e.g., classification, clustering).

In health, awareness about AI may provide knowledge about the problems it may introduce such as ethical and moral ones in healthcare education. Yet, a monotonic view of AI use in healthcare education may divert attention from real-world humanistic ethical and moral challenges that may be left unnoticed or become hidden with the use of AI. As such there may be two stories of responsible healthcare practices that need to be written in tandem,

namely responsible healthcare practice both in the presence and absence of AI. In healthcare, various ethical principles may need to be considered. Examples are Health Equity, Accountability, Avoiding Unfair Bias, Security and Privacy, Confidentiality, Safety, Transparency, Social Justice, and Autonomy. Using AI may help make diagnoses more accurate and informed, yet it may also allocate more responsibility and accountability to the AI, making the human expert less informed and responsible. This may in the long run impact patient's sense of trust and health care responsibility and create new challenges of its own. For example, new concerns may emerge such as now that the responsible AI made a false diagnosis, what counter or next steps need to be carried out and should the irresponsibility be mitigated.

In Science, AI trading may introduce new risks and variables into the financial ecosystem. A question thus becomes whether existing financial systems can adapt and transition to accept AI within their social and empirical frameworks. Whereas in sectors such as healthcare more knowledge of AI may lead to enhanced quality of human life, in some other areas such as finance and politics the dependency on AI may degrade and take away human's goals. When training data is unreliable, modeling may be achieved using a Named Entity Recognition model or NER. The NER itself, however, extracts information from the text and may be generalizable the best to the text itself and no other instances where data is unreliable.

In Social Science, connecting and filling the gap between the social and technical notions of responsible AI is difficult, yet the challenge does not end here. Humans and even AI programs may have different interpretations of responsibility within each of the social and technical notions. For example, the construction and thresholds of transparency may differ across individuals and contexts, making the enactment of responsible AI variables as well as difficult to validate. Governments and regulatory organizations can play an important role in establishing standards for sustainability. A consideration, however, is to know if and how the governments and regulatory organizations use their AI systems their own and the impact of distributed uses of types of artificial intelligence on governance and the quality of humans' lives.

NEW CONCEPTIONS AND DIRECTIONS

Our review of 177 studies informed efforts that have been persistently and least used. Here we aim to suggest areas of consideration for future work that may help improve research and insight into responsible AI.

PRINCIPLES

In this work, we coded principles to be of social, empirical, or combination nature. We should note that all principles, whether hard coded and static or dynamic and transient, are socially developed and established constructs. Principles are often set by a language and logic understandable by humans, and so are largely informed semantics set by humans. Moreover, principles are understood and used differently by humans and teams of entities (Oliver, D., & Jacobs, 2007). We attempted to separate principles by attributing those that are more driven by scientific theories as empirical and humans' needs and challenges as social. We followed

this lens and identified that overall, most of the work on responsible AI is more social, rather than empirical or combination nature. This finding points to a lack of empirical formulations of responsible AI and an abundance of human perspectives and opinions on responsible AI. Together, these two findings can highlight a challenging concern that awaits us with the increased use of AI. That is how can we enable the so many different social constructions of responsible AI to co-exist and further be able to link the social with empirical principles and characterize them in more theoretical terms? Principles of responsible AI may need to transcend social or theoretical perspectives and delve further into principles of behavior (Hull, 1943), principles that can be causal or dynamic (Dishop et al., 2020), and so on. Even within each of the social and empirical or combination principles, the rules of thumb may change particularly within and across different areas. For example, the statistical principles may come to be different or follow different protocols of validation under different circumstances within a business or when comparing business with health. Therefore, we find the notion of principle making and setting to be a complex, multifaceted, and nested concept that needs further discussion and classification.

RESEARCH DESIGN

In this work, we also coded research designs to be of qualitative, quantitative, or combination methods. We came to find prior research mostly employs qualitative methods. A key reason for this could be because responsible AI, even if denoted purely algorithmically or statistically, is understood through the impact it has on humans (e.g., stakeholders such as designers or users) and their needs (Deshpande & Sharp, 2022). As a result, responsible AI is often measured and evaluated based on humans' self-reported experiences and concerns (Sartori & Bocca, 2022). Future research designs thus may need to offer mechanisms for more exact, reproducible, and quantifiable, hence quantitative collection and analysis of responsible AI. Much of this may be achieved with future breakthroughs of portable, non-invasive, and accurate modeling of humans' brains and mathematical formulations of perceptions and experiences (Gazzaniga, 2013; S. J. Morse, 2006). But until then, we may need to focus on how to quantitatively chunk down complex and comprehensive notions of responsible AI and help make social consensus on what does or does not qualify as appropriate research design to implement, measure, or assess responsible AI.

RESEARCH CHALLENGES

In this work, we further coded research challenges to be of empirical, social, or combination types. Again, we found social challenges to be the most but interestingly, the empirical and combination challenges were not very far social challenges, suggesting all sorts of challenges exist in light of predominant social principles and qualitative research designs. The findings highlight that the study and execution of responsible AI have two important dimensions, namely how we perceive and use it socially and qualitatively as humans and how we come to be challenged both socially and empirically in the presence of AI in the loop. Since much of the logic and existence of AI is empirical, even its non-empirical applications lead to empirical challenges mainly because algorithms, numeric classification, and clustering are the drivers and backbone of what constitutes AI (Simon, 1995). A direction for future work,

perhaps in the opposite direction to probing and quantifying humans' brains, thinking, and emotions, may be to make the logic, explanation, and algorithms behind AI more qualitative by stripping away AI from its predominant quantitative semantics. This direction may be even more promising, given that AI's logic is likely to supersede that of humans. By making AI language and logic of nature that is natural to humans, even the most advanced notions of AI may still be readily digestible by humans. Yet, there is a lingering concern of deceit and mismatched explanations from the AI side.

RESEARCH FINDINGS

In this work, we further coded the findings of research on responsible AI to be descriptive, diagnostic, predictive, or prescriptive. We found descriptive studies to be dominant with a high lead in comparison to diagnostic, predictive, or prescriptive findings. This is concerning because it suggests that research is good in describing responsible AI but falls behind in identifying and troubleshooting where responsible AI falls short, predicting the presence or absence of responsible AI, and prescribing actions to maintain responsible AI in present or future times. This finding may highlight that we are in an infantry state of research on responsible AI and much more discussions and work is needed to characterize, execute, and enforce responsible AI.

CONCLUSION

Rooted in a grounded theory approach, the goal of this research was to review literature across diverse domains emerging from the studies (i.e., science, engineering, governance, social views, and science) and identify research trends related to responsible artificial intelligence (AI). Our review suggests that different areas devise different assortments of principle types, research designs, and data analyses, and discover different challenges and findings. Despite this diversity across the studied areas, research on responsible AI is inherently rooted and bounded by the social perceptions and needs of humans and empirical formulations and mechanisms of AI. A key challenge and goal of future work thus needs to align these two by either making humans more easily understand the empirical language of AI or conversely stripping AI away from a mathematical and empirical conception and reconstructing a new kind of AI that is solely social in a humanistic way, qualitative, and natural to humans' language, semantics, goals, and address responsible behavior in a universally acceptable way.

AVAILABILITY OF DATA AND MATERIAL

The author confirms that all data generated or analyzed during this study are included in this published article.

FUNDING

This work was supported by the Canada Research Chair Program; and the Canada Foundation for Innovation

ACKNOWLEDGMENTS

none.

CONFLICT OF INTEREST

None.

REFERENCES

Abeywickrama, D. B., Cirstea, C., Ramchurn, S. D., & IEEE. (2019). Model Checking Human-Agent Collectives for Responsible AI. In *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (Issues 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)). https://doi.org/10.1109/ro-man46459.2019.8956429 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Agarwal, S., & Mishra, S. (2021). Responsible AI: Implementing Ethical and Unbiased Algorithms. In *Responsible AI: Implementing Ethical and Unbiased Algorithms*. https://doi.org/10.1007/978-3-030-76860-7

Akbarighatar, P., Pappas, I., & Vassilakopoulou, P. (2023). A sociotechnical perspective for responsible AI maturity models: Findings from a mixed-method literature review. *International Journal of Information Management Data Insights*, *3*(2), 100193. https://doi.org/https://doi.org/10.1016/j.jjimei.2023.100193

Al-Dhaen, F., Hou, J., Rana, N. P., & Weerakkody, V. (2021). Advancing the Understanding of the Role of Responsible AI in the Continued Use of IoMT in Healthcare. *Information Systems Frontiers*, 1–20.

Al-Hwsali, A., Alsaadi, B., Abdi, N., Khatab, S., Alzubaidi, M., Solaiman, B., & Househ, M. (2023). Scoping Review: Legal and Ethical Principles of Artificial Intelligence in Public Health...21st International Conference on Informatics, Management, and Technology in Healthcare (ICIMTH), July 1-3, 2023, Athens, Greece. *Studies in Health Technology & Informatics*, *305*, 640–643. https://doi.org/10.3233/SHTI230579

Al Hashlamoun, N., Al Barghuthi, N., & Tamimi, H. (2023). Exploring the Intersection of AI and Sustainable Computing: Opportunities, Challenges, and a Framework for Responsible Applications. *2023 9th International Conference on Information Technology Trends, ITT 2023*, 220–225. https://doi.org/10.1109/ITT59889.2023.10184228

Alam, A. (2023). Developing a Curriculum for Ethical and Responsible AI: A University Course on Safety, Fairness, Privacy, and Ethics to Prepare Next Generation of AI Professionals. In *Lecture Notes on Data Engineering and Communications Technologies* (Vol. 171, pp. 879–894). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-981-99-1767-9_64

Aler Tubella, A., Mora-Cantallops, M., & Nieves, J. C. (2024). How to teach responsible AI in Higher Education: challenges and opportunities. *Ethics & Information Technology*, *26*(1), 1–14. https://doi.org/10.1007/s10676-023-09733-7

Alfrink, K., Keller, I., Doorn, N., Kortuem, G., & ACM. (2023). Contestable Camera Cars: A Speculative Design Exploration of Public AI That Is Open and Responsive to Dispute. *Proceedings Of The 2023 Chi Conference on Human Factors in Computing Systems, CHI 2023*, *CHI conference on Human Factors in*

Computing Systems (CHI). https://doi.org/10.1145/3544548.3580984 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Alibašić, H. (2023). Developing an Ethical Framework for Responsible Artificial Intelligence (AI) and Machine Learning (ML) Applications in Cryptocurrency Trading: A Consequentialism Ethics Analysis. *FinTech*, *2*(3), 430–443. https://doi.org/10.3390/fintech2030024

Anagnostou, M., Karvounidou, O., Katritzidaki, C., Kechagia, C., Melidou, K., Mpeza, E., Konstantinidis, I., Kapantai, E., Berberidis, C., Magnisalis, I., & Peristeras, V. (2022). Characteristics and challenges in the industries towards responsible AI: a systematic literature review. *Ethics & Information Technology*, *24*(3), 1–18. https://doi.org/10.1007/s10676-022-09634-1 WE - Social Science Citation Index (SSCI) WE - Arts & Humanities Citation Index (A&HCI)

Ansari, A., Hoffmann, A. L., Gurses, S., Sloane, M., Vasquez, M. A., & Pearl, Z. (2021). Technology, equity and social justice roundtable. In C. B., S. K.A., P. Z., D. R., & L. H.A. (Eds.), *2021 IEEE International Symposium on Society and Technology, ISTAS 2021* (Vols. 2021-Octob). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ISTAS52410.2021.9629208

Arora, A., Vats, P., Tomer, N., Kaur, R., Saini, A. K., Shekhawat, S. S., & Roopak, M. (2024). Data-Driven Decision Support Systems in E-Governance: Leveraging AI for Policymaking. *Lecture Notes in Networks and Systems*, *844*, 229–243. https://doi.org/10.1007/978-981-99-8479-4_17

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F., Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., … Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/https://doi.org/10.1016/j.inffus.2019.12.012

Arya, A., Roy, S., & Jonnala, S. (2023). An Ensemble-based approach for assigning text to correct Harmonized system code. *2023 International Conference on Artificial Intelligence and Smart Communication, AISC 2023*, 35–41. https://doi.org/10.1109/AISC56616.2023.10085512

Asif, R., Hassan, S. R., & Parr, G. (2023). Integrating a Blockchain-Based Governance Framework for Responsible AI. *Future Internet*, *15*(3), 97. https://doi.org/10.3390/fi15030097

Ayoubi, H., Tabaa, Y., & El Kharrim, M. (2023). Artificial Intelligence in Green Management and the Rise of Digital Lean for Sustainable Efficiency. *E3S Web of Conferences*, *412*. https://doi.org/10.1051/e3sconf/202341201053

Azafrani, R., & Gupta, A. (2023). Bridging the civilian-military divide in responsible AI principles and practices. *Ethics & Information Technology*, *25*(2), 1–5. https://doi.org/10.1007/s10676-023-09693-y

Baeza-Yates, R., & Villoslada, P. (2022). Human vs. Artificial Intelligence. In *2022 IEEE 4th International Conference on Cognitive Machine Intelligence* (Issue IEEE 4th International Conference on Cognitive Machine Intelligence (CogMI), pp. 40–48). https://doi.org/10.1109/CogMI56440.2022.00016 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Bani Ahmad, A. Y. A. (2024). Ethical implications of artificial intelligence in accounting: A framework for responsible ai adoption in multinational corporations in Jordan. *International Journal of Data and Network Science*, *8*(1), 401–414. https://doi.org/10.5267/j.ijdns.2023.9.014

Bao, Z. C., Zhang, W. S., Zeng, X. J., Zhao, H. W., Dong, C. H., Nie, Y. M., Liu, Y. E. R., Liu, Y. E. R., & Wu, J. Z. (2023). Software Architecture for Responsible Artificial Intelligence Systems: Practice in the Digitization of Industrial Drawings. *Computer*, *56*(4), 38–49. https://doi.org/10.1109/MC.2023.3240416 WE - Science Citation Index Expanded (SCI-EXPANDED)

Barletta, V. S., Caivano, D., Gigante, D., Ragone, A., Santa Barletta, V., Caivano, D., Gigante, D., Ragone, A., & ACM. (2023). A Rapid Review of Responsible AI frameworks: How to guide the development of ethical AI. *27TH International Conference on Evaluation and Assessment in Software Engineering, EASE 2023, 27th International Conference on Evaluation and Assessment in Software Engineering (EASE)*, 358–367. https://doi.org/10.1145/3593434.3593478 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Belle, V. (2019). The quest for interpretable and responsible artificial intelligence. *Biochemist*, *41*(5), 16–19. https://doi.org/10.1042/bio04105016

Benjamins, R., Barbado, A., & Sierra, D. (2019). Responsible AI by design in practice. *ArXiv Preprint*.

Bhattacharyya, R., Raj, N., Mukherjee, S., Choudhury, P., & Gaur, R. (2023). Responsible AI for Sustainable Agriculture: Forecasting Rice Yield to Combat Global Food Insecurity. *2023 14th International Conference on Computing Communication and Networking Technologies, ICCCNT 2023*. https://doi.org/10.1109/ICCCNT56998.2023.10306460

Birks, M., & Mills, J. (2022). *Grounded theory: A practical guide*.

Blockeel, H., Devos, L., Frénay, B., Nanfack, G., & Nijssen, S. (2023). Decision trees: from efficient prediction to responsible AI. *Frontiers in Artificial Intelligence*, *6*, 1124553. https://doi.org/10.3389/frai.2023.1124553

Borda, A., Molnar, A., Neesham, C., & Kostkova, P. (2022). Ethical Issues in AI-Enabled Disease Surveillance: Perspectives from Global Health. *Applied Sciences-Basel*, *12*(8). https://doi.org/10.3390/app12083890

Boulanin, V., & Lewis, D. A. (2023). Responsible reliance concerning development and use of AI in the military domain. *Ethics & Information Technology*, *25*(1), 1–5. https://doi.org/10.1007/s10676-023-09691-0 WE - Social Science Citation Index (SSCI) WE - Arts & Humanities Citation Index (A&HCI)

Brand, D. J. (2022). Responsible Artificial Intelligence in Government: Development of a Legal Framework for South Africa. *EJournal of EDemocracy & Open Government*, *14*(1), 130–150. https://doi.org/10.29379/jedem.v14i1.678

Brumen, B., Göllner, S., & Tropmann-Frick, M. (2023). Aspects and Views on Responsible Artificial Intelligence. In G. Nicosia, V. Ojha, E. LaMalfa, G. LaMalfa, P. Pardalos, G. DiFatta, G. Giuffrida, & R. Umeton (Eds.), *Machine Learning, Optimization, and Data Science, LOD 2022, PT I* (Vol. 13810, Issues 8th Annual International Conference on Machine Learning, Optimization and Data science (LOD), pp. 384–398). https://doi.org/10.1007/978-3-031-25599-1_29 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Buchholz, J., Lang, B., & Vyhmeister, E. (2022). The development process of Responsible AI: The case of ASSISTANT*. *IFAC-PapersOnLine*, *55*(10), 7–12. https://doi.org/https://doi.org/10.1016/j.ifacol.2022.09.360

Buijsman, S. (2023). Why and How Should We Explain AI? *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *13500 LNAI*, 196–215. https://doi.org/10.1007/978-3-031-24349-3_11

Cachat-Rosset, G., & Klarsfeld, A. (2023). Diversity, equity, and inclusion in artificial intelligence: An evaluation of guidelines. *Applied Artificial Intelligence*, *37*(1). https://doi.org/10.1080/08839514.2023.2176618

Cai, Z. N., Ren, B. J., Ma, R. H., Guan, H. B., Tian, M. K., & Wang, Y. (2023). Guardian: A Hardware-Assisted Distributed Framework to Enhance Deep Learning Security. *IEEE Transactions on Computational Social Systems*, *10*(6), 3012–3020. https://doi.org/10.1109/TCSS.2023.3262289

Calvo, A., Ortiz, N., Espinosa, A., Dimitrievikj, A., Oliva, I., Guijarro, J., & Sidiqqi, S. (2023). Safe AI: Ensuring Safe and Responsible Artificial Intelligence. *NIC Cybersecurity Conference (JNIC)*, 1–4.

Calvo, R. A., Peters, D., Vold, K., & Ryan, R. M. (2020). Supporting Human Autonomy in AI Systems: A Framework for Ethical Enquiry. In *Philosophical Studies Series* (Vol. 140, pp. 31–54). https://doi.org/10.1007/978-3-030-50585-1_2

Chang, Y. L., & Ke, J. (2023). Socially responsible artificial intelligence empowered people analytics: A novel framework towards sustainability. *Human Resource Development Review*, 15344843231200930.

Charalampakos, F., Tsouparopoulos, T., Papageorgiou, Y., Bologna, G., Panisson, A., Perotti, A., Koutsopoulos, I., & IEEE. (2023). Research Challenges in Trustworthy Artificial Intelligence and Computing for Health: The Case of the PRE-ACT project. *2023 Joint European Conference on Networks and Communications & 6g Summit, Eucnc/6g Summit, Joint European Conference on Networks and Communications / 6G Summit (EuCNC/6G Summit)*, 629–634.

https://doi.org/10.1109/EUCNC/6GSUMMIT58263.2023.10188239 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Chen, H., Chan-Olmsted, S., & Thai, M. (2023). Culture Sensitivity and Information Access: A Qualitative Study among Ethnic Groups. *Qualitative Report*, *28*(8), 2504–2522. https://doi.org/10.46743/2160-3715/2023.5981 WE - Emerging Sources Citation Index (ESCI)

Cheng, L., Varshney, K. R., & Liu, H. (2021). Socially responsible AI algorithms: Issues, purposes, and challenges. *Journal of Artificial Intelligence Research*, *71*, 1137–1181. https://doi.org/10.1613/JAIR.1.12814

Christos, S. C., Panagiotis, T., & Christos, G. (2020). Combined multi-layered big data and responsible AI techniques for enhanced decision support in Shipping. *2020 International Conference on Decision Aid Sciences and Application, DASA 2020*, 669–673. https://doi.org/10.1109/DASA51403.2020.9317030

Church, K., Schoene, A., Ortega, J. E., Chandrasekar, R., & Kordoni, V. (2022). Emerging trends: Unfair, biased, addictive, dangerous, deadly, and insanely profitable. *Natural Language Engineering*, *29*(2), 483–508. https://doi.org/10.1017/S1351324922000481

Clarke, R. (2019). Principles and business processes for responsible AI. *Computer Law & Security Review*, *35*(4), 410–422. https://doi.org/10.1016/j.clsr.2019.04.007 WE - Social Science Citation Index (SSCI)

Conrad, C. (1982). Grounded theory: An alternative approach to research in higher education. *The Review of Higher Education*, *5*(4), 239–249.

Constantinescu, M., Voinea, C., Uszkai, R., & Vica, C. (2021). Understanding responsibility in Responsible AI. Dianoetic virtues and the hard problem of context. *Ethics and Information Technology*, *23*(4), 803–814. https://doi.org/10.1007/s10676-021-09616-9

Contractor, D., McDuff, D., Haines, J. K., Lee, J., Hines, C., Hecht, B., ... & Li, H. (2022). Behavioral use licensing for responsible AI. *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 778–788.

Crockett, K., Colyer, E., Gerber, L., & Latham, A. (2023). Building Trustworthy AI Solutions: A Case for Practical Solutions for Small Businesses. *IEEE Transactions on Artificial Intelligence*, *4*(4), 778–791. https://doi.org/10.1109/TAI.2021.3137091

De, A., Gudipudi, S. S., Panchanan, S., & Desarkar, M. S. (2023). ComplAI: Framework for Multi-factor Assessment of Black-Box Supervised Machine Learning Models. *Proceedings of the ACM Symposium on Applied Computing*, 1096–1099. https://doi.org/10.1145/3555776.3577771

De Brito Duarte, R. (2023). Towards Responsible AI: Developing Explanations to Increase Human-AI Collaboration. *Frontiers in Artificial Intelligence and Applications*, *368*, 470–482. https://doi.org/10.3233/FAIA230126

Delecraz, S., Eltarr, L., Becuwe, M., Bouxin, H., Boutin, N., & Oullier, O. (2022). Responsible Artificial Intelligence in Human Resources Technology: An innovative inclusive and fair by design matching algorithm for job recruitment purposes. *Journal of Responsible Technology*, *11*, 100041. https://doi.org/https://doi.org/10.1016/j.jrt.2022.100041

Dennehy, D., Griva, A., Pouloudi, N., Dwivedi, Y. K., Mäntymäki, M., & Pappas, I. O. (2023). Artificial Intelligence (AI) and Information Systems: Perspectives to Responsible AI. *Information Systems Frontiers*, *25*(1), 1–7. https://doi.org/10.1007/s10796-022-10365-3

Deshmukh, J., & Srinivasa, S. (2022). Computational Transcendence: Responsibility and agency. *Frontiers in Robotics and AI*, *9*, 977303. https://doi.org/10.3389/frobt.2022.977303

Deshpande, A., & Sharp, H. (2022). Responsible AI Systems: Who are the Stakeholders? *2022 AAAI/ACM Conference on AI, Ethics, and Society*, 227–236.

Dholakia, A., Ellison, D., Hodak, M., & Dutta, D. (2023). Benchmarking Considerations for Trustworthy and Responsible AI (Panel). *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *13860 LNCS*, 110–119. https://doi.org/10.1007/978-3-031-29576-8_8

Dignum, V. (2019). *Responsible artificial intelligence: how to develop and use AI in a responsible way* (Vol. 1). Cham: Springer.

Dishop, C. R., Olenick, J., & DeShon, R. P. (2020). Principles for taking a dynamic perspective. *Handbook on the Temporal Dynamics of Organizational Behavior*, 26–43.

Domínguez Figaredo, D., Stoyanovich, J., Figaredo, D. D., & Stoyanovich, J. (2023). Responsible AI literacy: A stakeholder-first approach. *Big Data & Society*, *10*(2). https://doi.org/10.1177/20539517231219958

Dominique, B., El Maghraoui, K., Piorkowski, D., Herger, L., Maghraoui, K. E., Piorkowski, D., & Herger, L. (2023). FactSheets for Hardware-Aware AI Models: A Case Study of Analog In Memory Computing AI Models. In C. Ardagna, N. Atukorala, C. Chang, R. Chang, J. Fan, G. Fox, S. Helal, Z. Jin, Q. Lu, T. Seceleanu, & S. S. Yau (Eds.), *Proceedings - 2023 IEEE International Conference on Software Services Engineering, SSE 2023* (Issue 1st IEEE International Conference on Software Services Engineering (IEEE SSE), pp. 148–158). https://doi.org/10.1109/SSE60056.2023.00029 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Dong, T., Li, S., Chen, G., Xue, M., Zhu, H., & Liu, Z. (2023). RAI2: Responsible Identity Audit Governing the Artificial Intelligence. *30th Annual Network and Distributed System Security Symposium, NDSS 2023*. https://doi.org/10.14722/ndss.2023.241012

Dunne, C. (2011). The place of the literature review in grounded theory research. *International Journal of Social Research Methodology*, *14(2)*, 111–124.

El-Haddadeh, R., Fadlalla, A., & Hindi, N. M. (2021). Is There a Place for Responsible Artificial Intelligence in Pandemics? A Tale of Two Countries. *Information Systems Frontiers*, *25*(6), 2221–2237. https://doi.org/10.1007/s10796-021-10140-w

Falco, G. (2019). Participatory AI: Reducing AI bias and developing socially responsible AI in smart cities. *IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)*, 154–158.

Farina, L. (2022). Artificial Intelligence Systems, Responsibility and Agential Self-Awareness. In V. C. Muller (Ed.), *Philosophy and Theory of Artificial Intelligence 2021* (Vol. 63, Issue Philosophy and Theory of Artificial Intelligence (PT-AI), pp. 15–25). https://doi.org/10.1007/978-3-031-09153-7_2 WE - Conference Proceedings Citation Index - Science (CPCI-S) WE - Conference Proceedings Citation Index - Social Science & Humanities (CPCI-SSH)

Faßbender, J. (2021). Particles of a Whole: Design Patterns for Transparent and Auditable AI-Systems. *UbiComp/ISWC 2021 - Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*, 272–275. https://doi.org/10.1145/3460418.3479345

Feher, K. (2021). Expectation of smart mentality and citizen participation in technology-driven cities. *Smart Structures and Systems*, *27*(3), 435–445. https://doi.org/10.12989/sss.2021.27.3.435

Fenwick, M., Vermeulen, E. P. M., & Corrales, M. (2018). Business and regulatory responses to artificial intelligence: Dynamic regulation, innovation ecosystems and the strategic management of disruptive technology. In *Perspectives in Law, Business and Innovation* (pp. 81–103). https://doi.org/10.1007/978-981-13-2874-9_4

Ferreira, R. M. F. D., Grilo, A., & Maia, M. J. (2023). A Maturity Model for Industries and Organizations of All Types to Adopt Responsible AI—Preliminary Results. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *14115 LNAI*, 67–78. https://doi.org/10.1007/978-3-031-49008-8_6

Fosso Wamba, S., Queiroz, M. M., Wamba, S. F., Queiroz, M. M., Fosso Wamba, S., & Queiroz, M. M. (2021). Responsible Artificial Intelligence as a Secret Ingredient for Digital Health: Bibliometric Analysis, Insights, and Research Directions. *Information Systems Frontiers*, *25*(6), 1–16. https://doi.org/10.1007/s10796-021-10142-8

Fritz-Morgenthal, S., Hein, B., & Papenbrock, J. (2022). Financial Risk Management and Explainable, Trustworthy, Responsible AI. *Frontiers in Artificial Intelligence*, *5*, 779799. https://doi.org/10.3389/frai.2022.779799 WE - Emerging Sources Citation Index (ESCI)

GAON, A., & STEDMAN, I. A. N. (2019). A Call to Action: Moving Forward With the Governance of Artificial Intelligence in Canada. *Alberta Law Review*, *56*(4), 1137–1165. http://10.0.113.245/alr2547

Garg, M. (2023). Mental Health Analysis in Social Media Posts: A Survey. *Archives of Computational Methods in Engineering : State of the Art Reviews*, *30*(3), 1819–1842. https://doi.org/10.1007/s11831-022-09863-z

Gavrilova, M. L. (2023). Responsible Artificial Intelligence and Bias Mitigation in Deep Learning Systems. In E. Banissi, H. Siirtola, A. Ursyn, J. M. Pires, N. Datia, K. Nazemi, B. Kovalerchuk, R. Andonie, M. Nakayama, M. Temperini, F. Sciarrone, Q. V Nguyen, M. S. Mabakane, A. Rusu, U. Cvek, M. Trutschl, H. Mueller, R. Francese, F. Bouali, & G. Venturini (Eds.), *Proceedings of the International Conference on Information Visualisation* (Issues 27th International Conference on Information Visualisation (IV) / 19th International Conference Computer Graphics, Imaging and Visualization (CGiV), pp. 329–333). https://doi.org/10.1109/IV60283.2023.00062 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Gazzaniga, M. (2013). Understanding layers: From neuroscience to human responsibility. *().. Neurosciences and the Human Person: New Perspectives on Human Activities*, 1–14.

Gianni, R., Lehtinen, S., Nieminen, M., Gianni, R., Lehtinen, S., & Nieminen, M., Gianni, R., Lehtinen, S., & Nieminen, M. (2022). Governance of Responsible AI: From Ethical Guidelines to Cooperative Policies. *Frontiers in Computer Science*, *4*, 873437. https://doi.org/10.3389/fcomp.2022.873437

Gilbert, Y., Kayanja, E. W., Kalungi, J. E., Kyagaba, J. M., & Marvin, G. (2023). Explainable AI for Black Sigatoka Detection. *Lecture Notes in Networks and Systems*, *798 LNNS*, 181–196. https://doi.org/10.1007/978-981-99-7093-3_12

Giroux, M., Kim, J., Lee, J. C., & Park, J. (2022). Artificial intelligence and declined guilt: Retailing morality comparison between human and AI. *Journal of Business Ethics*, *178*(4), 1027–1041.

Golbin, I., Rao, A. S., Hadjarian, A., & Krittman, D. (2020). Responsible AI: A Primer for the Legal Community. In X. T. Wu, C. Jermaine, L. Xiong, X. H. Hu, O. Kotevska, S. Y. Lu, W. J. Xu, S. Aluru, C. X. Zhai, E. Al-Masri, Z. Y. Chen, & J. Saltz (Eds.), *2020 IEEE International Conference on Big Data (BIG DATA)* (Issue 8th IEEE International Conference on Big Data (Big Data), pp. 2121–2126). https://doi.org/10.1109/BigData50022.2020.9377738 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Gold, N. E. (2021). Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. *Genetic Programming and Evolvable Machines*, *22*(1), 137–139. https://doi.org/10.1007/s10710-020-09394-1

Grammatikos, N. A., Anagnostopoulou, E., Apostolou, D., & Mentzas, G. (2023). A Conversational Digital Assistant for STEM Education. *14th International Conference on Information, Intelligence, Systems and Applications, IISA 2023*. https://doi.org/10.1109/IISA59645.2023.10345918

Gupta, S., Kamboj, S., & Bag, S. (2021). Role of Risks in the Development of Responsible Artificial Intelligence in the Digital Healthcare Domain. *Information Systems Frontiers*, *25*(6), 2257–2274. https://doi.org/10.1007/s10796-021-10174-0

Gwagwa, A., Kazim, E., Kachidza, P., Hilliard, A., Siminyu, K., Smith, M., & Shawe-Taylor, J. (2021). Road map for research on responsible artificial intelligence for development (AI4D) in African countries: The case study of agriculture. *Patterns*, *2*(12), 100381. https://doi.org/https://doi.org/10.1016/j.patter.2021.100381

Hacker, P., & Passoth, J. H. (2022). Varieties of AI Explanations Under the Law. From the GDPR to the AIA, and Beyond. In A. Holzinger, R. Goebel, R. Fong, T. Moon, K. R. Muller, & W. Samek (Eds.), *XXAI - BEYOND EXPLAINABLE AI: International Workshop, Held in Conjunction with ICML 2020, July 18, 2020, Vienna, Austria, Revised and Extended Papers* (Vol. 13200, Issue International Workshop on Beyond Explainable Artificial Intelligence (xxAI), pp. 343–373). https://doi.org/10.1007/978-3-031-04083-2_17 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Hadley, E. (2022). Prioritizing Policies for Furthering Responsible Artificial Intelligence in the United States. *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*, 5029–5038. https://doi.org/10.1109/BigData55660.2022.10020551

Haidar, A. (2024). An Integrative Theoretical Framework for Responsible Artificial Intelligence. *International Journal of Digital Strategy, Governance, & Business Transformation (IJDSGBT)*, *13*(1), 1–23. https://doi.org/10.4018/IJDSGBT.334844

Harfouche, A. L., Petousi, V., & Jung, W. (2024). AI ethics on the road to responsible AI plant science and societal welfare. *Trends in Plant Science*. https://doi.org/10.1016/j.tplants.2023.12.016

Hernández, A. D., & Galanos, V. (2022). A toolkit of dilemmas: Beyond debiasing and fairness formulas for responsible AI/ML. *International Symposium on Technology and Society, Proceedings*, *2022-Novem*. https://doi.org/10.1109/ISTAS55053.2022.10227133

Herrmann, H. (2023). What's next for responsible artificial intelligence: a way forward through responsible innovation. *Heliyon*, *9*(3), e14379. https://doi.org/https://doi.org/10.1016/j.heliyon.2023.e14379

Heuillet, A., Couthouis, F., Diaz-Rodriguez, N., & Díaz-Rodríguez, N. (2022). Collective eXplainable AI: Explaining Cooperative Strategies and Agent Contribution in Multiagent Reinforcement Learning with Shapley Values. *IEEE Computational Intelligence Magazine*, *17*(1), 59–71. https://doi.org/10.1109/MCI.2021.3129959

Hu, P. (2023). SatAIOps: Revamping the Full Life-Cycle Satellite Network Operations. *Proceedings of IEEE/IFIP Network Operations and Management Symposium 2023, NOMS 2023*. https://doi.org/10.1109/NOMS56928.2023.10154334

Hull, C. L. (1943). *Principles of behavior: an introduction to behavior theory*.

Inuwa-Dutse, I. (2023). *FATE in AI: Towards Algorithmic Inclusivity and Accessibility*. arXiv preprint. https://doi.org/10.48550/arXiv.2301.01590

Islam, M. R., Sakib, M. K. H., Ulhaq, A., Akter, S., Zhou, J. L., Asirvathamt, D., & Asirvatham, D. (2023). *SIDVis*: Designing Visual Interactive System for Analyzing Suicide Ideation Detection. In E. Banissi, H. Siirtola, A. Ursyn, J. M. Pires, N. Datia, K. Nazemi, B. Kovalerchuk, R. Andonie, M. Nakayama, M. Temperini, F. Sciarrone, Q. V Nguyen, M. S. Mabakane, A. Rusu, U. Cvek, M. Trutschl, H. Mueller, R. Francese, F. Bouali, & G. Venturini (Eds.), *Proceedings of the International Conference on Information Visualisation* (Issues 27th International Conference on Information Visualisation (IV) / 19th International Conference Computer Graphics, Imaging and Visualization (CGiV), pp. 384–389). https://doi.org/10.1109/IV60283.2023.00071 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Iwasawa, M., Kobayashi, M., & Otori, K. (2023). Knowledge and attitudes of pharmacy students towards artificial intelligence and the ChatGPT. *Pharmacy Education*, *23*(1), 665–675. https://doi.org/10.46542/pe.2023.231.665675

Jaipuria, N., Stevo, K., Zhang, X. L., Gaopande, M. L., Garcia, I. C., Jain, J., Murali, V. N., & IEEE. (2022). deepPIC: Deep Perceptual Image Clustering For Identifying Bias In Vision Datasets. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2022* (Issue IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4792–4801). https://doi.org/10.1109/CVPRW56347.2022.00526 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Jawad, M. S., Mahdin, H., Mohammed Alduais, N. A., Hlayel, M., Mostafa, S. A., & Abd Wahab, M. H. (2021). Recent and Future Innovative Artificial Intelligence Services and Fields. *Proceedings - ISAMSR 2021: 4th International Symposium on Agents, Multi-Agents Systems and Robotics*, 29–32. https://doi.org/10.1109/ISAMSR53229.2021.9567891

Ji, Z. L., Ma, P. C., Wang, S., Li, Y. H., & IEEE. (2023). Causality-Aided Trade-off Analysis for Machine Learning Fairness. *2023 38TH IEEE/ACM International Conference on Automated Software Engineering, ASE, 38th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, 371–383. https://doi.org/10.1109/ASE56229.2023.00105

Johnson, M., Albizri, A., & Harfouche, A. (2021). Responsible Artificial Intelligence in Healthcare: Predicting and Preventing Insurance Claim Denials for Economic and Social Wellbeing. *Information Systems Frontiers*, *25*(6), 2179–2195. https://doi.org/10.1007/s10796-021-10137-5

Jörg, S., Ziethmann, P., & Breuer, S. (2023). MedAIcine: A Pilot Project on the Social and Ethical Aspects of AI in Medical Imaging. *Communications in Computer and Information Science*, *1832 CCIS*, 455–462. https://doi.org/10.1007/978-3-031-35989-7_58

Kamada, T. (2016). The issues of automated driving vehicle and the expectations for 3D integration technology. In *2015 International 3d Systems Integration Conference (3DIC 2015)* (Issue IEEE International 3D Systems Integration Conference (3DIC)).

Kapania, S., Siy, O., Clapper, G., Meena, S. P. A., Sambasivan, N., ACM, Sp, A. M., & Sambasivan, N. (2022). "Because AI is 100% right and safe": User Attitudes and Sources of AI Authority in India. *Proceedings of the 2022 Chi Conference on Human Factors in Computing Systems (CHI' 22)*, *CHI Conference on Human Factors in Computing Systems (CHI)*. https://doi.org/10.1145/3491102.3517533

Kapania, S., Taylor, A. S., Wang, D., & ACM. (2023). A hunt for the Snark: Annotator Diversity in Data Practices. *Proceedings of the 2023 Chi Conference On Human Factors in Computing Systems, CHI 2023*, *CHI conference on Human Factors in Computing Systems (CHI)*. https://doi.org/10.1145/3544548.3580645 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Karagianni, C., Paraschou, E., Yfantidou, S., & Vakali, A. (2023). MINDSET: A benchMarking suIte exploring seNsing Data for SElf sTates inference. *2023 IEEE 10th International Conference on Data Science and Advanced Analytics, DSAA 2023 - Proceedings*. https://doi.org/10.1109/DSAA60987.2023.10302638

Karpagam, G. R., Varma, A., Samrddhi, M., & Shri Shivathmika, V. (2022). Understanding, Visualizing and Explaining XAI Through Case Studies. *8th International Conference on Advanced Computing and Communication Systems, ICACCS 2022*, 647–654. https://doi.org/10.1109/ICACCS54159.2022.9785199

Knopp, M. I., Warm, E. J., Weber, D., Kelleher, M., Kinnear, B., Schumacher, D. J., Santen, S. A., Mendonça, E., & Turner, L. (2023). AI-Enabled Medical Education: Threads of Change, Promising Futures, and Risky Realities Across Four Potential Future Worlds. *JMIR Medical Education*, *9*(1), e50373. https://doi.org/10.2196/50373

Krafft, P. M., Young, M., Katell, M., Lee, J. E., Narayan, S., Epstein, M., Dailey, D., Herman, B., Tam, A., Guetler, V., Bintz, C., Raz, D., Jobe, P. O., Putz, F., Robick, B., Barghouti, B., & ACM. (2021). An Action-Oriented AI Policy Toolkit for Technology Audits by Community Advocates and Activists. In *Proceedings of the 2021 Acm Conference on Fairness, Accountability, and Transparency, FACCT 2021* (Issue ACM Conference on Fairness, Accountability, and Transparency (FAccT), pp. 772–781). https://doi.org/10.1145/3442188.3445938 WE - Conference Proceedings Citation Index - Science (CPCI-S) WE - Conference Proceedings Citation Index - Social Science & Humanities (CPCI-SSH)

Krarup, T., & Horst, M. (2023). European artificial intelligence policy as digital single market making. *BIG DATA & SOCIETY*, *10*(1). https://doi.org/10.1177/20539517231153811

Kuck, K. (2023). Generative Artificial Intelligence: A Double-Edged Sword. *2023 IEEE IFEES World Engineering Education Forum and Global Engineering Deans Council: Convergence for a Better World: A Call to Action, WEEF-GEDC 2023 - Proceedings*. https://doi.org/10.1109/WEEF-GEDC59520.2023.10343638

Kuennen, C. S. (2023). Developing Leaders of Character for Responsible Artificial Intelligence. *Journal of Character & Leadership Development (JCLD)*, *10*(3), 52–59. https://doi.org/10.58315/jcld.v10.273

Kumar, K. P., Thiruthuvanathan, M. M., K.K., S., & Chandra, D. R. (2023). Human AI: Explainable and responsible models in computer vision. In *Emotional AI and Human-AI Interactions in Social Networking* (pp. 237–254). https://doi.org/10.1016/B978-0-443-19096-4.00006-7

Kumar, P., Dwivedi, Y. K., & Anand, A. (2021). Responsible Artificial Intelligence (AI) for Value Formation and Market Performance in Healthcare: the Mediating Role of Patient's Cognitive Engagement. *Information Systems Frontiers*, *25*(6), 2197–2220. https://doi.org/10.1007/s10796-021-10136-6

Lee, S. U., Fernando, N., Lee, K., & Schneider, J.-G. (2024). A survey of energy concerns for software engineering. *Journal of Systems and Software*, *210*, 111944. https://doi.org/https://doi.org/10.1016/j.jss.2023.111944

Lescano, L. F., Flores, M. L., & Pico, M. P. (2023). Distributed Facial Recognition Facial Recognition in Visual Internet of Things (VIoT) An Intelligent Approach. *Journal of Intelligent Systems and Internet of Things*, *10*(2), 18–23. https://doi.org/10.54216/JISIoT.100202

Lewis, G. A., Echeverría, S., Pons, L., Chrabaszcz, J., & IEEE. (2022). Augur: A Step Towards Realistic Drift Detection in Production ML Systems. In *2022 IEEE/ACM 1ST International Workshop on Software Engineering for Responsible Artificial Intelligence (SE4RAI 2022)* (Issue 1st IEEE/ACM International Workshop on Software Engineering for Responsible Artificial Intelligence (SE4RAI), pp. 37–44). https://doi.org/10.1145/3526073.3527590 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Li, C., & Yang, L. (2021). Responsible AI: The Revolution in Governance Technology in China. *Proceedings - 2021 2nd International Conference on Artificial Intelligence and Education, ICAIE 2021*, 76–80. https://doi.org/10.1109/ICAIE53562.2021.00023

Li, K., Lau, B. P. L., Yuan, X., Ni, W., Guizani, M., & Yuen, C. (2023). Toward Ubiquitous Semantic Metaverse: Challenges, Approaches, and Opportunities. *IEEE Internet of Things Journal*, *10*(24), 21855–21872. https://doi.org/10.1109/JIOT.2023.3302159

Liao, Q. V, Zhang, Y., Luss, R., Doshi-Velez, F., & Dhurandhar, A. (2022). Connecting Algorithmic Research and Usage Contexts: A Perspective of Contextualized Evaluation for Explainable AI. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, *10*, 147–159. https://doi.org/10.1609/hcomp.v10i1.21995

Liao, W., Liu, Z., Dai, H., Xu, S., Wu, Z., Zhang, Y., Huang, X., Zhu, D., Cai, H., Li, Q., Liu, T., & Li, X. (2023). Differentiating ChatGPT-Generated and Human-Written Medical Texts: Quantitative Study. *JMIR Medical Education*, *9*, e48904. https://doi.org/10.2196/48904

Lin, Z. (2024). Towards an ai policy framework in scholarly publishing. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2023.12.002

Liu, J., Chen, H., Shen, J., & Choo, K. R. (2023). FairCompass: Operationalising Fairness in Machine Learning. *IEEE Transactions on Artificial Intelligence*, 1–10. https://doi.org/10.1109/TAI.2023.3348429

Liu, R., Gupta, S., & Patel, P. (2021). The Application of the Principles of Responsible AI on Social Media Marketing for Digital Health. *Information Systems Frontiers*, *25*(6), 2275–2299. https://doi.org/10.1007/s10796-021-10191-z

Lo, S. K., Liu, Y., Lu, Q. H., Wang, C., Xu, X. W., Paik, H.-Y. Y., & Zhu, L. M. (2023). Toward Trustworthy AI: Blockchain-Based Architecture Design for Accountability and Fairness of Federated Learning Systems. *IEEE Internet of Things Journal*, *10*(4), 3276–3284. https://doi.org/10.1109/JIOT.2022.3144450 WE - Science Citation Index Expanded (SCI-EXPANDED)

Lu, Q., Zhu, L., Xu, X., Whittle, J., Douglas, D., & Sanderson, C., Lu, Q. H., Zhu, L. M., Xu, X. W., Whittle, J., Douglas, D., Sanderson, C., & Soc, I. C. (2022). Software engineering for responsible AI: An empirical study and operationalised patterns. *Software Engineering in Practice, ACM/IEEE 44th International Conference on Software Engineering-Software Engineering in Practice (ICSE-SEIP)*, 241–242. https://doi.org/10.1145/3510457.3513063 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Lu, Q., Luo, Y., Zhu, L., Tang, M., Xu, X., & Whittle, J. (2023). Developing Responsible Chatbots for Financial Services: A Pattern-Oriented Responsible Artificial Intelligence Engineering Approach. *IEEE Intelligent Systems*, *38*(6), 42–51. https://doi.org/10.1109/MIS.2023.3320437

Lukkien, D. R. M., Nap, H. H., Buimer, H. P., Peine, A., Boon, W. P. C., Ket, J. C. F., Minkman, M. M. N., & Moors, E. H. M. (2023). Toward Responsible Artificial Intelligence in Long-Term Care: A Scoping Review on Practical Approaches. *Gerontologist*, *63*(1), 155–168. https://doi.org/10.1093/geront/gnab180

Mahoney, C., Gronvall, P., Huber-Fliflet, N., & Zhang, J. (2022). Explainable Text Classification Techniques in Legal Document Review: Locating Rationales without Using Human Annotated Training Text Snippets. *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*, 2044–2051. https://doi.org/10.1109/BigData55660.2022.10020626

Maki, H. A., Al Mubarak, M., & Bakir, A. (2023). Understanding Artificial Intelligence Through Its Applications and Concerns. In *Internet of Things: Vol. Part F1270* (pp. 135–152). https://doi.org/10.1007/978-3-031-35525-7_9

Mallinger, K., & Baeza-Yates, R. (2024). Responsible AI in Farming: A Multi-Criteria Framework for Sustainable Technology Design. *Applied Sciences-Basel*, *14*(1). https://doi.org/10.3390/app14010437 WE - Science Citation Index Expanded (SCI-EXPANDED)

Man Li, R. Y., & Crabbe, M. J. C. (2022). Artificial Intelligence Robot Safety: A Conceptual Framework and Research Agenda Based on New Institutional Economics and Social Media. In *Current State of Art in Artificial Intelligence and Ubiquitous Cities* (pp. 41–61). https://doi.org/10.1007/978-981-19-0737-1_3

Marcoux, A. M., & Martineau, J. T. (2022). From Principled to Applied AI Ethics in Organizations: A Scoping Review. *Communications in Computer and Information Science*, *1655 CCIS*, 641–646. https://doi.org/10.1007/978-3-031-19682-9_81

Maree, C., Modal, J. E., Omlin, C. W., & IEEE. (2020). Towards Responsible AI for Financial Transactions. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)* (Issue IEEE Symposium Series on Computational Intelligence (IEEE SSCI), pp. 16-21 WE-Conference Proceedings Citation Index).

Maruyama, Y. (2022). Categorical Artificial Intelligence: The Integration of Symbolic and Statistical AI for Verifiable, Ethical, and Trustworthy AI. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *13154 LNAI*, 127–138. https://doi.org/10.1007/978-3-030-93758-4_14

Marvin, G., Nakatumba-Nabende, J., Hellen, N., & Alam, M. G. R. (2022). Responsible Artificial Intelligence for Preterm Birth Prediction in Vulnerable Populations. *Proceedings of IEEE Asia-Pacific Conference on Computer Science and Data Engineering, CSDE 2022*. https://doi.org/10.1109/CSDE56538.2022.10089301

Marzouk, M., Zitoun, C., Belghith, O., & Skhiri, S. (2023). The Building Blocks of a Responsible Artificial Intelligence Practice: An Outlook on the Current Landscape. *IEEE Intelligent Systems*, *38*(6), 9–18. https://doi.org/10.1109/MIS.2023.3320438

Meerveld, H. W., Lindelauf, R. H. A., Postma, E. O., & Postma, M. (2023). The irresponsibility of not using AI in the military. *Ethics & Information Technology*, *25*(1), 1–6. https://doi.org/10.1007/s10676-023-09683-0 WE - Social Science Citation Index (SSCI) WE - Arts & Humanities Citation Index (A&HCI)

Merhi, M. I. (2023). An Assessment of the Barriers Impacting Responsible Artificial Intelligence. *Information Systems Frontiers*, *25*(3), 1147–1160. https://doi.org/10.1007/s10796-022-10276-3

Meske, C., Abedin, B., Klier, M., & Rabhi, F., Meske, C., Abedin, B., Klier, M., & Rabhi, F. (2022). Explainable and responsible artificial intelligence. *Electronic Markets*, *32*(4), 2103–2106. https://doi.org/10.1007/s12525-022-00607-2

Methnani, L., Aler Tubella, A., Dignum, V., Theodorou, A., Tubella, A. A., Dignum, V., & Theodorou, A. (2021). Let Me Take Over: Variable Autonomy for Meaningful Human Control. *Frontiers in Artificial Intelligence*, *4*, 737072. https://doi.org/10.3389/frai.2021.737072

Methnani, L., Brännström, M., & Theodorou, A. (2023). Operationalising AI Ethics: Conducting Socio-technical Assessment. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *13500 LNAI*, 304–321. https://doi.org/10.1007/978-3-031-24349-3_16

Mikalef, P., Conboy, K., Lundström, J. E., & Popovic, A. (2022). Thinking responsibly about responsible AI and "the dark side" of AI. *European Journal of Information Systems*, *31*(3), 257–268. https://doi.org/10.1080/0960085X.2022.2026621

Minkkinen, M., Niukkanen, A., & Mantymaki, M. (2024). What about investors? ESG analyses as tools for ethics-based AI auditing. *AI & SOCIETY*. https://doi.org/10.1007/s00146-022-01415-0

Minkkinen, M., Zimmer, M. P., & Mäntymäki, M. (2023). Co-Shaping an Ecosystem for Responsible AI: Five Types of Expectation Work in Response to a Technological Frame. *Information Systems Frontiers*, *25*(1), 103–121. https://doi.org/10.1007/s10796-022-10269-2

Morse, J. M. (2016). Tussles, tensions, and resolutions. In Developing grounded theory. *Routledge*, 13–22.

Morse, S. J. (2006). Moral and legal responsibility and the new neuroscience. *Neuroethics: Defining the Issues in Theory, Practice, and Policy*, 33–50.

Müftüoğlu, Z., Kızrak, M. A., & Yıldırım, T. (2022). Privacy-Preserving Mechanisms with Explainability in Assistive AI Technologies. In *Learning and Analytics in Intelligent Systems* (Vol. 28, pp. 287–309). https://doi.org/10.1007/978-3-030-87132-1_13

Murray-Rust, D., Lupetti, M. L., Nicenboim, I., & Hoog, W. V. (2023). Grasping AI: experiential exercises for designers. *AI & SOCIETY*. https://doi.org/10.1007/s00146-023-01794-y

Narayanan, S. (2023). Developing Responsible AI Business Model. In *CSR, Sustainability, Ethics and Governance* (pp. 205–217). https://doi.org/10.1007/978-3-031-09245-9_10

Nedunuri, U. (2023). Factors Influencing the Adoption of Responsible AI. *Proceedings - 2023 10th International Conference on Electrical and Electronics Engineering, ICEEE 2023*, 214–218. https://doi.org/10.1109/ICEEE59925.2023.00046

Oliver, D., & Jacobs, C. (2007). Developing guiding principles: an organizational learning perspective. *Journal of Organizational Change Management*, 813–828.

Ong, L. M., & Findlay, M. (2023). A Realist's Account of AI for SDGs: Power, Inequality and AI in Community. In *Philosophical Studies Series* (Vol. 152, pp. 43–64). https://doi.org/10.1007/978-3-031-21147-8_4

Onyejegbu, L. N., Okengwu, U. A., Oghenekaro, L. U., Musa, M. O., & Ugbari, A. O. (2023). AI-Based QOS/QOE Framework for Multimedia Systems. *Lecture Notes in Networks and Systems*, *559 LNNS*, 248–259. https://doi.org/10.1007/978-3-031-18461-1_16

Öz, B., Wang, R. J., Chandler, C., Karran, A. J., Coursaris, C., & Léger, P. M. (2022). Responsible Artificial Intelligence in Knowledge Work: User Experience Design Problems and Implications. In J. Y. C. Chen, G. Fragomeni, H. Degen, & S. Ntoa (Eds.), *HCI International 2022 - Late Breaking Papers: Interacting With Extended Reality and Artificial Intelligence* (Vol. 13518, Issues 24th International Conference on Human-Computer Interaction (HCII), pp. 461–470). https://doi.org/10.1007/978-3-031-21707-4_32 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., Falco, G., Fiore, S. M., Garibay, I., Grieman, K., Havens, J. C., Jirotka, M., Kacorri, H., Karwowski, W., Kider, J., Konstan, J., Koon, S., Lopez-Gonzalez, M., Maifeld-Carucci, I., & McGregor, S. (2023). Six Human-Centered Artificial Intelligence Grand Challenges. *International Journal of Human-Computer Interaction*, *39*(3), 391–437. http://10.0.4.56/10447318.2022.2153320

Pak, C. (2022). Responsible AI and algorithm governance: An institutional perspective. In *Human-Centered Artificial Intelligence: Research and Applications* (pp. 251–270). https://doi.org/10.1016/B978-0-323-85648-5.00018-9

Papagiannidis, E., Mikalef, P., Krogstie, J., & Conboy, K. (2022). From Responsible AI Governance to Competitive Performance: The Mediating Role of Knowledge Management Capabilities. In S. Papagiannidis, E. Alamanos, S. Gupta, Y. K. Dwivedi, M. Mantymaki, & I. O. Pappas (Eds.), *Role of Digital Technologies in Shaping The Post-Pandemic World* (Vol. 13454, Issues 21st International-Federation-of-Information-Processing (IFIP)-Working-Group-6.11 Conference on e-Business, e-Services and e-Society (I3E), pp. 58–69). https://doi.org/10.1007/978-3-031-15342-6_5 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Pisoni, G., & Molnár, B. (2023). Data Management and Enterprise Architectures for Responsible AI Services. *Lecture Notes in Networks and Systems*, *763 LNNS*, 879–884. https://doi.org/10.1007/978-3-031-42467-0_83

Porayska-Pomsta, K. (2023). A Manifesto for a Pro-Actively Responsible AI in Education. *International Journal of Artificial Intelligence in Education*, 1–11.

Prabhudesai, S., Hauth, J., Guo, D. K., Rao, A. R., Banovic, N., & Huan, X. (2023). Lowering the computational barrier: Partially Bayesian neural networks for transparency in medical imaging AI. *Frontiers in Computer Science*, *5*. https://doi.org/10.3389/fcomp.2023.1071174 WE - Emerging Sources Citation Index (ESCI)

Puaschunder, J. M. (2019). Artificial Intelligence evolution: On the virtue of killing in the artificial age. *Scientia Moralitas-International Journal of Multidisciplinary Research*, *4*(2), 51–72.

Pushkarna, M., Zaldivar, A., & Kjartansson, O. (2022). Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI. *ACM International Conference Proceeding Series*, 1776–1826. https://doi.org/10.1145/3531146.3533231

Raffoul, F. (2010). *The origins of responsibility*. Indiana University Press.

Ramirez-Amaro, K., Torre, I., Diehl, M., & Dean, E. (2023). The Importance of Human Factors for Trusted Human-Robot Collaborations. *ACM International Conference Proceeding Series*, 502–503. https://doi.org/10.1145/3623809.3623981

Rantanen, E. M., Lee, J. D., Darveau, K., Miller, D. B., Intriligator, J., & Sawyer, B. D. (2021). Ethics Education of Human Factors Engineers for Responsible AI Development. *Proceedings of the Human Factors and Ergonomics Society*, *65*(1), 1034–1038. https://doi.org/10.1177/1071181321651038

Reis, S., Coelho, L., Sarmet, M., Araujo, J., & Corchado, J. M. (2023). The Importance of Ethical Reasoning in Next Generation Tech Education. *2023 5th International Conference of the Portuguese Society for Engineering Education, CISPEE 2023*. https://doi.org/10.1109/CISPEE58593.2023.10227651

Rivas, P., Thompson, C., Tafur, B., Khanal, B., Ayoade, O., Jui, T. D., Sooksatra, K., Orduz, J., & Bejarano, G. (2023). AI ethics for earth sciences. In *Artificial Intelligence in Earth Science: Best Practices and Fundamental Challenges* (pp. 379–396). https://doi.org/10.1016/B978-0-323-91737-7.00007-4

Rizinski, M., Peshov, H., Mishev, K., Chitkushev, L. T., Vodenska, I., & Trajanov, D. (2022). Ethically Responsible Machine Learning in Fintech. *IEEE ACCESS*, *10*, 97531–97554. https://doi.org/10.1109/ACCESS.2022.3202889 WE - Science Citation Index Expanded (SCI-EXPANDED)

Robbins, S. (2020). AI and the path to envelopment: knowledge as a first step towards the responsible regulation and use of AI-powered machines. *AI & SOCIETY*, *35*, 391–400.

Rodier, C., Millar, J., Deisinger, W., & Hodgson, S. J. (2023). Art Critically Examining Generative AI. *2023 IEEE IFEES World Engineering Education Forum and Global Engineering Deans Council: Convergence for a Better World: A Call to Action, WEEF-GEDC 2023 - Proceedings*. https://doi.org/10.1109/WEEF-GEDC59520.2023.10343903

Roemmich, K., & Andalibi, N. (2021). Data Subjects' Conceptualizations of and Attitudes Toward Automatic Emotion Recognition-Enabled Wellbeing Interventions on Social Media. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2). https://doi.org/10.1145/3476049

Roy, S., & Salimi, B. (2023). Causal Inference in Data Analysis with Applications to Fairness and Explanations. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *13759 LNCS*, 105–131. https://doi.org/10.1007/978-3-031-31414-8_3

Rudd, J., & Igbrude, C. (2023). A global perspective on data powering responsible AI solutions in health applications. *AI and Ethics*, 1–11. https://doi.org/10.1007/s43681-023-00302-8

Ruttkamp-Bloem, E. (2023). Intergenerational Justice as Driver for Responsible AI. *Communications in Computer and Information Science*, *1976 CCIS*, 18–30. https://doi.org/10.1007/978-3-031-49002-6_2

Sætra, H. S., & Danaher, J. (2022). To each technology its own ethics: The problem of ethical proliferation. *Philosophy & Technology*, *35*(4), 93.

Sanderson, C., Douglas, D., Lu, Q. H., & IEEE. (2023). Implementing Responsible AI: Tensions and Trade-Offs Between Ethics Aspects. *2023 International Joint Conference on Neural Networks, IJCNN*, *2023-June*(International Joint Conference on Neural Networks (IJCNN)). https://doi.org/10.1109/IJCNN54540.2023.10191274 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Sanderson, C., Lu, Q., Douglas, D., Xu, X., Zhu, L., Whittle, J., & Sanderson, C., Lu, Q., Douglas, D., Xu, X., Zhu, L., & Whittle, J. (2022). Towards Implementing Responsible AI. *IEEE International Conference on Big Data (Big Data)*, 5076–5081. https://doi.org/10.1109/BigData55660.2022.10021121

Santoni de Sio, F., & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology*, *34*(4), 1057–1084.

Sartori, L., & Bocca, G. (2022). Minding the gap(s): public perceptions of AI and socio-technical imaginaries. *AI and Society*. https://doi.org/10.1007/s00146-022-01422-1

Satish, M., Babu, S. M., Kumar, P. P., Devi, S., & Reddy, K. P. (2023). Artificial Intelligence (AI) and the Prediction of Climate Change Impacts. *5th IEEE International Conference on Cybernetics, Cognition and Machine Learning Applications, ICCCMLA 2023*, 660–664. https://doi.org/10.1109/ICCCMLA58983.2023.10346636

Satornino, C. B., Du, S., & Grewal, D. (2024). Using artificial intelligence to advance sustainable development in industrial markets: A complex adaptive systems perspective. *Industrial Marketing Management*, *116*, 145–157. https://doi.org/https://doi.org/10.1016/j.indmarman.2023.11.011

Saxena, D., Wall, P. J., Lewis, D., & IEEE. (2023). Artificial Intelligence (AI) Ethics: A Critical Realist Emancipatory Approach. *2023 IEEE International Symposium on Technology and Society, ISTAS, 29th Annual IEEE International Symposium on Technology and Society (ISTAS)*. https://doi.org/10.1109/ISTAS57930.2023.10305995 WE - Conference Proceedings Citation Index - Science (CPCI-S) WE - Conference Proceedings Citation Index - Social Science & Humanities (CPCI-SSH)

Scepanovic, S., Bogucka, E. P., Quercia, D., & Nattero, C. (2023). Responsible AI for Earth Observation: Attitides among Experts. *International Geoscience and Remote Sensing Symposium (IGARSS)*, *2023-July*, 1934–1936. https://doi.org/10.1109/IGARSS52108.2023.10282983

Schiff, D., Rakova, B., Ayesh, A., Fanti, A., & Lennon, M. (2020). Principles to practices for responsible AI: closing the gap. *ArXiv Preprint*.

Shneiderman, B. (2021). Responsible AI: Bridging from ethics to practice. *Communications of the ACM*, *64*(8), 32–35.

Siala, H., & Wang, Y. C. (2022). SHIFTing artificial intelligence to be responsible in healthcare: A systematic review. *Social Science & Medicine*, *296*. https://doi.org/10.1016/j.socscimed.2022.114782 WE - Science Citation Index Expanded (SCI-EXPANDED) WE - Social Science Citation Index (SSCI)

Simon, H. A. (1995). Artificial intelligence: an empirical science. *Artificial Intelligence*, *77*(1), 95–127.

Sinde, R., Diwani, S., Leo, J., Kondo, T., Elisa, N., & Matogoro, J. (2023). AI for Anglophone Africa: Unlocking its adoption for responsible solutions in academia-private sector. *Frontiers in Artificial Intelligence*, *6*, 1133677. https://doi.org/10.3389/frai.2023.1133677

Sivarajah, U., Wang, Y. C., Olya, H., & Mathew, S. (2023). Responsible Artificial Intelligence (AI) for Digital Health and Medical Analytics. *Information Systems Frontiers*, *25*(6), 2117–2122. https://doi.org/10.1007/s10796-023-10412-7

Sothilingam, R. (2023). A Requirements-Driven Conceptual Modeling Framework for Responsible AI. *Proceedings of the IEEE International Conference on Requirements Engineering*, *2023-Septe*, 391–395. https://doi.org/10.1109/RE57278.2023.00061

Souza, R., Skluzacek, T. J., Wilkinson, S. R., Ziatdinov, M., & Da Silva, R. F. (2023). Towards Lightweight Data Integration Using Multi-Workflow Provenance and Data Observability. *Proceedings 2023 IEEE 19th*

International Conference on E-Science, e-Science 2023. https://doi.org/10.1109/e-Science58273.2023.10254822

STAHL, B. C. (2022). Responsible innovation ecosystems: Ethical implications of the application of the ecosystem concept to artificial intelligence. *International Journal of Information Management*, *62*, N.PAG-N.PAG. https://doi.org/10.1016/j.ijinfomgt.2021.102441

Stough, L. M., & Lee, S. (2021). Grounded theory approaches used in educational research journals. *International Journal of Qualitative Methods*, *20*, 16094069211052204.

Szabo, L., Raisi-Estabragh, Z., Salih, A., McCracken, C., Pujadas, E. R., Gkontra, P., Kiss, M., Maurovich-Horvath, P., Vago, H., Merkely, B., Lee, A. M., Lekadir, K., & Petersen, S. E. (2022). Clinician's guide to trustworthy and responsible artificial intelligence in cardiovascular imaging. *Frontiers in Cardiovascular Medicine*, *9*. https://doi.org/10.3389/fcvm.2022.1016032 WE - Science Citation Index Expanded (SCI-EXPANDED)

Szczekocka, E., Tarnec, C., & Pieczerak, J. (2022). Standardization on Bias in Artificial Intelligence as Industry Support. *Proceedings - 2022 IEEE International Conference on Big Data, Big Data 2022*, 5090–5099. https://doi.org/10.1109/BigData55660.2022.10020735

Tahaei, M., Constantinides, M., Quercia, D., Kennedy, S., Muller, M., Stumpf, S., Liao, Q. V, Baeza-Yates, R., Aroyo, L., Holbrook, J., Luger, E., Madaio, M., Blumenfeld, I. G., De-Arteaga, M., Vitak, J., & Olteanu, A. (2023). Human-Centered Responsible Artificial Intelligence: Current & Future Trends. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3544549.3583178

Tayebi, A., & Garibay, O. O. (2023). Improving Fairness via Deep Ensemble Framework Using Preprocessing Interventions. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *14050 LNAI*, 477–489. https://doi.org/10.1007/978-3-031-35891-3_29

Teixeira, S., Veloso, B., Rodrigues, J. C., & Gama, J. (2023). Ethical and Technological AI Risks Classification: A Human Vs Machine Approach. *Communications in Computer and Information Science*, *1752 CCIS*, 150–166. https://doi.org/10.1007/978-3-031-23618-1_10

Thornberg, R. (2012). Informed grounded theory. *Scandinavian Journal of Educational Research*, *56*(3), 243–259.

Timmermans, S., & Tavory, I. (2007). Advancing ethnographic research through grounded theory practice. *The Sage Handbook of Grounded Theory*, 493–512.

Trocin, C., Mikalef, P., Papamitsiou, Z., & Conboy, K. (2023). Responsible AI for Digital Health: a Synthesis and a Research Agenda. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-021-10146-4

Tutun, S., Harfouche, A., Albizri, A., Johnson, M. E., & He, H. Y. (2022). A Responsible AI Framework for Mitigating the Ramifications of the Organ Donation Crisis. *Information Systems Frontiers*, *25*(6), 2301–2316. https://doi.org/10.1007/s10796-022-10340-y

Upreti, K., Verma, A., Mittal, S., Vats, P., Haque, M., & Ali, S. (2023). A Novel Framework for Harnessing AI for Evidence-Based Policymaking in E-Governance Using Smart Contracts. *Communications in Computer and Information Science*, *1921 CCIS*, 231–240. https://doi.org/10.1007/978-3-031-45124-9_18

Wach, K., Duong, C. D., Ejdys, J., Kazlauskaite, R., Korzynski, P., Mazurek, G., Paliszkiewicz, J., & Ziemba, E. (2023). The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. *Entrepreneurial Business and Economics Review*, *11*(2), 7–30. https://doi.org/10.15678/EBER.2023.110201 WE - Emerging Sources Citation Index (ESCI)

Walsh, P., Bera, J., Sharma, V. S., Kaulgud, V., Rao, R. M., Ross, O., & Soc, I. C. (2021). Sustainable AI in the Cloud Exploring Machine Learning Energy Use in the Cloud. In *2021 36TH IEEE/ACM International Conference on Automated Software Engineering Workshops (ASEW 2021)* (Issue 36th IEEE/ACM International Conference on Automated Software Engineering (ASE), pp. 265–266). https://doi.org/10.1109/ASEW52652.2021.00058 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Wang, W. S., Chen, L., Xiong, M. R., & Wang, Y. C. (2021). Accelerating AI Adoption with Responsible AI Signals and Employee Engagement Mechanisms in Health Care. *Information Systems Frontiers*, *25*(6), 2239–2256. https://doi.org/10.1007/s10796-021-10154-4

Wang, Y., Xiong, M., & Olya, H. G. T. (2020). Toward an understanding of responsible artificial intelligence practices. *Proceedings of the Annual Hawaii International Conference on System Sciences*, *2020-Janua*, 4962–4971. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85087789376&partnerID=40&md5=65b91535aa6d5c0afb18152435cbd358

Weekes, T. R., & Eskridge, T. C. (2022). Responsible Human-Centered Artificial Intelligence for the Cognitive Enhancement of Knowledge Workers. In J. Y. C. Chen, G. Fragomeni, H. Degen, & S. Ntoa (Eds.), *HCI International 2022 - Late Breaking Papers: Interacting With Extended Reality and Artificial Intelligence* (Vol. 13518, Issues 24th International Conference on Human-Computer Interaction (HCII), pp. 568–582). https://doi.org/10.1007/978-3-031-21707-4_41 WE - Conference Proceedings Citation Index - Science (CPCI-S)

Weiss, A., Vrecar, R., Zamiechowska, J., & Purgathofer, P. (2023). It's Only a Bot! How Adversarial Chatbots can be a Vehicle to Teach Responsible AI. In *CSR, Sustainability, Ethics and Governance* (pp. 235–250). https://doi.org/10.1007/978-3-031-09245-9_12

Wellnhofer, E. (2022). Real-World and Regulatory Perspectives of Artificial Intelligence in Cardiovascular Imaging. *Frontiers in Cardiovascular Medicine*, *9*, 890809. https://doi.org/10.3389/fcvm.2022.890809

Wolfswinkel, J. F., Furtmueller, E., & Wilderom, C. P. (2013). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems*, *22*(1), 45–55.

Yang, Q., Liu, Y., Cheng, Y., Kang, Y., Chen, T., & Yu, H. (2020). Federated Learning. In *Synthesis Lectures on Artificial Intelligence and Machine Learning* (Vol. 13, Issue 3, pp. 1–207). https://doi.org/10.2200/S00960ED2V01Y201910AIM043

Yudkowsky, E. (2008). Artificial intelligence as a positive and negative factor in global risk. *Global Catastrophic Risks*, *1*(303), 184.

Authors' Note

All correspondence should be addressed to
Bahar Memarian
Simon Fraser University
Canada

Human Technology
ISSN 1795-6889
https://ht.csr-pub.eu

Appendix

Table 8
Overview of the reviewed studies paradigm, AI use, and sector

Area	Reference	Area: 1=business, 2=engineering, 3=governance, 4=health, 5=science, 6=social views and education)	Type: C=Conf, J=Journal, B=Book, BC=Book Chapter, LN=Lecture Notes, PP=Preprint	Principles: 1=empirical, 2=social	Design research: 1=quant, 2=qual, 3=mixed	Challenge: 1=empirical, 2=social, 3=both	Findings: 1=descriptive, 2=diagnostic, 3=predictive, 4=prescriptive
2	(Abeywickrama et al. , 2019)	2	J	2	3	2	2
5	(Agarwal & Mishra, 2021)	5	B	3	2	1	1
5	(Akbarighatar et al. , 2023)	5	J	2	2	3	1
5	(Al Hashlamoun et al. , 2023)	5	C	2	2	2	1
6	(Alam, 2023)	6	J	1	1	1	1
4	(Al-Dhaen et al., 2021)	4	J	3	2	2	3
6	(Aler Tubella et al. , 2024)	6	C	1	1	1	2
2	(Alfrink et al. , 2023)	2	C	3	3	2	2
4	(Al-Hwsali et al., 2023)	4	J	2	2	3	1
5	(Alibašić, 2023)	5	J	1	2	2	1
6	(Anagnostou et al. , 2022)	6	C	1	1	3	3
6	(Ansari et al. , 2021)	6	J	1	2	1	2
3	(Arora et al. , 2024)	3	LN	1	3	3	1
5	(Arya et al. , 2023)	5	C	1	1	1	1
1	(Asif et al. , 2023)	1	J	3	1	1	1
5	(Ayoubi et al. , 2023)	5	C	3	2	3	1
2	(Öz et al. , 2022)	2	J	2	2	3	1
3	(Azafrani & Gupta, 2023)	3	J	3	2	2	1
5	Baeza-Yates & Villoslada, 2022)	5	C	2	2	2	2
5	(Bani Ahmad, 2024)	5	J	1	2	2	1
2	(Bao et al. , 2023)	2	J	2	3	3	2
6	(Barletta et al. , 2023)	6	J	1	2	2	1
5	(Belle, 2019)	5	J	2	2	3	1
5	(Bhattacharyya et al., 2023)	5	C	1	1	1	1
5	(Blockeel et al., 2023)	5	J	1	3	3	1
4	(Borda et al. , 2022)	4	J	2	2	2	1

3	(Boulanin & Lewis, 2023)	3	J	2	2	2	2
3	(Brand, 2022)v	3	J	2	2	3	3
6	(Brumen et al. , 2023)	6	J	1	2	3	4
6	(Buchholz et al. , 2022)	6	J	1	3	1	2
6	(Buijsman, 2023)	6	C	1	3	2	2
6	(Cachat-Rosset & Klarsfeld, 2023)	6	C	2	1	1	1
4	(Cai et al. , 2023)	4	C	1	1	1	2
4	(Calvo et al., 2023)	4	C	1	1	1	2
6	(Calvo et al. , 2020)	6	J	2	1	2	1
6	(Chang, Y. L. , & Ke, 2023)	6	J	2	2	1	1
4	(Charalampakos et al. , 2023)	4	C	1	1	1	2
6	(H. Chen et al. , 2023)	6	C	2	2	1	2
4	(Cheng et al. , 2021)	4	J	1	1	3	1
5	(Christos et al. , 2020)	5	C	1	1	1	4
5	(Church et al., 2022)	5	J	2	2	3	1
6	(Constantinescu et al. , 2021)	6	C	2	2	1	4
6	(Contractor et al., 2022)	6	J	2	2	2	1
1	(Crockett et al., 2023)	1	J	2	3	3	1
2	(De Brito Duarte, 2023)	2	J	2	1	3	4
5	(De et al. , 2023)	5	PP	3	1	1	1
2	(Delecraz et al., 2022)	2	J	3	1	1	3
6	(Dennehy et al. , 2023)	6	J	2	2	2	1
5	(Deshmukh & Srinivasa, 2022)	5	J	2	1	1	1
6	(Dholakia et al. , 2023)	6	C	2	2	2	1
6	(Domínguez Figaredo et al. , 2023)	6	C	2	2	2	1
2	(Dominique et al. , 2023)	2	C	3	3	2	2
6	(Dong et al. , 2023)	6	J	2	2	2	1
4	(El-Haddadeh et al. , 2021)	4	J	3	3	3	2
2	(Faßbender, 2021)	2	C	1	2	2	1
2	(Falco, 2019)	2	C	2	2	2	1
6	(Farina, 2022)	6	J	2	2	2	1
1	(Fenwick et al. , 2018)	1	BC	2	2	3	3
1	(Ferreira et al. , 2023)	1	C	2	3	3	1
4	(Fosso Wamba et al. , 2021)	4	J	3	3	3	1
5	(Fritz-Morgenthal et al. , 2022)	5	J	3	2	2	4
4	(GAON & STEDMAN, 2019)	4	LN	3	2	3	4
4	(Garg, 2023)	4	J	3	1	3	3
5	(Feher, 2021)	5	B	3	3	3	1
2	(Gavrilova, 2023)	2	C	2	2	3	1
3	(Gianni et al. , 2022)	3	J	2	2	2	1
4	(Gilbert et al. , 2023)	4	LN	1	1	1	2
4	(Golbin et al. , 2020)	4	C	2	2	3	1

6	(Grammatikos et al., 2023)	6	J	2	2	2	1
4	(Gupta et al., 2021)	4	J	3	2	3	1
6	(Gwagwa et al., 2021)	6	C	2	2	2	1
3	(Hacker & Passoth, 2022)	3	J	3	2	3	4
1	(Hadley, 2022)	1	C	2	2	2	3
1	(Haidar, 2024)	1	J	1	2	3	4
5	(Harfouche et al., 2024)	5	J	3	2	3	1
6	(Hernández & Galanos, 2022)	6	C	2	2	2	1
5	(Herrmann, 2023)	5	J	3	2	2	1
5	(Heuillet et al., 2022)	5	C	1	1	1	3
2	(Hu, 2023)	2	C	1	3	1	2
6	(Inuwa-Dutse, 2023)	6	J	2	2	2	2
4	(Islam et al., 2023)	4	C	3	3	1	1
4	(Iwasawa et al., 2023)	4	J	3	1	2	1
4	(Jörg et al., 2023)	4	J	2	2	2	1
5	(Jaipuria et al., 2022)	5	C	1	1	1	1
4	(Jawad et al., 2021)	4	C	2	2	2	1
5	(Ji et al., 2023)	5	C	1	1	2	1
4	(Johnson et al., 2021)	4	J	3	1	1	1
2	(Kamada, 2016)	2	C	2	1	2	2
6	(Kapania et al., 2022)	6	J	2	2	2	3
5	(Kapania et al., 2023)	5	C	3	2	3	1
4	(Karagianni et al., 2023)	4	C	1	1	3	4
6	(Karpagam et al., 2022)	6	J	2	2	2	3
4	(Knopp et al., 2023)	4	J	2	2	2	4
3	(Krafft et al., 2021)	3	C	2	3	3	1
3	(Krarup & Horst, 2023)	3	J	3	3	2	2
2	(Kuck, 2023)	2	C	3	1	1	1
6	(Kuennen, 2023)	6	J	2	2	2	4
5	(K. P. Kumar et al., 2023)	5	BC	3	1	1	1
4	(P. Kumar et al., 2021)	4	J	1	3	2	1
2	(Lee et al., 2024)	2	J	1	3	3	1
2	(Lescano et al., 2023)	2	J	1	1	3	2
2	(Lewis et al., 2022)	2	J	3	1	1	3
3	(C. Li & Yang, 2021)	3	C	2	2	3	1
4	(K. Li et al., 2023)	4	C	3	2	3	1
5	(Q. V Liao et al., 2022)	5	C	3	3	3	1
4	(W. Liao et al., 2023)	4	J	1	1	2	2
5	(Lin, 2024)	5	J	3	2	1	1
5	(J. Liu et al., 2023)	5	C	1	3	3	1
4	(R. Liu et al., 2021)s	4	J	2	2	2	1

6	(Lo et al., 2023)	6	J	2	2	2	4
2	(Lu et al., 2023)	2	J	2	2	3	2
2	(Lu et al., 2022)	2	C	2	2	2	1
6	(Lukkien et al., 2023)	6	C	2	2	3	1
4	(Müftüoğlu et al., 2022)	4	J	2	3	3	1
5	(Mahoney et al., 2022)	5	C	1	1	2	4
1	(Maki et al., 2023)	1	J	3	2	2	1
6	(Mallinger & Baeza-Yates, 2024)	6	J	2	2	3	1
6	(Man Li & Crabbe, 2022)	6	J	2	2	3	1
1	(Marcoux & Martineau, 2022)	1	J	2	2	2	1
1	(Maree et al., 2020)	1	C	1	1	1	1
6	(Maruyama, 2022)	6	J	2	2	3	2
4	(Marvin et al., 2022)	4	C	1	1	2	2
6	(Marzouk et al., 2023)	6	J	2	2	3	2
3	(Meerveld et al., 2023)	3	J	2	2	2	2
5	(Merhi, 2023)	5	J	3	2	2	2
5	(Meske et al., 2022)	5	J	3	2	3	1
6	(sMethnani et al., 2023)	6	J	2	2	3	4
6	(Methnani et al., 2021)	6	PP	2	3	2	1
6	(Mikalef et al., 2022)	6	C	2	3	2	1
6	(Minkkinen et al., 2024)	6	J	2	3	2	1
6	(Minkkinen et al., 2023)	6	C	2	3	2	2
6	(Murray-Rust et al., 2023)	6	J	2	3	3	1
1	(Narayanan, 2023)	1	J	2	2	1	1
6	(Nedunuri, 2023)	6	J	2	3	3	4
6	(Ong & Findlay, 2023)	6	J	3	1	1	3
2	(Onyejegbu et al., 2023)	2	J	1	1	1	2
3	(Pak, 2022)	3	BC	1	1	1	1
1	(E. Papagiannidis et al., 2022)	1	J	1	1	1	3
5	(Pisoni & Molnár, 2023)	5	J	1	2	1	1
6	(Porayska-Pomsta, 2023)	6	J	3	2	1	1
4	(Prabhudesai et al., 2023)	4	J	1	1	1	2
5	(Pushkarna et al., 2022)	5	J	3	3	1	1
2	(Ramirez-Amaro et al., 2023)	2	J	3	3	1	3
5	(Rantanen et al., 2021)	5	J	3	2	2	1
6	(Reis et al., 2023)	6	J	3	2	2	2
5	(Rivas et al., 2023)	5	BC	3	2	3	1
5	(Rizinski et al., 2022)	5	J	1	3	3	3
6	(Rodier et al., 2023)	6	J	3	2	2	1
4	(Roemmich & Andalibi, 2021)	4	J	3	1	2	1
5	(Roy & Salimi, 2023)	5	BC	2	2	1	1

4	(Rudd & Igbrude, 2023)	4	J	2	2	3	1
6	(Ruttkamp-Bloem, 2023)	6	J	3	2	2	2
6	(Sanderson et al., 2023)	6	J	3	2	2	2
6	(Sanderson et al., 2022)	6	C	3	2	2	2
5	(Satish et al., 2023)	5	C	1	3	1	1
6	(Satornino et al., 2024)	6	J	3	2	2	4
6	(Saxena et al., 2023)	6	J	3	2	3	1
6	(Scepanovic et al., 2023)	6	J	3	2	3	1
6	(Sinde et al., 2023)	6	C	3	2	3	1
4	(Sivarajah et al., 2023)	4	J	1	2	2	4
5	(Sothilingam, 2023)	5	C	1	2	3	1
5	(Souza et al., 2023)	5	C	3	3	1	4
6	(STAHL, 2022)	6	C	3	2	3	1
4	(Szabo et al., 2022)	4	J	3	2	2	1
6	(Szczekocka et al., 2022)	6	J	3	3	1	1
6	(Tayebi & Garibay, 2023)	6	B	3	3	3	1
6	(Teixeira et al., 2023)	6	PP	2	2	3	1
4	(Trocin et al., 2023)	4	J	3	3	2	1
5	(Tutun et al., 2022)	5	J	3	1	2	1
3	(Upreti et al., 2023)	3	J	3	3	3	4
1	(Wach et al., 2023)	1	J	2	2	3	1
5	(Walsh et al., 2021)	5	C	1	1	1	3
4	(W. S. Wang et al., 2021)	4	J	3	3	3	1
4	(Weekes & Eskridge, 2022)	4	BC	2	3	1	1
5	(Weiss et al., 2023)	5	BC	2	2	2	1
4	(Wellnhofer, 2022)	4	J	3	2	2	1
6	(Yang et al., 2020)	6	J	2	3	2	1