International Journal of Information Management Data Insights 3 (2023) 100193

Contents lists available at ScienceDirect

International Journal of Information Management Data Insights

journal homepage: www.elsevier.com/locate/jjimei

A sociotechnical perspective for responsible AI maturity models: Findings from a mixed-method literature review

Pouria Akbarighatar [a,*], Ilias Pappas [a,b], Polyxeni Vassilakopoulou [a]

[a] Department of Information Systems, University of Agder, Universitetsveien 25, 4604, Kristiansand, Norway
[b] Department of Computer Science, Norwegian University of Science and Technology, Sem Sælandsvei 9, 7491, Trondheim, Norway

ARTICLE INFO

Keywords:
Artificial intelligence
Maturity model
Responsible AI capabilities
Topic modeling
Sociotechnical

ABSTRACT

As artificial intelligence (AI) is increasingly used in various industries, it becomes crucial for organizations to enhance their capabilities and maturity in adopting AI responsibly. This paper employs a mixed-method approach that combines topic modeling with manual content analysis to provide a comprehensive review of the literature on AI maturity and readiness. The review encompasses an extensive corpus of 1451 papers, identifying the main themes and topics within this body of literature. Based on these findings, a subset of papers was selected and further analyzed to identify AI capabilities utilizing a sociotechnical lens. This further analysis led to the identification of foundational and responsible AI (RAI) capabilities. These capabilities have been integrated in a sociotechnical framework of capabilities for AI maturity models providing valuable insights for organizations and AI service providers and a basis for further research.

1. Introduction

Far-reaching benefits of artificial intelligence (AI) applications in public and private organizations ranging from healthcare, transportation, manufacturing, finance, and education have become evident and undeniable (Benbya et al., 2021; Berente et al., 2021; Meadows et al., 2022). As AI becomes widespread and influences individual's lives and organizations' operations, this new ubiquitous technology's complex social and technical aspects must be comprehended and addressed (Benbya et al., 2021; Berente et al., 2021; Dwivedi et al., 2021). The potential misuse and unintentional consequences of technological development have spawned debates and warnings about ethical issues and social challenges related to AI (Floridi et al., 2018; Mikalef, et al., 2022). Using AI systems responsibly can facilitate a path toward creating sustainable societies (Pappas et al., 2018), while considering the potential adverse outcomes of AI use (Dignum, 2019; Duan et al., 2019; Vassilakopoulou, 2020). Therefore, it is beneficial to focus on developing frameworks to guide the responsible development and use of AI technology. In recent years, the literature that deals with various principles for responsible AI (RAI), including explainability, fairness, safety, reliability, transparency, and accountability, has advanced considerably (Dignum, 2019; Osoba & Welser, 2017; Kempton & Vassilakopoulou, 2021). Furthermore, high-tech organizations, such as Google, Microsoft, and IBM, have proposed sets of RAI principles to provide effective insight regarding the ethical or unintentional consequences (Google, 2022; IBM, 2021; Microsoft, 2020). These principles aim to guide the ethical development and deployment of AI technology.

The operationalization of these principles is challenging, as it is not easy to practically implement them in complex real-world environments (Lauer, 2021). Criticism has been raised toward the high level of abstraction and ambiguity that hinder practical implementation and the deployment of control mechanisms and governance arrangements (Krijger et al., 2022; Fukas et al., 2021). Hence, the adoption of principles may end up being fragmented and isolated, rather than a systemic and interconnected approach (Akbarighatar et al., 2023a). Furthermore, the principles are considered insufficient in effectively addressing the whole spectrum of potential negative consequences of AI (Munn, 2022). A recent report by the consulting company Accenture revealed that 94% of technology and service providers require assistance operationalizing responsible AI principles (Accenture, 2022). AI practitioners, industry experts, and academics in the field emphasize the urgent need to bridge this gap and develop practical RAI implementation (Munn, 2022; Lauer, 2021).

Practical RAI implementation entails attending to both humanistic and technical-business aspects. However, although the technical and business aspects of AI have received considerable attention from

* Corresponding author at: Department of Information Systems, University of Agder, Universitetsveien 25, 4604, Kristiansand, Norway.
E-mail address: Pouriaa@uia.no (P. Akbarighatar).

https://doi.org/10.1016/j.jjimei.2023.100193

Available online 16 August 2023

organizations, humanistic objectives and the interplay between technical and social structures are often overlooked (Akbarighatar et al., 2023a; Asatiani et al., 2021). Organizations tend to prioritize technical and business assessments, which are easier to perform during the development and deployment of AI systems (Asatiani et al., 2021; Krijger et al., 2022); however, this approach neglects the importance of considering the ethical, societal, and human impacts of AI (Zimmer et al., 2022). Organizations looking to implement RAI principles or willing to commit to operationalizing principles in their governance practices must recognize the importance of both technical and social aspects (Vassilakopoulou et al., 2022). By doing so, organizations can ensure that the deployment and development of AI systems align with their overall objectives while also upholding ethical and societal values. Taking a sociotechnical perspective can assist organizations in enhancing their AI maturity by recognizing the interplay between technical and social components and considering ethical values in addition to technical-business objectives (Fukas et al., 2021; Yablonsky, 2021).

One approach for moving from RAI principles to practice ensuring a balance between ethical values and technical-business objectives is to develop sociotechnical maturity models and frameworks. These can be developed across an axis of cohesion spanning humanistic and instrumental objectives as suggested by Sarker et al. (2019). Organizations can use these to assess their current status and improve their AI-related capabilities (Someh et al., 2022; Jöhnk et al., 2021). Maturity models (MMs) were initially introduced in strategic management to enable organizations to identify areas for improvement and develop roadmaps for achieving their strategic goals. However, they have since been applied across different domains, including business processes, digital transformations, and AI (Akbarighatar, 2022; Hujran et al., 2021; Krijger et al., 2022). They offer a systematic approach for assessing an organization's current state and for identifying the steps required to improve. In this paper we review the RAI maturity and readiness literature and use the findings for developing a cohesive framework that brings together humanistic and instrumental objectives (Sarker et al., 2019).

While there are various ways to perform a literature review, most of these start with researchers reading and coding the literature. However, using topic modeling techniques as a first step is a good method for obtaining an overview of the themes within extensive corpuses of literature (Kotsialos & Vassilakopoulou, 2023). Topic modeling techniques refer to text mining, which captures key concepts, trends, and hidden relationships (Kumar et al., 2021). This technique has been used in past years across various domains, for instance, for social media analysis by similar post content (Ahn et al., 2021; Rajendran & Sundarraj, 2021; Abramova et al., 2022) and for academic exploratory reviews (Asmussen & Møller, 2019; Zarindast et al., 2021; Larsen et al., 2019). For example, Asmussen and Møller (2019) proposed a framework for employing topic modeling for an exploratory literature review. They argued that the proposed framework could provide better reliability than conventional approaches and could facilitate covering large volumes of literature. Given the above, in this study, we used topic modeling as a first step and then the traditional literature review process to ensure comprehensiveness and integrity in literature analysis. This approach is particularly well-suited for a rapidly evolving and complex field like AI, with a growing research volume. The three overarching research questions for this review are:

RQ1) What are the main research themes regarding maturity and readiness for AI?
RQ2) Which of these research themes specifically relate to RAI?
RQ3) What capabilities are required to ensure the responsible use of AI to achieve humanistic and instrumental objectives?

The current study contributes to the literature in three ways. Firstly, it provides a systematic mapping of research themes and topics that researchers in this domain have previously explored. It presents a comprehensive, reliable, and quick overview of the existing research landscape. Secondly, it suggests a classification of AI-related capabilities to foundational and RAI, adopting a sociotechnical perspective by relating information to humanistic and instrumental objectives. The study sheds light on the importance of considering these capabilities in designing and developing AI systems. By identifying the capabilities, this study offers a direction for organizations to design and develop AI systems that align with their objectives and values while ensuring that they meet the ethical and social responsibilities associated with their use. Finally, these capabilities have been integrated into a sociotechnical framework of capabilities that can be used for developing a comprehensive AI maturity model and thereby providing a sound basis for further research.

In the remaining sections, we first present related literature on sociotechnical systems, responsible AI principles, and maturity models. We then elucidate the three phases of the research methodology. Subsequently, we present the results of topic modeling and content analysis. Finally, we discuss the results providing an integrative sociotechnical view and conclude by pointing to the limitations of this research and further research directions.

2. Background

2.1. Artificial intelligence through a sociotechnical perspective

In managing AI, technical components, which focus on technological and work activities, and social components, including human and organizational aspects, can be observed. The interaction between machines, humans, and organizational environments is essential to consider in the context of AI-based solutions. Sociotechnical system theory is a way of thinking developed to solve organizational problems through social-oriented approaches (stemming from psychological, sociological, and organizational disciplines) and technical-oriented approaches (such as computer science) (Beath et al., 2013). Information systems (IS) research has a long tradition of solving sociotechnical problems in different domains, such as algorithmic fairness and technostress (Dolata et al., 2022; Tarafdar et al., 2019). As AI becomes implemented in organizations, many technical and social challenges emerge at the organizational level. Fig. 1 presents the essence of the sociotechnical perspective as captured by Sarker and colleagues (Sarker et al., 2019). The figure provides the conceptual basis for our literature review. Sarker and colleagues make a distinction between technical and social components in sociotechnical systems. Technical components are human-created tools that achieve predefined goals and include technological and task-related aspects (Lee et al., 2015; Ryan et al., 2002; Sarker et al., 2019). Social components refer to relationships between individuals or social collectives and include organizational structure and actors (Lee et al., 2015; Ryan et al., 2002; Sarker et al., 2019). Building an AI-based system entails engaging with strategic, social, organizational, and technical cores (Beath et al., 2013; Berente et al., 2021). Hence, AI systems can be viewed as sociotechnical systems composed of interacting social and technical components.

Collectively, following a sociotechnical perspective entails acknowledging the need for a fit/harmony/joint optimization between social and technological components (Mumford, 2006). Asatiani and colleagues introduced the concept of sociotechnical envelopment of AI models within organizations and illustrated how the interactions between social and technical factors enable organizations to find harmony between efficiency (accuracy of the AI models) and expansibility and safety (Asatiani et al., 2021). Moreover, Zimmer and colleagues define responsible AI as a sociotechnical system in which both technical and social entities are required to be responsible (Zimmer et al. 2022). These considerations also show that a balance and harmony between the technical and social components are desired to bring about better instrumental and humanistic objectives. Those supporting a sociotechnical approach maintain that AI development and adoption entail

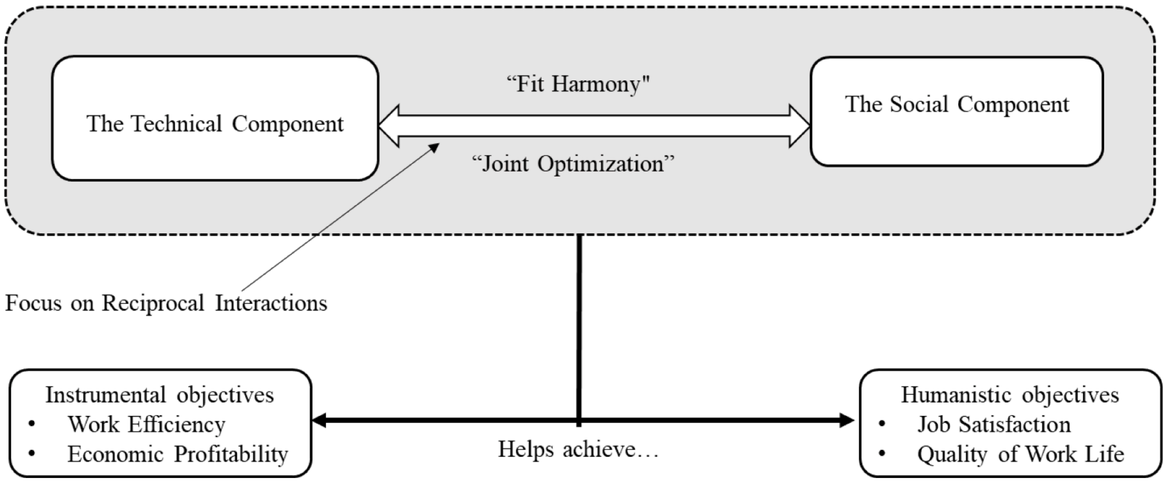


Fig. 1. A sociotechnical perspective on Information Systems, Source: (Sarker et al., 2019).

considerable interactions and coordination between AI artifacts, such as hardware, ML models, and social settings like cultural, economic, and psychological contexts.

This paper investigates the RAI capabilities that can facilitate instrumental and humanistic objectives, each of which has social and technical components. This categorization will serve as the foundation for analyzing the literature on AI maturity and readiness. A sociotechnical perspective in developing maturity models can be beneficial, because it allows organizations to consider potential technical and social issues and requirements in achieving both humanistic and instrumental objectives.

2.2. *Responsible AI principles*

To effectively manage the potential ethical challenges of AI, researchers, governmental agencies, and business organizations have released guidelines, frameworks, and principles for developing and implementing AI technology in a responsible way. These respond to increasing concerns about the potentially adverse impact of AI. For instance, Floridi and colleagues proposed five key principles, including beneficence, non-maleficence, autonomy, justice/fairness, and accountability (Floridi et al., 2018). Their fundamental proposition is that no technology should be harmful to humankind, and to do this, organizations should respect the principles suggested. Another example involves Google's principles, which include four core requirements: "Be socially beneficial", "Avoid creating or reinforcing AI bias", "Be built and tested for safety", and "Be accountable to people" (Google, 2022). Microsoft is another high-tech company that defines principles to ensure that AI systems are developed and deployed responsibly. These principles are privacy and security, transparency, fairness, reliability and safety, accountability, and inclusiveness (Microsoft, 2020). Liu and colleagues summarized the principles from three different RAI frameworks to help build a better understanding of the diversity in RAI dimensions (Liu et al., 2021). Their principles include fairness, inclusiveness, reliability-safety, transparency, privacy-security, beneficence, non-maleficence, and autonomy.

However, the usability of these frameworks and principles has been debated, pointing to challenges associated with operationalizing them and introducing them in practice (Akbarighatar et al., 2023b; Zimmer et al., 2022). Vakkuri and colleagues conducted two empirical studies in 2019 and 2020, highlighting a significant gap between the theoretical research on AI ethics and the practical implementation of ethical principles in AI development (Vakkuri et al., 2019; Vakkuri et al. 2020). Similarly, Munn argued that ethical principles for AI could be more effective, pointing out a gap between abstract principles and their actual implementation in technology (Munn 2022). These studies collectively highlight the ongoing challenge of bridging the divide between theoretical frameworks and principles and their practical implementation in the field of AI. The debates surrounding usability and implementation shed light on the complexities involved in ensuring that ethical considerations are effectively integrated into AI systems.

2.3. *Maturity and readiness models*

A maturity model is a structured mechanism for assessing the current effectiveness of capabilities and ongoing progress in a particular domain (Becker et al., 2009; Gold et al., 2001). It provides a systematic approach for evaluating the organization's level of maturity and offers guidance for further enhancing its capabilities (Becker et al., 2009). The capabilities are described in levels or stages of maturity, starting from low maturity and progressing to higher levels. There are two approaches to constructing maturity models. When using a top-down approach (Becker et al., 2009), a fixed number of maturity stages or levels are established, and then conditions and criteria are created to construct the model. With a bottom-up approach (Lehmkuhl et al., 2013), the starting point is a set of distinct features or factors categorized into capabilities. The bottom-up approach is more common in well-established domains (De Bruin et al., 2005). In this approach, the maturity levels are defined in a second step (Lasrado et al., 2015). Similarly, maturity models, readiness models, and frameworks have been suggested in information systems (IS) research as tools for assessing the organizational state of preparation for successful technology adoption (Molla & Licker, 2005). Overall, maturity models and readiness frameworks include characteristics and guidelines to enable organizations to contemplate context-related considerations and the facilities technology adoption process (Becker et al., 2009).

AI is not an easy-to-use or easy-to-deploy technology compared to other digital technologies. Before implementation and during operations, technical and humanistic challenges arise, and organizations must be prepared against these by fostering AI maturity and readiness (Jöhnk et al., 2021; Lokuge et al., 2019). Maturity and readiness assessment tools can assist organizations in identifying gaps in their AI capabilities, accelerating the adoption process, and ensuring that they are progressing toward their AI goals (Alshawi, 2007; Molla et al., 2009). However, there are few research papers on responsible AI maturity and readiness that take a sociotechnical perspective (Akbarighatar, 2022; Pumplun et al., 2020). For instance, (Krijger et al., 2022) proposed a maturity framework aimed at operationalizing AI ethics and applying a variety of ethical principles within organizational contexts. This framework consists of six dimensions, including awareness and culture, policy, tooling, development process, communication and training, and governance, as well as four levels. As the authors acknowledged in this study, the framework still needs to be broadly validated.

3. Methodology

This study follows an integrative mixed-method approach that first integrates BERTopic, as one of the novel topic modeling techniques and then conducts a systematic literature review including a selective manual content analysis (Fig. 2.). Mixing topic modeling (quantitative) and content analysis (qualitative) approaches makes it possible to develop an understanding of the literature in a practical, effective, and accurate way. The approach consists of three main phases: (i) a systematic search to gather related papers according to search criteria, (ii) conducting BERTopic over the collected papers, and (iii) a manual literature review including content analysis of selected papers within relevant themes identified through topic modeling. In the first phase, we collected the corpus of papers. In the second phase, we used BERTopic to capture the topics and primary themes. In the final phase, we incorporated the findings of the second phase to define a list of papers to be further analyzed. Combining topic modeling (for utility purposes) with content analysis (for accuracy purposes), we provide an effective and efficient approach to analyzing a large number of documents.

3.1. Literature search

A literature search was performed in Scopus, Web of Science, and the electronic library of the Association for Information Systems (AIS). Scopus and Web of Science provide access to a wide range of peer-reviewed literature, including academic journals, conference proceedings, and other documents in various academic fields. Moreover, we searched the AIS database, because it covers the latest advances in practice and academia in information systems research including conference publications that are not included in Scopus and Web of Science. We narrowed the time period to the past 15 years to enable us to gather the most up-to-date AI and RAI research available. We used a composite search string to search for literature in the three different databases (see Table 1). We decided to set "artificial intelligence" and "maturity model" as the primary terms for our query. Moreover, as some terms can be used interchangeably, we included "capability model," "readiness," and "machine learning" to supplement the primary terms. These are the most frequent terms related to AI and maturity models. The terms were used to search within the title, abstract, and keywords of existing studies. The search string was composed by linking together the primary terms and their alternatives using the Boolean operators "OR" and "AND."

The search was confined according to the following inclusion criteria:

1 Published in a journal or conference (book, book chapter, etc. were excluded)
2 Published from 2007 to 2022 (the past 15 years)
3 Written in English

Table 1
Search string used in Scopus, Web of Science, and AIS eLibrary.

Search term
("artificial intelligence" OR "machine learning") AND ("maturity model" OR "capability model" OR "readiness")

We selected papers using three steps: first, the initial search in the databases yielded 1726 academic articles. Out of these, 151 were removed as duplicates, meaning 1575 different articles remained to be evaluate based on the inclusion criteria in the final step. By applying these criteria, 124 articles were removed, and 1451 articles remained. After exporting to CSV files, the data were merged and sorted in Python programming language to perform text processing. In this way, we aimed to fully cover the literature on AI maturity, determining patterns and creating the basis for identifying what is relevant for responsible AI.

3.2. Topic modeling with BERTopic

As a text mining technique, topic modeling has been applied in different areas, such as bioinformatics, recommended systems, financial analysis, manufacturing applications, computer science, social network analysis, and exploratory literature review (Ahn et al., 2021; Asmussen & Møller, 2019; Rajendran & Sundarraj, 2021; Zarindast et al., 2021; Singh et al., 2022; Dwivedi et al., 2023). These techniques, which rely on statistical modeling, machine learning, and natural language processing, aim to extract topical patterns within a collection of unlabeled texts. Although topic modeling is used in numerous fields for primary research, there are few review papers utilizing topic modeling to identify themes or topics for categorizing research papers (Asmussen & Møller, 2019; Kotsialos & Vassilakopoulou, 2023; Mäntylä et al., 2018). Topic modeling is a good first step for a literature review, because it does not require pre-existing knowledge of the categories of the papers and does not require significant time resources (Asmussen & Møller, 2019).

Topic modeling can be automated and reduce pre-analysis and even post-analysis costs. However, the accuracy of this method is low compared to manual reading and human coding practices, as this usage focuses more on utility than accuracy. Egger and Yu evaluated the performance of four topic modeling techniques, including Latent Dirichlet Allocation (LDA), non-negative matrix factorization (NMF), Top2Vec, and BERTopic on Twitter posts and mapped out their weakness and strengths in a social context (Egger & Yu, 2022). One of the key advantages mentioned in this research for using BERTopic was the flexibility of using a wide range of embedded models. Another advantage of BERTopic is the availability of search functions that go from topic to document. This can be useful for delving deeply into a particular topic

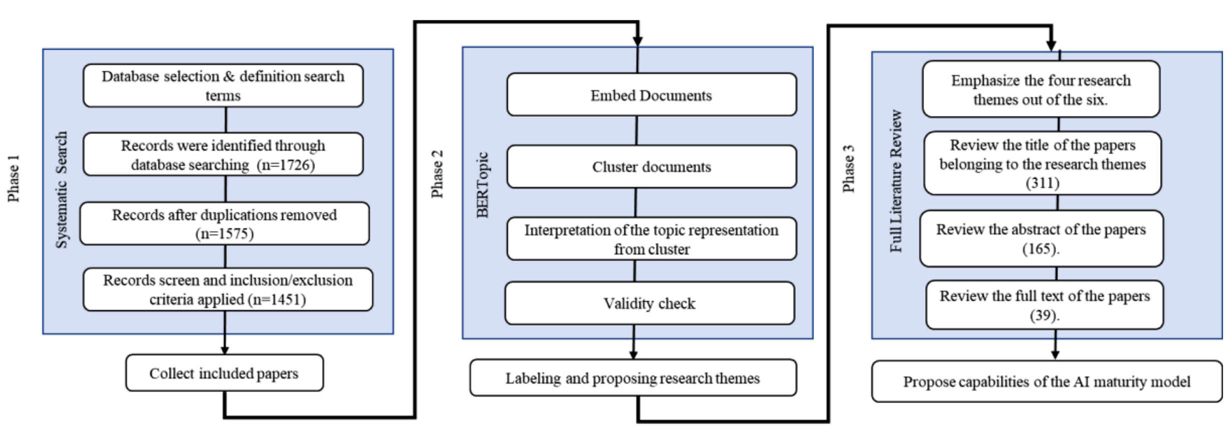


Fig. 2. Flowchart of methodology process.

facilitating qualitative content analysis. Moreover, with BERTopic, outliers can be detected automatically, and this can be beneficial for researchers because it rapidly eliminates nonrelated papers.

BERTopic has four main components. The first step starts with word embedding, perhaps one of the critical breakthroughs for the impressive performance of deep learning and NLP methods (Goldberg, 2017). In this method, the Sentence-BERT (SBERT) framework converts sentences and paragraphs to dense vector representations using pre-trained language models. One of the critical advantages of BERTopic is the possibility of using any other embedding technique, which allows continuous flow with the current state-of-the-art in embedding techniques (Egger & Yu, 2022). After embedding, we performed clustering using the UMAP approach, which is supported in BERTopic for dimension reduction, and it is the second component that is used in BERTopic. The result is transmitted to HDBSCAN (the third component) to cluster semantically similar sets of documents (Sánchez-Franco & Rey-Moreno, 2022). The last component represents the topics by extracting the most relevant words for each cluster. The class-based TF-IDF approach is performed to obtain the topics' representation; in other words, in this step, the importance of terms within a cluster is compared, and term representation is created (Grootendorst, 2020; Sánchez-Franco & Rey-Moreno, 2022). This approach allows the modeling of the importance of words in clusters instead of individual documents (Prasad et al., 2023; Sharevski et al., 2022).

In order to apply BERTopic in our corpus of papers, we prepared the text content for algorithmic analysis (preprocessing). To do this, first, we used Natural Language Toolkit (NLTK) in Python, which has a list of common stop words, and we removed them from the text. Moreover, we removed white spaces and converted all words to lowercase. In addition, to improve the clustering quality, we counted the frequency of one, two, and three-letter words to detect and add meaningless words to our stop word list and exclude them. While cleaning the data and rerunning the algorithms several times, the clustering and topics become more sensible and meaningful. Table 2 provides information about the collected documents before and after the cleaning steps. According to these numbers, the frequency of each word in documents was high. After cleaning and removing the stop words, we had 151,445 words; among these, only 13,695 unique words were detected. In other words, each word was repeated on average 11 times, and this represents the quality of our final dataset for topic modeling.

After preprocessing, the BERTopic library in Python was used, and the results were analyzed and discussed among all authors over two runs. Topics in the first run were not sufficient; as a result, we decided to add UMAP, TF-IDF model, and HDBSCAN to the BERTopic model and rerun it. In the second run, the quality of the results was improved. The calculation time of all papers for 40 topics took almost 3 minutes on a standard laptop. One outcome of the BERTopic modeling is a list of outlier documents and topics with the number of documents in each topic. The list is used in the labeling process. One of the critical advantages of this algorithm is that outlier documents are excluded from the clustering. These documents have little similarity to the other documents, and the model considers them as outliers. Another advantage is that BERTopic represents the number of documents on each topic, which allowed us to investigate and review documents on specific concepts. When the labeling process was complete, the results were used to analyze the content of identified topic clusters aiming to identify responsible AI capabilities.

Table 2
Word frequency before and after the data cleaning.

Number of abstracts	1,451
Total number of words before the cleaning process	267,062
Total number of unique words before the cleaning process	14,960
Total number of words after cleaning	151,445
Total number of unique words after the cleaning process	13,695
Total number of stop words used in the model	115,617

3.3. Qualitative labeling and selection of papers for content analysis

BERTopic automatically suggested topics that were assessed using a combination of reviewing the most frequent words in each topic and a title review to enhance their understandability. The proposed topics by the BERTopic model were used as input for qualitative labeling by the authors. To be able to review the title of papers on each topic, we allocated papers to topics. The sociotechnical systems model presented in Fig. 1 was used to guide the qualitative labels using key theory aspects: technology, tasks, organizational structures, and actors. During the qualitative labeling phase, the mind mapping software "EdrawMind" was used to facilitate the collaboration between the authors.

After labeling the topics, we consolidated them into research themes to present a comprehensive view of the literature. In total, we identified six themes. To perform a more focused analysis, we conducted a content analysis of the papers under four of these six themes. We excluded the themes that focused on technical aspects and specific AI applications, since our aim was to identify specific capabilities for RAI, which is consistent with the (Makadok, 2001) terminology applied to the AI domain. This approach allowed us to narrow our focus and provide a more targeted analysis of the relevant literature, enabling us to identify the required capabilities for developing responsible AI.

To drill down and select the papers for content analysis we employed a four-step procedure (Fig. 3) restricting our corpus of papers to four main themes (human resource aspects, ethical AI models aspects, organizational aspects, and data aspects) that were extracted from topic modeling. As a first step, we screened all retrieved articles' titles to evaluate eligibility based on inclusion criteria. The following inclusion criteria were defined: 1) Include a conceptualization or practical measurement of AI systems' maturity and readiness, 2) Point out principles or related considerations for the responsible use of AI, and 3) Include research motivations that cover both AI's social and technical components. Then, the abstracts of the remaining publications were screened and checked. In the final step, we reviewed the full text of all remaining papers to identify the ones to be further analyzed. Throughout these three steps of paper selection, the inclusion criteria were applied. Finally, after checking the full text, we performed a backward and forward search using the singled out papers. The list of papers for which we performed the content analysis is presented in Appendix 3.

The final result of topic modeling (main themes) was used to create a coding schema. In this step, we conducted a thorough full-text analysis of the included papers and applied sub-coding to the main themes. The resulting codes were then subjected to an exploratory examination, and the research team engaged in discussions to refine them. To facilitate the coding process, we used Microsoft Excel. Through an iterative approach, we developed a more nuanced understanding of the main themes and capabilities emerging from the codes, specifically regarding the required capabilities to develop responsible AI. This approach enabled us to comprehensively analyze the research questions and contribute to a broader understanding of the topic. After analyzing the included papers and adding some sub-coding to the main themes, the codes were examined in an explorative way. The research team discussed the included papers and refined the codes using Microsoft Excel.

4. Findings

4.1. Overview of literature streams and paper topics

Among the full set of papers in our literature corpus (1451), 60% were published in journals, and 40% were conference articles. More than 68% of these research papers had been published in the past two years, indicating the growth in this research field. One essential feature of the unsupervised machine learning method is that the dataset's hidden patterns and data grouping are discovered, which means some findings might be unexpected but valid. Through BERTopic, similar papers were divided into clusters named "topics." The construction of

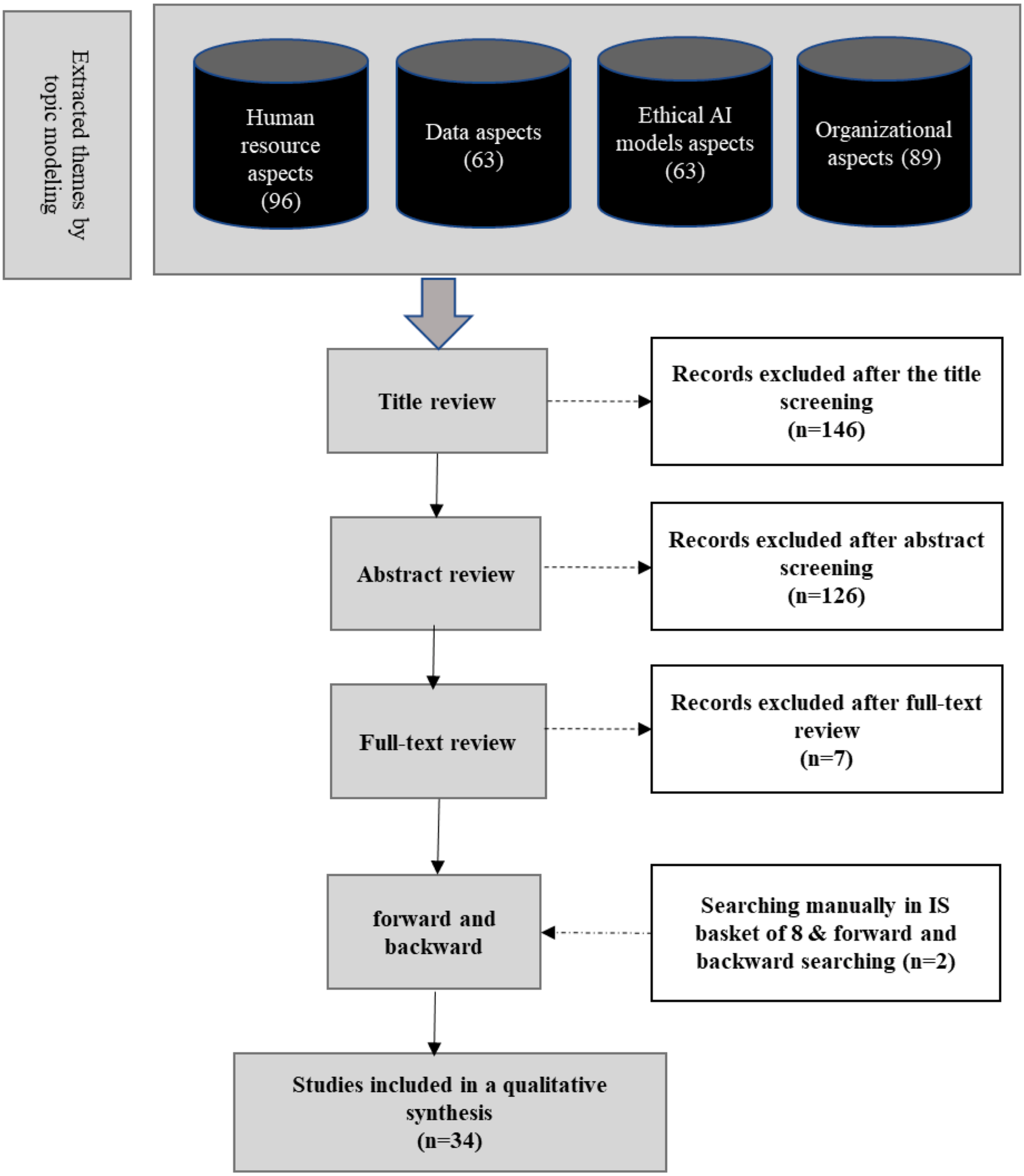


Fig. 3. Flowchart of the selection process of papers for content analysis.

main themes related to artificial intelligence involved several steps. First, similar topics were identified, and each topic's ten most frequently used words were selected. This allowed for creating a similarity matrix, confirming the topics' relationships. Next, the topics were labeled and renamed based on their similarities. Finally, the related topics were grouped to form the main themes of AI maturity and readiness. This process provides a comprehensive and structured approach to analyzing and understanding the various aspects of AI literature and is essential for scholars and researchers studying this rapidly-evolving field.

We identified six main themes in the literature corpus. The first main theme represents literature that focused on maturity related to specific applications of AI in different empirical domains (Table 3). The most prevalent were: "AI in healthcare", "robotics", and "cybersecurity". The other main themes including technological human resources, organizational, data, and ethical AI models aspects can be classified based on (Makadok, 2001) terminology applied to the AI domain. The technological aspect refers to assets, and the other four (i.e., human resources, organizational, data, and ethical AI models aspects) refer to different capabilities for using assets (Aral & Weill 2007). The technological aspect comprises four different types of assets: for data computation, for software development, for storage, and for networking. The role of these assets is to enable the functionality of AI-based systems. In Makadok

Table 3
Identified topics with the use of BERTopic.

Main Themes	Topic No	Topic Label	Number of papers	Main Themes	Topic No	Topic Label	Number of papers
Applications	T0	robotics	105	**Technological aspects**	T5	edge computing	25
	T3	process control	16		T7	network communication	39
	T4	AI for sustainability	19		T8	computing	11
	T6	autonomous vehicles	37		T10	computing chips	37
	T9	cybersecurity	86		T17	digital technologies	54
	T11	health care	43		T20	AI software development	14
	T12	imaging in health care	26		T24	natural language processing	43
	T13	image processing in health care	34		T25	infrastructure	25
	T14	covid19	12		T30	technologies development	79
	T15	digital healthcare	19		T31	AI-driven platforms	12
	T16	chatbots	16		T32	data analytics techniques	45
	T34	e-government services	14				**Sum: 384**
			Sum: 427	**Human resource aspects**	T22	data experts	30
Organizational aspects	T28	knowledge management	25		T26	data analytics capabilities	31
	T29	process management	22		T36	job involvement	13
	T35	innovation management	15		T37	human resource management	22
	T33	organizational maturity	27				**Sum: 96**
			Sum: 89	**Data aspects**	T27	data governance	30
Ethical AI models aspects	T18	human AI interactions	41		T38	data quality	16
	T19	ethical AI	22		T39	data management	17
			Sum: 63				**Sum: 63**

(2001) terminology, capabilities are considered practices, routines, skills, and competencies to use AI assets. We identified the themes of ethical AI, organizational, human resource, and data aspects and chose to evaluate deeply and discover the required capabilities.

A comprehensive overview of the research's main themes and discovered topics within each theme is provided in Table 3. Additionally, in Table 3, we provided detailed information for each theme, its topics, and documents per topic. Finally, in Appendix 1, we provided an overview of the five most frequent words within each topic and their frequency. After evaluating topics, the authors eliminated four topics (T1 on cancer., T2 on energy, T21 on user inputs and T23 on semantic ontologies) that were irrelevant to AI maturity. The topic named "-1" refers to all records with no topics allocated. In HDBSCAN, the algorithm assigns documents to the outlier cluster if they cannot be allocated to another cluster. In total, 188 documents were detected as outliers and excluded from clustering. As topic modeling findings in Table 3 show, only 63 papers related directly to ethical AI aspects; thus, the literature on this type of capability is in its infancy. Moreover, within the other themes related to capabilities in the third research stream (human resource, organizational, and data), there are aspects related to RAI. Hence, these papers need to be investigated more deeply. The similarity matrix, which is presented in Appendix 2, helped us identify similarities between topics, guiding us to the papers that needed to be analyzed further. As a result, we proceeded to conduct a content analysis on papers that belonged to these themes to gain a better insight into them.

4.2. Capabilities identified through content analysis

Transitioning from themes and topics to capabilities through content analysis entails a structured and rigorous approach. Initially, the research team read the full text of the final set of included articles, consisting of 34 papers. Through close examination, we identified relevant keywords and phrases that represented critical capabilities needed for different levels and conditions of AI maturity. To ensure a robust analysis, the team performed an extensive review of the identified terms to detect specific capabilities required for responsible AI development and use. The team then synthesized the information and classified the capabilities deductively, using the literature to ensure that the resulting capabilities were firmly grounded in empirical evidence.

The identified capabilities were further categorized into two main areas: foundational and responsible AI capabilities. Foundational capabilities are the technical and managerial skills necessary for developing and deploying AI systems, including data management, algorithm development, and system architecture. In addition, responsible AI capabilities encompass ethical, social, and legal aspects that must be considered in AI development and use, such as accountability, transparency, and fairness. Through this comprehensive qualitative coding method, we were able to provide a clear understanding of the capabilities needed for responsible AI development and use, ensuring that these capabilities were grounded in empirical evidence and widely applicable to the field. Appendix 4 provides details on examples of keywords and descriptions of the capabilities.

It is important to note that foundational and responsible capabilities are necessary to achieve RAI maturity. Organizations need a solid technical foundation to implement responsible AI practices. In contrast, organizations may develop AI systems without responsible AI capabilities that have unintended consequences or do not align with societal values. Therefore, both categories of capabilities must be developed and integrated into an organization's AI development and deployment processes. For example, the capability of managing innovation is essential for companies to achieve instrumental objectives, while other capabilities are related to accountability, fairness evaluation, and understandability of support organizations in attaining humanistic objectives. We emphasize the importance of developing and integrating foundational and responsible AI capabilities into an organization's AI development and deployment processes to achieve RAI maturity.

4.2.1. Foundational AI capabilities

The foundational capabilities are general but necessary; without such capabilities for organizational strategy and change management, AI adoption cannot be successful. In other words, these capabilities form the instrumental foundation for the responsible development and deployment of AI systems to achieve humanistic objectives. In the paragraphs that follow, we elaborate on each of the foundational AI capabilities identified in our study. Organizations must prioritize the development and implementation of these foundational capabilities to ensure the successful adoption of AI, as they provide the basis for future advanced AI capabilities that may be required. The present study posits that infrastructural and computational capabilities, denoted by technical and application-oriented themes, can be regarded as foundational capabilities. However, during the content analysis stage, the deliberate exclusion of technical and application-oriented themes was employed to delineate the boundaries of the capabilities under investigation. This focused approach aimed to specifically address the capabilities mentioned in the second and third research questions.

- Data quality

Data quality capability involves ensuring that data used to train and operate AI models are reliable and produce accurate outcomes. This means that capabilities throughout the AI lifecycle must be in place to check the data's age and accuracy of data labeling. For instance, ensuring that the data used for training AI models are up to date and correctly labeled to produce reliable results is paramount. Similarly, during operation, the data being fed to the AI model must be monitored to ensure it remains accurate and high quality (Dinter, 2012; Jöhnk et al., 2021).

- Data availability

The data availability capability involves ensuring that adequate amounts of appropriate data types are accessible for AI models (Lichtenthaler, 2020; Jöhnk et al., 2021). To accomplish this, tasks, technologies, and tools must be designed in a manner that allows for the management and continued availability of data that satisfies various AI types, ranging from traditional operational and predictive methods to other data science techniques. This design must also be flexible enough to accommodate the unique data requirements of different AI models.

- Financial

Financial capability pertains to an organization's ability to manage and allocate its financial resources effectively, which includes investing in AI. Given the high costs associated with AI investments, companies must carefully weigh the potential benefits against the costs. Regular cost-benefit analyses of AI investments during top management meetings allow organizations to assess the expected returns of their investments and ensure they align with their business objectives, as noted by Hradecky et al.(2022) and Çınar et al. (2021). Such analyses require comparing the anticipated benefits of an AI investment with the expenses involved in implementing it. These benefits encompass enhanced efficiency, reduced operational costs, heightened customer satisfaction, and competitive edge, while the costs consist of software, hardware, maintenance, training, and employee compensation expenses. By consistently evaluating the feasibility of AI investments from a financial perspective, firms can make informed decisions and ensure that their investments align with their business goals. This strategy enables companies to avoid unnecessary expenses and maximize the benefits of AI.

- Organizational strategies

The capability of organizational strategic planning is critical for successfully adopting and implementing AI. Coates and Martin (2019) and Sadiq et al.(2021) stress the significance of aligning the AI strategy with the organization's overall strategy. This alignment ensures that the benefits and outcomes of AI are directly linked to the organization's goals and objectives, as highlighted by Desouza et al. (2021) and Çınar et al. (2021). Organizations can maintain their agility and flexibility by achieving this alignment, allowing them to adjust their strategy quickly and respond proactively to opportunities and challenges. This capability is vital to effectively integrate AI into the organization and ensure its contribution to overall success.

- Innovation management

AI has the capacity to bring about significant changes in the way organizations manage innovation, as indicated by Haefner et al.(2021). To facilitate the contribution of innovative ideas to organizational plans and foster a reliable environment for the implementation of AI, it is crucial to establish a culture of innovation and protocols, as emphasized by Alsheiabani et al. (2019) and Holmström (2022). The integration of AI can provide tangible benefits to organizations, enhancing their competitive standing. By leveraging AI, companies can identify new avenues for innovation and streamline original processes, leading to increased efficiency and effectiveness. However, companies must establish explicit guidelines and ethical policies for the responsible and trustworthy use of AI in the management of innovation.

- Change management

Change management is a systematic approach to dealing with planned changes in an organization's goals, processes, or technologies (Pillai & Sivathanu, 2020; Facchini et al., 2019). Change management aims to implement strategies, prepare, support, and help individuals, teams, and organizations adapt to changes. The maturity level of this capability can be enhanced through the implementation of change management models, action plans, and top management support. The change management capability also relates to encouraging employees and removing barriers to achieving desired outcomes (Jöhnk et al., 2021; Holmström, 2022).

- Human resource management

Human resource management capabilities play a critical role in ensuring that employees possess the necessary skills and behaviors to effectively leverage AI solutions. It is important to have employees that are digitally literate and motivated to leverage AI solutions. Human resource management capability relates to the strategies adopted by AI firms to provide staff with opportunities to grow and develop skills related to the responsible use of AI (Pillai & Sivathanu, 2020; Chowdhury et al., 2022). Increasing the level of analytical competencies is a key task in this capability, encompassing skills related to the application of new technologies and analytical programming tools, as well as expertise in the specific application domain (Chowdhury et al., 2022; Saltz, 2017). By focusing on developing the necessary human resource management capabilities, organizations can better equip their employees with the knowledge and skills needed to maximize the potential benefits of AI solutions.

- Interdepartmental coordination

Interdepartmental coordination refers to the processes and activities involved in defining, documenting, and monitoring the work required for achieving AI goals and strategies across different departments in a business, according to Sadiq et al. (2021). Additionally, this ability shows the extent to which AI is utilized in various cases. In order to begin, use cases should be assessed, and the most promising opportunities for the organization across multiple departments must be identified. Successful interdepartmental coordination in AI implementation can lead to enhanced efficiency, productivity, and innovation across the organization.

- Performance of AI models

The capability of ensuring a sound AI performance is necessary for organizations that incorporate AI in their operations. In this regard, the ability to monitor, diagnose, and enhance the performance of AI models is a critical aspect for organizations. To achieve this objective, statistical metrics and data mining techniques are utilized to assess the performance of AI models. The performance evaluation of AI models is based on various factors, such as accuracy, sensitivity, specificity, and other key metrics. These metrics are instrumental in identifying and diagnosing potential issues in AI models, such as errors, bias, and inconsistency (Sternkopf & Mueller, 2018). Incorporating these metrics in AI maturity models is critical in ensuring effectiveness and enabling organizations to achieve their intended objectives of AI adoption.

4.2.2. RAI capabilities

Recently, the *MIT Sloan Management Review* magazine, in collaboration with Boston Consulting Group (BCG), reported that 41% of business leaders demonstrated that they already recognize benefits from their programs for responsible AI development and use and that AI maturity is strengthened when organizations have a robust RAI program (Renieris et al., 2022). Hence, identifying the capabilities of RAI can help reap more business benefits and diminish the risks associated with AI applications. Only strengthening foundational capabilities and increasing the use of AI solutions increases the risk of AI failures. As a result, organizations must pay close attention to both foundational and RAI capabilities.

RAI capabilities, like fairness and accountability aptitudes, affect both instrumental and humanistic objectives. We define RAI capabilities as the extent to which an organization providing AI services is able to mobilize and deploy tools, practices, strategies, and producers effectively to address AI-specific ethical issues. These capabilities are deployed in combination with foundational capabilities and relate to the use of AI assets. The following paragraphs describe the RAI capabilities identified in the literature.

- Continuous impact analysis

Continuous impact analysis refers to the ongoing process of periodically evaluating the effects of AI decisions. The aim is to determine whether the system could generate ethical or responsible outcomes that benefit all stakeholders (Shneiderman, 2020). This capability goes beyond merely assessing the accuracy of models based on historical patterns and data. It involves reviewing AI outcomes at the societal and individual human levels, considering different groups' unique characteristics and needs. The need for continuous impact analysis is further emphasized by the potential for AI to reinforce bias or unethical arrangements (Krijger et al., 2022; Coates & Martin, 2019;). For instance, AI systems may inadvertently discriminate against certain groups, perpetuate societal biases, or violate ethical norms. Therefore, continuous impact analysis can help detect and mitigate such risks, promoting fairness, accountability, and transparency.

- Employee's ethical awareness

The capability that relates to employees' ethical awareness involves ensuring the awareness of technical staff (including developers and data scientists) and also non-technical staff (such as domain experts) regarding ethical issues of AI. This is accomplished through training programs, knowledge sharing, and collaboration (Krijger et al., 2022; Jantunen et al., 2021). It is a critical RAI capability, as employees are involved in a multitude of decisions about AI and need to be alerted to possible negative consequences.

- Security and privacy

The capability to protect individuals' rights with respect to privacy and personal data is a crucial requirement for organizations. This involves implementing practices that ensure that privacy and security are respected. Privacy pertains to the personal information collected and how it is accessed, while security measures are taken to protect data and AI applications from potential harm, danger, or threats. By establishing these practices, organizations can ensure that they meet the necessary standards for protecting personal information and maintaining data security (Chen et al., 2021).

- Fairness evaluations

The ability to conduct fairness evaluations is essential in avoiding systemic discrimination against individuals based on factors, such as race, gender, or socioeconomic class, particularly in specific fields like recruitment, medical predictions, or finance allocations (Jantunen et al., 2021; Someh et al., 2022). To enhance the capability of fairness evaluations, practical actions such as conducting in-depth evaluations of training datasets are mentioned in the literature (Coates & Martin, 2019; Shneiderman, 2020). These evaluations help to ensure that fairness is maintained in different contexts and fields, and that discrimination is avoided.

- Understandability of AI models

The importance of clarity and explainability is emphasized in AI models. Meske et al. (2022) asserted that AI models should be structured in a manner that ensures transparency, interpretability, and explainability to effectively address the needs of end-users, decision-makers, and stakeholders who depend on their output. To achieve this, the authors propose nine quality criteria, with six specifically pertaining to the understandability of AI models. These criteria encompass generalizability, explainability power, interpretability, comprehensibility, plausibility, and effort. In this regard, understandability is vital in providing valid, reliable, and useful explanations of how AI models operate, which can help users make informed decisions based on the model's output.

Explainability, on the other hand, involves making AI models interpretable using technical methods and approaches, such as identifying which input affects which output and to what extent. For example, this capability can help identify which input affects which output and to what extent (Shneiderman, 2020; Someh et al., 2022). Transparency is closely related to understandability and can be achieved through formalized procedures that document and explain AI components, such as datasets, variables, and outcomes, in a precise and comprehensible way, such as through visual aids or simple language. This ensures that customers, users, and decision-makers at the managerial level can easily comprehend the workings of the AI model (Shneiderman, 2020; Fukas et al., 2021). By providing transparency, AI models can be made more trustworthy and accountable, which is essential for fostering user trust and promoting responsible AI development (Gillespie et al. 2023; Shneiderman, 2020).

- Accountability

The designers and deployers of AI systems need to be accountable for the operations of their systems, particularly when their decisions affect people's lives. There should be a clear and defined chain of accountability across various stages of the AI system's lifecycle, from design and development to deployment and maintenance, ensuring that those responsible can be traced back to any decision that affects individuals. This helps to ensure that humans maintain control over the AI system, and that accountability is properly structured and enforced (Krijger et al., 2022). AI systems should not be viewed as independent decision-makers, but rather as tools that support human decision-making. Therefore, the accountability of AI systems should be designed to ensure that humans are in the loop on the decisions made by the AI models.

5. Discussion

This paper maps out the literature on maturity models and readiness frameworks for RAI by synthesizing the insights from the systematic literature review. The literature review was performed using a mixed-method approach that includes topic modeling followed by content analysis. We opted for mixing quantitative (topic modeling) and qualitative (content analysis) methods to ensure reliable and accurate findings more efficiently. This mixed-method methodology, which leverages machine learning, allows faster and more reliable results providing a quick evaluation of the literature as a first step and a focused content analysis on a selected subset of the literature as a second step. Other researchers, including junior researchers, can easily apply this approach to get a solid overview of a field of interest. The study adopts the sociotechnical perspective to identify the essential foundational capabilities and RAI capabilities necessary for the responsible development and use of AI in organizations. This perspective serves as a theoretical foundation for translating RAI principles into practice and for striking a balance between humanistic and instrumental objectives. We categorized the identified capabilities into responsible and foundational AI to achieve humanistic and instrumental objectives through the implementation of the AI systems. We discuss specific implications for practice and theory next.

5.1. Sociotechnical perspective on AI maturity

We adopted a sociotechnical perspective as a means of developing a framework of RAI capabilities for organizations. We drew from the works of Sarker et al. (2019) and developed a framework that provides a holistic approach for addressing both instrumental and humanistic objectives of AI development (Fig. 4). While Asatiani et al. (2021) support

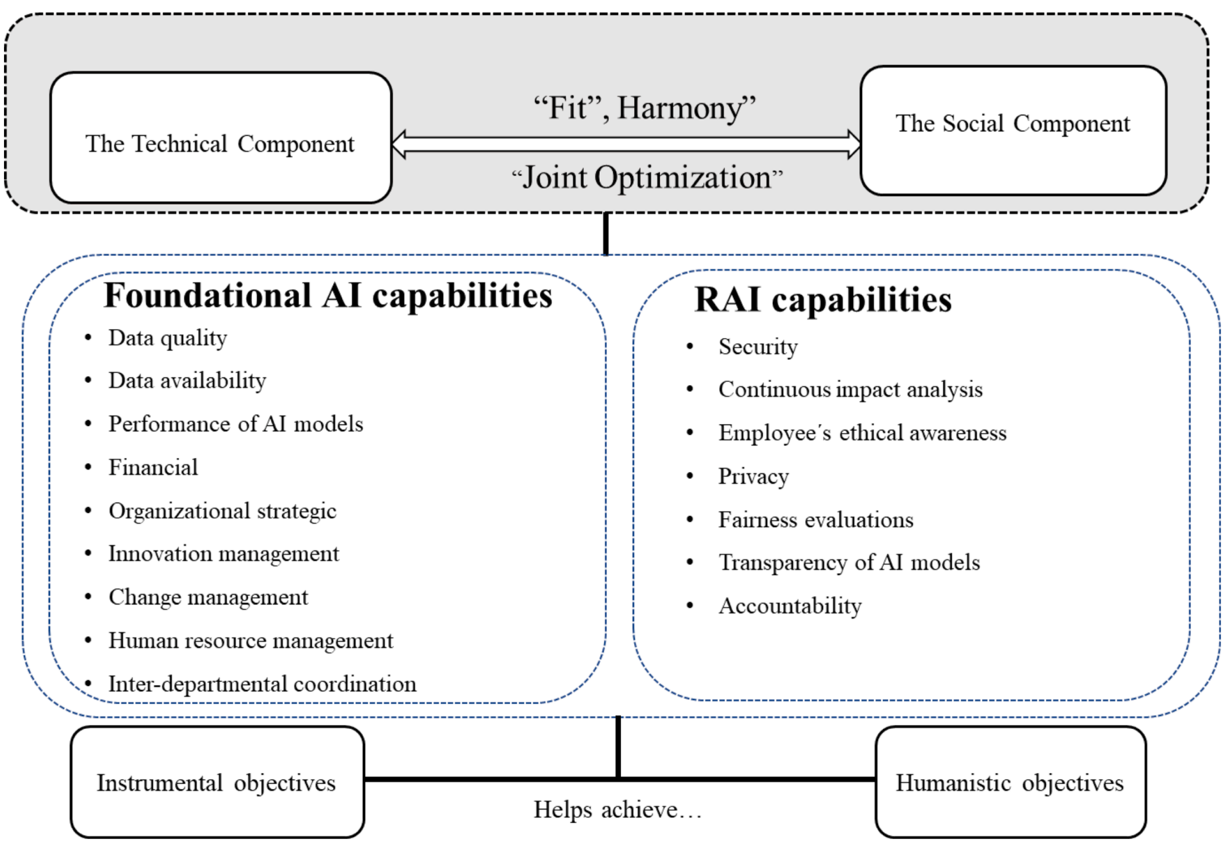


Fig. 4. Sociotechnical Framework of Capabilities for Organizations' AI Maturity Models.

this approach, they caution that addressing humanistic outcomes may pose more significant challenges for AI development than other technologies. Therefore, it is necessary to provide an approach that supports the humanistic and instrumental objectives to manage negative consequences and simultaneously provide reliable outcomes with high accuracy rates. To illustrate how the sociotechnical perspective can address these challenges and enhance the levels of maturity in terms of responsibility for AI systems, we use the example of explainability. AI systems require a balance between explainability and accuracy to ensure responsible outcomes when designing AI systems. Providing meaningful explanations enhances transparency, while high accuracy rates improve reliability. A balance between these factors helps manage unexpected negative consequences and improves the maturity levels of AI systems. It allows stakeholders to understand how AI systems function, providing transparency and accountability. By achieving this balance between explainability and accuracy, organizations can improve their maturity levels and adhere to responsible standards, effectively managing unexpected negative consequences.

Collectively, this paper and the proposed method contribute to the literature on RAI maturity models by identifying capabilities that are interconnected to each other and necessary for achieving humanistic and instrumental objectives through the AI development process (Fig. 4). The relationship between these capabilities is often nonlinear and fuzzy, and there are synergies among them. The sociotechnical perspective provides a unique approach to understanding holistically by the capabilities required for harnessing the power of AI while minimizing the risks for societies and individuals. This literature review adopts a sociotechnical perspective (Beath et al., 2013) to show how foundational AI capabilities and RAI capabilities are interwoven to arrive at a suitable level of maturity in their responsible AI operations. In an era of ongoing digital transformations, following responsible and sustainable practices is essential (Pappas et al., 2023a). In other words, while AI development and implementation in organizations are frequently motivated by instrumental objectives, humanistic objectives are also important for AI projects. This means that capabilities are not independent, and a combination of sociotechnical combinations of capabilities is needed to harmonize the instrumental and humanistic outcomes. For instance, data security and fairness evaluations relate to the ethical issues of AI use, but they also contribute to achieving the instrumental objective of increasing business value (Minkkinen et al., 2022). There are significant technical and social components for each of the capabilities identified (Asatiani et al., 2021). For example, for understandability and fairness, a combination of social and mathematical components is required to accomplish a high level of maturity. These examples show the reciprocal interactions between social and technical practices or components in the capabilities.

5.2. *Scope of future research for theoretical contributions and practical implications*

Our study makes a theoretical contribution to the ongoing discourse in IS research on driving and managing AI implementation in organizations. We provide a framework for developing organizational capabilities that facilitate the responsible use of AI systems, with a specific focus on the concept of responsible AI capabilities within the context of AI technology. Our findings emphasize the importance of such capabilities in effectively managing the AI adoption process and mitigating potential ethical risks, thus serving as a critical driver for successful AI implementation in organizations. In order to make significant theoretical contributions to the information systems research field, studies should deliberate efforts to address a range of significant questions related to various domains, sociotechnical factors, governance structures, and ontological and epistemological considerations. In addition, it is essential to critically examine and problematize the shortcomings of responsible AI (RAI) capabilities. and try to problematize lacking RAI capabilities. These questions can be addressed through the development of various theories, including, but not limited to, IS theories, ethical theories, political theories, organization theories, behavioral theories, and systems theories. Addressing these questions can lead to a more comprehensive and nuanced understanding of the role of information systems in various contexts and facilitate the development of effective and responsible practices for managing and utilizing these systems.

Table 4
Examples of research questions for future research.

No	Research areas	Emerging Research Questions
1	AI Capability Development and Maturity	How can foundational and responsible AI capabilities be further analyzed, operationalized, and linked to different maturity levels?
2		How can a maturity model be developed to attain both humanistic and instrumental outcomes while considering the fundamental and responsible capabilities of AI?
3		What practices or conditions are needed to improve the maturity level of proposed capabilities, and how can they be examined?
4		How can building a responsible AI office (RAIO) facilitate improving RAI maturity in organizations?
5	Responsible AI Implementation and Management	How do responsible AI capabilities affect the organization's performance and comparative advantages?
6		How can responsible AI capabilities enhance the truthfulness of AI systems?
7		What documentation and procedures are required in selecting datasets to train machine learning algorithms, and who is responsible for evaluating outcomes?
8		What are the organizational implications of building an RAIO, and which departments must lead the development of RAI capabilities?
9		What roles and duties can RAI offices take for improving the required capabilities, from analyzing the social impact of AI systems to developing responsible AI strategies and policies, to resolving technical issues?
10	Research Methodology and Validation	How can the prioritization and weighting of different factors and assessing their influence on the success of AI initiatives be investigated?
11		What are the practical knowledge and experiences of responsible AI project managers or researchers in the industry, and how can they be explored using qualitative methods and interviews?
12		How can the findings of the interviews be validated using quantitative approaches and data collection from a broader population of practitioners?
13		What is the relationship between the identified capabilities, and how can they reinforce or restrict one another?

Further research can empirically validate and potentially expand our findings, developing a maturity model by considering the foundational and responsible AI capabilities. In other words, each of these capabilities needs to be further analyzed, operationalized, and linked to different maturity levels. A mixed-method approach could be beneficial for further investigation. Qualitative methods and interviews with responsible AI project managers or researchers with related experience could help to explore practical knowledge in the industry. The findings of the interviews can be validated using quantitative approaches and data collection from a broader population of practitioners. Empirical research can also explore the relationship between the identified capabilities and how they can reinforce or restrict one another. Future research can also investigate the prioritization and weighting of different factors assessing their influence on the success of AI initiatives.

Our study offers valuable insights for practical implementation, particularly given the challenges that many organizations face in adopting responsible AI practices (Krijger et al., 2022; Vassilakopoulou et al., 2022). One of the practical benefits of foundational and responsible AI capabilities as a tool for implementation is that it provides a road map for improving the level of maturity. Our research findings offer guidance for practitioners, particularly top managers, regarding the organizational capabilities required to facilitate AI implementation and address potential ethical issues. These insights can serve as a starting point for practitioners to strategically allocate organizational resources and drive responsible AI programs. As a result, additional future research can examine which practices or conditions are needed for improving the level of maturity in proposed capabilities. For example, what documentation and procedures are required in selecting datasets to train ML algorithms? How can we document the methods and algorithms selected, and who is responsible for evaluating outcomes? The answers to these questions could vary based on the context of use of AI systems; hence, an adaptive approach is required.

Another future study can research organizational implications, for instance, investigating which departments in organizations must lead the development of RAI capabilities. This is an interesting area for further research that can be pursued through Interventionist research approaches such as action research and clinical research in close engagement with practice (Pappas et al., 2023b). Building a responsible AI office (RAIO), which companies like Salesforce have launched and reported their experiences, might facilitate improving RAI maturity in organizations. The roles and duties that RAIOs take for improving the required capabilities can vary widely, ranging from analyzing the social effect of AI systems to developing accountability strategies and policies to fixing technical issues. These are the areas needing more investigation.

Table 4 presents a categorization of research questions that can guide future investigations in the development of responsible AI. The research questions are organized into three categories: AI Capability Development and Maturity, Responsible AI Implementation and Management, and Research Methodology and Validation. Table 4 offers a roadmap for some possible opportunities for future research in RAI by providing examples of research questions.

5.2. Limitations

This systematic study has some limitations as indicated below. First, the review is limited to the past 15 years and explicitly focuses on conference and journal papers. Therefore, the study may have overlooked relevant books or articles published before 2007. The use of AI has significantly increased in recent years; however, AI technologies have been discussed in the literature for more than five decades; hence some relevant papers may have been published during early AI times. Furthermore, our study is based only on literature analysis, so we may have missed some relevant white papers from industry, which provide more practical insights. Despite the wide range of advantages of using topic modeling in literature reviews, some considerations must also be noted. This approach labels topics by the most frequent words in each cluster. There is a correlation between topics, making the process challenging. Some papers may discuss the same capabilities but belong to different topics. Finally, another limitation of the study is that only the abstracts of the papers were used to analyze and cluster them into topics. This limitation was addressed by the qualitative content analysis of the full text, which was added to ensure the accuracy and reliability of the review.

6. Conclusion

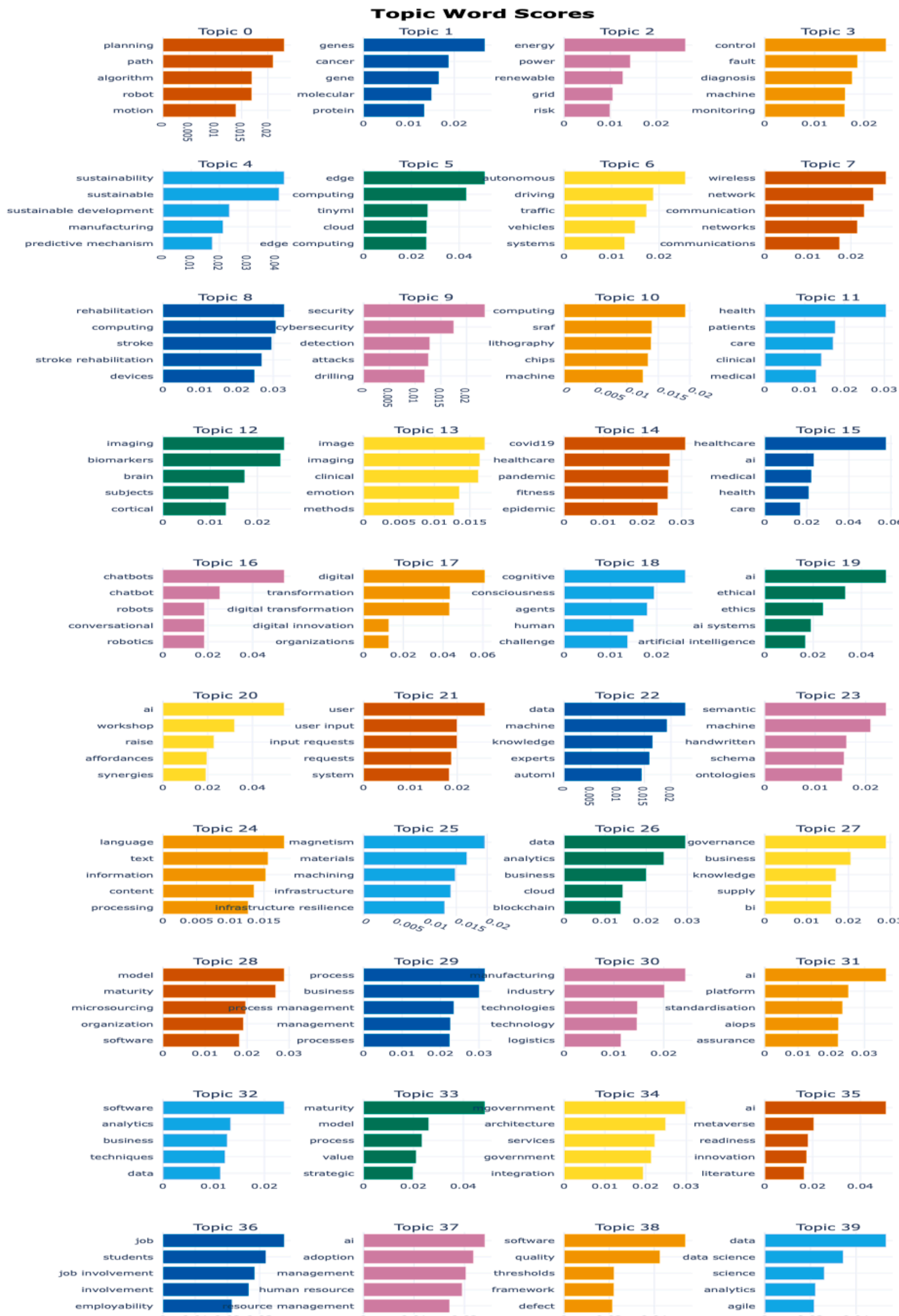
This paper presents the findings of a mixed-methods literature review. These findings were used to develop a framework for the responsible development and use of AI. The framework includes a comprehensive set of required capabilities. This approach has several main advantages. First, topic modeling reduces the need for manually reading all papers and enables the analysis of large numbers of papers quickly. The present study employed BERTopic, an unsupervised technique that clusters related documents into topics. The use of BERTopic represents a departure from traditional literature review approaches, which rely heavily on subjective decisions made by researchers. By adopting this approach, we provided a more transparent and objective analysis of various aspects of AI. This is especially important in a rapidly evolving and complex field like AI. Our findings highlight the significance of employing such techniques, as they offer a structured and comprehensive approach to analyzing AI-related literature. As such, this paper provides valuable insights for scholars and researchers seeking a deeper understanding of the subject matter.

The research method followed consisted of three main phases: systematic search, BERTopic, and content analysis. The corpus of literature reviewed was classified into six main themes using AI resource terminology. The first two main themes relate to specific AI application domains and to AI assets, while the four remaining main themes relate to capabilities (technological and application aspects). The capabilities relate to organizational, data, human resources, and ethical AI model aspects. The papers in these main themes were analyzed through content analysis to obtain a profound understanding of AI capabilities required for responsible AI development and use. The foundational and RAI capabilities identified in our review provide a framework for assessing companies' AI readiness and can be used as a basis for a comprehensive maturity model. Furthermore, we extend the current discourse on the development of responsible AI by proposing a series of research questions pertinent to AI capability development, implementation, and management. By building on the existing literature, our study contributes to advancing knowledge and understanding in the field of responsible AI and provides a resource for researchers, practitioners, and policymakers.

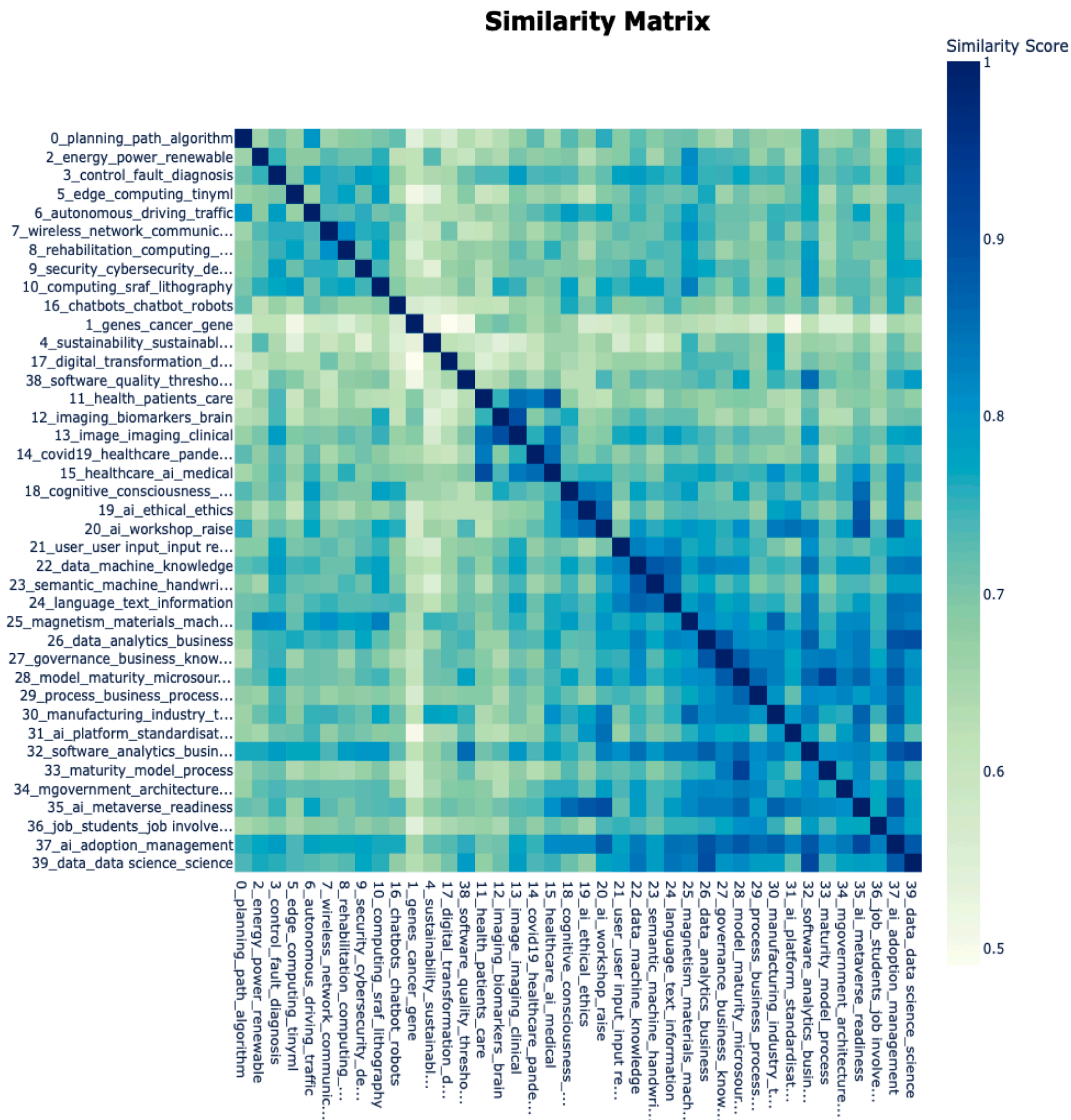
Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix 1. Identified topics by BERTopic



Appendix 2. Similarity Matrix between topics



Appendix 3. Papers Reviewed and related main themes.

No	Authors and Title	Topic number	Main themes	Outlet
1	(Alsheibani et al., 2019), Towards an Artificial Intelligence Maturity Model: From Science Fiction to Business Facts	T33	Organizational	PACIS
2	(Jantunen et al., 2021), Building a Maturity Model for Developing Ethically Aligned AI Systems.	T19	Ethical AI model	IRIS
3	(Fukas et al., 2021), Developing an Artificial Intelligence Maturity Model for Auditing	T35	Organizational	ECIS
4	(Schuster et al., 2021), Maturity Models for the Assessment of Artificial Intelligence in Small and Medium-Sized Enterprises.	T35	Organizational	Polish Chapter of Association for IS
5	(Russell et al., 2010), Organic Evolution and the Capability Maturity of Business Intelligence	T33	Organizational	AMCIS
6	(Sternkopf & Mueller, 2018), Doing Good with Data: Development of a Maturity Model for Data Literacy in Non-governmental Organizations	T27	Data aspects	HICSS
7	(Komatsu & Mantovani, 2021), Business Intelligence Maturity Level in Brazilian Companies	T27	Data aspects	AMCIS
8	(Saltz, 2017), Acceptance Factors for Using a Big Data Capability and Maturity Model	T39	Data aspects	ECIS
9	(Felch et al., 2019), Maturity Models in the Age of Industry 4.0 – Do the Available Models Correspond to the Needs of Business Practice?	T28	Organizational	HICSS
10	(Sadiq et al., 2021), Artificial Intelligence Maturity Model: A Systematic Literature Review	T33	Organizational	PeerJ Computer Science
11	(Chen et al., 2021), Establishment of a Maturity Model to Assess the Development of Industrial AI in Smart Manufacturing	T33	Organizational	Enterprise Information Management
12	(Pappel, et al., 2022), Maturity Model for Automatization of Service Provision and Decision-Making Processes in Municipalities	T29	Organizational	ICICT
13	(Bettoni et al., 2021), An AI Adoption Model for SMEs: A Conceptual Framework	T33	Organizational	INCOM
14	(Dphil & DIng, 2021), Industry 4.0- Artificial Intelligence (AI) Contribution to Capability Maturity	T35	Organizational	International Annual Conference of the American Society for Engineering Management
15	(Dinter, 2012), The Maturing of a Business Intelligence Maturity Model	T33	Organizational	AMCIS
16	(Desouza et al., 2021), Maturity Model for Cognitive Computing Systems in the Public Sector	T19	Ethical AI	HICSS
17	(Facchini et al., 2019), A Maturity Model for Logistics 4.0: An Empirical Analysis and a Roadmap for Future Research	T28	Organizational	Sustainability
18	(Williams & Lang, 2019), Digital Maturity Models for Small and Medium-sized Enterprises: A Systematic Literature Review	T33	Organizational	ISPIM
19	(Coates & Martin, 2019), An Instrument to Evaluate the Maturity of Bias Governance Capability in Artificial Intelligence Projects.	T19	Ethical AI	IBM-Journal of Research and Development
20	(Ellefsen et al., 2019), Striving for Excellence in AI Implementation: AI Maturity Model Framework and Preliminary Research Results	T28	Organizational	Scientific Journal of Logistics
21	(Lichtenthaler, 2020), Five Maturity Levels of Managing AI: From Isolated Ignorance to Integrated Intelligence	T35	Organizational	Journal of Innovation Management
22	(Jöhnk et al., 2021), Ready or Not, AI Comes— An Interview Study of Organizational AI Readiness Factors	T29	Organizational	Business and Information Systems Engineering
23	(Mikalef, et al., 2022), Thinking Responsibly about Responsible AI and "The Dark Side" of AI	T35	Organizational	European Journal of Information Systems
24	(Holmström, 2022), From AI to Digital Transformation: The AI Readiness Framework	T29	Organizational	Business Horizons
25	(Chowdhury et al., 2022), Unlocking the Value of Artificial Intelligence in Human Resource Management through AI Capability Framework	T37	Human Resource aspects	Human Resource Management
26	(Hradecky et al., 2022), Organizational Readiness to Adopt Artificial Intelligence in the Exhibition Sector in Western Europe	T35	Organizational	International Journal of Information Management
27	(Martínez-Plumed et al., 2021), Futures of Artificial Intelligence through Technology Readiness Levels	T35	Organizational	Telematics and Informatics
28	(Alsheibani et al., 2018), Artificial Intelligence Adoption: AI-Readiness at Firm-Level	T35	Organizational	PACIS
29	(Kinkel et al., 2022), Prerequisites for the Adoption of AI Technologies in Manufacturing – Evidence from a Worldwide Sample of Manufacturing Companies	T29	Organizational	Technovation
30	(Pillai & Sivathanu, 2020), Adoption of Artificial Intelligence (AI) for Talent Acquisition in IT/ITeS Organizations	T37	Human Resource aspects	Benchmarking
31	(Someh et al., 2022), Building an Artificial Intelligence Explanation Capability	T19	Ethical AI	MIS Quarterly Executive
32	(Çınar et al., 2021), A Framework for Industry 4.0 Readiness and Maturity of Smart Manufacturing Enterprises: A Case Study	T33	Organizational	Sustainability
33	(Shneiderman, 2020), Bridging the Gap between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-Centered AI Systems	-	Forward and backward searching	ACM Transactions on Interactive Intelligent Systems
34	(Krijger et al., 2022), The AI Ethics Maturity Model: A Holistic Approach to Advancing Ethical Data Science in Organizations.	-	Forward and backward searching	AI and Ethics

Appendix 4. Foundational and responsible AI capabilities with examples of keywords.

Category of the capability	Capabilities	Description	Examples	Related Topic
Foundational	Data quality	Ensure reliable and accurate outcomes in AI by maintaining data quality throughout the AI lifecycle through accurate data labeling during training and monitoring data quality during operations.	Data quality	T27
			Data verification and validation	T27
			Data redundancy	T27
			Data integrity	T39
	Data availability	Data availability in AI requires sufficient amounts of appropriate data types accessible for various AI models, achieved by designing flexible tasks, technologies, and tools for data management and continuous availability that cater to different AI requirements.	Data accessibility	T27
			Data durability	T27
			Storage facilities	T39
	Financial	Financial capability involves effective management and allocation of financial resources for AI investments, requiring regular cost-benefit analyses to ensure alignment with business objectives and maximize benefits while minimizing expenses.	Financial budgets	T33
			Cost-benefit analysis	T33
	Organizational strategies	Organizational strategic planning is vital for successful AI adoption and implementation, requiring alignment with overall strategy to link AI outcomes with business objectives and achieve agility.	Strategic alignment	T33
			Process and organization	T33
			Top management support	T33
	Innovation management	AI can facilitate innovation in organizations, but establishing a culture of innovation and ethical guidelines for its use is crucial to maximize its benefits.	Innovation strategy	T35
			Innovation culture	T35
			Collaborative work	T35
	Change management	Change management is a methodical approach to dealing with planned changes in an organization's goals, processes, or technologies.	Change management	T28
			Norms and institutions	T28
			Organizational changes	T28
			Change leadership	T28
	Human Resource management	Human resource management is crucial in equipping employees with digital literacy and analytical competencies to leverage AI.	Analytical competences	T37
			Recruitment and selection	T37
			Digital literacy	T37
	Interdepartmental coordination	Interdepartmental coordination involves defining, documenting, and monitoring work across different departments to achieve AI goals and strategies.	Interdepartmental collaboration	T29
			Cross-functional cooperation	T29
			Improved coordination	T29
	Performance of AI models	The ability to monitor and enhance the performance of AI models is critical for organizations that use AI in their operations.	Sensitivity	T33
			Accuracy	T33
			Specificity	T33
			Performance evaluation	T33
Responsible AI	Employees' ethical awareness	This capability is related to employees' ethical awareness regarding AI issues through training, knowledge sharing, and collaboration.	Training programs	T19
			Awareness and a culture promoting environment	Forward and backward searching
	Continuous impact analysis	Continuous impact analysis is the ongoing process of evaluating the effects of AI decisions to ensure ethical and responsible outcomes for all stakeholders.	Continuous evaluation	Forward and backward searching
			Constant review	Forward and backward searching
	Security and privacy	This capability involves implementing practices that ensure that privacy and security are respected.	Privacy rights	T19
			Information security	T19
			Privacy impact	Forward and backward searching
			Data protection	T19
	Fairness evaluations	Fairness evaluations are critical to prevent discrimination against individuals based on their race, gender, or socioeconomic status in fields, such as recruitment, medical predictions, and finance	Bias evaluations	T19
			Fair decisions	T19
			Fair metrics	Forward and backward searching
			Mitigate bias	T19
	Understandability of AI models	This capability requires the capability of understandability and expansibility, which provide valid and reliable explanations of how AI models operate and identify which input affects the output.	Transparency	T19
			Explainability	T19
			Understandable	Forward and backward searching
			Interpretable	T19
	Accountability	This capability ensures that humans maintain control over the AI system and are responsible for any decisions that affect individuals.	Accountability	T19
			Response mechanisms	Forward and backward searching
			Communication	Forward and backward searching

References

Abramova, O., Batzel, K., & Modesti, D. (2022). Collective response to the health crisis among German Twitter users: A structural topic modeling approach. *International Journal of Information Management Data Insights, 2*(2), Article 100126. https://doi.org/10.1016/j.jjimei.2022.100126

Accenture. (2022). *From AI compliance to competitive advantage*. Available at: https://www.accenture.com/us-en/insights/artificial-intelligence/ai-compliance-competitive-advantage.

Ahn, J., Son, H., & Chung, A. D. (2021). Understanding public engagement on Twitter using topic modeling: The 2019 Ridgecrest earthquake case. *International Journal of Information Management Data Insights, 1*(2), Article 100033. https://doi.org/10.1016/j.jjimei.2021.100033

Alshawi, M. (2007). Rethinking IT in construction and engineering: Organisational readiness. *Rethinking IT in construction and engineering: Organisational readiness*. Taylor & Francis.

Akbarighatar, P. (2022). Maturity and readiness models for responsible artificial intelligence (RAI): A systematic literature review. In *Proceedings of the 14th Mediterranean conference on information systems (MCIS 2022).12.* aisel.aisnet.org/mcis2022/12/.

Akbarighatar, P., Pappas, I., & Vassilakopoulou, P. (2023a). Justice as fairness: A hierarchical framework of responsible AI principles. In *Proceedings of the 31st European conference on information systems (ECIS 2023)*. Research-in-Progress Paper. 79. aisel.aisnet.org/ecis2023_rip/79/.

Akbarighatar, P., Pappas, I., & Vassilakopoulou, P. (2023b). Practices for responsible AI: Findings from interviews with experts. In *Proceedings of the American conference on information systems (AMCIS 2023*. aisel.aisnet.org/amcis2023/sig_odis/4/.

Alsheiabani, S., Cheung, Y., & Messom, C. (2019). Towards an artificial intelligence maturity model: From science fiction to business facts. In *Proceedings of the 23rd Pacific Asia conference on information systems (PACIS 2019)* (p. 46). aisel.aisnet.org/pacis2019/46.

Alsheibani, S., Cheung, Y., & Messom, C. (2018). Artificial intelligence adoption: AI-readiness at firm-level. In *Proceedings of the 22nd Pacific Asia conference on information systems (PACIS 2018)* (p. 37). aisel.aisnet.org/pacis2018/37.

Aral, S., & Weill, P. (2007). IT Assets, organizational capabilities, and firm performance: How resource allocations and organizational differences explain performance variation. *Organization Science, 18*(5), 763–780. https://doi.org/10.1287/orsc.1070.0306

Asatiani, A., Malo, P., Nagbøl, P. R., Penttinen, E., Rinta-Kahila, T., & Salovaara, A. (2021). Sociotechnical envelopment of artificial intelligence: An approach to organizational deployment of inscrutable artificial intelligence systems. *Journal of the Association for Information Systems (JAIS), 22*(2). https://doi.org/10.17705/1jais.00664, 325-252.

Asmussen, C. B., & Møller, C. (2019). Smart literature review: A practical topic modelling approach to exploratory literature review. *Journal of Big Data, 6*(1), 1–18. https://doi.org/10.1186/s40537-019-0255-7

Beath, C., Berente, N., Gallivan, M. J., & Lyytinen, K. (2013). Expanding the frontiers of information systems research: Introduction to the special issue. *Journal of the Association for Information Systems, 14*(4), 4. https://doi.org/10.17705/1jais.00330

Becker, J., Knackstedt, R., & Pöppelbuß, J. (2009). Developing maturity models for IT management. *Business & Information Systems Engineering, 1*(3), 213–222. https://doi.org/10.1007/s12599-009-0044-5

Benbya, H., Pachidi, S., & Jarvenpaa, S. (2021). Special issue editorial: Artificial intelligence in organizations: Implications for information systems research. *Journal of the Association for Information Systems, 22*(2), 10. https://doi.org/10.17705/1jais.00662

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS quarterly, 45*(3), 1433–1450. https://doi.org/10.25300/MISQ/2021/16274

Bettoni, A., Matteri, D., Montini, E., Gładysz, B., & Carpanzano, E. (2021). An AI adoption model for SMEs: A conceptual framework. *IFAC-PapersOnLine, 54*(1), 702–708. https://doi.org/10.1016/j.ifacol.2021.08.082

Chen, W., Liu, C., Xing, F., Peng, G., & Yang, X. (2021). Establishment of a maturity model to assess the development of industrial AI in smart manufacturing. *Journal of Enterprise Information Management, 35*(3), 701–728. https://doi.org/10.1108/JEIM-10-2020-0397

Chowdhury, S., Dey, P., Joel-Edgar, S., Bhattacharya, S., Rodriguez-Espindola, O., Abadie, A., & Truong, L. (2022). Unlocking the value of artificial intelligence in human resource management through AI capability framework. *Human Resource Management Review, 33*(1), Article 100899. https://doi.org/10.1016/j.hrmr.2022.100899

Çınar, Z. M., Zeeshan, Q., & Korhan, O. (2021). A framework for Industry 4.0 readiness and maturity of smart manufacturing enterprises: A case study. *Sustainability, 13*(12), 6659. https://doi.org/10.3390/su13126659

Coates, D. L., & Martin, A. (2019). An instrument to evaluate the maturity of bias governance capability in artificial intelligence projects. *IBM Journal of Research and Development, 63*(4/5). https://doi.org/10.1147/JRD.2019.2915062, 7-1.

De Bruin, T., Rosemann, M., Freeze, R., & Kulkarni, U. (2005). Understanding the main phases of developing a maturity assessment model. In *Proceedings of the 16th Australasian conference on information systems, ACIS 2005. Sydney* (pp. 8–19).

Desouza, K. C., Götz, F., & Dawson, G. S. (2021). Maturity model for cognitive computing systems in the public sector. In *Proceedings of the 54th annual Hawaii International conference on system sciences, (HICSS 2021)* (pp. 2173–2182).

Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way* (ed 1). Springer Nature.

Dinter, B. (2012). The maturing of a business intelligence maturity model. In *Proceedings of the American conference on information systems (AMCIS 2012)* (p. 37). aisel.aisnet.org/amcis2012/37.

Dolata, M., Feuerriegel, S., & Schwabe, G. (2022). A sociotechnical view of algorithmic fairness. *Information Systems Journal, 32*(4), 754–818. https://doi.org/10.1016/j.techfore.2020.120392

Dphil, A. V., & DIng, J. H. C. P. (2021). Industry 4.0-artificial intelligence (AI) contribution to capability maturity. *Proceedings of the International Annual Conference of the American Society for Engineering Management*, 1–10.

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of big data – Evolution, challenges and research agenda. *International Journal of Information Management, 48*, 63–71. https://doi.org/10.1016/j.ijinfomgt.2019.01.021

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., & Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management, 57*, Article 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Dwivedi, R., Nerur, S., & Balijepally, V. (2023). Exploring artificial intelligence and big data scholarship in information systems: A citation, bibliographic coupling, and co-word analysis. *International Journal of Information Management Data Insights, 3*(2), Article 100185. https://doi.org/10.1016/j.jjimei.2023.100185

Egger, R., & Yu, J. (2022). A topic modeling comparison LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts. *Frontiers in Sociology, 7*. https://doi.org/10.3389/fsoc.2022.886498

Ellefsen, A. P. T., Oleśków-Szłapka, J., Pawłowski, G., & Toboła, A. (2019). Striving for excellence in AI implementation: AI maturity model framework and preliminary research results. *LogForum, 15*(3). https://doi.org/10.17270/J.LOG.2019.354

Facchini, F., Oleśków-Szłapka, J., Ranieri, L., & Urbinati, A. (2019). A maturity model for logistics 4.0: An empirical analysis and a roadmap for future research. *Sustainability, 12*(1), 86. https://doi.org/10.3390/su12010086

Felch, V., Asdecker, B., & Sucky, E. (2019). Maturity models in the age of Industry 4.0–Do the available models correspond to the needs of business practice?. In *Proceedings of the 52nd Hawaii international conference on system sciences. aisel.aisnet.org/hicss*.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., & Schafer, B. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Fukas, P., Rebstadt, J., Remark, F., & Thomas, O. (2021). Developing an artificial intelligence maturity model for auditing. In *Proceedings of the 29th European Conference on Information Systems (ECIS 2021)* (p. 133). aisel.aisnet.org/ecis2021_rp/13.

Gillespie, N., Lockey, S., Curtis, C., Pool, J., & Akbari, A. (2023). *Trust in artificial intelligence: A global study*. The University of Queensland; KPMG Australia. https://doi.org/10.14264/00d3c94

Goldberg, Y. (2017). Neural network methods for natural language processing. *Synthesis Lectures on Human Language Technologies, 10*(1), 1–309.

Gold, A. H., Malhotra, A., & Segars, A. H. (2001). Knowledge management: An organizational capabilities perspective. *Journal of Management Information Systems, 18*(1), 185–214. https://doi.org/10.1080/07421222.2001.11045669

Google. (2022). *Google's AI principles*. Available at: https://ai.google/principles/ (Accessed 16 November 2022).

Grootendorst, M. (2020). BERTopic: Leveraging BERT and c-TF-IDF to create easily interpretable topics. *Zenodo, Version v0, 9*.

Haefner, N., Wincent, J., Parida, V., & Gassmann, O. (2021). Artificial intelligence and innovation management: A review, framework, and research agenda✰. *Technological Forecasting and Social Change, 162*, Article 120392. https://doi.org/10.1016/j.techfore.2020.120392

Holmström, J. (2022). From AI to digital transformation: The AI readiness framework. *Business Horizons, 65*(3), 329–339. https://doi.org/10.1016/j.bushor.2021.03.006

Hradecky, D., Kennell, J., Cai, W., & Davidson, R. (2022). Organizational readiness to adopt artificial intelligence in the exhibition sector in Western Europe. *International Journal of Information Management, 65*, Article 102497. https://doi.org/10.1016/j.ijinfomgt.2022.102497

Hujran, O., Alarabiat, A., Al-Adwan, A. S., & Al-Debei, M. (2021). Digitally transforming electronic governments into smart governments: SMARTGOV, an extended maturity model. *Information Development.* , Article 02666669211054188. https://doi.org/10.1177/02666669211054188

IBM. (2021). *AI ethics*. Available at https://www.ibm.com/artificial-intelligence/ethics (Accessed 16 November 2022).

Jantunen, M., Halme, E., Vakkuri, V., Kemell, K. K., Rebekah, R., Mikkonen, T., Nguyen Duc, A., & Abrahamsson, P. (2021). Building a maturity model for developing ethically aligned AI systems. *Selected Papers of IRIS, 12*, 5. aisel.aisnet.org/iris2021/5/.

Jöhnk, J., Weißert, M., & Wyrtki, K. (2021). Ready or not, AI comes—An interview study of organizational AI readiness factors. *Business & Information Systems Engineering, 63*(1), 5–20. https://doi.org/10.1007/s12599-020-00676-7

Kempton, AM., & Vassilakopoulou, P. (2021). Accountability, transparency and explainability in AI for healthcare. In *Proceedings of the 8th International conference on infrastructures in healthcare 2019. European society for socially embedded technologies (EUSSET)*.

Kinkel, S., Baumgartner, M., & Cherubini, E. (2022). Prerequisites for the adoption of AI technologies in manufacturing–Evidence from a worldwide sample of manufacturing companies. *Technovation, 110*, Article 102375. https://doi.org/10.1016/j.technovation.2021.102375

Krijger, J., Thuis, T., de Ruiter, M., Ligthart, E., & Broekman, I. (2022). The AI ethics maturity model: A holistic approach to advancing ethical data science in organizations. *AI and Ethics*, 1–13. https://doi.org/10.1007/s43681-022-00228-7

Komatsu, A. Y., & Mantovani, D. (2021). Business intelligence maturity level in Brazilian companies. In *Proceedings of the American conference on information systems (AMCIS 2021)* (p. 2). aisel.aisnet.org/amcis2021.

Kotsialos, A., & Vassilakopoulou, P. (2023). Fleet management enterprise systems and traffic control synergies: A literature review and research agenda. *Procedia Computer Science, 219*(c), 529–536. https://doi.org/10.1016/j.procs.2023.01.321

Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021). Applications of text mining in services management: A systematic literature review. *International Journal of Information Management Data Insights, 1*(1), Article 100008. https://doi.org/10.1016/j.jjimei.2021.100008

Larsen, K. R., Hovorka, D., Dennis, A., & West, J. D. (2019). Understanding the elephant: The discourse approach to boundary identification and corpus construction for theory review articles. *Journal of the Association for Information Systems, 20*(7), 15. https://doi.org/10.17705/1jais.00556

Lasrado, L. A., Vatrapu, R., & Andersen, K. N. (2015). Maturity models development in IS research: A literature review. In *Proceedings of the 38th information systems research seminar in scandinavia (IRIS 38)* (p. 6) (August 9-12).

Lauer, D. (2021). You cannot have AI ethics without ethics. *AI and Ethics, 1*(1), 21–25. https://doi.org/10.1007/s43681-020-00013-4

Lee, A. S., Thomas, M., & Baskerville, R. L. (2015). Going back to basics in design science: From the information technology artifact to the information systems artifact. *Information Systems Journal, 25*(1), 5–21. https://doi.org/10.1007/s12599-020-00676-7

Lehmkuhl, T., Baumöl, U., & Jung, R. (2013). Towards a maturity model for the adoption of social media as a means of organizational innovation. In *Proceedings of the 46th Hawaii international conference on system sciences*. IEEE (pp. 3067-2076).

Lichtenthaler, U. (2020). Five maturity levels of managing AI: From isolated ignorance to integrated intelligence. *Journal of Innovation Management, 8*(1), 39–50. https://doi.org/10.24840/2183-0606_008.001_0005

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*, 1–25. https://doi.org/10.1007/s10796-021-10191-z

Lokuge, S., Sedera, D., Grover, V., & Dongming, X. (2019). Organizational readiness for digital innovation: Development and empirical calibration of a construct. *Information and Management, 56*(3), 445–461. https://doi.org/10.1016/j.im.2018.09.001

Makadok, R. (2001). Toward a synthesis of the resource-based and dynamic-capability views of rent creation. *Strategic Management Journal, 22*(5), 387–401. https://doi.org/10.1002/smj.158

Mäntylä, M. V., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review, 27*, 16–32. https://doi.org/10.1016/j.cosrev.2017.10.002

Martínez-Plumed, F., Gómez, E., & Hernández-Orallo, J. (2021). Futures of artificial intelligence through technology readiness levels. *Telematics and Informatics, 58*, Article 101525. https://doi.org/10.1016/j.tele.2020.101525

Meadows, M., Merendino, A., Dibb, S., Garcia-Perez, A., Hinton, M., Papagiannidis, S., Pappas, I., & Wang, H. (2022). Tension in the data environment: How organisations can meet the challenge. *Technological Forecasting and Social Change, 175*, Article 121315. https://doi.org/10.1016/j.techfore.2021.121315

Meske, C., Bunde, E., Schneider, J., & Gersch, M. (2022). Explainable artificial intelligence: Objectives, stakeholders, and future research opportunities. *Information Systems Management, 39*(1), 53–63. https://doi.org/10.1080/10580530.2020.1849465

Microsoft, AI. (2020). *Responsible AI principles from Microsoft*. Available at https://www.microsoft.com/en-us/ai/responsible-ai (Accessed 16 November 2022).

Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, 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., Zimmer, M. P., & Mäntymäki, M. (2022). Co-shaping an ecosystem for responsible AI: Five types of expectation work in response to a technological frame. *Information Systems Frontiers*, 1–19. https://doi.org/10.1007/s10796-022-10269-2

Molla, A., Cooper, V. A., & Pittayachawan, S. (2009). IT and eco-sustainability: Developing and validating a green IT readiness model. In *ICIS 2009 Proceedings* (p. 141). aisel.aisnet.org/icis2009/141/.

Molla, A., & Licker, P. S. (2005). Perceived e-readiness factors in e-commerce adoption: An empirical investigation in a developing country. *International Journal of Electronic Commerce, 10*(1), 83–110. https://doi.org/10.1080/10864415.2005.11043963

Mumford, E. (2006). The story of socio-technical design: Reflections on its successes, failures and potential. *Information Systems Journal, 16*(4), 317–342. https://doi.org/10.1111/j.1365-2575.2006.00221.x

Munn, L. (2022). The uselessness of AI ethics. *AI and Ethics*, 1–9. https://doi.org/10.1007/s43681-022-00209-w

Osoba, O., & Welser, W. (2017). *An intelligence in our image: The risks of bias and errors in artificial intelligence* (ed 1). Rand Corporation.

Pappas, I. O., Vassilakopoulou, P., Kruse, L. C., & Purao, S. (2023b). Practicing Effective Stakeholder Engagement for Impactful Research. *IEEE Transactions on Technology and Society*. https://doi.org/10.1109/TTS.2023.3296991 (Early Access).

Pappas, I. O., Mikalef, P., Dwivedi, Y. K., Jaccheri, L., & Krogstie, J. (2023a). Responsible digital transformation for a sustainable society. *Information Systems Frontiers, 25*, 945–953. https://doi.org/10.1007/s10796-023-10406-5

Pappas, I. O., Mikalef, P., Giannakos, M. N., Krogstie, J., & Lekakos, G. (2018). Big data and business analytics ecosystems: Paving the way towards digital transformation and sustainable societies. *Information Systems and e-Business Management, 16*(3), 479–491. https://doi.org/10.1007/s10257-018-0377-z

Pappel, I., Gelashvili, T., & Pappel, I. (2022). Maturity model for automatization of service provision and decision-making processes in municipalities. In *Proceedings of sixth international congress on information and communication technology* (pp. 399–409). Springer.

Pillai, R., & Sivathanu, B. (2020). Adoption of artificial intelligence (AI) for talent acquisition in IT/ITeS organizations. *Benchmarking: An International Journal, 27*(9), 2599–2629. https://doi.org/10.1108/BIJ-04-2020-0186

Prasad, R., Udeme, A. U., Misra, S., & Bisallah, H. (2023). Identification and classification of transportation disaster tweets using improved bidirectional encoder representations from transformers. *International Journal of Information Management Data Insights, 3*(1), Article 100154. https://doi.org/10.1016/j.jjimei.2023.100154

Pumplun, L., Tauchert, C., & Heidt, M. (2020). A new organizational chassis for artificial intelligence -Exploring organizational readiness factors. In *Proceedings of 27th European conference on information systems (ECIS 2019)* (p. 106). aisel.aisnet.org/ecis2019_rp/106/.

Rajendran, D. P. D., & Sundarraj, R. P. (2021). Using topic models with browsing history in hybrid collaborative filtering recommender system: Experiments with user ratings. *International Journal of Information Management Data Insights, 1*(2), Article 100027. https://doi.org/10.1016/j.jjimei.2021.100027

Renieris, E. M., Kiron, D., & Mills, S. (2022). *To be a responsible AI leader, focus on being responsible*. MIT Sloan Management Review, Boston Consulting Group. September 2022.

Russell, S., Haddad, M., Bruni, M., & Granger, M. (2010). Organic evolution and the capability maturity of business intelligence. In *AMCIS 2010 proceedings* (p. 501). aisel.aisnet.org/amcis2010/501/.

Ryan, S. D., Harrison, D. A., & Schkade, L. L. (2002). Information-technology investment decisions: When do costs and benefits in the social subsystem matter? *Journal of Management Information Systems, 19*(2), 85–127. https://doi.org/10.1080/07421222.2002.11045725

Sadiq, R. B., Safie, N., Abd Rahman, A. H., & Goudarzi, S. (2021). Artificial intelligence maturity model: A systematic literature review. *PeerJ Computer Science, 7*, e661. https://doi.org/10.7717/peerj-cs.661

Saltz, J. (2017). Acceptance factors for using a big data capability and maturity model. *ECIS 2022 Proceedings, 13*. aisel.aisnet.org/ecis2017_rip/13/.

Sánchez-Franco, M. J., & Rey-Moreno, M. (2022). Do travelers' reviews depend on the destination? An analysis in coastal and urban peer-to-peer lodgings. *Psychology & Marketing, 39*(2), 441–459.

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The sociotechnical axis of cohesion for the IS discipline: Its historical legacy and its continued relevance. *Mis Quarterly, 43*(3), 695–720. aisel.aisnet.org/misq/vol43/iss3.

Schuster, T., Waidelich, L., & Volz, R. (2021). Maturity models for the assessment of artificial intelligence in small and medium-sized enterprises. In *PLAIS EuroSymposium on Digital Transformation* (pp. 22–36). Springer.

Sharevski, F., Huff, A., Jachim, P., & Pieroni, E. (2022). (Mis) perceptions and engagement on Twitter: COVID-19 vaccine rumors on efficacy and mass immunization effort. *International Journal of Information Management Data Insights, 2*(1), Article 100059. https://doi.org/10.1016/j.jjimei.2022.100059

Shneiderman, B. (2020). Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS), 10*(4), 1–31.

Singh, V., Chen, S.-S., Singhania, M., Nanavati, B., Kar, A. K., & Gupta, A. (2022). How are reinforcement learning and deep learning algorithms used for big data based decision making in financial industries–A review and research agenda. *International Journal of Information Management Data Insights, 2*(2), Article 100094. https://doi.org/10.1016/j.jjimei.2022.100094

Someh, I., Wixom, B. H., Beath, C. M., & Zutavern, A. (2022). Building an artificial intelligence explanation capability. *MIS Quarterly Executive, 21*(2), 5. aisel.aisnet.org/misqe/vol21/iss2/5/.

Sternkopf, H., & Mueller, R. M. (2018). Doing good with data: Development of a maturity model for data literacy in non-governmental organizations. In *Proceedings of the 51st Hawaii international conference on system sciences*.

Tarafdar, M., Cooper, C. L., & Stich, J. F. (2019). The technostress trifecta-techno eustress, techno distress and design: Theoretical directions and an agenda for research. *Information Systems Journal, 29*(1), 6–42. https://doi.org/10.1111/isj.12169

Vakkuri, V., Kemell, K.-K., Jantunen, M., & Abrahamsson, P. (2020). 'This Is Just a Prototype': How ethics are ignored in software startup- like environments. In *International conference on agile software development* (pp. 195–210). Springer.

Vakkuri, V., Kemell, K.-K., Kultanen, J., Siponen, M., & Abra- hamsson, P. (2019). *Ethically aligned design of autonomous systems: Industry viewpoint and an empirical study* (p. 10), 48550/arXiv.1906.07946.

Vassilakopoulou, P. (2020). Sociotechnical approach for accountability by design in AI systems. In *Proceedings of the 28th European conference on information systems (ECIS 2020)* (p. 12). aisel.aisnet.org/ecis2020_rip/12/.

Vassilakopoulou, P., Parmiggiani, E., Shollo, A., & Grisot, M. (2022). Responsible AI: Concepts, critical perspectives and an information systems research agenda. *Scandinavian Journal of Information Systems, 34*(2), 3. aisel.aisnet.org/sjis/vol34/iss2/3/.

Williams, C., Schallmo, D., Lang, K., & Boardman, L. (2019). Digital maturity models for small and medium-sized enterprises: A systematic literature review. In *ISPIM conference proceedings* (pp. 1–15). The International Society for Professional Innovation Management (ISPIM).

Yablonsky, S. (2021). AI-driven platform enterprise maturity: From human led to machine governed. *Kybernetes*, 2753–2789. https://doi.org/10.1108/K-06-2020-0384

Zarindast, A., Sharma, A., & Wood, J. (2021). Application of text mining in smart lighting literature–An analysis of existing literature and a research agenda. *International Journal of Information Management Data Insights, 1*(2), Article 100032. https://doi.org/10.1016/j.jjimei.2021.100032

Zimmer, M. P., Minkkinen, M., & Mäntymäki, M. (2022). Responsible artificial intelligence systems critical considerations for business model design. *Scandinavian Journal of Information Systems, 34*(2), 4. aisel.aisnet.org/sjis/vol34/iss2/4/.