AI and Ethics (2025) 5:3265–3279
https://doi.org/10.1007/s43681-024-00653-w

ORIGINAL RESEARCH

AI governance: a systematic literature review

Amna Batool[1] · **Didar Zowghi**[1] · **Muneera Bano**[1]

Received: 21 July 2024 / Accepted: 12 December 2024 / Published online: 14 January 2025

Abstract
As artificial intelligence (AI) transforms a wide range of sectors and drives innovation, it also introduces different types of risks that should be identified, assessed, and mitigated. Various AI governance frameworks have been released recently by governments, organizations, and companies to mitigate risks associated with AI. However, it can be challenging for AI stakeholders to have a clear picture of the available AI governance frameworks, tools, or models and analyze the most suitable one for their AI system. To fill the gap, we present the literature to answer key questions: WHO is accountable for AI systems' governance, WHAT elements are being governed, WHEN governance occurs within the AI development life cycle, and HOW it is implemented through frameworks, tools, policies, or models. Adopting the systematic literature review (SLR) methodology, this study meticulously searched, selected, and analyzed 28 articles, offering a foundation for understanding different facets of AI governance. The analysis is further enhanced by categorizing artifacts of AI governance under team-level governance, organization-level governance, industry-level governance, national-level governance, and international-level governance. The findings of this study on existing AI governance solutions can assist research communities in proposing comprehensive AI governance practices.

Keywords Artificial intelligence · AI governance · Responsible AI · Ethical AI

1 Introduction

Artificial intelligence (AI) has emerged as one of the most important technologies in many businesses and has grown to be an integral part of our society [1, 2]. However, the risks and negative effects of AI are growing with its widespread application in a variety of sectors, such as autonomous cars [3], healthcare [4], finance [5], and other areas. Various repositories[1],[2] of AI incidents contain over 3000 AI incidents, illustrating the significant challenges associated with AI deployment.

Ethical and Responsible AI focuses on the development and implementation of AI systems in alignment with principles of fairness, accountability, transparency, and inclusivity [6]. Responsible AI focuses on the development and deployment of AI to minimize the potential risks and negative consequences associated with it, such as bias, discrimination, and a lack of transparency [1]. Ethical AI underscores adherence to moral principles in the design and utilization of AI systems, making sure AI systems don't unfairly treat people, invade privacy, or disrespect human dignity [6, 7]. Both ethical and responsible AI concepts aim to build trust with users and stakeholders and are important for the fair and lasting progress of AI technology [6].

Addressing AI risks goes beyond technical concerns and includes developing robust AI governance solutions and techniques, which are essential for directing the ethical and responsible use of AI technology. This balance between innovation and ethical behavior is crucial for avoiding unforeseen effects [8]. AI governance encompasses a set of regulations, methods, procedures, and technological mechanisms used to ensure that an organization's development

[1] https://incidentdatabase.ai.

[2] https://www.aiaaic.org/aiaaic-repository.

Didar Zowghi and Muneera Bano have contributed equally to this work.

✉ Amna Batool
amna.batool@data61.csiro.au

Didar Zowghi
didar.zowghi@data61.csiro.au

Muneera Bano
muneera.bano@data61.csiro.au

[1] CSIRO's Data61, Melbourne, Australia

and deployment of AI technologies align with its strategies, principles, and goals [8]. Consequently, governmental and international organizations, including the European Union AI Act,[3] the Organization for Economic Co-operation and Development [OECD],[4] Australian government,[5] National Institute of Standards and Technology (NIST),[6] International Organization for Standardization (ISO),[7] and professional bodies e.g., the Institute of Electrical and Electronics Engineers [IEEE] [9] have published their ethical AI principles and guidelines [7, 10] for the governance of AI.

There are existing systematic literature reviews (SLRs) on AI governance that have provided valuable insights into national and global strategies, advanced AI governance terminologies, and ethical guidelines. For example, Attard-Frost et al. [11] conducted a semi-systematic review that highlights Canada's pioneering national AI strategy and examines 84 governance initiatives, finding gaps in ethics statements and AI education while offering practical recommendations for national AI governance improvements. Another review by Mass [12] addresses the terminological ambiguities in the field of *Advanced AI Governance*, categorizing key definitions and concepts to enhance clarity and facilitate constructive dialogue among stakeholders. Similarly, Kluge Correa et al. [13] conducted a meta-analysis of 200 global governance policies and ethical guidelines, identifying 17 prevalent principles and stressing the importance of these principles in future regulatory frameworks.

Despite these comprehensive reviews, there is a gap in the literature concerning a detailed and layered analysis of AI governance across multiple governance levels using a systematic set of questions. In this paper, therefore, we fill this gap and present a systematic literature review that provides the state of the art on AI Governance. We have employed 4 specific questions to extract relevant information from the research literature on AI governance: who is governing (i.e., stakeholders), what should be governed (e.g., data and/or system), when is it being governed (i.e., at what stage of the AI development life cycle), and how is AI being governed (i.e., frameworks, models, tools, ethics: ethics, in the context of AI governance, refers to the set of moral principles and values that aim to guide the design, deployment, and use of AI systems to ensure responsible and fair outcomes [14, 15]. By embedding ethical principles such as transparency, accountability, and non-discrimination, AI governance structures aim to protect human rights and promote equitable outcomes. This ensures that AI technologies are not only effective but also aligned with societal values and ethical standards [7]). A comprehensive analysis of literature using these questions is presented in this paper, and the categorization of key elements is covered under different layers of governance from the study by Lu et al. [1]: team-level, organization-level, industry-level, national-level, and international-level. By doing so, we provide a more detailed and layered understanding of AI governance, offering a nuanced perspective that complements and extends existing reviews.

This study aims to summarize and synthesize current AI governance solutions (i.e., frameworks, tools, models, and policies), examine challenges in existing AI governance solutions, and offer insights based on the four specific questions (i.e., who, what, when, how). The main contributions of this study are:

- A comprehensive analysis of 28 research papers selected from the academic literature.
- An exploration of the challenges and limitations of existing AI governance solutions.
- The categorization of key elements presented under five levels of governance.This paper is organized as follows Sect. 2 presents the background and related work. Section 3 presents the research methodology along with research questions and data extraction. Section 4 presents the data analysis carried out on 28 selected studies and the categorization of AI governance solutions under five AI governance levels. Section 5 presents the discussion and Sect. 6 presents the threats to validity and study limitations. Section 7 concludes the study with future work.

2 Background and related work

As AI technologies advance at an unprecedented pace, the need for effective governance mechanisms becomes increasingly apparent [16]. In this section, we present a comprehensive review of existing literature on AI governance, with a focus on systematically analyzing and synthesizing the current state of knowledge in this area.

Correa et al. [17] conducted a review of 200 guidelines and recommendation documents on AI governance sourced from two repositories: the *AI Ethics Guidelines Global Inventory* by AlgorithmWatch and the *Linking Artificial Intelligence Principles* (LAIP) guidelines. The authors defined these guidelines as policy frameworks, tools, AI principles, and recommendations, serving as a guide for discussions on how AI can be regulated. Another study by Wang et al. [16] undertook a systematic literature review on AI governance in child social care, including a total of 440

[3] https://www.ey.com/en_ch/forensic-integrity-services/the-eu-ai-act-what-it-means-for-your-business.

[4] https://oecd.ai/en/ai-principles.

[5] https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles.

[6] https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

[7] https://www.iso.org/standard/81230.html.

articles using the PRISMA methodology. They categorized articles into four levels: individual, family, community, and governance. The study presented a conceptual framework that interlinks the AI-Technological layer, AI-Ethics layer, AI-Regulatory layer, and AI-Implementation layer, aiming to enhance child resilience, well-being, and public services.

There are emerging risk management frameworks, mandatory laws, and standards including the European Union's AI Act (EU AI Act),[8] the National Institute of Standards and Technology's AI Risk Management Framework (NIST AI RMF),[9] etc., been discussed for different purposes such as AI risk management, information security management [18], etc. A review by Barraza de la Paz et al. [19] presents an overview of emerging versions of the NIST Cyber Security Framework (CSF), ISO/IEC 27001:2022, and MAGERIT frameworks. The comparison of these frameworks in terms of their approaches to the identification, assessment, and treatment of risks has been highlighted. The authors discussed that ISO/IEC 27001:2022 focuses on a structured, process-oriented approach to managing information security by identifying assets, evaluating risks, and implementing controls. It is effective but can be costly and resource-intensive, especially for IoT and IoS systems. The NIST CSF is user-friendly, flexible, and self-assessing, making it suitable for various risk management needs. It offers publications like NIST SP 800–30, 800–37, and 800–39 for different aspects of information security. Lastly, the MAGERIT, developed by the Spanish government, uses a life cycle approach for public sector information security, covering asset identification, threat and vulnerability assessment, control implementation, and monitoring. Its comprehensive approach is ideal for large organizations but may be too complex for smaller ones. These frameworks discussed are particularly for information security management.

Kreutz and Jahankhani [20] conducted a systematic literature review (SLR) in which they explored 40 AI security challenges encompassing various aspects, contexts, and dimensions of AI security, along with 17 security defenses. Their comparative analysis of these AI security challenges and defenses with ISO 27001/27002 security controls revealed that most AI security challenges are either not addressed or only partially addressed by the current ISO standards. To address these gaps, they proposed six new security controls focused on real-time AI-enhanced defenses, AI lifecycle security, and AI governance as the core, along with AI explainability, AI privacy protections, and diversity. Moreover, they suggested ten modifications to existing controls to enhance capabilities and make information security management systems (ISMS) more dynamic.

[8] https://www.ey.com/en_ch/forensic-integrity-services/the-eu-ai-act-what-it-means-for-your-business.

[9] https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

There are existing studies that have researched AI governance targeting a specific domain. A study by Alsaigh et al. [21] employed a deep journalism methodology to uncover 15 parameters or themes for AI explainability and governance in smart energy systems. Their study aims to support various stakeholders, including governments, industry, academics, and the energy sector, in comprehending the AI landscape within the energy sector. The intended outcome is an enhancement of design, operation, utilization, and risk management practices for energy systems. Similarly, another study by Stogiannos et al. [22] executed a scoping review of AI governance frameworks to facilitate organizations in adopting AI in medical imaging and radiotherapy within the UK. The authors included a total of 35 studies on AI governance frameworks published in the last 5 years (2017–2022). Their study proposed a generic AI governance framework for medical imaging and radiotherapy, encompassing elements such as rigorous validation, ongoing monitoring, compliance with accreditation bodies, and adherence to ethical principles. Another study by Abbas et al. [23] carried out a meta-analytic systematic review on AI governance in higher education using the PRISMA methodology. The study employed a thematic analysis approach to explore trends and interconnections spanning AI, higher education, ethics, digital transformation, privacy, policy, and sustainability.

We have conducted a systematic literature review which differs from the existing review studies on the relevant topic in the following aspects:

- Focus of review: In this systematic literature review (SLR), an extensive analysis of AI governance elements has been presented using 4 specific questions (Who, What, When, How) to extract rich data about AI governance.
- Review timeline: This SLR covers publications from 2013 to 2023 (10 years).
- Type of included studies: This SLR followed the guidelines for the Systematic Literature Review established by Kitchenham et al. [24].
- Scope of subject domain: This SLR provides a comprehensive exploration of diverse sectors like media and communication, healthcare, transportation, robotics, academia, industry, etc. The existing reviews particularly focused on a single domain, either healthcare [25, 26], robotics [27], or energy sector [28].
- AI Governance and Stakeholders Classification: This SLR has classified the AI governance solutions identified from the question *How it is being governed* and the stakeholders found from the questions *Who is governing* into different levels, i.e., team, organization, industry, national, and international. The existing literature

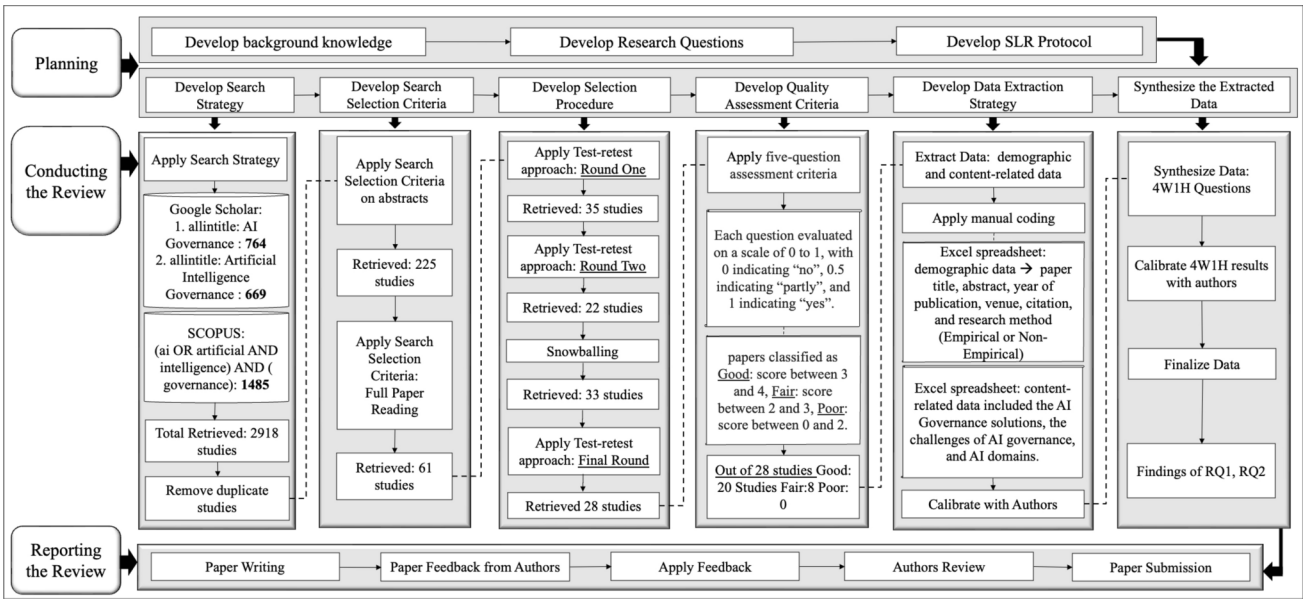

Fig. 1 Research methodology: planning, conducting the review, reporting the review

focused on a few of these levels but not on national or international levels [1].This SLR aims to contribute to a more nuanced and thorough understanding of AI governance, guiding future research towards comprehensive and responsible practices in the dynamic landscape of AI technology.

3 Systematic review methodology

This systematic literature review has been conducted following the guidelines established by Kitchenham et al. [24]. Figure 1 presents the steps followed, i.e., planning, conducting the review, and reporting the review. The complete SLR protocol is available as an online material.[10]

3.1 Planning

In this SLR, the relevant formal/academic literature selected are primary studies. This research was guided by the following two research questions:

- RQ1: What frameworks, models, tools, and policies (ethical AI principles, guidelines, policies) for AI governance are offered in the literature?
- RQ2: What are the limitations and challenges of AI governance discussed in the literature?

[10] https://docs.google.com/document/d/1GsHUo84XwjP1Oa4ne2d7El8G84VSl_g4/edit?usp=sharing&ouid=106380959154063406110&rtpof=true&sd=true.

Table 1 Search string in digital libraries and their corresponding results

Database	Search string	Search within	Time frame	Search time	No of papers returned
SCOPUS	(AI OR artificial AND intelligence) AND (governance)	Title and abstract	2013–2023	10/07/2023 (10:00 am)	1485
Google scholar	Allintitle: AI governance	Title only	2013–2023	21/07/2023 (9:00 am)	764
Google scholar	Allintitle: Artificial intelligence governance	Title only	2013–2023	21/07/2023 (9:00 am)	669
					Total: 2918

3.2 Search strategy

The keywords used to search for the articles in search engines and digital libraries include "AI", "artificial intelligence", and "governance". At the start, additional search terms like AI algorithms, regulations, or policy were added, that showed excessive unrelated results. We aimed to strike a balance between coverage and relevance. The final keywords used to collect the data from digital libraries are "AI", "artificial intelligence", and "governance" in Google Scholar and Scopus databases as shown in table 1. The choice of Google Scholar and Scopus was driven by their broad coverage of both peer-reviewed and grey literature, which we felt was important for capturing a comprehensive set of AI governance discussions. While Google Scholar does introduce some noise, its vast inclusion of academic

and non-academic sources (like conference papers and preprints) was useful for obtaining a wider perspective.

The search terms resulted in a total of 2918 papers. The search selection criteria have been applied to the abstracts of 2918 papers to retrieve the most relevant papers. The inclusion criteria of the year of publication and relevance were applied. The articles published in English between 2013 to 2023 in Google Scholar, and Scopus were considered while determining inclusion criteria. Moreover, the papers relevant to the research focus were considered for inclusion, while book chapters and reports were excluded. All other papers that did not fit these requirements were omitted. Consequently, publications focusing on such topics as environmental governance, climate change, ecology, etc., were excluded. The inclusion and exclusion criteria applied to abstracts resulted in 225 articles. The same selection criteria have been applied to full papers (225), which resulted in the selection of 61 papers after mostly excluding non-empirical papers, opinion papers, and those not fitting into AI Governance.

The selection procedure, i.e., the test-retest approach, was followed by inviting two experts of SLR to review the selected set of 61 papers in three rounds. The first round of review resulted in 35 studies after reviewing the research methodology used in each study. The second round of the selection procedure resulted in 22 studies after reviewing the full papers. There were secondary studies that were removed in both rounds. To make sure the systematic literature review was thorough, experts were consulted to decide when to stop the selection process. They advised stopping the selection and moving on to the next steps.

The snowballing approach (forward and backward) applied to 22 sets of studies resulted in 33 studies. The forward and backward snowballing helped in determining more primary studies. However, the final round of the selection procedure, in which two experts reviewed each set of papers, resulted in 28 papers, as again, there were a few secondary studies that were omitted. The selected set of 28 papers for this SLR is available in Appendix A.

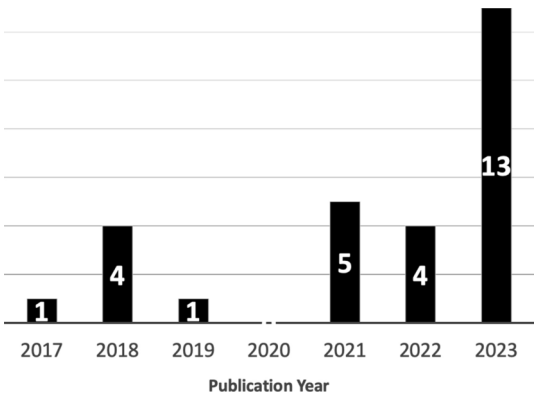

Fig. 2 Publication year of selected studies

3.3 Quality assessment

We have used the five-question criteria proposed by Liu et al. [29] to evaluate the quality of our selected studies. These questions evaluate the study's empirical or non-empirical nature, limitations and future work description, findings and contributions clarity, research design suitability, and clarity. Every question was rated from 0 to 1, where 0 meant "no," 0.5 meant "partially," and 1 meant "yes." The papers were categorized as Good if the score was between 3 and 4, Fair if it was between 2 and 3, and Poor if it was between 0 and 2. The total quality score was determined by adding the scores of the five questions. Out of the 28 selected papers, 20 were deemed "Good" quality, 8 were "Fair" quality, and 0 were "Poor" quality, demonstrating the robustness of this review.

3.4 Data extraction

Demographic and content-related information from the 28 chosen papers on AI governance was extracted using an Excel spreadsheet. Demographic data encompassed paper title, abstract, publication year, venue, and citation count. The demographic analysis of the publication years of 28 papers indicated that the highest number of studies, i.e., 13 (A1, A2, A7-A10, A14, A17, A19-A22, and A26), were published in 2023, and the lowest number of studies, i.e., 2, were published in 2017 (A27) and 2019 (A4) as shown in Fig. 2. The analysis of the citation count of 28 studies found that 3 studies (A13, A24, and A27) have received more than 200 to 500 citations. On the other hand, the remaining 25 studies got less than 100 citations. The Content-related data involved extracting AI governance solutions (frameworks, models, tools, policies, and guidelines), answers to 4 questions (who, what, when, how), and the AI governance challenges discussed in the studies. The first author performed manual coding for data extraction, which was cross-checked in weekly meetings with the other authors.

4 Data synthesis and analysis

This section presents the analysis of 28 selected studies by answering the 4 specific questions i.e., who is governing, what should be governed, when is it being governed, and how is AI being governed, and also presents the categorization of stakeholders based on *Who is governing?* and artifacts based on *How is it governed?* under five levels of AI governance i.e., team, organization, industry, national, and

international. Appendix B has the terminology to define the terms used in this SLR.

4.1 Analysis

Our analysis approach consists of questions, Who, What, When, and How, which have been explored for the selected set of 28 studies using thematic analysis. This analysis involved a comprehensive examination of the literature to identify recurring themes related to AI governance. Each question has been designed to cover one of the four pillars of an AI ecosystem, i.e., humans, data, systems, and processes, as proposed by Zowghi and da Rimini [30]. The human pillar covers the question, i.e., Who is governing? The data and system pillars cover the question, What is being governed? and the process pillar covers the question, When is it being governed? In an AI ecosystem, the human pillar plays an important role, as without human involvement, data, systems, or processes cannot be governed, and all four of these pillars are key elements of governance in an AI ecosystem [30]. The following are the questions:

- Who is governing?: focuses on the stakeholders (key AI oversight roles) who should be involved in governing to oversee if AI complies with standards or regulations for ethical and responsible AI.
- What is being governed?: focuses on data and system to ensure what element either data or/and system is being governed in AI governance. The *data* element emphasizes the importance of responsible data acquisition, processing, and utilization [30]. Effective AI governance requires a focus on data quality, privacy, and security to mitigate biases and promote fair and transparent AI decision-making [30]. The *system* encompasses the AI systems, which should be governed in consideration of their robustness, interpretability, and accountability to guarantee their responsible and reliable operation [30]. While there are other elements such as “algorithms" and “applications" that are indeed significant categories, particularly in the EU AI Act [31], this study chose not to categorize them separately because these elements are inherently embedded within broader system governance. The governance of AI systems often encompasses algorithmic transparency and oversight as part of its robustness and accountability measures. Similarly, applications are sector-specific manifestations of AI systems, and their governance largely falls under the operational controls and ethical considerations of the broader system. By focusing on data and systems, our analysis captures the foundational elements that underpin AI governance across multiple domains, providing a more holistic view of the ethical and operational challenges inherent in AI development and deployment.
- When is it being governed?: represents the three development stages of AI to which the particular governance solutions should be applied [30]. The three stages of AI development are the pre-development (planning, approval, and data collection), during-development (design, build models, verify, and validate), and post-development (monitor, and use), in alignment with established frameworks and guidelines from organizations such as NIST [32], OECD [33].
- How is it governed?: explores the artifacts of governance solutions for AI, such as ethical policies, principles and guidelines, tools, frameworks, and models.

Table 2 Who is governing?

Humans (Stakeholders)		
Who is governing?		
P-ID	Key oversight roles	Levels
A2	AI ethics committees by government	National level
A3	A global governance coordinating committee GGCC	International level
A11	Multi-disciplinary steering committee along with project-specific sub-committees	Organization and team level
A13	Public managers	Organization level
A23	World Health Organization and International Telecommunication Union state members	International level
A28	Hospital administrators	Organization level

4.1.1 Comprehensive analysis of questions

Who is governing?: Table 2 presents the answers to the question: who is governing? Six out of the 28 studies answered the question, and the stakeholders (key oversight roles) lie at various levels. Study [A2] mentions that national-level key oversight roles (AI ethics committee developed by the governments) should govern the higher-level ethics committee organized by the companies to ensure that the risks associated with medical AI are minimized and that it is developed and deployed following AI standards and regulations [14]. The key factor discussed in the study [A2] is that multi-participation and international cooperation is essential in medical AI governance because the risks associated with medical AI are global. These risks can only be mitigated through international communication, which can be achieved if governments take responsibility by establishing an AI ethics committee after collecting multiple opinions [14]. Similarly, study [A3] has emphasized that international-level stakeholders i.e., the Global Governance Coordinating Committee (GGCC) should take the responsibility of governing regional regulatory bodies.

This ensures that they develop AI free of ethical and societal risks. The study [A3] indicates that the GGCC is an extended version of the Governance Coordinating Committee (GCC). It can propose a better governance mechanism for national and regional bodies to develop their AI systems free of risks [34]. Another study [A23] has also given importance to international communication and developed a Focus Group for Healthcare AI (FG-AI4H). In this group, international-level stakeholders i.e., members of the World Health Organization and the International Telecommunication Union can govern health and care service providers to ensure that healthcare AI is developed and deployed following AI standards and policies [35].

Three studies (A2, A3, A23) have focused on involving national and international level stakeholders in AI governance. This approach assists in formulating optimal AI governance mechanisms, engaging individuals from various national and international bodies to take responsibility for governing AI practitioners at the team, organization, and national levels. The remaining three studies (A11, A13, and A28) focused on involving a multi-disciplinary committee, public managers, and administration staff at the organizational level to govern faculty staff, firm managers, and healthcare providers and insurers. All three studies (A11, A13, and A28) have mentioned the key roles of governance in healthcare AI. Notably, no study among these 6 has focused on industry-level governance. As discussed earlier, the classification of stakeholders under different levels i.e., team, organization, industry, national and international has been done based on the study by Lu et al. [1].

AI governance varies significantly across regions, reflecting distinct geopolitical priorities. In Europe, the focus is on human rights and data protection, with frameworks like the General Data Protection Regulation (GDPR) [36] and the EU AI Act [31] emphasizing fairness, transparency, and accountability. European regulations aim to ensure AI systems uphold individual rights and do not contribute to inequalities or biases. In contrast, the United States adopts a market-driven approach, with sector-specific regulations and voluntary guidelines such as the NIST AI Risk Management Framework [37], emphasizing innovation, trustworthiness, and economic growth, with less centralized oversight. In the Asia-Pacific (APAC) region, there is considerable diversity: China follows a government-driven approach, using AI for social governance and state security, while countries like Singapore and Japan focus on responsible AI development, balancing ethics with technological advancement. Singapore's Model AI Governance Framework highlights accountability and transparency, while Japan's human-centric AI principles emphasize inclusivity and ethical use [38]. Australia, similarly, takes a proactive approach with its AI Ethics Framework, focusing on transparency, fairness, and preventing discrimination, aligning with ethical principles while promoting innovation [39]. These regional differences reflect underlying geopolitical biases, with Europe prioritizing ethical standards, the U.S. focusing on innovation and market freedom, and APAC countries balancing state control and economic development. Australia's framework, which emphasizes both ethical standards and innovation, offers a middle ground. Understanding these geographical nuances adds depth to the analysis of "who" governs AI and how governance structures are shaped by regional objectives.

What is governed?: *Data*: Nine studies have emphasized the need to govern data. Study [A1] highlighted that governing data is important to reduce the several challenges of security and privacy when collecting data using AI in the media and communication sector. Another study [A2] reveals that the unstructured nature and lack of standardization in medical data impact the quality of AI models due to which governing data is essential. The study [A14] suggested that data should be governed, particularly focusing on ethical and responsible data stewardship, governance, and accountability. The findings of the study [A21] emphasized that data governance significantly influences policy choices across different governance approaches. Another study [A23] focused on governing data is crucial for ensuring the reliability and effectiveness of AI systems in healthcare. The study [A27] highlighted the importance of governing data by arranging a workshop where the authors identified challenges like inefficiencies, biases, lack of transparency, and unfairness, due to which the authors emphasized the governance of data is important. Similarly, a study [A28] explored the significance of data governance, particularly in the context of implementing Responsible AI in healthcare. The remaining two studies [A12 and A13] emphasized governing data as well as systems which are covered later in this section.

System: Six studies have emphasized the need to govern the system. Study [A3] highlighted the need for governance of AI systems to mitigate various risks including ethical concerns related to military applications, ensuring safety in AI systems, addressing privacy issues, etc., associated with its rapid development. Similarly, another study [A5] also highlighted several risks associated with the deployment of autonomous systems in the Long-Term Care (LTC) sector in Singapore. These risks encompass safety, privacy, data security, and liability concerns, due to which the study emphasized the need for the governance of AI systems. Another study [A11] highlighted the importance of governing AI systems in healthcare to mitigate several risks including concerns related to patient safety, accountability, clinician trust, and the successful adoption of AI in healthcare. The findings of paper [A25] indicated that reducing ethical

issues especially the transparency in AI systems is essential and this can be done by governing AI systems using different governance mechanisms such as Ethical Considerations and Challenges of Learning Algorithms (ECCOLA) a tool for the governance of ethical and trustworthy AI systems. Overall, these studies focused on various risks associated with AI which emphasized that governing AI system is essential.

Data and System: There are 2 studies (A12, A13), common to both pillars: data and system, which emphasized the importance of governing both data and systems. Study [A12] highlighted that the rapid advancement of AI-driven technology in healthcare surpasses the global capacity for AI systems and data governance. Similarly, study [A13] emphasized that governing both data and system is essential to mitigate the seven dimensions of challenges i.e., social, economic, ethical, political/legal/policy-related, organizational/managerial, data, and technological challenges, associated with the adoption of AI in healthcare.

When it is being governed?: Table 3 presents the studies that answer the question, i.e., when it is being governed. As mentioned earlier, there are three stages of the AI development life cycle under the process pillar i.e., pre-development, during-development, and post-development [30].

Eight studies emphasize that AI governance solutions must cover different AI development stages. Five of these studies [A2], [A12], [A19], [A23], and [A25] emphasized that governance solutions should cover all stages of an AI development life cycle to ensure ethical development and deployment of AI. Study [A12] highlighted that there is a need for comprehensive policy frameworks covering the entire AI life cycle, including design, building, deployment, testing, and monitoring. Study [A19] highlighted a gap, namely, the lack of information robustness in ECCOLA. Consequently, the study suggests that governance measures should be implemented at all stages of an AI development life cycle to reduce the issues of information robustness.

A study [A7] suggested that involving humans at the AI stage, i.e., during development, for population health enhances governance practices and can result in the reduction of ethical issues in AI in population health. The study [A11] has highlighted the ethical challenges (patient safety, accountability, clinician trust, etc.) and mentioned that these challenges can be reduced if governance practices occur at the pre-development and post-development stages of an AI development life cycle. These studies emphasized the necessity of governance, particularly in addressing various ethical concerns, and indicated the importance of a comprehensive governance structure across the different stages of the AI development life cycle to ensure ethical and responsible AI deployment.

Table 3 When it is being governed?

Process		
Pre-development	During development	Post-development
A2	A2	A2
–	A7	–
A11	–	A11
A12	A12	A12
–	A13	–
A19	A19	A19
A23	A23	A23
A25	A25	A25

Table 4 How is it governed?

P-ID	Framework/models/tools/ethics (principles, policies, guidelines)	Type
A2	Ethical AI Governance: Ethical values (top-level) lead to Ethical Principles which guide the Ethical Norms	Ethical principles
A12	AI Governance through Policies: Design Phase Policies, Testing and Validation Policies, Deployment Policies	Ethical policie
A11, A13	AI Governance through Guidelines: (1) Avoid "Vision Lock-In" (2) Use Adaptive Governance (3) Prioritize Data and Management Guidelines (4) Focus on Governance of AI - Value stream structure to operationalize ethical guidelines	Ethical guidelines
A1	Media-AI Governance Framework: Multi-level framework: automating data processing and capture, automating content creation, automating content moderation, and automating communication	Framework
A20	Theoretical AI Governance Framework	Framework
A10	Dimensional AI Governance Model: AI dimensions: structural, relational, and procedural	Model
A19	Extension of ECCOLA - A tool developed to embed AI Ethics	Tool
A25	ECCOLA: A tool developed to embed AI Ethics	Tool

How is it governed?: Table 4 presents the studies answered to the question of How is it governed. Nine studies have proposed different solutions (frameworks, models, tools, ethical principles, policies, or guidelines) for the governance of AI.

Ethics (Principles, Policies, Guidelines): There are 4 studies [A2], [A11-A13] proposed different ethical policies, guidelines, and principles to govern AI in different domains. The study [A2] proposed comprehensive governance countermeasures for trustworthy medical AI, emphasizing ethical, legal, and regulatory aspects. The proposed solution includes ethical values as a top-level design that leads to ethical principles that guide the ethical norms, emphasizing that AI technology must comply with medical ethics. The study [A12] discovered the range of policies span various

phases of the AI life cycle within the Global Digital Health Partnership. These policies include; *Design Phase Policies*: to introduce hard governance mechanisms. *Testing and Validation Policies*: should be flexible, appropriate, and adaptable for proving the efficacy of AI-driven technologies. *Deployment Policies*: focusing on policy development to address gaps in translating research into clinical practice. *Monitoring and Oversight Procedures*: to enhance national oversight procedures to improve collective intelligence at an international level.

The study [A11] highlighted the crucial need for effective governance in the rapidly growing adoption of AI applications in healthcare, focusing on ethical, regulatory, and trust concerns. The authors emphasized the importance of comprehensive and adaptive governance at the hospital level to address these issues and facilitate the deployment and adoption of predictive solutions. The next study [A13] used a qualitative approach in exploring the challenges of adopting AI in public healthcare, and based on the insights gained, it proposed sets of guidelines focusing on the governance of AI. The study highlighted that rather than replacing human decision-makers with AI, the focus should be on governing AI effectively.

Frameworks: There are two studies [A1], [A20] that have proposed frameworks to govern AI in different domains.

The study [A1] proposed a four-level AI governance framework to identify risks associated with AI in the media and communication (MC) sector: automating data capture and processing, content generation, content mediation, and communication. The next study [A20] mentioned that the fairness factors impact user experience, influencing aspects like unbiased treatment, transparency, and efficiency in work processes, due to which the study has introduced a theoretical framework in which AI technology governance encompassing accountability and fairness factors are explained.

Model: Only one study [A10] proposed an AI governance model that encapsulates structural, procedural, and relational components as key elements. The primary benefits of the model highlighted in the study include gaining a competitive advantage, reducing costs, and establishing reliable AI systems, all crucial for business success, especially in competitive markets. However, the study mentioned a few limitations including potential bias in data collection from companies with limited use of sensitive data, a snapshot approach that may not fully capture evolving practices, and the lack of sector diversity, raising concerns about generalizability.

Tools: The study [A25] proposed the ECCOLA (Ethical Considerations and Challenges of Learning Algorithms) tool to govern AI through ethical policies bridging the gap between AI ethics research and industry practices. It has three primary goals: raise awareness of AI ethics, offer a modular method suitable for diverse Software Engineering contexts, and be adaptable to agile development processes for the development of AI. The study [A19] highlighted a gap in ECCOLA i.e., lack of information robustness, requiring the need for additional governance measures to govern AI effectively. Due to this, the study suggested aligning it with Generally Accepted Recordkeeping principles (GARP) information governance (IG) practices. The analysis of the study revealed that ECCOLA in alignment with GARP IG practices improves its adaptability and reduces the gaps in information robustness.

In summary, studies [A2], [A11], [A12], [A13] under the category of *ethics* have explored various ethical governance countermeasures, encompassing guidelines and policies to ensure the ethical development of AI. However, except for study [A11], these studies lack real-world implementation details, potentially which could limit their applicability across diverse contexts. The ethical AI principles emphasized in two studies [A1] and [A20] under the category of AI governance frameworks are fairness and transparency. However, the frameworks introduced in the studies exhibit a potential limitation in terms of proper structure and validation, raising questions about the comprehensive evaluation of key findings to justify the frameworks' validity.

Three studies [A2], [A11], and [A13] have answered all 4 questions (who, what, when, and how). Comprehensive details on who should be accountable for the AI governance system, what elements need to be governed, when governance must occur within the AI development life cycle, and how it should be implemented through frameworks, tools, policies, or models are outlined in studies [A2], [A11], and [A13].

4.2 AI governance levels

Utilizing the categories suggested by Lu et al. [1], the stakeholders found under the question *Who is governing?* in this SLR is classified under the five groups, as illustrated in Table 2 and explained in Sect. 4.1.1 (Who is governing). AI governance solutions found under the question *How is it governed?* are grouped under team-level, organization-level, industry-level, national-level, and international-level governance. Figure 3 shows the categorization of the AI governance solutions under these levels. Unfortunately, none of the solutions lies under team-level and national-level governance therefore, three layers are added to Fig. 3.

The highest number (i.e., 7) of AI governance solutions are under organizational-level governance. The 3 studies [A2], [A11], and [A13], which answered all questions, lie under organizational-level governance. This reveals that these solutions are comprehensive and clear on who should be responsible for regulating the AI systems, what elements

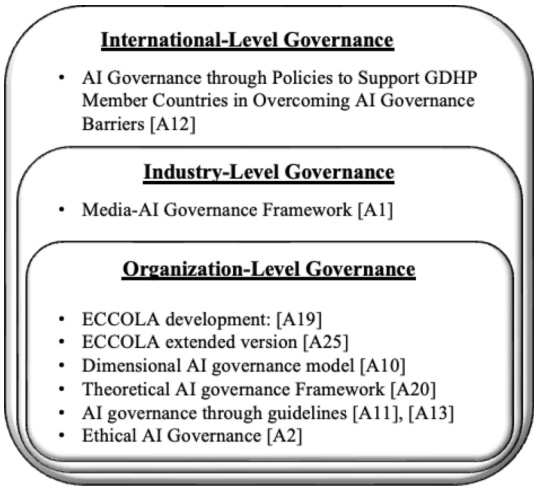

Fig. 3 AI Governance Levels

Table 5 Ethical AI principles targeted in AI governance solutions

P-ID	AI governance solutions	Principles
A2	Ethical AI Governance	Fairness privacy and security
A12	AI Governance through policies	Not specified
A11, A13	AI Governance through guidelines	Interpretability, accuracy, and fairness
A1	Media-AI Governance framework	Transparency
A20	Theoretical AI Governance framework	Fairness
A10	Dimensional AI Governance model	Not specified
A19	Extension of ECCOLA	Privacy and robustness
A25	ECCOLA development	Robustness

should be governed at each development stage, and how to implement them through frameworks, tools, or policies. There is one AI governance solution in the study [A12] that lies under international-level governance.

It is important to recognize that, alongside organizational governance, the collaboration between national and international governance plays a pivotal role [34]. Coordinating national bodies with international efforts enables the formulation of best AI governance solutions, ensuring responsible behavior and ethical standards in the global AI landscape [34, 40].

5 Discussion

The analysis of AI governance frameworks, models, tools, and policies offered in the literature reveals important trends and gaps in addressing ethical challenges associated with AI technologies.

RQ1: What governance frameworks, models, tools, and policies (ethical policies, guidelines, principles) for AI are offered in the literature?

Table 4 presents AI governance frameworks, models, tools, and policies (ethical principles, policies, guidelines) offered in the literature. The comprehensive analysis of each solution has already been covered in Sect. 4.1.1. As 9 studies have proposed various AI governance solutions, including frameworks, tools, and ethics principles, each study has been examined to observe which ethical AI principles as mentioned in the specific study they target to make AI ethical and responsible across different domains. The studies on governance solutions have primarily focused on just one or two ethical AI principles, with the most frequently focused principle being fairness, followed by privacy as shown in Table 5. While it is understandable that very few studies have focused on ethical principles while proposing AI governance frameworks, models, tools, etc., it is essential to introduce more structured and robust governance solutions that make AI responsible and ethical [1].

RQ2: What are the limitations and challenges of AI governance discussed in the literature?

AI governance solutions are required to mitigate the risks associated with AI, such as bias, discrimination, lack of transparency, and unintended harmful impacts [39] [30]. However, there are challenges associated with AI governance solutions as well, which are discussed in the literature. There are AI governance solutions proposed in the literature that have not covered various ethical aspects including fairness, transparency, privacy, and trust, which makes these solutions non-robust and can not be used to make AI responsible and ethical [A6], [A15], [A22], and [A24]. Here are key findings on ethical challenges discovered in the literature:

1. **Less attention to ethical and responsible AI principles**: The literature notes a perceived lack of attention given to ethical and responsible AI principles in existing governance efforts. Many existing governance frameworks prioritize compliance and risk management, often focusing on technical and operational aspects such as performance, accuracy, and scalability. However, this technical focus can overshadow essential ethical concerns like fairness, transparency, and inclusiveness. For example, in some cases, ethical guidelines are merely “suggestions" with no enforceable regulations to ensure compliance, leading to ethical principles being either overlooked or poorly integrated into AI systems. Additionally, the lack of a unified, standardized approach

to responsible AI across different sectors and regions makes it challenging to adopt ethical AI principles universally. This inconsistency creates gaps, where organizations might adhere to the minimal legal requirements but fail to implement more robust ethical standards, potentially leading to bias, discrimination, or violations of privacy. The absence of enforceable ethical norms allows companies to bypass these principles, particularly in competitive markets where time-to-market and profitability are prioritized over ethical considerations [A2, A6, A15, A22, A24].

2. **Lack of human involvement:**: The lack of human involvement in current AI governance approaches is discussed as a challenge that makes these approaches less human-centric. Many AI governance frameworks focus on automating decision-making processes, often sidelining the role of human judgment and accountability in these decisions. As AI systems increasingly handle sensitive and high-stakes tasks, such as hiring, law enforcement, and healthcare, the absence of human oversight can lead to unintended and harmful outcomes. Furthermore, human involvement is crucial to ensure that AI systems remain adaptable and responsive to changing ethical, social, and legal contexts. Without humans in the loop, AI systems might follow rigid rules and algorithms, which could be inappropriate or harmful when applied to evolving real-world situations [A6-A7, A24].
3. **Inadequacy of current governance approaches**: Some studies suggest that current AI governance frameworks may be insufficient to address all AI ethical concerns. While some frameworks emphasize risk management and regulatory compliance, they often fall short in addressing broader concerns related to trust, privacy, and accountability. This opacity can lead to "black box" models where the logic behind AI decisions is inaccessible, undermining trust in AI systems. Additionally, AI technologies are developed and deployed across borders, but there is no global standard or agreement on how to govern AI systems ethically. This lack of coherence makes it difficult to establish a comprehensive governance system that can mitigate the risks posed by AI across different regions and industries [A6, A24, A15, A22].By highlighting these challenges, it becomes evident that AI governance solutions should be fundamentally based on ethical and responsible AI principles for fostering trust, addressing societal concerns, and ensuring the long-term success of AI technologies [41–43]. As AI continues to integrate into high-stakes areas like healthcare, law enforcement, and employment, governance frameworks must evolve to prioritize ethical principles such as fairness, transparency, and accountability, which are currently under-addressed. For instance, the EU AI Act, which seeks to classify AI systems based on risk levels and impose strict regulations on high-risk applications, exemplifies this movement toward a more responsible AI landscape [31]. It is anticipated that frameworks like these will increasingly focus on human oversight, ensuring that AI systems not only meet technical standards but also align with societal values. Ultimately, the future of AI governance can involve balancing innovation with responsibility, with adaptable frameworks that can respond to evolving AI technologies while safeguarding ethical principles.

6 Threats to validity and limitations

6.1 Limitations

Although we have rigorously adhered to the comprehensive search strategy, ensuring a comprehensive selection of our studies, there's still a possibility that certain papers might not have been incorporated into our data collection. This may result from their inaccessibility or non-existence on electronic platforms, of which we might be unaware or we have not targeted different digital libraries and focused on only Google Scholar and Scopus. In the creation of our search strings, the key terms "AI regulations", "AI ethics", "AI regulatory frameworks", and "AI governance models" are omitted to minimize a large number of unrelated results. While we recognize this could have excluded certain relevant papers from our sample, we performed snowballing to recover any missing studies.

6.2 Internal validity

Our systematic literature review (SLR) includes primary studies published over the last ten years, up until 2023. As 2024 has already begun, we plan to incorporate studies from this year in our next research cycle. By focusing exclusively on primary studies, we acknowledge that our review might exclude various AI governance frameworks or models currently in use by organizations, which are often detailed in secondary sources such as white papers, or industry reports. Limiting our scope to the last ten years helps to manage the influence of rapidly evolving AI governance practices, ensuring a more consistent comparison of studies over a defined period. To maintain internal validity, we have defined clear inclusion and exclusion criteria, addressed potential biases by focusing on primary studies, and considered future updates to include newer studies. These measures ensure that our findings are robust, reliable,

and attributable to the primary studies reviewed, thereby enhancing the internal validity of our SLR.

6.3 External validity

The findings of this SLR provide a comprehensive summary of current AI governance solutions. However, the generalizability of these findings is limited by the scope of the selected studies, which may not capture the full spectrum of AI governance practices across diverse contexts. To manage this challenge, we tried to include a diverse range of primary studies and incorporate an ongoing analysis of grey literature to broaden our perspective.

7 Conclusion and future work

In conclusion, this systematic literature review of 28 selected studies achieves its objectives by providing a comprehensive summary of current AI governance solutions, including frameworks, tools, models, and ethical policies. The challenges within existing AI governance solutions are also presented in this SLR, emphasizing the prevalent focus on ethical considerations such as fairness, transparency, privacy, etc. The analysis reveals a critical gap in addressing the who, what, and when aspects of AI governance in a holistic manner. Only 3 studies have provided all the answers to questions (who is governing, what is being governed, when it is being governed, and how it is being governed). The analysis has also revealed that the highest number of governance solutions are classified under organizational-level governance. While ethical principles play a vital role, there is a call for enhanced clarity, especially in stakeholder involvement at different AI development stages. The findings emphasize the imperative for AI governance solutions to address not only risks but also crucial ethical aspects, promoting responsible and effective governance in the dynamic landscape of artificial intelligence. This indicates the need for future AI governance solutions to closely align with ethical AI principles, fostering a holistic and principled approach to the governance of artificial intelligence.

As part of future work, an extended analysis delving into grey literature is ongoing. The decision to extend our SLR to grey literature is because we have identified a gap in the existing academic literature. Specifically, the primary studies included in our systematic literature review (SLR) did not adequately cover standards of AI governance, such as ISO standards, the NIST AI governance framework, etc. We have completed the collection and analysis of the grey literature on AI governance. This exploration aims to extract additional insights, contributing to a more nuanced understanding of AI governance and enriching the discourse on AI governance practices, their implications, and ethical considerations.

Appendix A: List of 28 included studies

A1: Pierson, J., Kerr, A., Robinson, S.C., Fanni, R., Steinkogler, V.E., Milan, S. and Zampedri, G., 2023. Governing artificial intelligence in the media and communications sector. Internet policy review, 12(1), pp.28–28.

A2: Zhang, J. and Zhang, Z.M., 2023. Ethics and governance of trustworthy medical artificial intelligence. BMC medical informatics and decision making, 23(1), p.7.

A3: Wallach, W. and Marchant, G.E., 2018. An agile ethical/legal model for the international and national governance of AI and robotics. Association for the Advancement of Artificial Intelligence.

A4: Arthur, K.N.A. and Owen, R., 2022. A micro-ethnographic study of big data-based innovation in the financial services sector: Governance, ethics and organisational practices. In Business and the ethical implications of technology (pp. 57–69). Cham: Springer Nature Switzerland.

A5: Tan, S.Y. and Taeihagh, A., 2021. Governing the adoption of robotics and autonomous systems in long-term care in Singapore. Policy and society, 40(2), pp.211–231.

A6: Dickinson, H., Smith, C., Carey, N. and Carey, G., 2021. Exploring governance tensions of disruptive technologies: the case of care robots in Australia and New Zealand. Policy and Society, 40(2), pp.232–249.

A7: Couture, V., Roy, M.C., Dez, E., Laperle, S. and Bélisle-Pipon, J.C., 2023. Ethical implications of artificial intelligence in population health and the public's role in its governance: perspectives from a citizen and expert panel. Journal of Medical Internet Research, 25, p.e44357.

A8: Hohma, E., Boch, A., Trauth, R. and Lütge, C., 2023. Investigating accountability for Artificial Intelligence through risk governance: A workshop-based exploratory study. Frontiers in Psychology, 14, p.1073686.

A9: O'Shaughnessy, M.R., Schiff, D.S., Varshney, L.R., Rozell, C.J. and Davenport, M.A., 2023. What governs attitudes toward artificial intelligence adoption and governance?. Science and Public Policy, 50(2), pp.161–176.

A10: Papagiannidis, E., Enholm, I.M., Dremel, C., Mikalef, P. and Krogstie, J., 2023. Toward AI governance: Identifying best practices and potential barriers and outcomes. Information Systems Frontiers, 25(1), pp.123–141.

A11: Liao, F., Adelaine, S., Afshar, M. and Patterson, B.W., 2022. Governance of Clinical AI applications to

facilitate safe and equitable deployment in a large health system: Key elements and early successes. Frontiers in Digital Health, 4, p.931439.

A12: Morley, J., Murphy, L., Mishra, A., Joshi, I. and Karpathakis, K., 2022. Governing data and artificial intelligence for health care: developing an international understanding. JMIR formative research, 6(1), p.e31623.

A13: Sun, T.Q. and Medaglia, R., 2019. Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. Government Information Quarterly, 36(2), pp.368–383.

A14: Okun, S., Hanger, M., Browne-James, L., Montgomery, T., Rafaloff, G. and van Delden, J.J., 2023. Commitments for Ethically Responsible Sourcing, Use, and Reuse of Patient Data in the Digital Age: Cocreation Process. Journal of Medical Internet Research, 25, p.e41095.

A15: Mökander, J., Sheth, M., Gersbro-Sundler, M., Blomgren, P. and Floridi, L., 2022. Challenges and best practices in corporate AI governance: Lessons from the biopharmaceutical industry. Frontiers in Computer Science, 4, p.1068361.

A16: Bratton, B., 2021. AI urbanism: a design framework for governance, program, and platform cognition. AI and SOCIETY, 36(4), pp.1307–1312.

A17: Al Zaabi, K., Hammadi, K.A. and El Khatib, M., 2023. Highlights on Program Governance through AI and Blockchain. International Journal of Business Analytics and Security (IJBAS), 3(1), pp.92–102.

A18: SchÃ¤fer, M., Schneider, J., Drechsler, K. and vom Brocke, J., 2022. AI Governance: Are Chief AI Officers and AI Risk Officers needed?. In ECIS.

A19: Agbese, M., Alanen, H.K., Antikainen, J., Erika, H., Isomaki, H., Jantunen, M., Kemell, K.K., Rousi, R., Vainio-Pekka, H. and Vakkuri, V., 2023. Governance in ethical and trustworthy AI systems: Extension of the ECCOLA method for AI ethics governance using GARP. e-Informatica Software Engineering Journal, 17(1).

A20: Malik, N., Kar, A.K., Tripathi, S.N. and Gupta, S., 2023. Exploring the impact of fairness of social bots on user experience. Technological Forecasting and Social Change, 197, p.122913.

A21: Tan, E., Jean, M.P., Simonofski, A., Tombal, T., Kleizen, B., Sabbe, M., Bechoux, L. and Willem, P., 2023. Artificial intelligence and algorithmic decisions in fraud detection: An interpretive structural model. Data and policy, 5, p.e25.

A22: Heintz, F., CsÃ¡ki, C., Bernardini, F., and Dobbe, R., 2023. Trustworthy AI Governance - Panel Discussion. CEUR Workshop Proceedings, vol. 3449, 2023 Joint of Ongoing Research, Practitioners, Posters, Workshops, and Projects at EGOV-CeDEM-ePart, EGOV-CeDEM-ePart-JP 2023, Budapest, 4–7 September 2023. CEUR-WS.

A23: SalathÃ©, M., Wiegand, T. and Wenzel, M., 2018. Focus group on artificial intelligence for health. arXiv preprint arXiv:1809.04797.

A24: Cath, C., 2018. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2133), p.20180080.

A25: Vakkuri, V., Kemell, K.K., Jantunen, M., Halme, E. and Abrahamsson, P., 2021. ECCOLA-A method for implementing ethically aligned AI systems. Journal of Systems and Software, 182, p.111067.

A26: Arora, A., Alderman, J.E., Palmer, J., Ganapathi, S., Laws, E., McCradden, M.D., Oakden-Rayner, L., Pfohl, S.R., Ghassemi, M., McKay, F. and Treanor, D., 2023. The value of standards for health datasets in artificial intelligence-based applications. Nature Medicine, 29(11), pp.2929–2938.

A27: Danaher, J., Hogan, M.J., Noone, C., Kennedy, R., Behan, A., De Paor, A., Felzmann, H., Haklay, M., Khoo, S.M., Morison, J. and Murphy, M.H., 2017. Algorithmic governance: Developing a research agenda through the power of collective intelligence. Big data and society, 4(2), p.2053951717726554.

A28: Johnson, M., Albizri, A. and Harfouche, A., 2023. Responsible artificial intelligence in healthcare: Predicting and preventing insurance claim denials for economic and social wellbeing. Information Systems Frontiers, 25(6), pp.2179–2195.

Appendix B: Terminology

For the sake of clarity, it is essential to define the key terms used in this SLR before moving on to answering the questions and classification.

- AI Governance Framework: In this paper, an AI governance framework refers to structured approaches designed to systematically guide the development, deployment, and management of artificial intelligence systems, addressing ethical, regulatory, and operational considerations.
- AI Governance Model: In this SLR, an AI governance model, represents a set of steps for overseeing and managing the ethical, legal, and operational aspects of artificial intelligence (AI) systems.
- AI Governance Tool: In this paper, an AI governance tool refers to a set of specialized instruments, or structured set of steps, developed to assess, monitor, and

enhance the ethical and responsible use of artificial intelligence (AI).

- Ethical and Responsible AI Governance: In this study, Ethical and responsible AI governance represents a foundational set of values, and guidelines intended to guide the development, deployment, and utilization of artificial intelligence technologies in a manner that aligns with societal, moral, and legal considerations. These principles ensure responsible and ethical AI practices, emphasizing transparency, accountability, fairness, and the protection of individual rights.

Author contributions All authors contributed to this manuscript.

Funding Open access funding provided by CSIRO Library Services. Open access funding will be provided by CSIRO Library Services. No funding was received to assist with the preparation of this manuscript. No funding was received for conducting this study. No funds or grants were received.

Data availability Data related to this SLR can be accessed from the links given in the footnotes.

Declarations

Conflict of interest All the authors are full-time paid employee of Australian National Science Agency (CSIRO's Data61). The authors have no Conflict of interest to declare that are relevant to the content of this article. The authors have no financial or proprietary interests in any material discussed in this article. On behalf of all authors, the corresponding author states that there is no Conflict of interest.

References

1. Lu, Q., Zhu, L., Xu, X., Whittle, J., Zowghi, D., Jacquet, A.: Responsible AI pattern catalogue: a collection of best practices for AI governance and engineering. ACM Comput. Surv. **56**, 1–35 (2022)
2. Shams, R.A., Zowghi, D., Bano, M.: Ai and the quest for diversity and inclusion: a systematic literature review. AI and Ethics, 1–28 (2023)
3. Lütge, C., Poszler, F., Acosta, A.J., Danks, D., Gottehrer, G., Mihet-Popa, L., Naseer, A.: Ai4people: ethical guidelines for the automotive sector-fundamental requirements and practical recommendations. Int. J. Technoethics (IJT) **12**(1), 101–125 (2021)
4. Reddy, S., Allan, S., Coghlan, S., Cooper, P.: A governance model for the application of AI in health care. J. Am. Med. Inform. Assoc. **27**(3), 491–497 (2020)
5. Lee, J.: Access to finance for artificial intelligence regulation in the financial services industry. Eur. Bus. Organ. Law Rev. **21**, 731–757 (2020)
6. Bano, M., Zowghi, D., Shea, P., Ibarra, G.: Investigating responsible AI for scientific research: an empirical study. arXiv preprint arXiv:2312.09561 (2023)
7. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat. Mach. Intell. **1**(9), 389–399 (2019)
8. Mäntymäki, M., Minkkinen, M., Birkstedt, T., Viljanen, M.: Putting ai ethics into practice: the hourglass model of organizational ai governance. arXiv preprint arXiv:2206.00335 (2022)
9. Koene, A., Smith, A.L., Egawa, T., Mandalh, S., Hatada, Y.: Ieee p70xx, establishing standards for ethical technology. In: Proceedings of KDD, ExCeL London UK (2018)
10. Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., Srikumar, M.: Principled artificial intelligence: mapping consensus in ethical and rights-based approaches to principles for AI. Berkman Klein Center Research Publication (2020-1) (2020)
11. Attard-Frost, B., Brandusescu, A., Lyons, K.: The governance of artificial intelligence in Canada: findings and opportunities from a review of 84 AI governance initiatives. Gov. Inf. Q. **41**(2), 101929 (2024)
12. Maas, M.M.: Concepts in advanced AI governance: a literature review of key terms and definitions (2023)
13. Kluge Corrêa, N., Galvão, C., Santos, J.W., Del Pino, C., Pontes Pinto, E., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E.: Worldwide AI ethics: a review of 200 guidelines and recommendations for ai governance. arXiv e-prints, 2206 (2022)
14. Zhang, J., Zhang, Z.-M.: Ethics and governance of trustworthy medical artificial intelligence. BMC Med. Inf. Decis. Making **23**(1), 7 (2023)
15. Pierson, J., Kerr, A., Robinson, C., Fanni, R., Steinkogler, V., Milan, S., Zampedri, G.: Governing artificial intelligence in the media and communications sector. Internet Policy Rev. **12**(1), 28 (2023)
16. Wang, X., Oussalah, M., Niemilä, M., Ristikari, T., Petri, V.: Towards AI-governance in psychosocial care: a systematic literature review analysis. J. Open Innov. Technol. Mark. Complex. **9**, 100157 (2023)
17. Corrêa, N.K., Galvão, C., Santos, J.W., Del Pino, C., Pinto, E.P., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E., et al.: Worldwide AI ethics: a review of 200 guidelines and recommendations for AI governance. Patterns **4**(10), 100857 (2023)
18. Laux, J.: Institutionalised distrust and human oversight of artificial intelligence: towards a democratic design of ai governance under the European Union AI act. AI Soc., 1–14 (2023)
19. Paz, J.V., Rodríguez-Picón, L.A., Morales-Rocha, V., Torres-Argüelles, S.V.: A systematic review of risk management methodologies for complex organizations in industry 4.0 and 5.0. Systems **11**(5), 218 (2023)
20. Kreutz, H., Jahankhani, H.: Impact of artificial intelligence on enterprise information security management in the context of iso 27001 and 27002: A tertiary systematic review and comparative analysis. Cybersecur. Artif. Intell. Transf. Strategies Disrupt. Innov., 1–34 (2024)
21. Alsaigh, R., Mehmood, R., Katib, I.: Ai explainability and governance in smart energy systems: a review. Front. Energy Res. **11**, 1071291 (2023)
22. Stogiannos, N., Malik, R., Kumar, A., Barnes, A., Pogose, M., Harvey, H., McEntee, M.F., Malamateniou, C.: Black box no more: a scoping review of AI governance frameworks to guide

procurement and adoption of AI in medical imaging and radiotherapy in the UK. Br. J. Radiol. **96**(1152), 20221157 (2023)
23. Abbas, A., Mahrishi, M., Mishra, D.: Artificial intelligence (ai) governance in higher education: a meta-analytic systematic review. Available at SSRN 4657675 (2023)
24. Kitchenham, B., Brereton, O.P., Budgen, D., Turner, M., Bailey, J., Linkman, S.: Systematic literature reviews in software engineering-a systematic literature review. Inf. Softw. Technol. **51**(1), 7–15 (2009)
25. Guan, J.: Artificial intelligence in healthcare and medicine: promises, ethical challenges and governance. Chin. Med. Sci. J. **34**(2), 76–83 (2019)
26. Ho, C., Soon, D., Caals, K., Kapur, J.: Governance of automated image analysis and artificial intelligence analytics in healthcare. Clin. Radiol. **74**(5), 329–337 (2019)
27. Fontes, C., Corrigan, C., Lütge, C.: Governing AI during a pandemic crisis: initiatives at the EU level. Technol. Soc. **72**, 102204 (2023)
28. Niet, I., Est, R., Veraart, F.: Governing AI in electricity systems: Reflections on the EU artificial intelligence bill. Front. Artif. Intell. **4**, 690237 (2021). https://doi.org/10.3389/frai.2021.690237
29. Liu, Y., Lu, Q., Zhu, L., Paik, H.-Y., Staples, M.: A systematic literature review on blockchain governance. J. Syst. Softw. **197**, 111576 (2023)
30. Zowghi, D., Rimini, F.: Diversity and inclusion in artificial intelligence. In: Lu, Q., *et al.* (eds.) Responsible AI: Best Practices for Creating Trustworthy AI Systems. Addison-Wesley, Sydney (2023). https://research.csiro.au/ss/guidelines-for-diversity-and-inclusion-in-artificial-intelligence/
31. European Union: High-Level Summary of the Artificial Intelligence Act. Accessed: 2024-10-22 (2024). https://artificialintelligenceact.eu/high-level-summary/
32. AI, N.: Artificial intelligence risk management framework (ai rmf 1.0) (2023)
33. OECD: Artificial Intelligence in Society. Accessed: December 10, 2023. https://www.oecd-ilibrary.org/science-and-technology/artificial-intelligence-in-society_eedfee77-en
34. Wallach, W., Marchant, G.E.: An agile ethical/legal model for the international and national governance of ai and robotics. Association for the Advancement of Artificial Intelligence (2018)
35. Salathé, M., Wiegand, T., Wenzel, M.: Focus group on artificial intelligence for health. arXiv preprint arXiv:1809.04797 (2018)
36. GDPR: What is GDPR? https://gdpr.eu/what-is-gdpr/. Accessed: 2024-07-10 (2018)
37. AI RMF Knowledge Base. https://airc.nist.gov/AI_RMF_Knowledge_Base/AI_RMF. Accessed: December 10, 2023
38. (PDPC), P.D.P.C.: Model AI Governance Framework. Accessed: 2024-07-03 (2020). https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf
39. Australia's National Artificial Intelligence Centre: Responsible AI. https://www.csiro.au/en/work-with-us/industries/technology/national-ai-centre. Accessed: November 2023 (2023)
40. Liebig, L., Güttel, L., Jobin, A., Katzenbach, C.: Subnational AI policy: shaping AI in a multi-level governance system. AI Soc. **39**, 1477–1490 (2022)
41. Sidorova, A., Saeed, K.: Incorporating stakeholder enfranchisement, risks, gains, and AI decisions in AI governance framework (2022)
42. Wirtz, B.W., Weyerer, J.C., Sturm, B.J.: The dark sides of artificial intelligence: an integrated AI governance framework for public administration. Int. J. Public Adm. **43**(9), 818–829 (2020)
43. Knight, S., Shibani, A., Vincent, N.: Ethical AI governance: mapping a research ecosystem. AI and Ethics, 1–22 (2024)