Towards AI ethics-led sustainability frameworks and toolkits: Review and research agenda

Cumming, Douglas; Saurabh, Kumar; Rani, Neelam; Upadhyay, Parijat

DOI:
10.1016/j.josfa.2024.100003

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (Harvard):
Cumming, D, Saurabh, K, Rani, N & Upadhyay, P 2024, 'Towards AI ethics-led sustainability frameworks and toolkits: Review and research agenda', *Journal of Sustainable Finance and Accounting*, vol. 1, 100003. https://doi.org/10.1016/j.josfa.2024.100003

Journal of Sustainable Finance and Accounting 1 (2024) 100003

Contents lists available at ScienceDirect

Journal of Sustainable Finance and Accounting

journal homepage: www.sciencedirect.com/journal/josfa

Towards AI ethics-led sustainability frameworks and toolkits: Review and research agenda

Douglas Cumming[a,b,*], Kumar Saurabh[c], Neelam Rani[c], Parijat Upadhyay[c]

[a] College of Business, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA
[b] Birmingham Business School, University of Birmingham, University House, 116 Edgbaston Park Rd, Birmingham B15 2TY, United Kingdom
[c] Indian Institute of Management, Shillong, Meghalaya 793014, India

ARTICLE INFO

JEL Classification:
G30

Keywords:
AI
AI ethics
AI ethics toolkits
AI ethics-led sustainability frameworks
MAXQDA
Sustainability

ABSTRACT

Artificial intelligence (AI) is instrumental in building human skills, accessing knowledge, creating businesses, addressing societal concerns–including environmental issues–and much more. However, unfair, inequitable, and biased data usage for AI deployments does exist and raises ethical and sustainability debates and concerns. AI deployment frameworks are majorly developed by standard societies/groups, technology organisations, analyst groups and federal/government agencies. The paper explores the central themes of AI ethics and sustainability frameworks in declarative standards and statements published by various institutions. The paper offers a thematic analysis of the literature on AI ethics-led sustainability frameworks using MAXQDA software and identifies common principles. We show that there are an established 28 AI ethics-led sustainability frameworks that agencies and groups have disseminated. As well, 6 practical AI ethics toolkits/products are evaluated to translate common AI ethics-led sustainability framework recommendations to deploy AI ethics-led sustainability toolkits programmatically. The research findings validate that beneficence, non-maleficence, justice, explainability, autonomy, privacy, and biasedness need severe attention and postulating algorithmic trust based on AI ethics-led sustainability frameworks. The paper contributes to the unique AI ethics-led sustainability body of knowledge to become a helpful resource for both praxis and researchers.

1. Introduction

Artificial Intelligence (AI) has been experiencing unprecedented growth in numerous fields, including healthcare, fintech, government, logistics, information technology, education, defence, manufacturing, automobile industry (Allen et al., 2021; Hermann, 2021; Jobin et al., 2019; Morley et al., 2020). AI influences industries in institutional, political, and cultural contexts, which impacts individual and social/economic transformations (Davenport & Kirby, 2016; Krishnan, 2019; Hunkenschroer & Luetge, 2022). As a result, AI design, development, and deployment pose many challenges to the individual- well-being and autonomy, labour productivity, societal/economic- impacts on safety, reliability, inclusion, and unbiased decision-making processes (Bertino et al., 2019; Damioli et al., 2021; Goodman & Flaxman, 2017). AI now functions in partnership with humans in day-to-day activities. Therefore, AI needs to consider human relationships, emotions and consequential decision-making processes (Webster & Ivanov, 2020; Urquhart et al., 2019; Malkin et al., 2019). Various malicious prospects exist for AI systems to be biased, discriminant, and vulnerable to data breaches (Floridi et al., 2019; Jobin et al., 2020). It results in compromised privacy and security, leading to reputation and confidentiality risks (McGrath & Gupta, 2018; Rogerson et al., 2019; McLaren, 2003).

AI deployment concerns have ignited various government institutions, research agencies, analysts' groups and technology giants to formulate and draft AI ethics-led sustainability frameworks suitable to their conditions and ecosystems (Hoffman & Masucci, 2018; Jerome, 2021; Hleg, 2019; Yeung, 2020). These institutions have published declarative standards and statements for ethical AI deployments in public. This work emphasises the urgent need to comprehend and study available AI ethics-led sustainability frameworks to design, develop, and deploy ethical AI systems. We are interested in exploring the cohesions of AI ethics-led sustainability frameworks documented/framed by many institutions and agencies. This article highlights the current philosophical and practical ethical interests of AI ethics-led

* Corresponding author at: College of Business, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA.
E-mail addresses: cummingd@fau.edu (D. Cumming), kumar.phd20@iimshillong.ac.in (K. Saurabh), neelam@iimshillong.ac.in (N. Rani), parijat@iimshillong.ac.in (P. Upadhyay).

https://doi.org/10.1016/j.josfa.2024.100003
Received 16 March 2024; Received in revised form 31 March 2024; Accepted 1 April 2024
Available online 25 April 2024

sustainability frameworks when developing effective unities for AI deployments. The principal objective of the paper is to investigate the following research questions:

RQ1. What ethical aspects are explored in the existing literature in conceptualising and designing AI ethics-led sustainability frameworks?

RQ2. What are the central themes of AI ethics-led sustainability frameworks in declarative standards and statements published by various institutions?

RQ3. What practical and mature AI ethics-led sustainability toolkits are available from different technology and research organisations to translate the general AI ethics-led sustainability frameworks/ guidelines into simple AI ethics design, architecture, and code while developing AI systems?

The paper aims to begin by answering the formulated research questions and highlighting how AI ethics-led sustainability frameworks influence current AI deployments. The paper primarily underlines AI ethics-led sustainability frameworks' most common and cohesive elements: explainability, fairness, governance, privacy, and inclusiveness. The paper leverages a PRISMA-based systematic literature review of the AI ethics body of knowledge using 30 relevant shortlisted research papers.

The paper is organised as follows. Section 1 introduces AI ethics and AI ethics-led sustainability frameworks as subjects and presents the foundation of the paper. It also represents the definitions and foundations of the AI ethics-led sustainability frameworks. Existing AI ethics-led sustainability frameworks (from federal/government agencies, standards societies/groups, technology giants, and analysts' groups) have also been extensively surveyed to identify common principles. Section 2 of the article highlights the PRISMA-based AI ethics-led sustainability framework's literature review covering the current state of the literature, article selection methods and database metrics. It aims to analyse AI ethics-led sustainability framework principles using software based on thematic code analysis. Section 3 covers the consolidated view of commonly applied methods, theories, and the industry context with objective(s) and previously published work findings. It discusses the unique market analysis of the practical AI ethics toolkits to translate the AI ethics-led sustainability frameworks to reality and become a handy tool for managers. Finally, Section 4 presents the conclusive research and managerial implications for designing, developing, and deploying successful AI ethics-led sustainability frameworks and toolkits.

2. AI ethics-led sustainability frameworks

Today's society and ecosystems are becoming the testimony of AI-led transformations to execute many tasks that often require human intelligence. AI, which touches human lives, raises questions about ethical principles of autonomy, biasedness, fairness, and accountability, issues found in various sector-based literature (Jobin et al., 2019). Ethics as a concept is wide-ranging and spans utilitarian, deontological, and virtue-based theories. Utilitarianism refers to consequentialism-focused ethical choice, which defines the actions of right and wrong deeds based on outcomes. Deontological ethics is the study of 'deon' (Greek Word meaning duties), where the alternatives are chosen based on moral values of accountability, commitment, and rules rather than the consequential action (Ferrell et al., 1989; May & Pauli, 2002). Virtue ethics underscore the importance of the character or act of a virtuous individual facing ethical circumstances (Burmeister, 2017; Yang & Lee, 2010). AI ethics specifies the utilitarian, deontological, and virtue-based guidelines necessary to manage the moral responsibilities and duties of AI developers, promoters, and architects. Established theories are used to develop AI ethics-led sustainability frameworks and principles to code the actions and responsibilities of human behaviour performed by machines that lead to artificial moral machines (AMAs). When machines perform actions, it is not easy to address the issues and ethical binaries of right/wrong or good/bad (Siau & Wang, 2020).

AI and its applications in the research and managerial domain are nothing new; they have existed for many decades since the 1950s. Similarly, researchers and practitioners from the same timeframes contested the lack of associated risks of defined ethical consequences in the AI domain (Floridi et al., 2021; Yang et al., 2018). The latest topics of ethical interest entail: i) algorithmic biasedness (acts of AI systems that promote racial, social, and gender discrimination), ii) explainability (non-transparent, unexplainable algorithmic outcomes by the AI system), iii) autonomy (violations of rights and freedoms, struggle to identify the responsible entity), iv) privacy (individual data invasion driving AI data lakes and workflows), v) beneficence (machine obligation to protect a user's moral rules and rights), vi) non-maleficence (obligation to not harming the AI system users), and vi) justice (Floridi et al., 2021; Siau & Wang, 2020). Various AI ethics guidance manuscripts are available from bodies and agencies like analyst groups, technology companies, federal agencies, and standard/research organisations. These documents cover algorithmic biasedness, explainability, autonomy, privacy, beneficence, non-maleficence, and justice that lay the foundations of principles and frameworks (Whittlestone et al., 2019). Integrating the practical and theoretical aspects of the AI ethics principles and frameworks is broad and not specific to consequential scenarios. The principles vary in demographics, countries, culture, and societal settings (Carrillo, 2020; Morley et al., 2021). Despite having many principles and framework documents available, there is a dearth of toolkits to implement the guidelines. This makes the AI ethics document very abstract and out of reach for real AI developers and architects to make tangible design decisions (Floridi et al., 2019). The researchers envisage a balanced approach to understanding ethics, its principles, interaction, and outcomes with AI machines between humans and machines. AI ethics toolkits/simulators must utilise the early warning system to forecast the anticipated dangers and ensure open engagement and a transparent consensus-based approach (Schif et al., 2020).

2.1. AI ethics-led sustainability frameworks and common principles

As AI is adopted and diffused in numerous areas, with some areas presenting a risk, the burden rests on researchers and practitioners to lay down principles and frameworks of AI ethics. The central theme of this section is to provide a glimpse of available well-accepted AI ethics-led sustainability frameworks with their associated principles, goals, and objectives. Based on the literature survey, authors consolidated and grouped all surveyed 28 AI ethics-led sustainability frameworks under 4 categories (Fig. 1) coming from institutions like i) standard societies/ groups, ii) technology organisations, iii) analyst groups, and iv) federal/government agencies (by country).

Researchers, practitioners, the analyst community, activists, and lawmakers are putting forward their best efforts to produce manuals, process documents, write regulations, and create standards in AI ethics-led

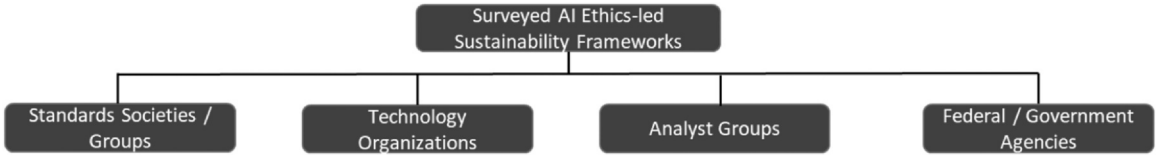

Fig. 1. Surveyed AI ethics-led sustainability frameworks.

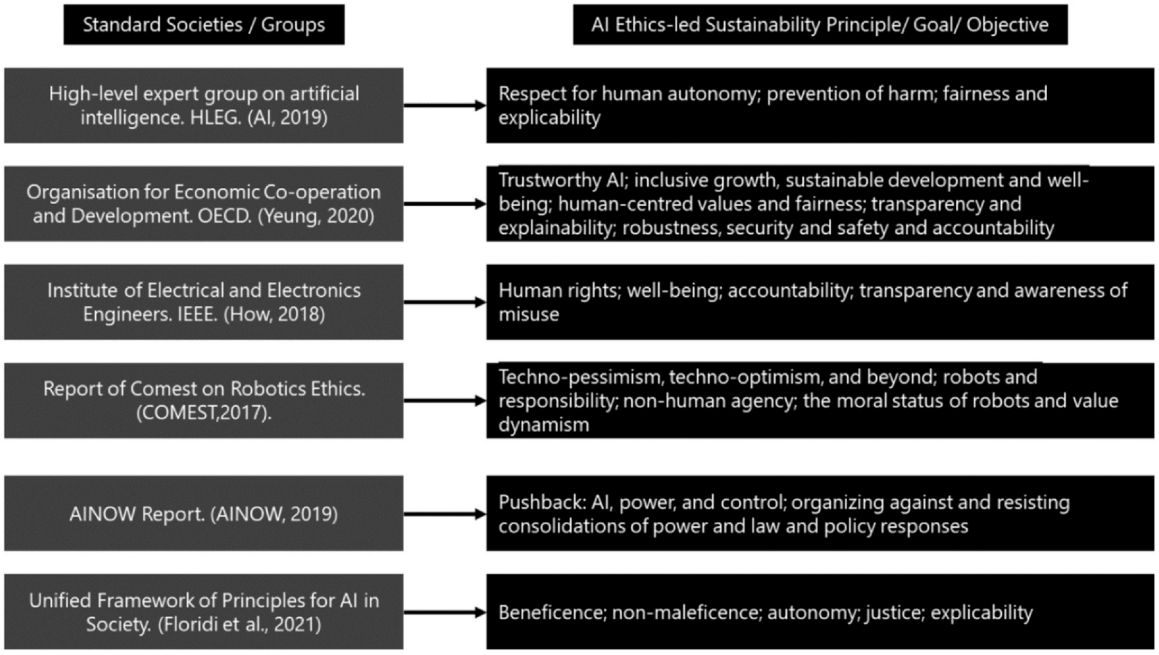

Fig. 2. Standards societies/groups: AI ethics-led sustainability principles. (Yeung, 2020; How, 2018; COMEST, 2017; AINOW, 2019; Floridi et al., 2021).

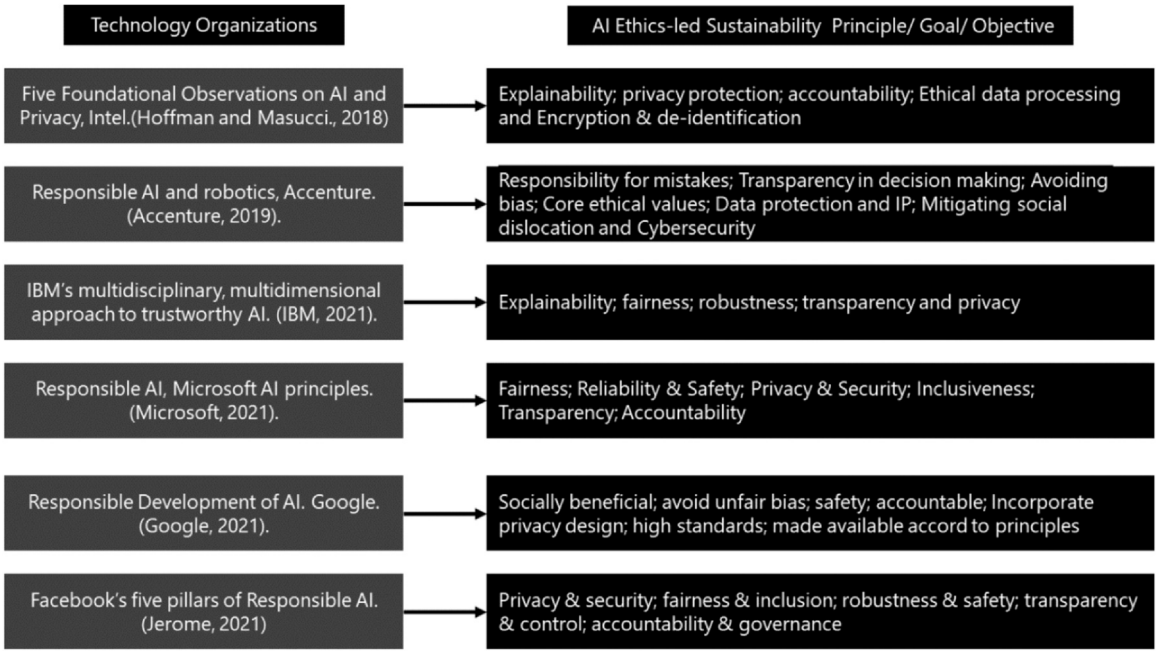

Fig. 3. Technology organisations: AI ethics-led sustainability principles. (Hoffman and Masucci, 2018; Accenture, 2019; IBM, 2021; Microsoft, 2021; Google, 2021; Jerome, 2021)

sustainability frameworks. Technical/standards institutions like the High-Level Expert Group on Artificial Intelligence (HLEG), the Organisation for Economic Co-operation and Development (OECD), the Institute of Electrical and Electronics Engineers (IEEE), the World Commission on the Ethics of Scientific Knowledge and Technology (COMEST), and AINOW (2019) contributed significantly to the AI ethics led sustainability frameworks. All the reports from these institutions possess a common thread of human autonomy, fairness and explicability, safety, accountability, power and control, beneficence, etc., as core principles (Fig. 2). Technology giants like Intel, Facebook (Meta), Accenture (2019), IBM, Microsoft, and Google (2021) are not far behind in the race and released their respective declarations to adhere to the common principles of explainability, fairness, inclusion, accountability, governance, data protection, and privacy design (Fig. 3). There is a need for standard AI guideline documents emphasising humanity, society, environmental well-being, values, safety, reliability, inclusivity, non-discrimination and abidance of laws and regulations across the regions. Analyst groups like KPMG, Delloite, Forrester, BCG, Gartner and PWC reciprocated the vitality of integrity, explainability, fairness, resilience, bias prevention, interpretability, explainability, robustness, and security in the same voice across the surveyed reports (Fig. 4). Countries like Singapore, China, Australia, Finland, Japan, Canada, the USA, the UK, and India can access the ethical framework through manuals and vision documents (Fig. 5).

28 AI ethics-led sustainability frameworks from standard societies/groups, technology organisations, analyst groups and federal/government agencies align well with academic literature. The section relates

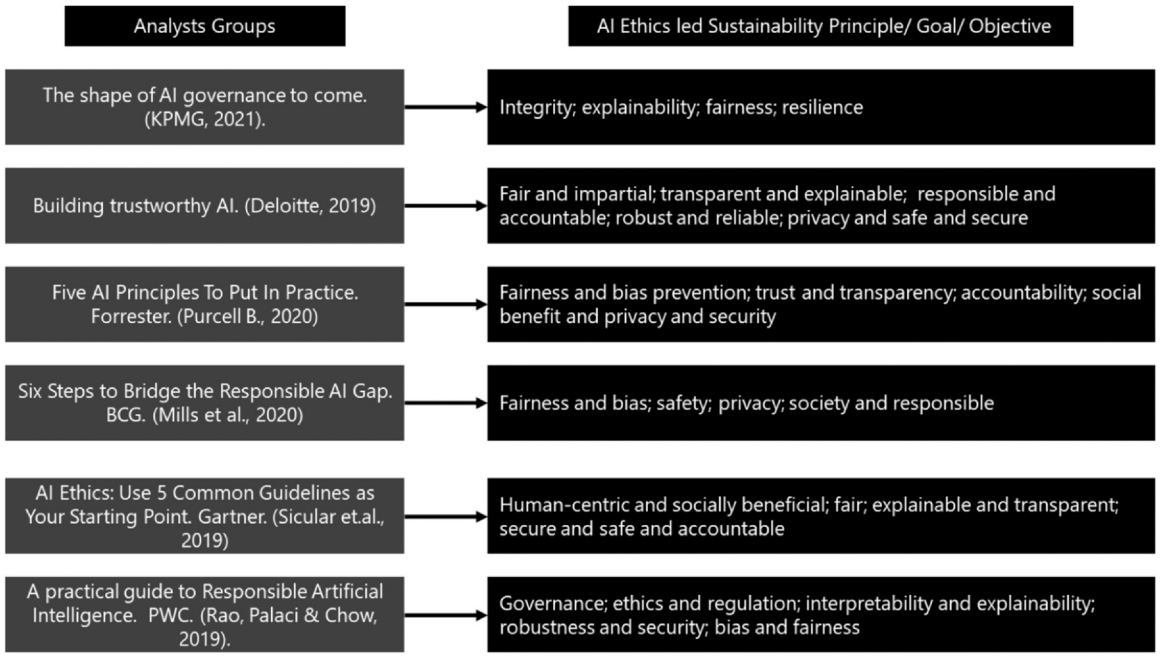

Fig. 4. Analysts groups: AI ethics-led sustainability principles. (KPMG, 2021; Deloitte, 2019; Orucell B, 2020; Mills et al., 2020; Sicular et al., 2019; Rao, Palaci & Chow, 2019).

Fig. 5. Federal/Government Agencies (by country): AI ethics-led sustainability principles. (Remilina, & Seah, 2019; BAAI, 2019; Dawson et al., 2019; FAAI, 2017; JSAI, 2017; AMA, 2018; Declaration, 2018; Lords, 2018; NIST, 2021; NITI, 2021).

Search string 1:

("ethics" OR "ethical" OR "Moral Reasoning" OR "virtues" OR "utilitarian" or "deontology") AND ("robotics" OR "robot" OR "machine learning" OR "ML" OR "deep learning" OR "artificial neural networks" OR "information system" OR "artificial intelligence" OR "AI" OR "sustainability")

Search string 2:

<Search string 1> AND ("concepts" OR "industry" OR "process management" OR "adoption" OR "challenges" OR "governance" OR "integration" OR "business model" OR "interoperability" OR "digital transformation" OR "sustainability")

Fig. 6. Search strings.

the AI ethics-led sustainability work in the academic world with the praxis propounded frameworks. Resseguier and Rodrigues (2020) emphasised that the following state and federal laws are the foundations of AI ethics and sustainability frameworks. Saltelli and Rommetveit (2020) underpin that fairness and equitability concerns are the most critical issues when management functions are coded as algorithms, which lead to harm and control of humanity. The work of Jobin et al. (2020) precisely covers most of the principles laid down by the discussed frameworks, which establishes AI ethics-led sustainability guidelines. It constitutes a framework characteristic cluster of "beneficence, dignity, justice, fairness, privacy, non-maleficence, transparency, freedom and autonomy, sustainability, trust, solidarity and responsibility. Siau and Wang (2020) establish the AI ethics-led sustainability principles as moral duties and virtues specified and obligated to the creators of the AI applications. Leslie's (2019) work states that widely accepted standards in AI product deployment help humans to choose between right and wrong, which leads to establishing moral conduct and eventually establishes AI ethics-led sustainability values and principles. Hagendorff's (2020) paper emphasises the lack of mechanisms of normative claims while strengthening AI ethics-led sustainability principles in general. Müller's (2020) research highlights AI ethics-led sustainability principles as the discipline of studying the ethics of technology related explicitly to sustainable AI systems. Chauhan and Gullapalli (2021) refer to AI ethics-led sustainability concerns regarding gender, discrimination, racial aggravation, and patient privacy issues, specifically in healthcare.

3. Methodology

The section illustrates the methodology adopted to accomplish the study based on a literature search and eventually conducting thematic code analysis of the shortlisted articles.

3.1. Literature search

The literature search uses PRISMA to produce transparent, fair, and inclusive results. Steps were adapted based on specific attributes from similar prior work noted in the literature (Briner and Denyer, 2012; Moher et al. 2009). The literature search was conducted from November 2022– January 2023. Literature survey foundations are laid using rich and respected databases like Wiley, Science Direct, Scopus, Google Scholar and Science Direct, followed by the trailing search of the shortlisted articles to reinforce the process. Literature from several databases was included based on a search using keywords like 'artificial intelligence + ethics', 'AI + ethics + sustainability', 'AI + ethics + organisation', and 'AI + ethics + organisation + people + process + sustainability' with various permutations and combinations etc.

3.1.1. Article selection

We conducted four steps based on bibliographic searches: identification, screening, eligibility, and inclusion (see Figs. 6 and 7). Besides keyword-based searches, we performed a bibliographic trail search using references while following the inclusion, exclusion, and quality assessment criteria for screening the relevant search, results-based articles. The inclusion criteria follow the robust steps. Literature published after 01/01/2016 is considered an element of the survey as we could not find relevant literature on AI ethics-led sustainability frameworks before the mentioned date. We ensured that the publication type was a journal article with full text available, the publication language was in English only, and the citation was available in previously selected literature. The keywords and theme of the search were related to AI Ethics, humans, society, and organisation. Search criteria consider the essence of the AI ethics-led sustainability framework of beneficence, non-maleficence, autonomy, and justice for humans and IT in the visited and shortlisted literature. The AI ethics-led sustainability literature is taken from the trustful knowledge repositories of technology companies, analyst groups, standards and federal institutions' website data, and collaterals. Similarly, the exclusion criteria followed strictly that non-peer-reviewed articles, books, reviews, dissertations, white papers, technical papers, and non-English papers should not be shortlisted for the literature survey. We omitted the repeated research articles in search criteria outputs, and AI papers did not directly target ethics-led sustainability in humans and society. The quality assessments of the exercise ensured the importance of choosing a reputable publisher, as well as an established database and research question targeted in the article.

We discussed among ourselves (authors) reaching a consensus on the final set of articles under consideration. We refined the search criteria using various database-based search refinement features to search the relevant articles. Initially, abstracts were analysed to cluster the paper based on themes, and finally, specific full-text-required articles were retrieved. We omitted some full-text articles at this stage as they were inappropriate for our research. We used the thematic content analysis method to cluster the articles based on the inclusion criteria. Our consensus rate was approximately 80% to include relevant articles and cluster them into relevant themes. According to eligibility, we included 1416 articles (1173 papers from databases + 243 papers from other sources) related to our anticipated AI ethics-led sustainability work theme. The papers were searched based on various criterion keywords like a) AI ethics concepts, 2) AI ethics industry and specific processes, 3) AI ethics adoption/acceptance, 4) AI ethics challenges, 5) AI ethics and governance, 6) AI ethics integration, 7) AI ethics business models, 8) AI ethics interoperability, 9) AI ethics-led sustainability and 10) AI ethics and digital transformation. After the first step, we

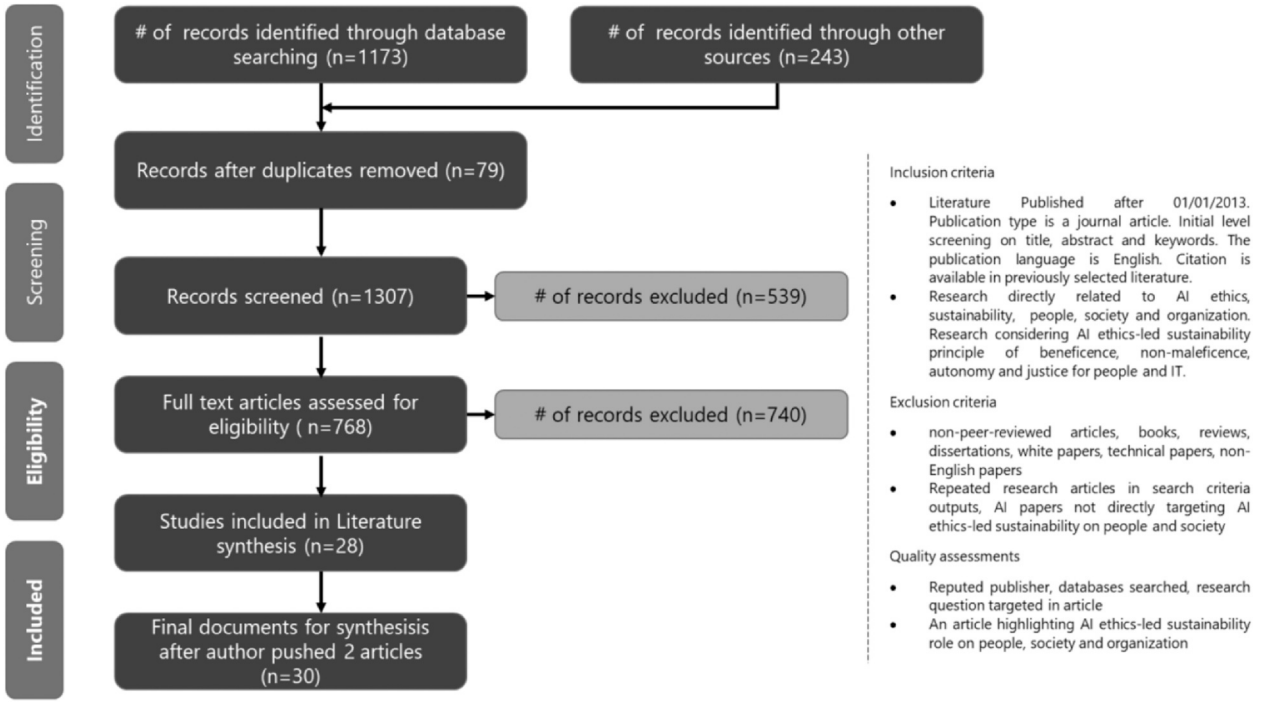

Fig. 7. PRISMA flow diagram.
Source: Moher et al. (2009).

identified and removed 79 records from our study. In the next phase, we excluded 539 articles based on the exclusion, inclusion, and quality assessment criteria. We filtered and excluded 768 papers from our study in the next iteration. Finally, 30 articles were included for literature synthesis.

The final 30 articles shortlisted for our study were considered the body of knowledge adapted from the literature on human, social, and ethical issues during AI ethics-led sustainability framework design and deployments. Table 1 depicts the bibliographic journal sources of AI ethics-led sustainability literature from which the papers are identified with their frequencies (given in a percentage). Table 2 represents the objectives and findings related to human aspects of AI ethics papers with their author details. Table 3 presents the industry-wide details of the study with author details. Table 4 talks about the commonly applied methods adopted by the authors in studying the personal aspects of AI ethics-led sustainability. Table 5 summarises the theories used by the authors in specific papers.

3.2. Code and thematic analysis

This section has performed AI ethics literature content and thematic analyses software. We clustered and mapped surveyed papers for developed AI ethics themes according to our requirements.

3.2.1. Data processing; data analysis

The first step was to import all papers shortlisted to MAXQDA Analytics Pro software during our literature survey. The principal aim of the data analysis stage was to examine the ethical themes of AI that should be practised while deploying any AI systems by coding a

Table 1
Bibliographic Sources on AI Ethics-Led Sustainability Literature.
Source: Author's own compilation.

S.No.	Bibliographic Source	Articles considered for review	Total %
1	Journal of Business Ethics	6	20
2	AI and Ethics	3	10
3	Ethics and Information Technology	2	6.67
4	Business Ethics, the Environment and Responsibility	2	6.67
5	Science and engineering ethics	2	6.67
6	Journal of Information, Communication and Ethics in Society.	2	6.67
7	Technology in Society	1	3.33
8	Forecasting and Social Change	1	3.33
9	Minds and Machines	1	3.33
10	Journal of Business Research	1	3.33
11	Journal of the Academy of Marketing Science	1	3.33
12	Minds and Machines	1	3.33
13	Nature Machine Intelligence	1	3.33
14	International Journal of Information Management	1	3.33
15	Philosophy & Technology	1	3.33
16	MIT Sloan Management Review	1	3.33
17	Technological Forecasting and Social Change	1	3.33
18	Futures	1	3.33
19	Business Horizons	1	3.33

Table 2
Objective(s) and findings of previously published works.

Author (s)	Objective (s)	Finding (s)
Part 1		
Neubert and Montañez (2020)	To investigate the relevance of virtue for the ethical design. Enumerate the ethical challenges faced by the companies like Google.	Paper establishes the ethical framework to support virtue as a decision-making exercise. Examine the people side of how employees can be attracted and retained who develop AI offerings.
Hermann (2021)	To scrutinize the ethical AI marketing challenges. Identify and reconcile AI-led social good perspectives in marketing.	Beneficence and non-maleficence require special attention because of AI advancements. AI Humanization & Intelligence is closely associated with ethical interdependencies challenges.
Zhu (2020)	The essay investigates technology's role in ethics and society. Deontology and consequentialism ethics is evaluated based of Confucian ethics.	Work explores the issues pertaining to ethical reasoning based on intellectual resources. Teaching aids are provided for students to impart knowledge on ethics to a society driven by technology like AI.
Hunkenschroer and Luetge (2022)	To review the recruitment process through the lens of AI ethics. Ethics review of 51 articles in the recruitment space driven by AI.	Findings synthesize the ethical risks and opportunities. A deeper investigation of moral questions and identifying the gaps in AI-led recruitment.
Rodriguez-Lluesma et al. (2021)	To study social change in digitalization of work To explore precondition solidary-economy networks	Argument paper on DT functions on social relation DT social relation with exchange value, correspondence, intrinsic extra-economic purpose and communication for reciprocal services
Wamba et al. (2021)	The Paper's objective is to develop an AI-led society for well-being. To explore available AI ethics principles, policies and frameworks.	Research underpins the rich set of questions visualizing AI-led good society. Social requirements like health, hunger, education, inclusion, justice, and security are highlighted using AI ethics principles.
Meszaros and Ho (2021)	To examine the adherence to GDPR requirements in AI research. To differentiate the requirements of AI research in academic and business settings.	Public interest is the topmost priority both for academic and commercial research. To lay down the limits of Privacy and innovation with respect to GDPR exemptions.
Ashok et al. (2022)	Review and present the conceptual model of digital ethics. Highlight the societal impact based on the Digital transformation archetypes.	Cognitive and information domain implications are discussed based on the 12 propositions. Autonomy, Privacy, fairness and intelligence implications are widely discussed under the governance domain.
Part 2		
Borenstein and Howard (2021)	The aim is to investigate the impact of AI on human life. To understand the ethical concerns of design, deployment and usage of AI technologies.	Paper emphasized training AI communities to reflect how it impacts people and society. AI ethics-based pedagogy is required to train professionals, developers and designers.
Etzioni and Etzioni (2016)	Examine the autonomy-based ethics concerns of intelligent instruments. To ensure that smart instruments do not engage in unethical practices.	The study facilitates the notion of AI systems accountable for audit, governance and operational activities. The article determines the ethical principles for bots and smart instruments according to people's preferences.
Hagendorff (2020)	To understand the ethical guidelines proposed in recent times. To propose the normative guidelines for AI technologies to get the best out of the disruptive potentials.	Paper provides the 22 AI ethics guidelines as a semi systematic evaluation. Paper provides the recommendations of AI ethics effectiveness as values are implemented.
Hickok (2021)	To advocate the ethical AI development by businesses, government and communities. To propose the study of currently available AI ethics principles and frameworks.	The Paper addresses the abstractions and arguments of AI ethics-based development. Findings provide the lessons learnt guidelines for standards, codes, law, people voices and experiences based on various cases.
Lauer (2021)	To examine whether AI ethics is a black box for industries. To understand the nuances of AI ethics outside small research communities of Maths experts.	Paper findings provide reasons for AI deployments are so elusive. Paper presents the examples and lessons from industry and research to promote an ethical culture in people and society.
Morley et al. (2020)	Debate on the ethical principles rather than practices. Awareness of the AI ethics impacts utility and society to the developers and research community.	Paper focused explicitly on machine learning, but results apply to mostly all AI branches. The research method provides a typology of findings applicable to future research.
Stahl et al. (2021)	Evaluate the equilibrium point of technology and economic benefits countering legal and social AI ethics issues. Conceptually capture both advantages and disadvantages.	Paper findings propose ML applications and social and metaphysical questions as AI discourses. It helps to visualize human rights and legislation based on AI ethics steering.
Spinello (2021)	Debate when code is simulated as law surrogate. Argue software code based normative forces	Amplify common good perspectives for right to a) physical security b) privacy Paper proposes exceptional access scheme to mitigate risk
Part 3		
ÓhÉigeartaigh et al. (2020)	Highlight cooperation barriers between North America, Europe and East Asia based on the AI development. Understand the reason for the optimistic cross-cultural cooperation.	The Paper Presents the cultural and regional arguments related to the trust and disagreements. An agreement is not always required on AI principles and standards for cooperation between the regions to address practical issues.
Toth et al. (2022)	To study ethical AI robot applications based on ethical theories. To address ethical issues related to morality and accountability.	Research posits a conceptual framework to interpret the ethical implications. Debate on the associated accountability for a set of actors.

(*continued on next page*)

Table 2 (*continued*)

Author (s)	Objective (s)	Finding (s)
Wu et al. (2020)	To understand the role of Governance and ethics plays in AI-led SDG. Organizations, people-centric societies, and R&D laid down the AI ethics guidelines and principles.	Paper findings propose the outcome of governance led by AI. The governance relationship between SDG and society is well established with future directions.
Davenport and Katyal (2018)	How AI-powered products and services impact consumers. How to manage the ethical decisions for AI applications and services.	Finding provides the ethical issues faced by 2000 startups in and outside the US. Seven actions were identified for AI-powered companies in the research.
Cubric (2020)	To provide a systematic literature review of AI adoption with issues in business and management. To enlist drivers and social implications of AI ethics in the context of people and society.	Paper conducts the tertiary study of 30 reviews papers for 15 years. Paper findings present the multi-industry review of AI adoption in the context of social and people-centric considerations.
Feher and Katona (2021)	Perform socio-cultural AI consequences with multiple contexts and settings to Industry 4.0 and AI governance. Topic modelling based on 607 is evaluated to identify 15 fields of action.	Finding points the AI technology usage in society and culture. Socio-cultural AI consequences are researched in the context of academic and policy research
Hongladarom (2021)	To understand the national AI ethics guidelines for Thailand. To envisage Thailand's political and cultural consequences in the light of AI.	Research highlights the contradictory expressions of AI modernization. Modernization is the outcome of the policy document while maintaining Thailand's Traditional Values.
Part 4		
Davenport et al. (2020)	Understand how AI is transforming marketing strategies. Customer behaviours need AI ethics evaluation based on intelligence levels and task types.	Paper proposes the multidimensional AI framework To address marketing strategies. Privacy, ethics and bias are better handled when humans are augmented with the combination of AI ecosystems.
Floridi et al. (2020)	Check the potential of using AI for social good. What salient point makes AI social a good enabler and reproduces initial success?	The Paper indicates 7 points factors for AI as the enabler of social good. Comprehensive 27 cases are discussed as AI projects where AI became the tool for social good with respect to people and organization.
Fröding and Peterson (2020)	Evaluate how the near-future AI system will behave in a friendly manner. Possibilities of coexistence of man and machines together to ensure reciprocity	Paper establishes the concept of "as-if friendship" with the chance of abuse on humans and society. The virtue alignment approach is outlined in the Paper to Ai ethics to enhance the value-based approach.
Timmers (2019)	Critically comment on the cases of sovereignty when cybersecurity risks are raised because of Digital transformation and cybersecurity events, Understand the role of AI ethics when autonomy is developed at the system level and challenges sovereignty.	Paper touched on the controversial topic of ethics as an accelerator of strategy. Commentary contributes to understanding AI's ethics while working towards autonomy and sovereignty.
Jobin et al. (2019)	Investigate the agreements on the corpus of guidelines and framework of AI ethics. Debate the constitutional elements if of AI ethics realization.	Paper presents the five converging principles of AI ethics. It establishes the importance, actors, interpretation, substantive divergence and implementation guidelines.
Munoko et al. (2020)	Review the audit potentials ethically using the AI technology. Discuss unintended consequences in accounting activities while performing aunting based on AI technologies. A	Paper reported ethical issues based on past audit experiences where AI is used as a technology enabler. Policy and governance are also discussed while working on ethical AI audit tools.
Watson et al. (2021)	Investigate the potential of AI as a leadership tool. AI possibilities to meet strategic organizational leadership requirements fulfilled by AI technology.	Paper key findings include the role of digital knowledge, networking, data focus and agility at a senior level based 33 Suite level interviews. Requirement of C-suite reskilling, recruitment, talent retention to develop the intrapreneurial culture.

literature survey. We linked, equated, and interplayed the imported content to collect data and perform our analysis. Selective coding and the theoretical background are taken as the foundation for coding the articles. The imported documents were considered as discrete sections to establish the connection between the codes. Finally, the authors selected the main code to relay all the considered codes for a comprehensive examination and to synthesise the core of the AI ethics-led sustainability study. The complete coding method was iterative while

Table 3
Industry-Wide Literature Context.
Source: Author's own compilation.

Industry	Authors	Percentage
Generic	Rodriguez-Lluesma et al. (2021); Etzioni and Etzioni (2016); Hagendorff (2020); Hickok (2021); Morley et al. (2020); Spinello (2021); Floridi et al. (2020); Fröding and Peterson (2020); Timmers, (2019); Jobin et al. (2019); Munoko et al. (2020); Ashok et al. (2022); ÓhÉigeartaigh et al. (2020); Toth et al. (2022); Davenport and Katyal (2018); Feher and Katona (2021); Stahl et al. (2021); Davenport et al. (2020); Watson et al. (2021); Cubric (2020)	66.67
Government	Meszaros and Ho (2021); Wu et al. (2020); Hongladarom (2021)	10.00
Information Technology	Neubert and Montañez (2020); Wamba et al. (2021)	6.67
Education	Zhu (2020); Borenstein and Howard (2021)	6.67
Healthcare	Lauer (2021)	3.33
Human Resources	Hunkenschroer snd Luetge (2022)	3.33
Marketing	Hermann (2021)	3.33

Table 4
Commonly Applied Methods.
Source: Author's own compilation.

Methodology	Authors	Percentage
Literature Review	Hermann (2021); Hunkenschroer and Luetge (2022); Wamba et al. (2021); Hagendorff (2020); Hickok (2021); Morley et al. (2020); Jobin et al. (2019); Wu et al. (2020); Hongladarom (2021); Cubric (2020); Davenport and Katyal (2018); Spinello (2021)	40.00
Conceptual Paper	Zhu (2020); Etzioni and Etzioni (2016); Fröding and Peterson (2020); Timmers (2019); Munoko et al. (2020); Ashok et al. (2022); ÓhÉigeartaigh et al. (2020); Toth et al. (2022); Neubert and Montañez (2020); Toth et al. (2022); Rodriguez-Lluesma et al. (2021)	36.67
Case Method	Meszaros and Ho (2021); Lauer (2021); Davenport et al. (2020); Floridi et al. (2020)	13.33
Empirical Study	Stahl et al. (2021)	3.33
Interview Study	Watson et al. (2021)	3.33
Topic Modeling	Feher and Katona (2021)	3.33

Table 5
Commonly applied theories.
Source: Author's own compilation.

Theory	Authors	Percentage
Ethical frameworks/Ethics Guidelines	Hermann (2021); Hunkenschroer and Luetge (2022); Hagendorff (2020); Munoko et al. (2020); Davenport and Katyal (2018); Hongladarom (2021)	20.00
AI technologies and theories/AI Ethics Lab/AI Ethics Guidelines/ AI Ethics principles/AI Fairness	Wamba et al. (2021); Hickok (2021); ÓhÉigeartaigh et al. (2020); Borenstein and Howard (2021); Stahl et al. (2021)	16.67
GDPR Law/OECD standard/Beijing Academy of Artificial Intelligence (BAAI, 2019)/European Commission 2019/High-Level Expert Group	Meszaros and Ho (2021); Morley et al. (2020); Timmers (2019); Jobin et al. (2019)	13.33
Virtue/Deontology/ Nichomachean Ethics	Neubert and Montañez (2020); Zhu (2020); Fröding and Peterson (2020)	10.00
Restricted Access/limited control	Spinello (2021)	3.33
Conceptual framework- Level of Abstraction (LoA)	Floridi et al. (2020)	3.33
Deep Neural Network, satisfiability modulo theory (SMT)	Wu et al. (2020)	3.33
Relational theory of society	Rodriguez-Lluesma et al. (2021)	3.33
Descriptive and normative ethical theory	Toth et al. (2022)	3.33
TAO model	Feher and Katona, (2021)	3.33
A communitarian approaches. A libertarian approach	Etzioni and Etzioni (2016)	3.33
Medical ethics and systems theory	Lauer (2021)	3.33
Tertiary study	Cubric (2020)	3.33
Marketing theories	Davenport et al. (2020)	3.33
Ontological framework	Ashok et al. (2022)	3.33
AI leadership Tool	Watson et al. (2021)	3.33

tagging the initially developed codes (People Centric Ethics Theory, People Centric Ethical AI intentions, People Centric Ethical AI Behaviour) to the final theme (People Centric Ethical AI Deployment) based on the research interest area. The themes considered in our research worked like one central data element to depict the qualitative study analyses. The study uses thematic analyses to investigate the literature-based evidence of theory, intention, behaviour, and deployment reflections of people-centric ethical AI deployment. The themes were developed based on data analysis, literature fields, and extracted content interpretation (Holstein & Gubrium, 1997; Braun & Clarke, 2006).

Table 6 presents the main themes (T) /sub-themes (ST) based on the content available in the imported literature. We worked on 4 themes and 12 subthemes in our study. The most ubiquitous subtheme indicated by all content analysis is 'People Centric Ethical AI Deployment," "People Centric Ethics Theory\Common Good & Duty," "People Centric Ethical AI Intensions\AI Privacy," and "People Centric Ethical AI Intensions\AI Governance.".

Table 6
Frequency Analysis [Code. seg. (all documents)].
Source: Author's own compilation.

Themes and Subthemes	Frequency	Percentage
People Centric Ethical AI Deployment	632	16.59
People Centric Ethics Theory\Common Good & Duty	577	15.15
People Centric Ethical AI Intensions\AI Privacy	420	11.03
People Centric Ethical AI Intensions\AI Governance	328	8.61
People Centric Ethical AI Intensions	285	7.48
People Centric Ethical AI Intensions\AI Explainability	220	5.78
People Centric Ethical AI Behaviour	211	5.54
People Centric Ethical AI Behaviour\Non-maleficence	199	5.22
People Centric Ethical AI Intensions\AI Fairness	188	4.94
People Centric Ethical AI Behaviour\AI Autonomy	171	4.49
People Centric Ethics Theory\Rights Based	162	4.25
People Centric Ethical AI Behaviour\AI Justice	162	4.25
People Centric Ethical AI Behaviour\Beneficence	87	2.28
People Centric Ethics Theory\Virtue	77	2.02
People Centric Ethics Theory	53	1.39
People Centric Ethics Theory\Utilitarian / Deontological	37	0.97

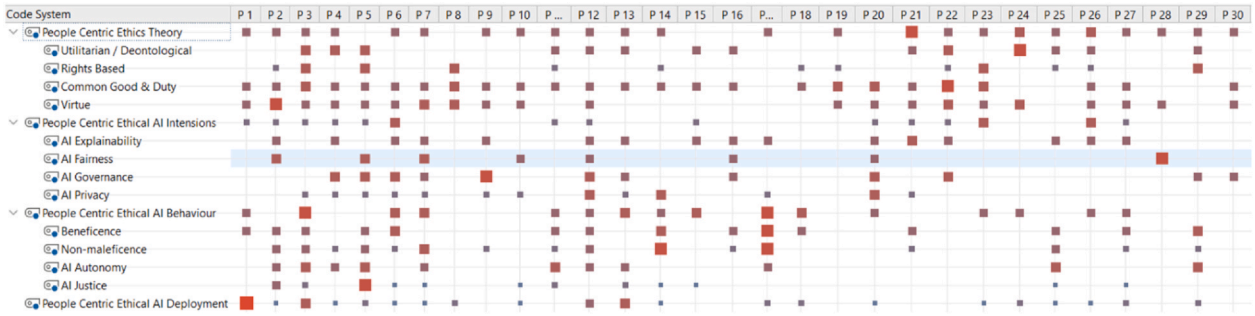

Fig. 8. Code matrix browser.

Fig. 8 represents the code matrix browser to aid in visualising the codes assigned across the documents. The output results help to investigate the code relationship, frequency, intention, and reflection—all helpful in developing any anticipated framework. Fig. 8 is designed based on the developed themes and subthemes and their individual frequencies in each imported document (research paper). The size of the square presents the significance of the code in the document. The frequency of the code in any document is proportional and decides the size of the square. Code distribution and frequency analysis can be used to comprehend the meaning and significance of the relative code.

Fig. 9 represents the code maps to visualise the intended research codes on the workspace. We could comprehend the relationship based on the codes and their categories. The maps do not represent any direction but can be used to organise data and explain code dependency.

4. Discussion

The literature survey confirms that the existing literature balanced both the theoretical and practical aspects of the AI ethics-led sustainability literature. Half (50%) of the papers banked on the established theories form the foundation of the research (Table 5). The literature survey indicates the use of well-accepted ethics theories on virtue, Nichomachean ethics, and deontology (Neubert & Montañez, 2020; Zhu, 2020; Fröding & Peterson, 2020), relational theory of society Rodriguez-Lluesma et al. (2021), Ontological framework (Ashok et al.,

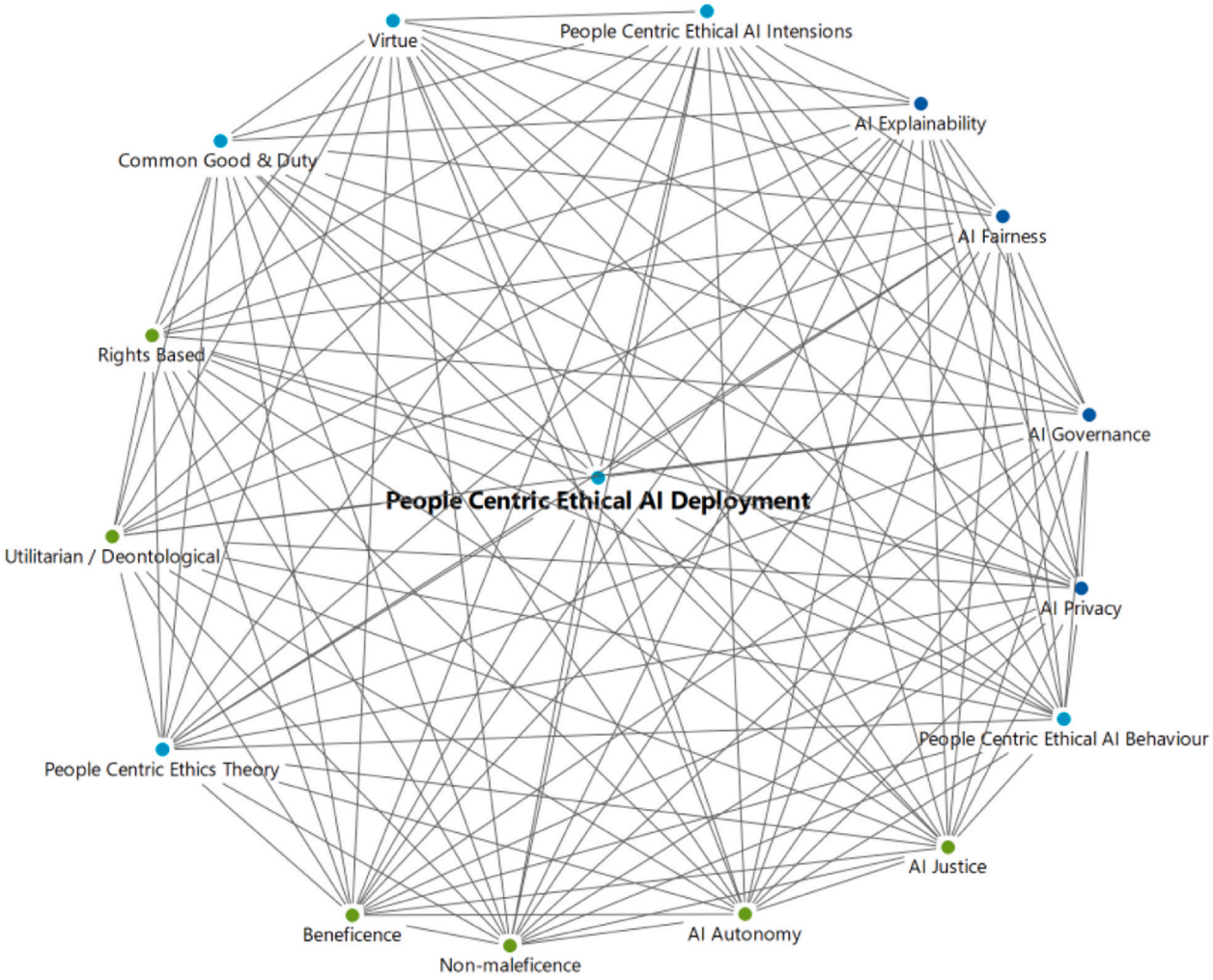

Fig. 9. Code maps.

2022), etc. to establish the theoretical genesis. The literature survey also indicates the practical usage of the ethical aspects of AI deployment based on institutional guidelines and manuals establishing the law conception of ethics. HLEG, Association for Computing Machinery (ACM), Beijing Academy of Artificial Intelligence (BAAI, 2019), IEEE, OECD, General Data Protection Regulation (GDPR), COMEST, Japanese Society for Artificial Intelligence (JSAI), American Medical Association (AMA, 2018), etc. proposed the ethics guidelines and frameworks for trustworthy AI deployments (Hagendorff, 2020; Mökander and Floridi, 2021; Jobin et al., 2019). Authors covered diverse topics from medical ethics and systems theory (Lauer, 2021), marketing ethics (Davenport et al., 2020), tertiary study (Cubric, 2020), and technological diversification (Ceipek et al., 2019) to indicate the widespread ethics concerns in AI deployment. At the same time, established theories based on comprehensive studies using utilitarian/deontological business ethics, institutional theory, triple bottom line (TBL), transition theory, and corporate social responsibility (CSR) were missing while performing the AI ethics-led sustainability keywords search during the literature review. The output of code frequencies very well validates the same phenomenon of less usage of prominent business ethics theories like utilitarian/deontological, with less than 1% with a frequency of just 37 (Table 6).

The research observes a minimal AI ethics-led sustainability context about specific industries. Most of the context illustrations are not industry specific rather than focused on ethical consequences while deploying AI systems. 65 plus percent of the shortlisted papers targeted generic industry-based segmentation (Table 3). The government sector is also an apparent and common industry that emerges from our review, covering AI deployment use cases and contextualised and researched ethics requirements (10%, Table 3). The research finds specific studies in the areas of information technology (6.67%), healthcare (3.33%), education (6.67%) and management domains of human resources and marketing (3.33%). The research emphasises the transforming relationship between humans and machines, the coexistence of human and AI ethical machines, and the relevance of virtue for ethical design (Neubert & Montañez, 2020). The research investigates various contexts of trustworthy human-AI partnerships, protects information privacy perspectives and the ethical concerns of design, deployment, and usage of AI). technologies (Hunkenschroer & Luetge, 2022; Spinello, 2021; Stahl et al., 2021). A relational theory of society is used to study social change in the digitalisation of work and explore precondition solidary-economy networks (Rodriguez-Lluesma et al., 2021). Zhu's (2020) essay investigates technology's role in ethics and society with deontology and consequentialism ethics based on Confucian ethics from China. Contextualised frameworks from HLEG and OECD emphasise the need for ethical AI deployment audit practices for humanitarian relief and efficient European law-making (Jobin et al., 2019). Singapore, China, Australia, Finland, Japan, the USA, Canada, the UK, and India have all documented the ethical guidelines for successful, trustworthy, responsible frameworks and principles for AI deployment (Fig. 5).

Most papers highlight the importance of AI ethics-led sustainability in AI project deployments. Ethics, having philosophical roots, inspires a good number of papers focused on the context of virtues like love, prudence, social issues, metaphysical questions, and AI deployment discourse (Stahl et al., 2021; Neubert & Montañez, 2020; Hickok, 2021). Researchers indicate the role of governance and ethics, especially where it plays a role in the AI-led technological diversifications documented by organisations, societies, R&D, and AI ethics-led sustainability guidelines and principles (Ceipek et al., 2019). Spinello's (2021) paper amplifies common good perspectives for the right to a) physical security and b) privacy and proposes an exceptional access scheme to mitigate risk. Past studies debate the ethical principles rather than practices and awareness of the impact of AI ethics-led sustainability on utility and society for the developers and research community (Jobin et al., 2019; Floridi et al., 2020). There is a need to consolidate such debates and open issues for ethical developers, regulators, and decision-makers. Most of the papers present the human side angle of ethics while deploying AI projects. Toth et al. (2022) highlight the study of machines' and robots' rights to argue how moral agency pursues the rights of machines. Research envisages deploying ethical systems, recognising moral issues, cultivating ethical intentions, and behaviour that recognises and interprets principles, rules, and recommendations to formulate better AI ethics-led sustainability practices (Wamba et al., 2021). However, the literature shows ignorance of technological adoption, diffusion, transitioning, institutional culture, and the climate of ethical AI toolkits and deployment in the practical space (Table 7). Demographics (e.g., age, gender, location, etc.) as AI ethics-led sustainability control variables in toolkits-based training datasets can reveal a fresh outlook for the adoption and diffusion of technology missing in the literature.

The research in Table 7 covers 6 widely accepted, practical *AI ethics toolkits* from different technological and research organisations with their core principles, the technology used, objectives, and adoption maturity. IBM's AI Fairness 360 is an open-source toolkit to investigate and remove the bias and discrimination in the datasets in ML models. The overall objective is to check for undesirable bias and discrimination in the model datasets with the help of the latest algorithms (Bellamy et al., 2019). Google Model Cards is an open-source condition-based document that supports ML models based on benchmarked performance evaluations. Model Cards' short documents primarily target ML models for vision, face detection, and natural language processing (Mitchell et al. (2019)). Aequitas is an open-source toolkit used to audit ML models to identify bias and discrimination in the datasets. It is used to develop impartial decision-making models while deploying predictive systems. Aequitas's primary goal is to recommend equitable action-based and informed decision-making processes following Equal, proportional, false positive, or false negative parity metrics (Saleiro et al. (2018)). Deon from DrivenData is a command line interface (CLI) tool for the ethical aspects of AI/ML projects. The tool supports multiple formats (new/existing) to append checklist files in the data science project. The idea is to remind the developers about the ethical conversation with actionable checklists in ML models (DrivenData, 2019). Ethical OS from The Institute for the Future, Digital Intelligence Lab, Omidyar Network is the Open Source/Enterprise level risk-based checklist comprising 8 zones that can harm humans. It raises the 14 scenarios to catalyse the conversation of social harm and proposes 7 strategies to safeguard the future. The objective is to develop a toolkit to respond to the unpleasant consequences while working on AI/ML projects and products (Lilley et al., 2020). The XAI toolkit is the product of The Institute for Ethical AI & ML Institute, which comprises ML libraries that provide AI explainability as a central theme. It helps to perform ML data model analysis and evaluation following AI ethics principles. The objective is to provide a library-based toolkit for data study, design, synthesise, simulate and evaluate the model for forecasting and monitoring purposes (Xai, 2021).

The research is considered valued and of quality if it is balanced in applied methods from qualitative (e.g., literature review, case method, thematic analysis, ethnography, etc.) to quantitative methods (e.g., cluster analysis, parametric and non-parametric studies, path and path and mediation analysis, etc.). Conceptual model/ framework-based AI ethics-led sustainability principles papers contributed significantly to our survey with a percentage of 36.67% (Table 4). Systematic literature /bibliometric review is the second most applied method to investigate the impact of AI deployments on ethical consequences, with a frequency percentage of 40.00% (Table 4). User Stories/Use case method is also an obvious choice to gather first-hand end-user experiences of AI deployments and their nuisances [like Genderify service to identify individual's gender, Tay chatbot to forecast criminality, etc. (Lauer, 2021)] from the live cases with a frequency percentage of 6.67%

Table 7
Practical AI ethics-led sustainability toolkits.
Source: Author's own compilation.

S.No	Toolkit	Company / Community	Solution Type (Opensource/ Proprietary)	Details	Followed AI Ethics Principles	Programming /Scripting Language	Objective	Research Adoption	Industry Adoption
1	AI Fairness 360	IBM	Open Source	AI Fairness 360 is open source toolkit to investigate and remove the bias and discrimination in the datasets in ML models.	Explainability; fairness; robustness; transparency and privacy	R, Python	Objective is to check for undesirable bias and discrimination in the model datasets with the help of latest algorithms.	Yes	Yes
2	Google Model Cards	Google	Open Source/ Commercial	Model cards are condition-based documents that support ML models based on benchmarked performance evaluations.	Socially beneficial; avoid unfair bias; safety; accountable; Incorporate privacy design; high standards; made available accord to principles	MediaPipe platform-Android, iOS, Python, JavaScript	Model cards short documents are primarily targeted to benchmark people-centric ML models for vision, face detection and natural language processing.	Yes	Yes
3	Aequitas	Data Science Policy, University of Chicago	Open Source	It is an open source-based toolkit to audit ML models identifying bias and discrimination in the datasets. It is used for developing impartial decision-making models while deploying the predictive systems.	Understand model bias; level of bias with parity; visualize bias metrics; swifter comprehension	R, Python	Aequitas primary goal is to recommend equitable action based on informed decision-making process following Equal, proportional, false positive, false negative parity metrics.	Yes	Yes
4	deon	Drivendata	Open Source	It is a command line interface (CLI) tool to ethical aspects of AI/ML projects. The tool supports (new/existing) multiple formats to append checklists files in the data science project.	Deon toolkit is based on following concepts. 1. Data collection based on consent, balanced personally identifiable information (pii), bias collection and mitigation 2. Data storage based on security, forgotten rights and retention plan 3. Analysis based on missing context, bias, honest depiction, privacy, auditability 4. Modelling based on proxy discrimination, fairness, metric selection, explainability, bias communication 5. Deployment based on redress, roll back, concept drift, unintended use	Rendered versions: Markdown, HTML, Jupyter Notebook, RST, Text	The overall idea is to remind the developers about the ethical conversation with actionable checklists in ML models.	Yes	Yes
5	EthicalOS	The Institute for the Future, Digital Intelligence Lab, Omidyar Network	Open Source/ Enterprise	This is risk-based checklist comprising 8 zones that can be harmful to human. It raises the 14 scenarios to catalyse the conversation of social harm and proposes 7 strategies to safeguard the future.	Trust, misinformation and propaganda; addiction and dopamine economy; privacy & data governance; environmental & societal wellbeing; governmental surveillance; data control and data as valuta; implicit trust and usability insights; evil and criminal act; diversity, non-discrimination & fairness; economical unfairness; machine ethics and algorithmic bias	Not available	Objective is to develop the toolkit to respond to the unpleasant consequences while working on AI/ML projects and products.	Yes	Yes

(*continued on next page*)

Table 7 (*continued*)

S.No	Toolkit	Company / Community	Solution Type (Opensource/ Proprietary)	Details	Followed AI Ethics Principles	Programming /Scripting Language	Objective	Research Adoption	Industry Adoption
6	XAI	The Institute for Ethical AI & ML	Open Source	XAI is a Machine Learning library that is designed with AI explainability in its core. XAI contains various tools that enable for analysis and evaluation of data and models	Human augmentation, Bias evaluation, Explainability by justification, Reproducible operations, Displacement strategy, Practical accuracy, Trust by privacy, Data risk awareness	R, Python	Objective is to provide library-based toolkit for data analysis, model performance evaluation and monitoring.	Yes	Yes

(Table 4). Topic Modeling (3.33%), case methods (13.33%), and multidimensional empirical studies (3.33%) are the other methods that the researchers used to study the ethical aspects of AI deployment. One crucial observation is our study's lack of quantitative analysis as the outcome.

Therefore, one whole series of factor analyses–path and mediation methods, parametric and non-parametric studies, and partial least squares–was missing entirely in the surveyed papers. Interview-based thematic analysis was also unavailable in plenty (3.33%), which shows the dearth of researchers and practitioners in the domain. This imbalance affects the research quality as a whole. While conducting the missed quantitative studies, proper sample identification and data collection techniques-based rigour should be applied to strengthen the research.

4.1. Theoretical implication

Current AI deployments influence the broader set of ethical aspects and interfaces to environmental, social, governance (ESG) and sustainable development goals (SDG). AI deployments provide an excellent opportunity to link ESG and SDG criteria with ethical aspects while deploying AI projects. Human-centric economics and resource and cognitive impacts can be further strengthened and investigated using established utilitarian/ deontological business ethics, TBL, CSR, equity and human rights, behavioural systems, and institutional and transition theories. Developed nations have laid down the country- or region-specific principles and frameworks targeting techno-socio needs; therefore, people-centric ethical AI studies are required extensively in developing countries. The literature survey reveals that a global AI ethics-led sustainability standard agency must regulate and standardise principles and frameworks, pointing to various guidelines by governmental agencies, analyst' groups, technology giants, and traditional agencies. AI ethics-led sustainability principles and framework survey Figs. 2–5 can be used as a baseline for global standardised people-centric ethics guidelines. From the previous studies, debate and open issues of AI ethics and its impact on utility and society need consolidation for ethical developers, regulators, and decision-makers. Demographic characteristics (e.g., age, gender, location, etc.) considered as AI ethics-led sustainability control variables in toolkits-based training datasets can reveal a fresh outlook. The research outcome recommends conducting quantitative studies with empirical evidence to strengthen the ethical impacts of AI deployment. Research attention is required to conduct factor analysis, path and mediation methods, parametric and non-parametric studies, and partial least square with rigorous sample identification and data collection methods. Interview-based thematic analysis is also sought to maintain the balance. While ethics-based empirical studies are available for reference, ethics theory-based constructs can be studied and tested for sound synthesis. Human-centric ethical aspects can be studied for vertical industries, geographies, and micro/macro business enterprises per the context. The literature survey also recommends introducing AI ethics-led sustainability principles into the early education system to appreciate its importance for unbiased, inclusive, transparent software architectural paradigms and for designing future social engineering models.

4.2. Practical implications

The intent to explore the ethical aspect of the AI project deployment cannot be completed without the engagement effort of the managers and leaders. The discussed/surveyed toolkits and framework/ principles can be introduced and practised well when the managers develop, adhere to, and execute good AI ethics-led sustainability programs. The practitioner's approach is required to embrace the multidimensional spirit of ethical design, critique, deploying toolkits (ethical), prioritising

ethical modelling, and safeguarding the interest of humans while deploying AI projects. Managers can mobilise the workforce by emphasising the sensitivity of human interests and charting the consequences if they are not handled appropriately for AI cybernetics projects. Managers themselves or brand ambassadors can serve as torchbearers for ethical characteristics to bridge the gaps of social, normative, moral, and virtue-based theories and the practical aspects of explainability, fairness, governance, privacy, and inclusiveness. At the ground level, practitioners can make developers and designers better understand cybernetics, safety, and reliability with the help of training, certification, and mandatory policies. The responsibility of the managers becomes more prominent in developing understandable instruments, toolkits, and collaterals to balance the autonomy, risks, and ethics introduced by data scientists between humans and machines. AI processes are envisaged to implement ethical preferences, heed social and moral values, adhere to compliance and principles, and practice frameworks. Managers must propagate the AI systems' viable assumptions, capabilities, and limitations for the informed acceptance of the desired systems to be deployed, ensuring safe, inclusive, and fair AI projects.

5. Conclusion

AI advantages are visible in every walk of life. Different industries become diffused by AI use cases empowering technical and social settings. Conversely, constant pressure exists to handle technology's ethical, social, and sustainable use in AI deployments. The paper explores the ethical aspects that must be practised while deploying any AI system. The findings baseline that algorithmic biasedness, explainability, autonomy, privacy, beneficence, non-maleficence, and justice lay the foundations of AI ethics-led sustainability frameworks.

The research establishes the ethical codes to be considered while conceptualising, designing, implementing, and managing AI ethics-led sustainability frameworks. The AI ethics toolkits discussed in this paper are still part of the research labs and are available at the centre of excellence (COEs). This research must be part of the mainstream practices and requires discussion and research. The AI ethics-led sustainability guideline documents (principles/frameworks) sources not listed in the scholarly databases require considerable work to consolidate and categorise as reviewed literature. One of the paper's findings suggests conducting empirical tests proven in the literature survey. Seeing the sector wise (vertical/domain) analysis of the AI projects with respect to ethics would be constructive. Proposed thematic codes bundled with AI practitioners and researchers can reveal new perspectives altogether. Most AI ethics toolkits are positioned from the open-source world; solution-centric scouting and evaluation are required to safeguard intellectual interests while pursuing ethical AI deployments. The discussion section also gives natural options to 'rethink' and address current literature, open issues and debates. Research frames the multiple lenses of literature, practices, platform/toolkits, and thematic analysis on ethical theories to recommend responsible and accountable ethical AI deployment. The findings can be a rich glossary of choosing ethical practices coupled with the right fit prebuilt/customised AI ethics toolkit to reduce valuable experimentation cycle time.

Further research could investigate these issues in different institutional settings around the world. AI ethics might vary depending on legal and cultural settings in different countries (e.g., Mollah et al., 2021; Rabbani et al., 2022). Further research of AI ethics in different contexts could help implement practice and policy in ways tailored to different country settings.

Research involving human participants and/or animals

Not applicable.

Informed consent

Not applicable.

CRediT authorship contribution statement

Douglas Cumming: Writing – review & editing, Supervision, Formal analysis, Conceptualization. **Kumar Saurabh:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Neelam Rani:** Writing – review & editing, Supervision, Formal analysis, Conceptualization. **Parijat Upadhyay:** Writing – review & editing, Supervision, Formal analysis, Conceptualization.

Declaration of Competing Interest

The authors have nothing to disclose – no competing interests.

Declaration of Generative AI and AI-assisted technologies in the writing process

The authors have nothing to disclose – AI was not used in writing.

References

Accenture, U.K. (2019). Responsible ai and robotics, an ethical framework. ⟨https://www.accenture.com/gb-en/company-responsible-ai-robotics⟩.

AINOW, (2019). AINOW report. ⟨https://ainowinstitute.org/AI_Now_2019_Report.pdf⟩.

Allen, F., Gu, X., & Jagtiani, J. (2021). A survey of fintech research and policy discussion. *Review of Corporate Finance, 1*(3-4), 259–339.

AMA, (2018). Augmented intelligence, technology – medical, augmented intelligence in health care h-480.940. American Medical Association (AMA) USA. ⟨https://policysearch.ama-assn.org/policyfinder/detail/augmented%20intelligence?uri=%2FAMADoc%2FHOD.xml-H-480.940.xml⟩.

Ashok, M., Madan, R., Joha, A., & Sivarajah, U. (2022). Ethical framework for artificial intelligence and digital technologies. *International Journal of Information Management, 62*, Article 102433.

BAAI, (2019). Beijing AI principles. Beijing: Beijing academy of artificial intelligence 2019. ⟨https://www.baai.ac.cn/blog/beijing-ai-principles⟩.

Bellamy, R. K., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., & Zhang, Y. (2019). AI fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development, 63*(4/5), 1–4.

Bertino, E., Kundu, A., & Sura, Z. (2019). Data transparency with blockchain and AI ethics. *Journal of Data and Information Quality, 11*(4), 1–8.

Borenstein, J., & Howard, A. (2021). Emerging challenges in AI and the need for AI ethics education. *AI and Ethics, 1*(1), 61–65.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Briner, R. B., & Denyer, D. (2012). Systematic review and evidence synthesis as a practice and scholarship tool.

Burmeister, O. K. (2017). Professional ethics in the information age. *Journal of Information, Communication and Ethics in Society, 15*(4), 348–356.

Carrillo, M. R. (2020). Artificial intelligence: From ethics to law. *Telecommunications Policy, 44*(6), Article 101937.

Ceipek, R., Hautz, J., Mayer, M. C., & Matzler, K. (2019). Technological diversification: A systematic review of antecedents, outcomes and moderating effects. *International Journal of Management Reviews, 21*(4), 466–497.

Chauhan, C., & Gullapalli, R. R. (2021). Ethics of AI in pathology: Current paradigms and emerging issues. *The American Journal of Pathology, 91*(10), 1673–1683.

COMEST, U. (2017). Report of COMEST on robotics ethics. World commission on the ethics of scientific knowledge and technology (COMEST). http://unesdoc.unesco.org/images/0025/002539 E, 253952.

Cubric, M. (2020). Drivers, barriers and social considerations for AI adoption in business and management: A tertiary study. *Technology in Society, 62*, Article 101257.

Damioli, G., Van Roy, V., & Vertesy, D. (2021). The impact of artificial intelligence on labor productivity. *Eurasian Business Review, 11*, 1–25.

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science, 48*(1), 24–42.

Davenport, T.H., & Katyal, V. (2018). Every leader's guide to the ethics of AI. MIT Sloan Management Review, The USA: Massachusetts Institute of Technology. https://sloanreview.mit.edu/article/every-leaders-guide-to-the-ethics-of-ai.

Davenport, T. H., & Kirby, J. (2016). *Only Humans Need Apply: Winners and Losers in the Age of Smart Machines*. New York: Harper Business.

Dawson, D., Schleiger, E., Horton, J., McLaughlin, J., Robinson, C., Quezada, G., & Hajkowicz, S. (2019). *Artificial intelligence: Australia's ethics framework*⟨https://apo.org.au/node/229596⟩.

Declaration, M. (2018). The Montreal declaration for the responsible development of artificial intelligence launched. ⟨https://monoskop.org/images/b/b2/Report_

Montreal_Declaration_for_a_Responsible_Development_of_Artificial_Intelligence_2018.pdf〉.
Delloite, (2019). Building trustworthy AI, A comprehensive approach to conduct, data protection and ethics. Delloite. 〈https://www2.deloitte.com/content/dam/Deloitte/ch/Documents/audit/deloitte-ch-en-audit-building-trustworthy-ai.pdf〉.
DrivenData, (2019). Deon: An ethics checklist for data. 〈https://deon.drivendata.org/〉.
Etzioni, A., & Etzioni, O. (2016). AI assisted ethics. *Ethics and Information Technology, 18*(2), 149–156.
FAAI, (2017). Turning Finland into a leading country in the application of artificial intelligence. Publications of the Ministry of Economic Affairs and Employment Ministry. 〈https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf?sequence=1&isAllowed=y〉.
Feher, K., & Katona, A. I. (2021). Fifteen shadows of socio-cultural AI: A systematic review and future perspectives. *Futures, 132*, Article 102817.
Ferrell, O. C., Larry, G., & Fraedrich, John (1989). A synthesis of ethical decision models for marketing. *Journal of Macro Marketing, 9*(2), 55–64.
Floridi, L. (2019). Translating principles into practices of digital ethics: Five risks of being unethical. *Philosophy & Technology, 32*(2), 185–193.
Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., & Vayena, E. (2021). An ethical framework for a good ai society: Opportunities, risks, principles, and recommendations. *Ethics, Governance, and Policies in Artificial Intelligence Springer, Cham,* 19–39.
Floridi, L., Cowls, J., King, T. C., & Taddeo, M. (2020). How to design AI for social good: Seven essential factors. *Science and Engineering Ethics, 26*(3), 1771–1796.
Fröding, B., & Peterson, M. (2020). Friendly AI. *Ethics and Information Technology, 23*, 207–214.
Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation. *AI Magazine, 38*(3), 50–57.
Google, (2021). Responsible development of AI. Google. 〈https://ai.google/static/documents/responsible-development-of-ai.pdf〉.
Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines, 30*(1), 99–120.
Hermann, E. (2021). Leveraging artificial intelligence in marketing for social good—an ethical perspective. *Journal of Business Ethics,* 1–19.
Hickok, M. (2021). Lessons learned from AI ethics principles for future actions. *AI and Ethics, 1*(1), 41–47.
Hleg, A.I. (2019). Ethics guidelines for trustworthy AI. B-1049 Brussels. 〈https://www.aepd.es/sites/default/files/2019–12/ai-ethics-guidelines.pdf〉.
Hoffman, D. & Masucci R. (2018). Intel's AI privacy policy. White Paper. Intel. 〈https://blogs.intel.com/policy/files/2018/10/Intels-AI-Privacy-Policy-White-Paper-2018.pdf〉.
Holstein, J. A., & Gubrium, J. F. (1997). Active interviewing. In D. Silverman (Vol. Ed.), *Qualitative research: theory, method and practice: 5*, (pp. 113–129). London: Sage.
Hongladarom, S. (2021). The Thailand national AI ethics guideline: An analysis. *Journal of Information, Communication and Ethics in Society*.
How, J. P. (2018). Ethically aligned design. *IEEE Control Systems Magazine, 38*(3), 3–4.
Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-enabled recruiting and selection: A review and research agenda. *Journal of Business Ethics,* 1–31.
IBM, (2021). IBM's multidisciplinary, multidimensional approach to trustworthy AI. 〈https://www.ibm.com/artificial-intelligence/ethics〉.
Jerome, P. (2021). Facebook's five pillars of Responsible AI. *Facebook*. 〈https://ai.facebook.com/blog/facebooks-five-pillars-of-responsible-ai/〉.
Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence, 1*(9), 389–399.
Jobin, A., Ienca, M., & Vayena, E. (2020). The global landscape of AI ethics guidelines. *Nature, 1*(9), 389–399.
JSAI, (2017). The Japanese Society for Artificial Intelligence Ethical Guidelines. 〈http://ai-elsi.org/wp-content/uploads/2017/05/JSAI-Ethical-Guidelines-1.pdf〉.
KPMG, (2021). The shape of AI governance to come. KPMG. 〈https://assets.kpmg/content/dam/kpmg/xx/pdf/2021/01/the-shape-of-ai-governance-to-come.pdf〉.
Krishnan, M. (2019). Against interpretability: A critical examination of the interpretability problem in machine learning. *Philosophy & Technology, 31*(2), 487–502.
Lauer, D. (2021). You cannot have AI ethics without ethics. *AI and Ethics, 1*(1), 21–25.
Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. 〈https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3403301〉.
Lilley, M., Currie, A., Pyper, A., & Attwood, S. (2020). *Using the ethical OS toolkit to mitigate the risk of unintended consequences. International conference on human-computer interaction.* Cham: Springer77–82.
Lords, H.O. (2018). AI in the UK: Ready, willing and able? 〈https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf〉.
Malkin, N., Deatrick, J., Tong, A., Wijesekera, P., Egelman, S., & Wagner, D. (2019). Privacy attitudes of smart speaker users. *Proceedings on Privacy Enhancing Technologies, 2019*(4), 250–271.
May, D. R., & Pauli, K. P. (2002). The role of moral intensity in ethical decision making: A review and investigation of moral recognition, evaluation, and intention. *Business & Society, 41*(1), 84–117.
McGrath, J., & Gupta, A. (2018). Writing a moral code: Algorithms for ethical reasoning by humans and machines. *Religions, 9*(8), 240–252.
McLaren, B. M. (2003). Extensionally defining principles and cases in ethics: An AI model. *Artificial Intelligence, 150*(1-2), 145–181.
Meszaros, J., & Ho, C. H. (2021). AI research and data protection: Can the same rules apply for commercial and academic research under the GDPR? *Computer Law & Security Review, 41*, Article 105532.
Microsoft, (2021). Responsible AI, Microsoft AI principles. Microsoft. 〈https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1:primaryr6〉.
Mills, S., Baltassis, E., Santinelli, M., Carlisi, C., Duranton, S., & Gallego, A. (2020). Six steps to bridge the responsible AI gap. 〈https://www.bcg.com/publications/2020/six-steps-for-socially-responsible-artificial-intelligence〉.
Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., & Gebru, T. (2019). Model cards for model reporting. *In Proceedings of the Conference on Fairness, Accountability, and Transparency,* 220–229.
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. Prisma Group. (2009). Reprint-preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Physical Therapy, 89*(9), 873–880.
Mökander, J., & Floridi, L. (2021). Ethics-based auditing to develop trustworthy AI. *Minds and Machines, 31*(2), 323–327.
Mollah, S., Skully, M., & Liljeblom, E. (2021). Strong boards and risk-taking in Islamic banks. *Review of Corporate Finance, 1*(1-2), 135–180.
Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). Ethics as a service: a pragmatic operationalisation of AI ethics. *Minds and Machines, 31*(2), 239–256.
Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics, 26*(4), 2141–2168.
Morley, J., Machado, C. C., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A mapping review. *Social Science & Medicine*Article 113172.
Müller, V.C. (2020). Ethics of artificial intelligence and robotics. Stanford Encyclopaedia of Philosophy. 〈https://plato.stanford.edu/entries/ethics-ai/?fbclid=IwAR3zBI5BYERCGCdEBZhAvLHExNJhPUJA9SYkvwteRUdmXBgB3ILfUk6y81o〉.
Munoko, I., Brown-Liburd, H. L., & Vasarhelyi, M. (2020). The ethical implications of using artificial intelligence in auditing. *Journal of Business Ethics, 167*(2), 209–234.
Neubert, M. J., & Montañez, G. D. (2020). Virtue as a framework for the design and use of artificial intelligence. *Business Horizons, 63*(2), 195–204.
NIST, (2021). Artificial Intelligence, Benchmarks & Metrics. National Institute of Standards and Technology (NIST). 〈https://www.nist.gov/artificial-intelligence/benchmarks-metrics〉.
NITI Aayog, (2021). Responsible ai #aiforall. Approach document for India part 1 – principles for responsible AI. NITI Aayog. 〈https://www.niti.gov.in/sites/default/files/2021–02/Responsible-AI-22022021.pdf〉.
ÓhÉigeartaigh, S. S., Whittlestone, J., Liu, Y., Zeng, Y., & Liu, Z. (2020). Overcoming barriers to cross-cultural cooperation in AI ethics and governance. *Philosophy & Technology, 33*(4), 571–593.
Rabbani, M. R., Sarea, A., Khan, S., & Abdullah, Y. (2022). Ethical concerns in artificial intelligence (AI): The role of RegTech and Islamic finance. In A. M. A. Musleh Al-Sartawi (Ed.). *Artificial intelligence for sustainable finance and sustainable technology. ICGER 2021. Lecture notes in networks and systems*Cham: Springer. https://doi.org/10.1007/978-3-030-93464-4_38
Resseguier, A., & Rodrigues, R. (2020). AI ethics should not remain toothless! A call to bring back the teeth of ethics. *Big Data & Society, 7*(2) 2053951720942541.
Rodriguez-Lluesma, C., García-Ruiz, P., & Pinto-Garay, J. (2021). The digital transformation of work: A relational view. *Business Ethics, The Environment & Responsibility, 30*(1), 157–167.
Rogerson, S., Miller, K. W., Winter, J. S., & Larson, D. (2019). Information systems ethics–challenges and opportunities. *Journal of Information, Communication and Ethics in Society, 17*(1), 87–97.
Saleiro, P., Kuester, B., Hinkson, L., London, J., Stevens, A., Anisfeld, A., & Ghani, R. (2018). Aequitas: A bias and fairness audit toolkit. https://arxiv.org/abs/1811.05577.
Sareen, S., Saltelli, A., & Rommetveit, K. (2020). Ethics of quantification: illumination, obfuscation and performative legitimation. *Palgrave Communication, 6*(1), 1–5.
Schiff, D., Biddle, J., Borenstein, J., & Laas, K. (2020). What's next for AI ethics, policy, and governance? A global overview. *In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society,* 153–158.
Siau, K., & Wang, W. (2020). Artificial intelligence (AI) ethics: Ethics of AI and ethical AI. *Journal of Database Management, 31*(2), 74–87.
Spinello, R. A. (2021). The ethical consequences of "going dark". *Business Ethics, the Environment & Responsibility, 30*(1), 116–126.
Stahl, B. C., Andreou, A., Brey, P., Hatzakis, T., Kirichenko, A., Macnish, K., & Wright, D. (2021). Artificial intelligence for human flourishing–beyond principles for machine learning. *Journal of Business Research, 124*, 374–388.
Timmers, P. (2019). Ethics of AI and cybersecurity when sovereignty is at stake. *Minds and Machines, 29*(4), 635–645.
Tóth, Z., Caruana, R., Gruber, T., & Loebbecke, C. (2022). The dawn of the AI robots: towards a new framework of AI robot accountability. *Journal of Business Ethics, 178*(4), 895–916.
Tóth, Z., Caruana, R., Gruber, T., & Loebbecke, C. (2022). The dawn of the AI robots: Towards a new framework of AI robot accountability. *Journal of Business Ethics,* 1–22.
Urquhart, L., Reedman-Flint, D., & Leesakul, N. (2019). Responsible domestic robotics: Exploring ethical implications of robots in the home. *Journal of Information, Communication and Ethics in Society, 17*(2), 246–272.
Wamba, S. F., Bawack, R. E., Guthrie, C., Queiroz, M. M., & Carillo, K. D. A. (2021). Are we preparing for a good AI society? A bibliometric review and research agenda. *Technological Forecasting and Social Change, 164*, Article 120482.

Watson, G. J., Desouza, K. C., Ribiere, V. M., & Lindič, J. (2021). Will AI ever sit at the C-suite table? The future of senior leadership. *Business Horizons, 64*(4), 465–474.
Webster, C., & Ivanov, S. (2020). *Robotics, artificial intelligence, and the evolving nature of work. Digital transformation in business and society*. Cham: Palgrave Macmillan127–143.
Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019). The role and limits of principles in AI ethics: towards a focus on tensions. *In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society,* 195–200.
Wu, F., et al. (2020). Towards a new generation of artificial intelligence in China. *Nature Machine Intelligence, 2*(6), 312–316.
Xai, (2021). eXplainableAI framework. The Institute for Ethical AI & ML. https://ethical.institute/xai.html.
Yang, G. Z., Bellingham, J., Dupont, P. E., Fischer, P., Floridi, L., Full, R., & Wood, R. (2018). The grand challenges of science robotics. *Science Robotics, 3*(14), Article eaar7650.
Yang, K., & Lee, H. J. (2010). Gender differences in using mobile data services: utilitarian and hedonic value approaches. *Journal of Research in Interactive Marketing, 4*(2), 142–156.
Yeung, K. (2020). Recommendation of the council on artificial intelligence (OECD). *International Legal Materials, 59*(1), 27–34.
Zhu, Q. (2020). Ethics, society, and technology: A Confucian role ethics perspective. *Technology in Society, 63*, Article 101424.