AI and Ethics
https://doi.org/10.1007/s43681-023-00387-1

REVIEW

Publics' views on ethical challenges of artificial intelligence: a scoping review

Helena Machado[1] · **Susana Silva**[2] · **Laura Neiva**[3]

Received: 8 October 2023 / Accepted: 16 November 2023

Abstract
This scoping review examines the research landscape about publics' views on the ethical challenges of AI. To elucidate how the concerns voiced by the publics are translated within the research domain, this study scrutinizes 64 publications sourced from PubMed® and Web of Science™. The central inquiry revolves around discerning the motivations, stakeholders, and ethical quandaries that emerge in research on this topic. The analysis reveals that innovation and legitimation stand out as the primary impetuses for engaging the public in deliberations concerning the ethical dilemmas associated with AI technologies. Supplementary motives are rooted in educational endeavors, democratization initiatives, and inspirational pursuits, whereas politicization emerges as a comparatively infrequent incentive. The study participants predominantly comprise the general public and professional groups, followed by AI system developers, industry and business managers, students, scholars, consumers, and policymakers. The ethical dimensions most commonly explored in the literature encompass human agency and oversight, followed by issues centered on privacy and data governance. Conversely, topics related to diversity, nondiscrimination, fairness, societal and environmental well-being, technical robustness, safety, transparency, and accountability receive comparatively less attention. This paper delineates the concrete operationalization of calls for public involvement in AI governance within the research sphere. It underscores the intricate interplay between ethical concerns, public involvement, and societal structures, including political and economic agendas, which serve to bolster technical proficiency and affirm the legitimacy of AI development in accordance with the institutional norms that underlie responsible research practices.

Keywords Artificial intelligence · Ethics · Public involvement · Publics' views · Responsible research

1 Introduction

Current advances in the research, development, and application of artificial intelligence (AI) systems have yielded a far-reaching discourse on AI ethics that is accompanied by calls for AI technology to be democratically accountable and trustworthy from the publics'[1] perspective [1–5]. Consequently, several ethics guidelines for AI have been released in recent years. As of early 2020, there were 167 AI ethics guidelines documents around the world [6]. Organizations such as the European Commission (EC), the Organization for Economic Co-operation and Development (OECD),

✉ Helena Machado
hmachado@ics.uminho.pt

[1] Department of Sociology, Institute for Social Sciences, University of Minho, Braga, Portugal

[2] Department of Sociology and Centre for Research in Anthropology (CRIA), Institute for Social Sciences, University of Minho, Braga, Portugal

[3] Institute for Social Sciences, Communication and Society Research Centre (CECS), University of Minho, Braga, Portugal

[1] In this article, we will employ the term "publics" rather than the singular "public" to delineate our viewpoint concerning public participation in AI. Our option is meant to acknowledge that there are no uniform, monolithic viewpoints or interests. From our perspective, the term "publics" allows for a more nuanced understanding of the various groups, communities, and individuals who may have different attitudes, beliefs, and concerns regarding AI. This choice may differ from the terminology employed in the referenced literature.

Published online: 19 December 2023

and the United Nations Educational, Scientific and Cultural Organization (UNESCO) recognize that public participation is crucial for ensuring the responsible development and deployment of AI technologies,[2] emphasizing the importance of inclusivity, transparency, and democratic processes to effectively address the societal implications of AI [11, 12]. These efforts were publicly announced as aiming to create a common understanding of ethical AI development and foster responsible practices that address societal concerns while maximizing AI's potential benefits [13, 14]. The concept of human-centric AI has emerged as a key principle in many of these regulatory initiatives, with the purposes of ensuring that human values are incorporated into the design of algorithms, that humans do not lose control over automated systems, and that AI is used in the service of humanity and the common good to improve human welfare and human rights [15]. Using the same rationale, the opacity and rapid diffusion of AI have prompted debate about how such technologies ought to be governed and which actors and values should be involved in shaping governance regimes [1, 2].

While industry and business have traditionally tended to be seen as having no or little incentive to engage with ethics or in dialogue, AI leaders currently sponsor AI ethics [6, 16, 17]. However, some concerns call for ethics, public participation, and human-centric approaches in areas such as AI with high economic and political importance to be used within an instrumental rationale by the AI industry. A growing corpus of critical literature has conceived the development of AI ethics as efforts to reduce ethics to another form of industrial capital or to coopt and capture researchers as part of efforts to control public narratives [12, 18]. According to some authors, one of the reasons why ethics is so appealing to many AI companies is to calm critical voices from the publics; therefore, AI ethics is seen as a way of gaining or restoring trust, credibility and support, as well as legitimation, while criticized practices are calmed down to maintain the agenda of industry and science [12, 17, 19, 20].

Critical approaches also point out that despite regulatory initiatives explicitly invoking the need to incorporate human values into AI systems, they have the main objective of setting rules and standards to enable AI-based products and services to circulate in markets [20–22] and might serve to avoid or delay binding regulation [12, 23]. Other critical studies argue that AI ethics fails to mitigate the racial, social, and environmental damage of AI technologies in any meaningful sense [24] and excludes alternative ethical practices [25, 26]. As explained by Su [13], in a paper that considers the promise and perils of international human rights in AI governance, while human rights can serve as an authoritative source for holding AI developers accountable, its application to AI governance in practice shows a lack of effectiveness, an inability to effect structural change, and the problem of cooptation.

In a value analysis of AI national strategies, Wilson [5] concludes that the publics are primarily cast as recipients of AI's abstract benefits, users of AI-driven services and products, a workforce in need of training and upskilling, or an important element for thriving democratic society that unlocks AI's potential. According to the author, when AI strategies articulate a governance role for the publics, it is more like an afterthought or rhetorical gesture than a clear commitment to putting "society-in-the-loop" into AI design and implementation [5, pp. 7–8]. Another study of how public participation is framed in AI policy documents [4] shows that high expectations are assigned to public participation as a solution to address concerns about the concentration of power, increases in inequality, lack of diversity, and bias. However, in practice, this framing thus far gives little consideration to some of the challenges well known for researchers and practitioners of public participation with science and technology, such as the difficulty of achieving consensus among diverse societal views, the high resource requirements for public participation exercises, and the risks of capture by vested interests [4, pp. 170–171]. These studies consistently reveal a noteworthy pattern: while references to public participation in AI governance are prevalent in the majority of AI national strategies, they tend to remain abstract and are often overshadowed by other roles, values, and policy concerns.

Some authors thus contended that the increasing demand to involve multiple stakeholders in AI governance, including the publics, signifies a discernible transformation within the sphere of science and technology policy. This transformation frequently embraces the framework of "responsible innovation",[3] which emphasizes alignment with societal imperatives, responsiveness

[2] The following examples are particularly illustrative of the multiplicity of organizations emphasizing the need for public participation in AI. The OECD Recommendations of the Council on AI specifically emphasizes the importance of empowering stakeholders considering essential their engagement to adoption of trustworthy [7, p. 6]. The UNESCO Recommendation on the Ethics of AI emphasizes that public awareness and understanding of AI technologies should be promoted (recommendation 44) and it encourages governments and other stakeholders to involve the publics in AI decision-making processes (recommendation 47) [8, p. 23]. The European Union (EU) White Paper on AI [9, p. 259] outlines the EU's approach to AI, including the need for public consultation and engagement. The Ethics Guidelines for Trustworthy AI [10, pp. 19, 239], developed by the High-Level Expert Group on AI (HLEG) appointed by the EC, emphasize the importance of public participation and consultation in the design, development, and deployment of AI systems.

[3] "Responsible Innovation" (RI) and "Responsible Research and Innovation" (RRI) have emerged in parallel and are often used interchangeably, but they are not the same thing [27, 28]. RRI is a policy-driven discourse that emerged from the EC in the early 2010s, while RI emerged largely from academic roots. For this paper, we will not consider the distinctive features of each discourse, but instead focus on the common features they share.

to evolving ethical, social, and environmental considerations, and the participation of the publics as well as traditionally defined stakeholders [3, 28]. When investigating how the conception and promotion of public participation in European science and technology policies have evolved, Macq, Tancoine, and Strasser [29] distinguish between "participation in decision-making" (pertaining to science policy decisions or decisions on research topics) and "participation in knowledge and innovation-making". They find that "while public participation had initially been conceived and promoted as a way to build legitimacy of research policy decisions by involving publics into decision-making processes, it is now also promoted as a way to produce better or more knowledge and innovation by involving publics into knowledge and innovation-making processes, and thus building legitimacy for science and technology as a whole" [29, p. 508]. Although this shift in science and technology research policies has been noted, there exists a noticeable void in the literature in regard to understanding how concrete research practices incorporate public perspectives and embrace multistakeholder approaches, inclusion, and dialogue.

While several studies have delved into the framing of the publics' role within AI governance in several instances (from Big Tech initiatives to hiring ethics teams and guidelines issued from multiple institutions to governments' national policies related to AI development), discussing the underlying motivations driving the publics' participation and the ethical considerations resulting from such involvement, there remains a notable scarcity of knowledge concerning how publicly voiced concerns are concretely translated into research efforts [30, pp. 3–4, 31, p. 8, 6]. To address this crucial gap, our scoping review endeavors to analyse the research landscape about the publics' views on the ethical challenges of AI. Our primary objective is to uncover the motivations behind involving the publics in research initiatives, identify the segments of the publics that are considered in these studies, and illuminate the ethical concerns that warrant specific attention. Through this scoping review, we aim to enhance the understanding of the political and social backdrop within which debates and prior commitments regarding values and conditions for publics' participation in matters related to science and technology are formulated and expressed [29, 32, 33] and which specific normative social commitments are projected and performed by institutional science [34, p. 108, [35, p. 856].

2 Methods

We followed the guidance for descriptive systematic scoping reviews by Levac et al. [36], based on the methodological framework developed by Arksey and O'Malley [37]. The steps of the review are listed below:

2.1 Stage 1: identifying the research question

The central question guiding this scoping review is the following: What motivations, publics and ethical issues emerge in research addressing the publics' views on the ethical challenges of AI? We ask:

- What motivations for engaging the publics with AI technologies are articulated?
- Who are the publics invited?
- Which ethical issues concerning AI technologies are perceived as needing the participation of the publics?

2.2 Stage 2: identifying relevant studies

A search of the publications on PubMed® and Web of Science™ was conducted on 19 May 2023, with no restriction set for language or time of publication, using the following search expression: ("AI" OR "artificial intelligence") AND ("public" OR "citizen") AND "ethics". The search was followed by backwards reference tracking, examining the references of the selected publications based on full-text assessment.

2.3 Stage 3: study selection

The inclusion criteria allowed only empirical, peer-reviewed, original full-length studies written in English to explore publics' views on the ethical challenges of AI as their main outcome. The exclusion criteria disallowed studies focusing on media discourses and texts. The titles of 1612 records were retrieved. After the removal of duplicates, 1485 records were examined. Two authors (HM and SS) independently screened all the papers retrieved initially, based on the title and abstract, and afterward, based on the full text. This was crosschecked and discussed in both phases, and perfect agreement was achieved.

The screening process is summarized in Fig. 1. Based on title and abstract assessments, 1265 records were excluded because they were neither original full-length peer-reviewed empirical studies nor focused on the publics' views on the ethical challenges of AI. Of the 220 fully read papers, 54 met the inclusion criteria. After backwards reference tracking, 10 papers were included, and the final review was composed of 64 papers.

2.4 Stage 4: charting the data

A standardized data extraction sheet was initially developed by two authors (HM and SS) and completed by two coders (SS and LN), including both quantitative and qualitative data (Supplemental Table "Data Extraction"). We used MS Excel to chart the data from the studies.

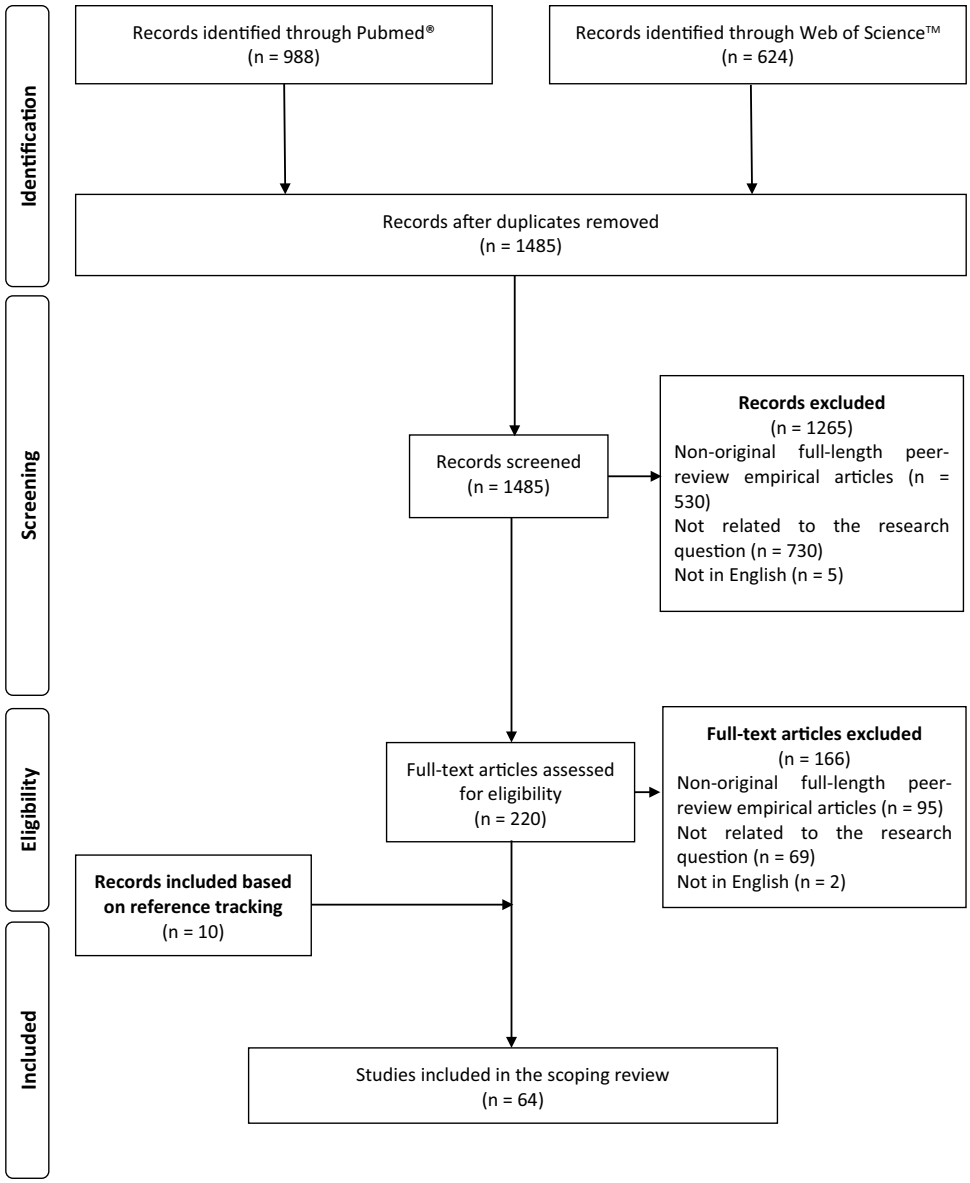

Fig. 1 Flowchart showing the search results and screening process for the scoping review of publics' views on ethical challenges of AI

The two coders independently charted the first 10 records, with any disagreements or uncertainties in abstractions being discussed and resolved by consensus. The forms were further refined and finalized upon consensus before completing the data charting process. Each of the remaining records was charted by one coder. Two meetings were held to ensure consistency in data charting and to verify accuracy. The first author (HM) reviewed the results.

Descriptive data for the characterization of studies included information about the authors and publication year, the country where the study was developed, study aims, type of research (quantitative, qualitative, or other), assessment of the publics' views, and sample. The types of research participants recruited as publics were coded into 11 categories: developers of AI systems; managers from industry and business; representatives of governance bodies; policymakers; academics and researchers; students; professional groups; general public; local communities; patients/consumers; and other (specify).

Data on the main motivations for researching the publics' views on the ethical challenges of AI were also gathered. Authors' accounts of their motivations were synthesized into eight categories according to the coding framework proposed by Weingart and colleagues [33] concerning public engagement with science and technology-related issues: education (to inform and educate the public about AI, improving public access to scientific knowledge); innovation (to promote innovation, the publics are considered to be a valuable source of knowledge and are called upon to contribute to knowledge production, bridge building and including knowledge outside 'formal' ethics); legitimation (to promote public trust in and acceptance of AI, as well as of policies

supporting AI); inspiration (to inspire and raise interest in AI, to secure a STEM-educated labor force); politicization (to address past political injustices and historical exclusion); democratization (to empower citizens to participate competently in society and/or to participate in AI); other (specify); and not clearly evident.

Based on the content analysis technique [38], ethical issues perceived as needing the participation of the publics were identified through quotations stated in the studies. These were then summarized in seven key ethical principles, according to the proposal outlined by the EC's Ethics Guidelines for Trustworthy AI [39]: human agency and oversight; technical robustness and safety; privacy and data governance; transparency; diversity, nondiscrimination and fairness; societal and environmental well-being; and accountability.

2.5 Stage 5: collating, summarizing, and reporting the results

The main characteristics of the 64 studies included can be found in Table 1. Studies were grouped by type of research and ordered by the year of publication. The findings regarding the publics invited to participate are presented in Fig. 2. The main motivations for engaging the publics with AI technologies and the ethical issues perceived as needing the participation of the publics are summarized in Tables 2 and 3, respectively. The results are presented below in a narrative format, with complimentary tables and figures to provide a visual representation of key findings.

There are some methodological limitations in this scoping review that should be taken into account when interpreting the results. The use of only two search engines may preclude the inclusion of relevant studies, although supplemented by scanning the reference list of eligible studies. An in-depth analysis of the topics explored within each of the seven key ethical principles outlined by the EC's Ethics Guidelines for Trustworthy AI was not conducted. This assessment would lead to a detailed understanding of the publics' views on ethical challenges of AI.

3 Results

3.1 Study characteristics

Most of the studies were in recent years, with 35 of the 64 studies being published in 2022 and 2023. Journals were listed either on the databases related to Science Citation Index Expanded (n = 25) or the Social Science Citation Index (n = 23), with fewer journals indexed in the Emerging Sources Citation Index (n = 7) and the Arts and Humanities Citation Index (n = 2). Works covered a wide range of fields, including health and medicine (services, policy, medical informatics, medical ethics, public and environmental health); education; business, management and public administration; computer science; information sciences; engineering; robotics; communication; psychology; political science; and transportation. Beyond the general assessment of publics' attitudes toward, preferences for, and expectations and concerns about AI, the publics' views on ethical challenges of AI technologies have been studied mainly concerning healthcare and public services and less frequently regarding autonomous vehicles (AV), education, robotic technologies, and smart homes. Most of the studies (n = 47) were funded by research agencies, with 7 papers reporting conflicts of interest.

Quantitative research approaches have assessed the publics' views on the ethical challenges of AI mainly through online or web-based surveys and experimental platforms, relying on Delphi studies, moral judgment studies, hypothetical vignettes, and choice-based/comparative conjoint surveys. The 25 qualitative studies collected data mainly by semistructured or in-depth interviews. Analysis of publicly available material reporting on AI-use cases, focus groups, a post hoc self-assessment, World Café, participatory research, and practice-based design research were used once or twice. Multi or mixed-methods studies relied on surveys with open-ended and closed questions, frequently combined with focus groups, in-depth interviews, literature reviews, expert opinions, examinations of relevant curriculum examples, tests, and reflexive writings.

The studies were performed (where stated) in a wide variety of countries, including the USA and Australia. More than half of the studies (n = 38) were conducted in a single country. Almost all studies used nonprobability sampling techniques. In quantitative studies, sample sizes varied from 2.3 M internet users in an online experimental platform study [40] to 20 participants in a Delphi study [41]. In qualitative studies, the samples varied from 123 participants in 21 focus groups [42] to six expert interviews [43]. In multi or mixed-methods studies, samples varied from 2036 participants [44] to 21 participants [45].

3.2 Motivations for engaging the publics

The qualitative synthesis of the motivations for researching the publics' views on the ethical challenges of AI is presented in Table 2 and ordered by the number of studies referencing them in the scoping review. More than half of the studies (n = 37) addressed a single motivation. Innovation (n = 33) and legitimation (n = 29) proved to have the highest relevance as motivations for engaging the publics in the ethical challenges of AI technologies, as articulated in 15 studies. Additional motivations are rooted in education (n = 13), democratization (n = 11), and inspiration (n = 9).

Table 1 Main characteristics of the empirical studies exploring the publics' views on ethical challenges of AI (n = 64)

Publication	Country	Aim	Sample	Assessment of public views
Quantitative studies (n = 27)				
Awad et al. 2018 [40]	233 countries, dependencies, or territories	To quantify societal expectations about the ethical principles that should guide machine behaviour, and how they varied between individuals and countries	2.3 M internet users who chose to visit the website and contribute to the data	Online experimental platform
Kaur and Rampersad 2018 [62]	Australia	To investigate the key factors influencing the perceptions of benefits, concerns, trust and adoption of driverless cars	101 Flinders University's staff and students based at Tonsley	Online survey
Kallioinen et al. 2019 [59]	24 countries (Germany, Armenia, Australia, Russia, and other)	To further understand the ethical issues of introducing self-driving cars	Virtual reality experiment: 184 participants. Simplified animation-based experiment: 368 participants	Moral judgement studies
Liyanage et al. 2019 [41]	Australia, Belgium, Canada, Croatia, Ireland, Italy, New Zealand, Spain, UK, and USA	To seek consensus on the perceptions, issues, and challenges of AI in primary care	Health informatics experts and clinicians involved in primary health care (20 in round 1, 12 in round 2, 13 in round 3)	Delphi study (two online surveys + an online panel discussion)
Awad et al. 2020 [88]	USA	To indicate how blame should be allocated when considering hypothetical cases in which a pedestrian was killed by a car operated under shared control of a primary and a secondary driver	Study 1: 786 participants. Study 2: 779 participants. Study 3: 973 participants. Study 4: 375 participants. Study 5 (representative sample): 2000 participants	Hypothetical vignettes that describe a crash
Esmaeilzadeh 2020 [52]	USA	To examine the perceived benefits and risks of AI medical devices with clinical decision support (CDS) features from consumers' perspectives	427 participants	Online survey
Chen and Wen 2021 [69]	Taiwan	To explore how Taiwanese people's perceptions of AI are affected by their institutional trust, attitudes toward the government and corporations	1009 participants	Online survey
Nichol et al. 2021 [87]	African countries (Ethiopia, Ghana, Kenya, Nigeria, Rwanda, South Africa, Uganda, Zambia, Zimbabwe), and USA	To understand experts' views about the ethical implications of ongoing research funded by the National Institutes of Health that uses machine learning to predict HIV/AIDS risk in sub-Saharan Africa based on publicly available Demographic and Health Surveys data, and to inform an ethical framework and recommendations for researchers	22 experts in informatics, African public health and HIV/AIDS and bioethics	Modified Delphi (three online scenario-based semi-structured surveys with closed and open-ended questions)

Table 1 (continued)

Publication	Country	Aim	Sample	Assessment of public views
Ploug et al. 2021 [73]	Denmark	To elicite the public's preferences for the performance and explainability of AI decision making in health care and determined whether these preferences depend on respondent characteristics, including trust in health and technology and fears and hopes regarding AI	1027 participants: representative sample of the adult Danish population	Choice-based conjoint survey
Tosoni et al. 2021 [95]	Canada	To acquire an in-depth understanding of the contemporary and specific consent needs of cancer patients at a large academic hospital to inform its own institutional consent policies	222 participants	Survey
Zheng et al. 2021 [74]	Various cities and countries, mainly in Zhejiang Province	To assess medical workers' and other professional technicians' familiarity with, attitudes toward, and concerns about AI in ophthalmology	562 health care workers or medical students, mainly members of the Zhejiang Society of Mathematical Medicine	Electronic questionnaire
Criado and de Zarate-Alcarazo 2022 [56]	Spain	To understand CIOs' (Chief Information Officers) interpretation of AI in the public sector	73 participants	Survey, with closed and open questions
De Graaf et al. 2022 [91]	USA, EU (e.g., Italy, Spain, Germany), and Asia (e.g. China, South Corea, Singapore)	To investigate layman's attitudes toward granting particular rights to robots, and the reasons for their willingness to grant them those rights	439 participants (USA = 200, EU = 97, Asia = 142)	Online survey
Ehret 2022 [51]	Germany, UK, India, Chile, and China	To what extent economic and especially the labor market consequences affect preferences for public policies governing AI: when the public demands imposing restrictions on (or even prohibiting) emerging AI technologies; to assess how the balance between normative and economic concerns varies across countries	932 participants	Comparative conjoint survey experiment

Table 1 (continued)

Publication	Country	Aim	Sample	Assessment of public views
Gray et al. 2022 [97]	Australia	To explore health care education experts' ideas and plans for preparing the health workforce to work with AI and identify critical gaps in curriculum and educational resources across a national health care system	39 senior people from different health workforce subgroups	Survey
Guerouaou et al. 2022 [92]	France	To initiate the data-driven study of expressive deep-fakes ethics (specifically here, vocal deep-fakes) and to quantify societal expectations about the principles that should guide their deployment	303 participants	Online questionnaire (24 text vignettes describing potential applications of vocal deep-fakes)
Ikkatai et al. 2022 [81]	Japan	To investigate public attitudes toward AI ethics using four scenarios in Japan	1029 participants	Online questionnaire
Isbanner et al. 2022 [57]	Australia	To understand Australians' normative judgments regarding the use of AI, especially in the fields of health care and social services, and determine the attributes of health care and social service AI systems that Australians consider most important	4448 participants: 1950 in a web-based panel; 2498 from an omnibus survey (for a subset of questions)	Survey
Kieslich et al. 2022 [48]	Germany	To investigate how ethical principles (explainability, fairness, security, accountability, accuracy, privacy, and machine autonomy) are weighted in comparison to each other, given the use of artificial intelligence in tax fraud detection	1099 participants (online access panel representative of the German population above 18 years of age, which at least occasionally uses the Internet)	Online survey (eight compositions of AI systems that varied only in the different fulfillment of the ethical principles)
Kopecky et al. 2022 [100]	Czech Republic	To assess whether the public visibility of the choice of an AV type choice make this choice more altruistic, and which type of situation makes it more difficult to choose altruistically: when choosing for society as a whole, when choosing only for oneself, or when choosing only for one's offspring	2769 participants	Online questionnaire

Table 1 (continued)

Publication	Country	Aim	Sample	Assessment of public views
Lam et al. 2022 [101]	UK, rest of Europe, North America, and South America	To define the term digital surgery and the ethical issues surrounding its clinical application, and to identify barriers and research goals for future practice	Round 1: 52 international experts + 20 members of the public. Round 2: 44 participants. Round 3: 24 surgeons, 8 academics, 3 from healthcare industry, 3 from healthcare policy/digital law/ethics	Delphi exercise
Ma et al. 2022 [75]	USA	To provide relevant government and industry bodies with a better understanding of the general public's receptiveness towards sex robots for more informed policy making, and to raise awareness of the significant social and ethical implications should sex robots become widely accepted and adopted	497 participants	Survey[a]
Sartori and Bocca 2022 [68]	Italy	To investigate the perceptions and attitudes of the general public towards AI (levels of knowledge, awareness, trust, hopes and fears)	5391 students, professors and other employees affiliated to the University of Bologna	Survey, with closed and open-ended questions
Willems et al. 2022a [63]	Austria	To hypothesize and test the way citizens trade-off the perceived usefulness of AI-driven applications with overall privacy concerns to engage in AI-driven public services	1048 participants: representative sample of Austrian citizens for place of residence, and gender over different age categories	Online vignette experiment
Willems et al. 2022b [49]	Austria and Germany	To explore and test whether robot appearance influences citizens' perceptions of the ethical trade-offs made in robot-supported public services	Study 1: 156 students at the University of Hamburg. Study 2: 1339 participants (representative sample of Austrian citizens for place of residence, and gender over different age categories)	Laboratory experiment involving eye-tracking recording (study 1) + Online vignette experiment (study 2)
Choung et al. 2023 [71]	USA	To investigate the importance assigned by the general population to ethical requirements of AI and the influence of these requirements on trust	525 current and potential consumers of AI products	Online survey
Hartwig et al. 2023 [72]	Japan, and USA	To investigate public attitudes toward AI research ethics in Japan and the US using a set of dilemma scenarios	Japan: 1108 respondents. USA: 1063 respondents	Online survey

Table 1 (continued)

Publication	Country	Aim	Sample	Assessment of public views
Qualitative studies (n = 25)				
Jenkins and Draper 2015 [42]	UK, France, and the Netherlands	To discuss some of the ethical issues relevant to the use of social-robots to care for older people in their homes	123 participants: people between the ages of 62 and 95, and their formal and informal carers	21 focus groups
Cresswell et al. 2018 [65]	USA, Norway, Italy, Switzerland, France, UK, Sweden, the Netherlands, and Australia	To understand the emerging role of robotics in health care and identify existing and likely future challenges to maximize the benefits associated with robotics and related convergent technologies	21 stakeholders involved in conceiving, procuring, developing, and using robotics in a range of national and international health care settings	Semi-structured interviews
Laïï et al. 2020 [53]	France	To obtain both an overview of how French health professionals perceive the arrival of AI in daily practice, the perception of the other actors involved in AI, and what influences their views	40 French stakeholders with diverse backgrounds	Semi-structured interviews
McCradden et al. 2020a [46]	Canada	To understand the perspectives of the general public regarding the use of health data in AI research	41 purposively sampled members of the public	Six focus groups
McCradden et al. 2020b [93]	Canada	To investigate current perspectives on ethical issues surrounding AI in health care	30 participants: 18 adult patients with meningioma, 7 caregivers, 5 healthcare providers	Interviews
Aitken et al. 2021 [70]	UK	To explore the value of engaging with the public as an approach to pursue socially-minded data-intensive innovation in banking	23 participants (students, senior citizens, young professionals, people in community centres)	Five focus groups
Bastian et al. 2021 [61]	The Netherlands and Switzerland	To investigate how media practitioners perceive the impact of ANRs (Algorithmic News Recommenders) on their professional norms and media organizations' missions, and how these norms and missions can be integrated into ANR design	17 media practitioners from two quality newspapers: 7 journalists, 3 data scientists, 5 product managers and owners, 2 user-experience researchers	Semi-structured interviews
Rogers et al. 2021 [94]	Not specified (AI applications are being used in Australia and the USA)	To analyze the ethical implications of specific healthcare AI applications, and to provide a detailed analysis of AI-based systems for clinical decision support	Publicly available materials reporting on the development, evidence-generation and deployment phases of Painchek® and Idx-DR: 8 academic articles, 2 regulatory documents, 3 websites, 2 media report/release	Inductive analysis of AI-use case studies

Table 1 (continued)

Publication	Country	Aim	Sample	Assessment of public views
Valles-Peris et al. 2021 [54]	Spain	To analyze the opinions, values, and concerns of patients concerning AI and robotic systems in healthcare, based on the study of their care experience when hospitalized	13 patients with COVID-19	In-depth semi-structured interviews
Alfrink et al. 2022 [98]	The Netherlands	To investigate how those who design, develop and govern urban AI systems (experts) understand transparency, and how users of those same systems (citizens) experience a transparent AI system	10 documents of 'UI for Smart EV Charging' project + 9 interviews with citizens-users of urban AI systems	Practice-based design research + Participatory action research
Allahabadi et al. 2022 [96]	Italy (and potential other countries not specified)	To evaluate the trustworthiness of an AI system for predicting a multiregional score conveying the degree of lung compromise in COVID-19 patients, experimentally deployed in a public hospital in the time of COVID-19 pandemic	58 experts (deep learning/and medical image recognition, ethics, healthcare ethics, radiology, clinicians, law/data protection/data privacy, social science, policy makers, representatives of patients, coordinators of the assessment)	*Post-hoc* self-assessment: Z-Inspection® process, based on the method of evaluating new technologies according to which ethical issues must be discussed through the elaboration of sociotechnical scenarios
Duke 2022 [64]	Israel	To analyze how aware AI developers are to the risks they are creating with new AI technologies, and what their attitudes are towards such risks	6 Israeli healthcare imaging AI start-ups that are developing clinician assistant tools, and not tools for direct patient use	Content analysis of online material (texts derived from blog and press/media sections)
Hallowell et al. 2022 [55]	Africa (1 participant), Europe (8 participants), Australia (5 participants), and USA (6 participants)	To explore under what conditions we could place trust in medical AI tools, which employ machine learning, and the perceived ethical issues arising from the use of computational phenotyping tools in healthcare	20 stakeholders (clinical geneticists, data scientists, bioinformaticians, industry and patient support group spokespersons) who design and/or work with computational phenotyping algorithms	In-depth interviews
Kuberkar et al. 2022 [58]	India	To explore potential risks and opportunities in adopting AI for citizen services	14 participants (professional experience of more than 10 years + detailed understanding of AI technology + researched or developed or using oneor more smart city services)	Expert interview technique

Table 1 (continued)

Publication	Country	Aim	Sample	Assessment of public views
Lee et al. 2022 [78]	Not specified	To understand how young people underrepresented in STEM make meaning of the role of AI in their lives and society and how their relationship to the technology evolves when they create their own AI-based tools and media. To analyze the curriculum and pedagogy behind three ethics-centered AI learning activities housed within an after-school multimedia production organization	Groups of youth ages 14–24 enrolled in Interactive and from communities underrepresented in STEM: Black, Latin, female and gender non-binary students	Ethnographic and participatory research
Papyshev and Yarime 2022 [103]	Russia	To analyze the institutional dynamics that led the Russian government to develop an unenforceable ethics based regulatory regime for AI and their consequences on facilitating innovation and managing risks	50 participants with actual participation in the policymaking process or participation in expert or consulting groups on AI governance in Russia (policymakers, governmental officials, representatives of AI companies, academics) + policy documents, governmental websites, reports from international organizations	Semi-structured interviews + Document analysis[b]
Street et al. 2022 [80]	Australia	To capture older Australians' knowledge about the possibilities and challenges of smart technologies	84 participants	World Cafés
Tzouganatou 2022 [43]	Finland	To critically investigate the role of AI development in relation to opening up born-digital archives online, by considering privacy and ethics issues. To discuss openness and transparency in AI, allowing possibilities for a socially inclusive, participatory culture through AI, tapping into the potential of human approach in AI development	6 interviewees with GLAM professionals (Galleries, Libraries, Archives and Museums), social innovators, service designers and open knowledge activists	Expert interviews
Amann et al. 2023 [66]	Germany and Switzerland	To explore the views of stroke survivors, family members of stroke survivors, and healthcare professionals specialized in stroke regarding the use of stroke related medical AI	34 participants (stroke survivors, family members of stroke survivors, healthcare professionals specialized in stroke)	Semi-structured interviews

Table 1 (continued)

Publication	Country	Aim	Sample	Assessment of public views
Aquino et al. 2023a [67]	Australia and internationally (not specified)	To examine the issue of deskilling from the perspective of diverse group of professional stakeholders with knowledge and/or experiences in the development, deployment and regulation of healthcare AI	72 participants (clinicians, regulators, developers/data scientists, researchers, entrepreneurs, healthcare administrators, consumer representatives)	Semi-structured interviews
Aquino et al. 2023b [86]	Australia and internationally (not specified)	To capture the range of strategies stakeholders endorse in attempting to mitigate algorithmic bias, and to consider the ethical question of responsibility for algorithmic bias	72 participants with specialist AI expertise and/or professional or clinical expertise (clinicians, regulators, developers/data scientists, researchers, entrepreneurs, healthcare administrators, consumer representatives)	Semi-structured interviews
Dempsey et al. 2023 [77]	USA (North Carolina)	To explore (a) police officer views of law enforcement in the twenty-first century United States, (b) police officer views on artificial intelligence technologies, including self-driving vehicles, and examine (c) their combined societal and ethical implications	20 law enforcement professionals	Semi-structured interviews
Henriksen and Blond 2023 [84]	Scandinavia (Denmark, Sweden, and Norway)	To study how AI is performed as developers enact two predictive systems along with stakeholders in public sector accounting and public sector healthcare: Which individuals are to be put first and centered? Whose abilities and agency are to be augmented? Who exactly is benefitting from the implementation of AI, and how?	Documents (e.g. works of the AI developers and discussions that the AI company had with and about users and potential customers or buyers) + 14 interviewees (e.g. managers, business developers, data scientists, data modelers)	Ethnography conducted in an AI company (participant observation, document analysis, and semi-structured interviews)
Nichol et al. 2023 [85]	USA	To examine developers' perceptions of moral awareness and responsibility	40 health care MLPA (machine learning predictive analytics) developers (data scientists, software engineers, project managers and executive leaders, among others)	Semi-structured interviews

Table 1 (continued)

Publication	Country	Aim	Sample	Assessment of public views
Tlili et al. 2023 [50]	Not specified	To examine the concerns of using chatbots, specifically ChatGPT, in education	19 interviewees (people who have been using ChatGPT in education and posting their experiences through blogs publicly) + experience of 3 educators (use of ChatGPT for a whole week to test similar and different teaching/learning scenarios, and then see the obtained results accordingly)	Instrumental case study: interviews + investigation of user experiences[c]
Other (multi- or mixed-methods studies) (n = 12)				
Bourla et al. 2018 [99]	France	To analyze psychiatrists' perspectives on new technologies by assessing the factors affecting their acceptability of 3 clinical decision support systems: (1) smartphone-based EMA, (2) connected wristband-based digital phenotype, and (3) ML-based prediction magnetic resonance imaging or blood test	515 psychiatrists working in France	Web-based survey, with closed and open-ended questions
Liljamo et al. 2018 [44]	Finland	To reveal whether people are ready for automated vehicles, which user groups are presumable early adopters, and what concerns people have that hinder the adoption of these vehicles	2036 individuals: 18–64 year olds living in Finland	Public survey with closed and open-ended questions
Blease et al. 2019 [89]	UK	To explore general practitioners' opinions about the potential impact of future technology on key tasks in primary care	720 participants	Web-based survey with closed and a single open-ended questions
Wang et al. 2019 [82]	USA	To involve residents of a local continuing care senior housing community in conversations about technologies that might facilitate their continued independent living status. To assess their privacy attitudes and preferences. To identify whether residents would be interested in co-designing technologies moving forward and if so, how to foster next steps	31 retirement community residents between the ages of 67 and 94 years	Two focus groups and a survey

Table 1 (continued)

Publication	Country	Aim	Sample	Assessment of public views
Blease et al. 2020 [90]	22 countries worldwide, representing North America, South America, Europe, and Asia–Pacific	To explore psychiatrists' opinions about the potential impact innovations in artificial intelligence and machine learning on psychiatric practice	791 participants (members of the Sermo.org)	Web-based survey with closed and open-ended questions
Vrščaj et al. 2020 [60]	The Netherlands	To assess adolescents' attitudes regarding the 'car of the future' as presented by car manufacturers	Bachelor students (17–25 years): 221 in the qualitative study + 251 in the quantitative study	Survey, with open-ended and closed questions
Rhim et al. 2020 [76]	Korea and Canada	To conduct a cross-cultural comparison of ethical decision-making processes in AV Moral Dilemma scenarios to design human-centered AV morality by gaining an in-depth understanding of human drivers' perspectives	70 participants: 33 Koreans + 37 Canadians	In-depth interviews + Moral thought experiment (crash cases scenarios and moral dilemma vignettes)
Blease et al. 2021 [47]	Switzerland	To explore postgraduate clinical psychology students' familiarity and formal exposure to topics related to artificial intelligence and machine learning (AI/ML) during their studies, and their opinions on the impact of AI/ML on their job. To investigate whether more education may be required so that trainee clinical psychologists/psychotherapists might ethically harness and advise patients about AI/ML-enabled tools	37 clinical psychology students enrolled in a two-year Masters' program at a Swiss University	Online survey, with open-ended and closed questions
Karaca et al. 2021 [102]	Turkey	To develop a valid and reliable psychometric measurement tool for the assessment of the perceived readiness of medical students on AI technologies and its applications in medicine	94 experts involved either using or developing AI in healthcare Medical students enrolled in two public universities: 544 (Exploratory Factor Analysis) + 321 (Confirmatory Factor Analysis)	Item generation: literature review, expert opinions, and examination of relevant curriculum examples. Validity and reliability: online survey

Table 1 (continued)

Publication	Country	Aim	Sample	Assessment of public views
Zhang et al. 2022 [83]	Not specified	To examine whether and to what extent the Developing AI Literacy (DAILy) curriculum helped middle school students develop three core domains of AI: technical concepts, processes, knowledge and skills; ideas about AI's ethical and societal implications; attitudes toward AI and AI careers	25 middle school students in a summer STEM program (urban students who have a grade of C or lower in science/ technology courses, are from low-income families, and may become the first person to go to college in their family), 19 of them purposefully selected for interviewing after the workshop	Three quantitative instruments investigate students' learning of AI with the DAILy workshops implemented online: AI Concept inventory, Attitudes toward AI survey; and AI Career Futures survey. Students' final presentations + observation notes + semi-structured interviews
Couture et al. 2023 [45]	Canada (and potential other countries not specified)	To explore the perspectives and attitudes of citizens and experts regarding the ethics of AI in population health, the engagement of citizens in AI governance, and the potential of a digital app to foster citizen engagement	21 participants: 11 citizens and 10 experts (AI experts, scholars, policymakers)	Web-based survey, with close ended and open questions
Kong et al. 2023 [79]	China	To design, implement and evaluate an AI literacy programme based on a multi-dimensional conceptual framework, which developed participants' conceptual understanding, literacy, empowerment and ethical awareness	36 university students with diverse academic backgrounds	Tests, surveys, focus group, and reflective writings (self-reflections), pre- and post-course activities

[a]This scoping review does not consider an additional study that examined Google search volume derived from the USA, from 2015 to 2019, related to search of items such as sex toys and prominent adult content websites

[b]Findings from media reports were not considered in this scoping review

[c]The exclusion criteria disallowed an initial study focusing on the analysis of 2330 tweets from 1530 twitter users

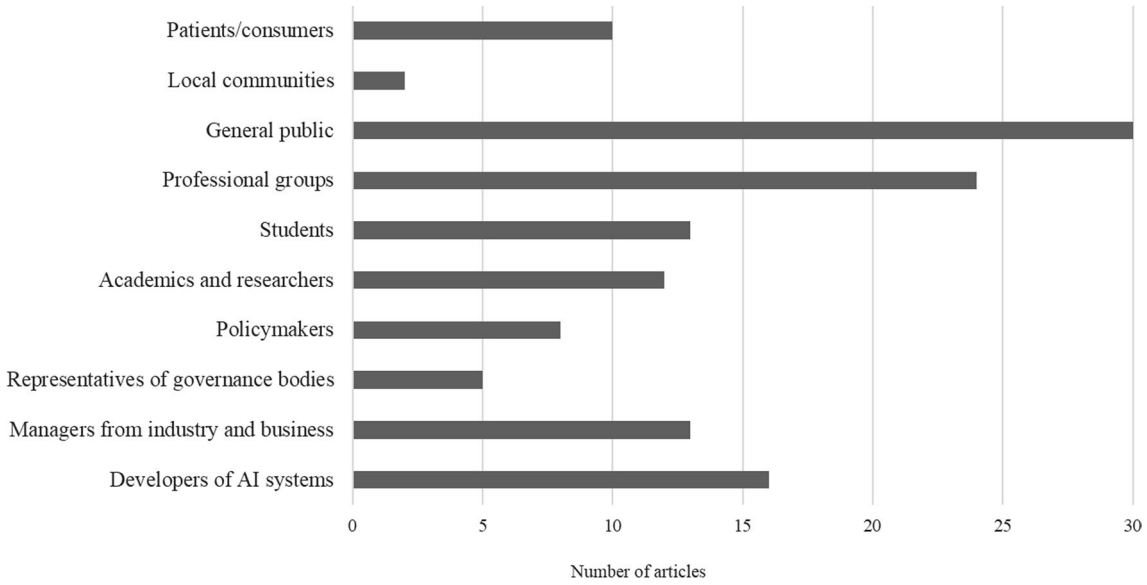

Fig. 2 Publics invited to engage with issues framed as ethical challenges of AI

Politicization was mentioned in five studies. Although they were not authors' motivations, few studies were found to have educational [46, 47], democratization [48, 49], and legitimation or inspirations effects [50].

To consider the publics as a valuable source of knowledge that can add real value to innovation processes in both the private and public sectors was the most frequent motivation mentioned in the literature. The call for public participation is rooted in the aspiration to add knowledge outside "formal" ethics at three interrelated levels. First, at a societal level, by asking what kind of AI we want as a society based on novel experiments on public policy preferences [51] and on the study of public perceptions, values, and concerns regarding AI design, development, and implementation in domains such as health care [46, 52–55], public and social services [49, 56–58], AV [59, 60] and journalism [61]. Second, at a practical level, the literature provides insights into the perceived usefulness of AI applications [62, 63] and choices between boosting developers' voluntary adoption of ethical standards or imposing ethical standards via regulation and oversight [64], as well as suggesting specific guidance for the development and use of AI systems [65–67]. Finally, at a theoretical level, literature expands the social-technical perspective [68] and motivated-reasoning theory [69].

Legitimation was also a frequent motivation for engaging the publics. It was underpinned by the need for public trust in and social licences for implementing AI technologies. To ensure the long-term social acceptability of AI as a trustworthy technology [70, 71] was perceived as essential to support its use and to justify its implementation. In one study [72], the authors developed an AI ethics scale to quantify how AI research is accepted in society and which area of ethical, legal, and social issues (ELSI) people are most concerned with. Public trust in and acceptance of AI is claimed by social institutions such as governments, private sectors, industry bodies, and the science community, behaving in a trustworthy manner, respecting public concerns, aligning with societal values, and involving members of the publics in decision-making and public policy [46, 48, 73–75], as well as in the responsible design and integration of AI technologies [52, 76, 77].

Education, democratization, and inspiration had a more modest presence as motivations to explore the publics' views on the ethical challenges of AI. Considering the emergence of new roles and tasks related to AI, the literature has pointed to the public need to ensure the safe use of AI technologies by incorporating ethics and career futures into the education, preparation, and training of both middle school and university students and the current and future health workforce. Improvements in education and guidance for developers and older adults were also noticed. The views of the publics on what needs to be learned or how this learning may be supported or assessed were perceived as crucial. In one study [78], the authors developed strategies that promote learning related to AI through collaborative media production, connecting computational thinking to civic issues and creative expression. In another study [79], real-world scenarios were successfully used as a novel approach to teaching AI ethics. Rhim et al. [76] provided AV moral behavior design guidelines for policymakers, developers, and the publics by reducing the abstractness of AV morality.

Studies motivated by democratization promoted broader public participation in AI, aiming to empower citizens both

Table 2 Motivations for researching publics' views on ethical challenges of AI

Innovation (33 articles)[a]: illustrative quotes
Public engagement presents valuable opportunities to incorporate diverse views and perspectives and to enable critical reflection on organisational practices and/or the direction of innovation. (…) Wider public engagement can play a valuable role in contesting framings and opening up discourses around data ethics and responsible AI to a wider range of perspectives and considerations. (…) This can add real value to innovation processes in both the private and public sector [70]
Our analysis can contribute to underline the need to place the concerns, values, and opinions of citizens at the centre of the development and implementation processes of AI and robotics systems in health care, as the only way to ensure that these technologies seek to respond to individual and collective well-being and good living [54]
We need to gauge social expectations about how autonomous vehicles should solve moral dilemmas [40]
This paper complements existing normative guidelines and ethical frameworks by empirically probing stakeholders' views on ethically relevant issues relating to the use of medical AI in a concrete clinical context, namely, stroke medicine [66]
Understanding the scenarios when users are most willing to adopt driverless cars will assist in early implementation programs among adopting target groups and settings [62]
Legitimation (29 articles)[b]: illustrative quotes
Citizens must support the use of such technology to justify their implementation in public services [63]
Attention to understanding and incorporating the values of older citizens will be important for the acceptance and effectiveness of smart technologies for supporting independent and full lives for older citizens [80]
When people's skepticism of the discipline and integrity of government, private sectors, and science community remains high and unaddressed, then public distrust of AI will not diminish, because the masses would be constantly agitated over potential abuse of it by those social institutions in power [69]
The study has social implications in terms of ensuring that proper guidelines are developed for using AI technology for citizen services, thereby bridging the ever-critical trust gap between citizens and city administration [58]
A top-down design methodology (…) fails to engage users in the design process. This has frequently created significant mismatches between the needs and preferences of the users and the products that are developed to fulfill their needs. (…) These mismatches can hinder meaningful adoption and sustained usage, and risk leaving priority needs of end-users unmet [82]
Education (13 articles)[c]: illustrative quotes
To be ready for new roles and tasks, medical students and physicians will need to understand the (…) ethical and medico-legal issues [102]
The rapid expansion of artificial intelligence (AI) necessitates promoting AI education at the K-12 level. (…) Students must learn three core domains of AI: technical concepts and processes, ethical and societal implications, and career futures in the AI era [83]
At present, there is no obvious consensus among educators about what needs to be learned or how this learning may be supported or assessed [97]
There is a need to increase technology literacy of older adults along with aging literacy of technologists [82]
Regulatory approaches (…) might have limited success without education and guidance for ML developers about the extent of their responsibilities and how to implement them [85]
Democratization (11 articles)[d]: illustrative quotes
We brought together notions of an informed public, rich collaborative discussion in a hospitable space, and collective knowledge made visible. (…) We aimed for diversity but also purposely recruited participants from low socioeconomic backgrounds, since these individuals, in general, have less input into public policy and fewer choices than individuals from high socioeconomic groups [80]
This (…) study (…) aims to broaden participation in AI (…), and to prepare students to investigate and address ethical issues in AI as critical consumers and potential future creators of AI technologies [83]
A user-centered design approach was used to identify older adults' perspectives regarding AAL and AI technologies and gauge interest in participating in a co-design process [82]
Before we allow our cars to make ethical decisions, we need to have a global conversation to express our preferences to the companies that will design moral algorithms, and to the policymakers that will regulate them [40]
We (…) explore the ways in which citizens can be supported to participate in AI governance through a digital app. (…) The goal was to estimate whether a digital app would be able to contribute to four variables: (1) to inform, (2) to raise public awareness, (3) to rigorously measure citizen attitudes on the issues of AI, and (4) to support collective decision-making [45]
Inspiration (9 articles)[e]: illustrative quotes
The process of embedding values into ANRs can be addressed only through close collaboration between all those involved in the decision to implement, use, and design the technology [61]
To guide and inspire future empirical and design research on fostering AI literacy among educated citizens of diverse backgrounds [79]
To promote developers' moral awareness—the appreciation that there is an ethical aspect to the decisions that they make [85]
Politicization (5 articles)[f]: illustrative quotes
To strengthen collective wisdom in promoting STEM engagement, particularly amongst historically marginalized populations [78]
To serve social equity and sustainable development goals [79]
To address (…) the region's [SSA countries] history of human rights abuses… [87]

[a][40, 46, 48, 49, 51–70, 75, 86, 87, 91, 92, 94, 96, 99, 101]

[b][40, 41, 44, 46, 48, 52, 55, 58, 62, 63, 65, 69–77, 80, 82, 87, 88, 93, 95, 96, 100, 101]

[c][50, 76, 78, 79, 82, 83, 85, 89, 90, 93, 97, 99, 102]

[d][40, 43, 45, 78–83, 88, 96]

[e][46, 48, 61, 78, 79, 84, 85, 96, 98]

[f][78, 79, 86, 87, 103]

Table 3 Ethical issues perceived as needing the participation of the publics

Human agency and oversight (55 articles)[a]: illustrative quotes
Attributing the same kind of agency humans have to robots and AI systems is the source of a distorted portrayal of future technical possibilities [68]
Few of these developers, however, described taking action to prevent or mitigate harms, possibly because of lack of knowledge about how to do so, or perceiving lack of agency [85]
While our interviewees saw CP technology as an expert tool (a reliable and competent tool) they did not advocate that these tools should replace trusted human experts [55]
Even though patients intensely value their face-to-face relationships with hospital staff, from their experience of the oversaturation of the healthcare system and the lived experience of lack of care resources, robots in healthcare are accepted as a possible solution [54]
So the technology would have to be sophisticated enough to judge situations properly but people need to have some kind of control over it to say, "This is not what I need." [70]
They emphasized that AI cannot replace human judgment. (…) Four types of citizen involvement were proposed (…): (1) information, (2) consultation, (3) decision-making, and (4) involvement in AI design. (…) In other words, participants expect that citizens play an engaged normative role in AI governance [45]
Community and stakeholder engagement that includes Africans, ideally in relevant countries, were seen as key to minimising risks at several stages of the research process, including data access, protocol oversight and dissemination and implementation of findings. (…) When asked which specific stakeholders would be critically important to include in stakeholder engagement, over half of participants believed African data scientists, African ethicists, representatives from a national Ministry of Health and representatives from African universities were necessary to include. Interestingly, responses were split (roughly in half) on whether African religious leaders and healers, African health workers and African patients and families were critically important to include or not important to include in stakeholder engagement. There were divergent opinions as to the necessity of representatives from local communities as well [87]
Privacy and data governance (43 articles)[b]: illustrative quotes
It was clear that participants were aware of risks relating to data sharing and/or data misuse. The extent to which participants were interested in using new services underpinned by Open Banking depended on how confident they felt that they would remain in control of who had access to their data. (…) There was concern that while current regulations might limit data sharing or data reuse to purposes which were perceived to be acceptable and legitimate, in the future such regulation might change to allow further possibilities which are less acceptable (including potential government access of personal data). Moreover, there was concern that data might be used, or reused inappropriately or shared with third parties [70]
Given the number of data breaches and the extensive use of personal data by corporations, it is not surprising that privacy rose to the top three in perceived importance [71]
The term 'regulation' (…) appears roughly in two types of instances: mundanely, when it comes to mentioning/boasting that the company achieved clearance by a regulatory body; and in order to criticize/complain about the burden of regulation [64]
Panellists agreed that determining appropriate access to data was an important issue and that governance processes at present were overcomplicated and obstructive. (…) Panellists agreed that there are currently no guidelines concerning data ownership, and the international legal requirements concerning data sharing are unclear. (…) Moreover, data sharing across international boundaries is problematic and there is no consensus on data sharing formats. (…) Preservation of privacy and confidentiality is essential, not only to safeguard patient autonomy but also to ensure patient trust [101]
There must be regulations that mandate data protection such as the right to be forgotten, the
right to privacy, and seeking unambiguous consent, which is very important, for data collection. So explicit consent should be obtained from citizens for each AI solution before using their data [58]
Diversity, non-discrimination and fairness (39 articles)[c]: illustrative quotes
There are several ways through which we may introduce bias, which means that decisions might penalize or reward one group over another. (…) These classes could be based on ethnicity, culture, physical disability, or socioeconomic status. (…) It is possible that sample or training data collected for AI solution only considers data from certain classes and does not represent the entire population (…). Also, it should not propagate a wrong message, which negatively impacts the peace and harmony within the society. (…) Using AI algorithms that are developed in some other countries with different social norms could act as a deterrent for citizens who are used to the local social norms and moral values [58]
The participants were concerned about potential biases in the design and the use of AI, which could lead to increased social inequality. For instance, they referred to the risks that AI exacerbates unequal access to health care, contributes to the discrimination of subpopulations, and enhances social inequalities as well as economic equity issues. Some participants underlined the digital divide (disparity in access to technologies) and the issues of fairness and diversity of data. (…) One member of the panel mentioned the necessity of collecting data on more vulnerable populations to make sound decisions regarding population health [45]
Discriminatory bias substantially impacts prohibition preferences in Germany (…) and Chile

Table 3 (continued)

(…), but less so in China (…) or India (…) [51]
Panellists agreed the following important issues affect public trust: (…) the fear of AI reinforcing biases in datasets. (…) They highlighted issues surrounding inequality of power and differing motives between hospitals and commercial companies [101]
Open-ended responses (…) reflected issues of re-identification, stigma, discrimination against individuals, families or geographically defined and/or socially defined groups (…). Another theme that emerged was concern about data on Africans being used by non-African researchers [87]
Societal and environmental well-being (39 articles)[d]: illustrative quotes
I believe that AI will lead to (…) unemployment; loss of control to machines; increased data collection and mass surveillance; more jobs; longer lives; more quality of life; peace and political stability. I believe that AI will cause unintentional harm to humans [73]
Some also considered that overreliance may lead to a loss of expertise and competences in future generations of clinicians and, consequently, dependence [66]
Many of the participants demonstrated cautious and conditional acceptance of smart technologies, while identifying concerns about social isolation (…). The principles of beneficence and non-maleficence obligate those providing care (here, arguably, including technology developers) to support the wellbeing of others (…). Other themes related to concerns about the impact of smart technologies on social norms and how we related to each other in community [80]
Indeed, they made references to the ecological burden and greenhouse gas emissions because of the power demand for overarchiving and electronic medical devices. (…) Another societal issue raised by some respondents is related to the transformation of the health sector. Participants mentioned the risk that the introduction of AI could affect the "access to healthcare and caregivers." Another response highlighted the participation of such technologies in the "commercialization of healthcare." From the perspective of HCPs, respondents noted that "AI encourages the bureaucratization of work" and that the technology will contribute to the "transformation of healthcare professions." [45]
Technical robustness and safety (38 articles)[e]: illustrative quotes
The primary care informatics community needs to be proactive and to guide the ethical and rigorous development of AI applications so that they will be safe and effective [41]
Safety was the main concern for participants in all of the four moral dilemma vignettes [76]
We need to be reassured that AI tools are subject to stringent reliability and quality assurance checks [55]
Panellists agreed that institutions are not equipped and under resourced to perform appropriate cybersecurity [101]
The poorer the performance of the AI, and the less mature the systems, the more harmful this could be to patients. However even for high-performing AI it remained a problem: we tend to believe machines more than we would humans and we tend to follow their advice even if it is wrong, so that's a tendency that humans have, and so even if a model is 99% accurate, how are you going to deal with the 1% of cases where it's not going to perform? [67]
Transparency (35 articles)[f]: illustrative quotes
The process of data collection (…) should be transparent. This will build more trust in the system if citizens know that their representation is properly done in the training dataset [58]
While highly educated, most participants lacked understanding of the granularity of data that can be captured with pervasive sensing technology and the associated analytics used by digital platforms to identify patterns. The mystery of AI, including what it is and how it works, contributed to fears of data loss or being harmed from decisions made from their personal data [82]
Machine learning, an essential part of modern AI, has been criticized widely in the media for its "black box" approach to solutions, which may have contributed to concerns about transparency. Further, transparency is closely associated with accountability because most users are willing to share some data if corporations are more transparent of potential risks and rewards [71]
We find that citizens experience transparency as burdensome; experts hope transparency ensures acceptance, while citizens are mostly indifferent to AI; and with absent means of control, citizens question transparency's relevance. The tensions we identify suggest transparency cannot be reduced to a product feature, but should be seen as a mediator of debate between experts and citizens. (…) We can also see that not only decisions, but motivations for them must be made transparent [98]
Accountability (31 articles)[g]: illustrative quotes
Some participants indicated media as a key mechanism for accountability. Some participants indicated skepticism that institutions and companies could be held accountable [93]
The potential psychological impact of AI is shown (…) when young people begin to question whether they may be the problem or cause for not getting the results they expect or want. (…) Donald makes the critical move to hold accountable the companies and designers who fail to factor the full range of human experiences into the technology they create, developing systems that reflect and reify inequalities [78]
Thus, if AI tools become crucial in medical decisions, physicians stated that they were not prepared (would not agree?) to be held criminally responsible if a medical error was made by an AI tool. (…) Paradoxically, they [healthcare industrial partners] said that the question concerning responsibility in case of injury was not yet relevant. (…) In addition, those in industry were quite clear about their not being ready to be held responsible for their AI tools if such a tool induced harm to a patient because of an unpredictable evolution of the tool due to a "black box" phenomenon. (…) Members of regulatory agencies are beginning to take an interest in the subject but appear to be currently overwhelmed [53]

Table 3 (continued)

It is currently unclear who holds responsibility for data integrity under law. (…) Panellists agreed that there is a lack of regulation concerning litigation and liability, both for failing digital surgical systems and for surgeons who elect to not follow systems such as AI decision support tools. Additionally, if a surgeon were to follow AI decision support, which resulted in a negative outcome, it is unclear how liability would be adjudicated [101]
As with other IT systems, someone must take responsibility for the end-to-end AI process deployed for a specific purpose. This also means setting up clear boundaries for the AI based application from conceptualization to deployment. The data collection process and how the algorithms are selected before moving the system into production deployment should be supervised. A monitoring mechanism is required so that no bias enters, and the models continue to work as intended. Hence, someone should be responsible for AI processes, policies, and protocols. Somebody must be responsible for determining if the output and performance are as per the given framework [58]

[a][40–48, 50, 53–62, 64–68, 70, 71, 73–78, 80–87, 89–101, 103]

[b][41–48, 50–53, 55–58, 60–66, 70–72, 78–82, 85–87, 90, 91, 93–97, 101, 102]

[c][40, 41, 43, 45, 47, 48, 50–52, 55–61, 64–67, 71–73, 78–83, 85–87, 90, 93, 94, 96–98, 101]

[d][41, 42, 44–47, 49–53, 55–57, 59, 60, 62, 65, 66, 68, 70–76, 80, 81, 83, 85, 87, 89, 90, 94, 96, 97, 99, 100]

[e][41, 44, 45, 47, 48, 50–52, 55–58, 60, 62, 64–67, 70, 71, 73, 76, 79–81, 83–90, 93, 94, 96, 99, 101]

[f][41, 43, 45, 46, 48, 51, 53, 55–58, 61, 64–66, 69, 71–73, 78, 79, 81, 82, 84, 87, 92–98, 101–103]

[g][41, 45, 48, 52, 53, 55–58, 60, 61, 65, 66, 68, 71, 76–79, 81, 85–89, 91, 93, 94, 96, 101, 103]

to express their understandings, apprehensions, and concerns about AI [43, 78, 80, 81] and to address ethical issues in AI as critical consumers, (potential future) developers of AI technologies or would-be participants in codesign processes [40, 43, 45, 78, 82, 83]. Understanding the publics' views on the ethical challenges of AI is expected to influence companies and policymakers [40]. In one study [45], the authors explored how a digital app might support citizens' engagement in AI governance by informing them, raising public awareness, measuring publics' attitudes and supporting collective decision-making.

Inspiration revolved around three main motivations: to raise public interest in AI [46, 48]; to guide future empirical and design studies [79]; and to promote developers' moral awareness through close collaboration between all those involved in the implementation, use, and design of AI technologies [46, 61, 78, 84, 85].

Politicization was the less frequent motivation reported in the literature for engaging the publics. Recognizing the need for mitigation of social biases [86], public participation to address historically marginalized populations [78, 87], and promoting social equity [79] were the highlighted motives.

3.3 The invited publics

Study participants were mostly the general public and professional groups, followed by developers of AI systems, managers from industry and business, students, academics and researchers, patients/consumers, and policymakers (Fig. 2). The views of local communities and representatives of governance bodies were rarely assessed.

Representative samples of the general public were used in five papers related to studies conducted in the USA [88], Denmark [73], Germany [48], and Austria [49, 63]. The remaining random or purposive samples from the general public comprised mainly adults and current and potential users of AI products and services, with few studies involving informal caregivers or family members of patients (n = 3), older people (n = 2), and university staff (n = 2).

Samples of professional groups included mainly healthcare professionals (19 out of 24 studies). Educators, law enforcement, media practitioners, and GLAM professionals (galleries, libraries, archives, and museums) were invited once.

3.4 Ethical issues

The ethical issues concerning AI technologies perceived as needing the participation of the publics are depicted in Table 3. They were mapped by measuring the number of studies referencing them in the scoping review. Human agency and oversight (n = 55) was the most frequent ethical aspect that was studied in the literature, followed by those centered on privacy and data governance (n = 43). Diversity, nondiscrimination and fairness (n = 39), societal and environmental well-being (n = 39), technical robustness and safety (n = 38), transparency (n = 35), and accountability (n = 31) were less frequently discussed.

The concerns regarding human agency and oversight were the replacement of human beings by AI technologies and deskilling [47, 55, 67, 74, 75, 89, 90]; the loss of autonomy, critical thinking, and innovative capacities [50, 58, 61, 77, 78, 83, 85, 90]; the erosion of human judgment and oversight [41, 70, 91]; and the potential for (over)dependence on technology and "oversimplified" decisions [90] due to the lack of publics' expertise in judging and controlling AI technologies [68]. Beyond these ethical challenges, the following contributions of AI systems to empower human beings were noted: more fruitful and empathetic social relationships [47, 68, 90]; enhancing human capabilities and

quality of life [68, 70, 74, 83, 92]; improving efficiency and productivity at work [50, 53, 62, 65, 83] by reducing errors [77], relieving the burden of professionals and/or increasing accuracy in decisions [47, 55, 90]; and facilitating and expanding access to safe and fair healthcare [42, 53, 54] through earlier diagnosis, increased screening and monitoring, and personalized prescriptions [47, 90]. To foster human rights, allowing people to make informed decisions, the last say was up to the person themselves [42, 43, 46, 55, 64, 67, 73, 76]. People should determine where and when to use automated functions and which functions to use [44, 54], developing "job sharing" arrangements with machines and humans complementing and enriching each other [56, 65, 90]. The literature highlights the need to build AI systems that are under human control [48, 70] whether to confirm or to correct the AI system's outputs and recommendations [66, 90]. Proper oversight mechanisms were seen as crucial to ensure accuracy and completeness, with divergent views about who should be involved in public participation approaches [86, 87].

Data sharing and/or data misuse were considered the major roadblocks regarding privacy and data governance, with some studies pointing out distrust of participants related to commercial interests in health data [55, 90, 93–95] and concerns regarding risks of information getting into the hands of hackers, banks, employers, insurance companies, or governments [66]. As data are the backbone of AI, secure methods of data storage and protection are understood as needing to be provided from the input to the output data. Recognizing that in contemporary societies, people are aware of the consequences of smartphone use resulting in the minimization of privacy concerns [93], some studies have focused on the impacts of data breaches and loss of privacy and confidentiality [43, 45, 46, 60, 62, 80] in relation to health-sensitive personal data [46, 93], potentially affecting more vulnerable populations, such as senior citizens and mentally ill patients [82, 90] as well as those at young ages [50], and when journalistic organizations collect user data to provide personalized news suggestions [61]. The need to find a balance between widening access to data and ensuring confidentiality and respect for privacy [53] was often expressed in three interrelated terms: first, the ability of data subjects to be fully informed about how data will be used and given the option of providing informed consent [46, 58, 78] and controlling personal information about oneself [57]; second, the need for regulation [52, 65, 87], with one study reporting that AI developers complain about the complexity, slowness, and obstacles created by regulation [64]; and last, the testing and certification of AI-enabled products and services [71]. The study by De Graaf et al. [91] discussed the robots' right to store and process the data they collect, while Jenkins and Draper [42] explored less intrusive ways in which the robot could use information to report back to carers about the patient's adherence to healthcare.

Studies discussing diversity, nondiscrimination, and fairness have pointed to the development of AI systems that reflect and reify social inequalities [45, 78] through non-representative datasets [55, 58, 96, 97] and algorithmic bias [41, 45, 85, 98] that might benefit some more than others. This could have multiple negative consequences for different groups based on ethnicity, disease, physical disability, age, gender, culture, or socioeconomic status [43, 55, 58, 78, 82, 87], from the dissemination of hate speech [79] to the exacerbation of discrimination, which negatively impacts peace and harmony within society [58]. As there were cross-country differences and issue variations in the publics' views of discriminatory bias [51, 72, 73], fostering diversity, inclusiveness, and cultural plurality [61] was perceived as crucial to ensure the transferability/effectiveness of AI systems in all social groups [60, 94]. Diversity, nondiscrimination, and fairness were also discussed as a means to help reduce health inequalities [41, 67, 90], to compensate for human preconceptions about certain individuals [66], and to promote equitable distribution of benefits and burdens [57, 71, 80, 93], namely, supporting access by all to the same updated and high-quality AI systems [50]. In one study [83], students provided constructive solutions to build an unbiased AI system, such as using a dataset that includes a diverse dataset engaging people of different ages, genders, ethnicities, and cultures. In another study [86], participants recommended diverse approaches to mitigate algorithmic bias, from open disclosure of limitations to consumer and patient engagement, representation of marginalized groups, incorporation of equity considerations into sampling methods and legal recourse, and identification of a wide range of stakeholders who may be responsible for addressing AI bias: developers, healthcare workers, manufacturers and vendors, policymakers and regulators, AI researchers and consumers.

Impacts on employment and social relationships were considered two major ethical challenges regarding societal and environmental well-being. The literature has discussed tensions between job creation [51] and job displacement [42, 90], efficiency [90], and deskilling [57]. The concerns regarding future social relationships were the loss of empathy, humanity, and/or sensitivity [52, 66, 90, 99]; isolation and fewer social connections [42, 47, 90]; laziness [50, 83]; anxious counterreactions [83, 99]; communication problems [90]; technology dependence [60]; plagiarism and cheating in education [50]; and becoming too emotionally attached to a robot [65]. To overcome social unawareness [56] and lack of acceptance [65] due to financial costs [56, 90], ecological burden [45], fear of the unknown [65, 83] and/or moral issues [44, 59, 100], AI systems need to provide public benefit sharing [55], consider discrepancies between public discourse about AI and the utility of the tools in real-world

settings and practices [53], conform to the best standards of sustainability and address climate change and environmental justice [60, 71]. Successful strategies in promoting the acceptability of robots across contexts included an approachable and friendly looking as possible, but not too human-like [49, 65], and working with, rather than in competition, with humans [42].

The publics were invited to participate in the following ethical issues related to technical robustness and safety: usability, reliability, liability, and quality assurance checks of AI tools [44, 45, 55, 62, 99]; validity of big data analytic tools [87]; the degree to which an AI system can perform tasks without errors or mistakes [50, 57, 66, 84, 90, 93]; and needed resources to perform appropriate (cyber)security [62, 101]. Other studies approached the need to consider both material and normative concerns of AI applications [51], namely, assuring that AI systems are developed responsibly with proper consideration of risks [71] and sufficient proof of benefits [96]. One study [64] highlighted that AI developers tend to be reluctant to recognize safety issues, bias, errors, and failures, and when they do so, they do so in a selective manner and in their terms by adopting positive-sounding professional jargon as AI robustness.

Some studies recognized the need for greater transparency that reduces the mystery and opaqueness of AI systems [71, 82, 101] and opens its "black box" [64, 71, 98]. Clear insights about "what AI is/is not" and "how AI technology works" (definition, applications, implications, consequences, risks, limitations, weaknesses, threats, rewards, strengths, opportunities) were considered as needed to debunk the myth about AI as an independent entity [53] and for providing sufficient information and understandable explanations of "what's happening" to society and individuals [43, 48, 72, 73, 78, 102]. Other studies considered that people, when using AI tools, should be made fully aware that these AI devices are capturing and using their data [46] and how data are collected [58] and used [41, 46, 93]. Other transparency issues reported in the literature included the need for more information about the composition of data training sets [55], how algorithms work [51, 55, 84, 94, 97], how AI makes a decision [57] and the motivations for that decision [98]. Transparency requirements were also addressed as needing the involvement of multiple stakeholders: one study reported that transparency requirements should be seen as a mediator of debate between experts, citizens, communities, and stakeholders [87] and cannot be reduced to a product feature, avoiding experiences where people feel overwhelmed by explanations [98] or "too much information" [66].

Accountability was perceived by the publics as an important ethical issue [48], while developers expressed mixed attitudes, from moral disengagement to a sense of responsibility and moral conflict and uncertainty [85]. The literature has revealed public skepticism regarding accountability mechanisms [93] and criticism about the shift of responsibility away from tech industries that develop and own AI technologies [53, 68], as it opens space for users to assume their own individual responsibility [78]. This was the case in studies that explored accountability concerns regarding the assignment of fault and responsibility for car accidents using self-driving technology [60, 76, 77, 88]. Other studies considered that more attention is needed to scrutinize each application across the AI life cycle [41, 71, 94], to explainability of AI algorithms that provide to the publics the cause of AI outcomes [58], and to regulations that assign clear responsibility concerning litigation and liability [52, 89, 101, 103].

4 Discussion

Within the realm of research studies encompassed in the scoping review, the contemporary impetus for engaging the publics in ethical considerations related to AI predominantly revolves around two key motivations: innovation and legitimation. This might be explained by the current emphasis on responsible innovation, which values the publics' participation in knowledge and innovation-making [29] within a prioritization of the instrumental role of science for innovation and economic return [33]. Considering the publics as a valuable source of knowledge that should be called upon to contribute to knowledge innovation production is underpinned by the desire for legitimacy, specifically centered around securing the publics' endorsement of scientific and technological advancements [33, 104]. Approaching the publics' views on the ethical challenges of AI can also be used as a form of risk prevention to reduce conflict and close vital debates in contention areas [5, 34, 105].

A second aspect that stood out in this finding is a shift in the motivations frequently reported as central for engaging the publics with AI technologies. Previous studies analysing AI national policies and international guidelines addressing AI governance [3–5] and a study analysing science communication journals [33] highlighted education, inspiration and democratization as the most prominent motivations. Our scoping review did not yield similar findings, which might signal a departure, in science policy related to public participation, from the past emphasis on education associated with the deficit model of public understanding of science and democratization of the model of public engagement with science [106, 107].

The underlying motives for the publics' engagement raise the question of the kinds of publics it addresses, i.e., who are the publics that are supposed to be recruited as research participants [32]. Our findings show a prevalence of the general public followed by professional groups and developers of AI systems. The wider presence of the general public indicates

not only what Hagendijk and Irwin [32, p. 167] describe as a fashionable tendency in policy circles since the late 1990s, and especially in Europe, focused on engaging 'the public' in scientific and technological change but also the avoidance of the issues of democratic representation [12, 18]. Additionally, the unspecificity of the "public" does not stipulate any particular action [24] that allows for securing legitimacy for and protecting the interests of a wide range of stakeholders [19, 108] while bringing the risk of silencing the voices of the very publics with whom engagement is sought [33]. The focus on approaching the publics' views on the ethical challenges of AI through the general public also demonstrates how seeking to "lay" people's opinions may be driven by a desire to promote public trust and acceptance of AI developments, showing how science negotiates challenges and reinstates its authority [109].

While this strategy is based on nonscientific audiences or individuals who are not associated with any scientific discipline or area of inquiry as part of their professional activities, the converse strategy—i.e., involving professional groups and AI developers—is also noticeable in our findings. This suggests that technocratic expert-dominated approaches coexist with a call for more inclusive multistakeholder approaches [3]. This coexistence is reinforced by the normative principles of the "responsible innovation" framework, in particular the prescription that innovation should include the publics as well as traditionally defined stakeholders [3, 110], whose input has become so commonplace that seeking the input of laypeople on emerging technologies is sometimes described as a "standard procedure" [111, p. 153].

In the body of literature included in the scoping review, human agency and oversight emerged as the predominant ethical dimension under investigation. This finding underscores the pervasive significance attributed to human centricity, which is progressively integrated into public discourses concerning AI, innovation initiatives, and market-driven endeavours [15, 112]. In our perspective, the importance given to human-centric AI is emblematic of the "techno-regulatory imaginary" suggested by Rommetveit and van Dijk [35] in their study about privacy engineering applied in the European Union's General Data Protection Regulation. This term encapsulates the evolving collective vision and conceptualization of the role of technology in regulatory and oversight contexts. At least two aspects stand out in the techno-regulatory imaginary, as they are meant to embed technoscience in societally acceptable ways. First, it reinstates pivotal demarcations between humans and nonhumans while concurrently producing intensified blurring between these two realms. Second, the potential resolutions offered relate to embedding fundamental rights within the structural underpinnings of technological architectures [35].

Following human agency and oversight, the most frequent ethical issue discussed in the studies contained in our scoping review was privacy and data governance. Our findings evidence additional central aspects of the "techno-regulatory imaginary" in the sense that instead of the traditional regulatory sites, modes of protecting privacy and data are increasingly located within more privatized and business-oriented institutions [6, 35] and crafted according to a human-centric view of rights. The focus on secure ways of data storage and protection as in need to be provided from the input to the output data, the testing and certification of AI-enabled products and services, the risks of data breaches, and calls for finding a balance between widening access to data and ensuring confidentiality and respect for privacy, exhibited by many studies in this scoping review, portray an increasing framing of privacy and data protection within technological and standardization sites. This tendency shows how forms of expertise for privacy and data protection are shifting away from traditional regulatory and legal professionals towards privacy engineers and risk assessors in information security and software development. Another salient element to highlight pertains to the distribution of responsibility for privacy and data governance [6, 113] within the realm of AI development through engagement with external stakeholders, including users, governmental bodies, and regulatory authorities. It extends from an emphasis on issues derived from data sharing and data misuse to facilitating individuals to exercise control over their data and privacy preferences and to advocating for regulatory frameworks that do not impede the pace of innovation. This distribution of responsibility shared among the contributions and expectations of different actors is usually convoked when the operationalization of ethics principles conflicts with AI deployment [6]. In this sense, privacy and data governance are reconstituted as a "normative transversal" [113, p. 20], both of which work to stabilize or close controversies, while their operationalization does not modify any underlying operations in AI development.

Diversity, nondiscrimination and fairness, societal and environmental well-being, technical robustness and safety, transparency, and accountability were the ethical issues less frequently discussed in the studies included in this scoping review. In contrast, ethical issues of technical robustness and safety, transparency, and accountability "are those for which technical fixes can be or have already been developed" and "implemented in terms of technical solutions" [12, p. 103]. The recognition of issues related to technical robustness and safety expresses explicit admissions of expert ignorance, error, or lack of control, which opens space for politics of "optimization of algorithms" [114, p. 17] while reinforcing "strategic ignorance" [114, p. 89]. In the words of the sociologist Linsey McGoey, strategic ignorance refers to "any actions which mobilize, manufacture or exploit unknowns

in a wider environment to avoid liability for earlier actions" [115, p. 3].

According to the analysis of Jobin et al. [11] of the global landscape of existing ethics guidelines for AI, transparency comprising efforts to increase explainability, interpretability, or other acts of communication and disclosure is the most prevalent principle in the current literature. Transparency gains high relevance in ethics guidelines because this principle has become a pro-ethical condition "enabling or impairing other ethical practices or principles" [Turilli and Floridi 2009, [11], p. 14]. Our findings highlight transparency as a crucial ethical concern for explainability and disclosure. However, as emphasized by Ananny and Crawford [116, p. 973], there are serious limitations to the transparency ideal in making black boxes visible (i.e., disclosing and explaining algorithms), since "being able to see a system is sometimes equated with being able to know how it works and governs it—a pattern that recurs in recent work about transparency and computational systems". The emphasis on transparency mirrors Aradau and Blanke's [114] observation that Big Tech firms are creating their version of transparency. They are prompting discussions about their data usage, whether it is for "explaining algorithms" or addressing bias and discrimination openly.

The framing of ethical issues related to accountability, as elucidated by the studies within this scoping review, manifests as a commitment to ethical conduct and the transparent allocation of responsibility and legal obligations in instances where the publics encounters algorithmic deficiencies, glitches, or other imperfections. Within this framework, accountability becomes intricately intertwined with the notion of distributed responsibility, as expounded upon in our examination of how the literature addresses challenges in privacy and data governance. Simultaneously, it converges with our discussion on optimizing algorithms concerning ethical concerns on technical robustness and safety by which AI systems are portrayed as fallible yet eternally evolving towards optimization. As astutely observed by Aradau and Blanke [114, p. 171], "forms of accountability through error enact algorithmic systems as fallible but ultimately correctable and therefore always desirable. Errors become temporary malfunctions, while the future of algorithms is that of indefinite optimization".

5 Conclusion

This scoping review of how publics' views on ethical challenges of AI are framed, articulated, and concretely operationalized in the research sector shows that ethical issues and publics formation are closely entangled with symbolic and social orders, including political and economic agendas and visions. While Steinhoff [6] highlights the subordinated nature of AI ethics within an innovation network, drawing on insights from diverse sources beyond Big Tech, we assert that this network is dynamically evolving towards greater hybridity and boundary fusion. In this regard, we extend Steinhoff's argument by emphasizing the imperative for a more nuanced understanding of how this network operates within diverse contexts. Specifically, within the research sector, it operates through a convergence of boundaries, engaging human and nonhuman entities and various disciplines and stakeholders. Concurrently, the advocacy for diversity and inclusivity, along with the acknowledgement of errors and flaws, serves to bolster technical expertise and reaffirm the establishment of order and legitimacy in alignment with the institutional norms underpinning responsible research practices.

Our analysis underscores the growing importance of involving the publics in AI knowledge creation and innovation, both to secure public endorsement and as a tool for risk prevention and conflict mitigation. We observe two distinct approaches: one engaging nonscientific audiences and the other involving professional groups and AI developers, emphasizing the need for inclusivity while safeguarding expert knowledge. Human-centred approaches are gaining prominence, emphasizing the distinction and blending of human and nonhuman entities and embedding fundamental rights in technological systems. Privacy and data governance emerge as the second most prevalent ethical concern, shifting expertise away from traditional regulatory experts to privacy engineers and risk assessors. The distribution of responsibility for privacy and data governance is a recurring theme, especially in cases of ethical conflicts with AI deployment. However, there is a notable imbalance in attention, with less focus on diversity, non-discrimination, fairness, societal, and environmental well-being, compared to human-centric AI, privacy, and data governance being managed through technical fixes. Last, acknowledging technical robustness and safety, transparency, and accountability as foundational ethics principles reveals an openness to expert limitations, allowing room for the politics of algorithm optimization, framing AI systems as correctable and perpetually evolving.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s43681-023-00387-1.

Acknowledgements The authors would like to express their gratitude to Rafaela Granja (CECS, University of Minho) for her insightful support in an early stage of preparation of this manuscript, and to the AIDA research netwrok for the inspiring debates.

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by HM, SS, and LN. The first draft of the manuscript was written by HM and SS. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by FCT|FCCN (b-on). Helena Machado and Susana Silva did not receive funding to assist in the preparation of this work. Laura Neiva received funding from FCT—Fundação para a Ciência e a Tecnologia, I.P., under a PhD Research Studentships (ref.2020.04764.BD), and under the project UIDB/00736/2020 (base funding) and UIDP/00736/2020 (programmatic funding).

Data availability This manuscript has data included as electronic supplementary material. The dataset constructed by the authors, resulting from a search of publications on PubMed® and Web of Science™, analysed in the current study, is not publicly available. But it can be available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., Floridi, L.: Artificial intelligence and the 'good society': the US, EU, and UK approach. Sci. Eng. Ethics **24**, 505–528 (2017). https://doi.org/10.1007/s11948-017-9901-7
2. Cussins, J.N.: Decision points in AI governance. CLTC white paper series. Center for Long-term Cybersecurity. https://cltc.berkeley.edu/publication/decision-points-in-ai-governance/ (2020). Accessed 8 July 2023
3. Ulnicane, I., Okaibedi Eke, D., Knight, W., Ogoh, G., Stahl, B.: Good governance as a response to discontents? Déjà vu, or lessons for AI from other emerging technologies. Interdiscip. Sci. Rev. **46**(1–2), 71–93 (2021). https://doi.org/10.1080/03080188.2020.1840220
4. Ulnicane, I., Knight, W., Leach, T., Stahl, B., Wanjiku, W.: Framing governance for a contested emerging technology: insights from AI policy. Policy Soc. **40**(2), 158–177 (2021). https://doi.org/10.1080/14494035.2020.1855800
5. Wilson, C.: Public engagement and AI: a values analysis of national strategies. Gov. Inf. Q. **39**(1), 101652 (2022). https://doi.org/10.1016/j.giq.2021.101652
6. Steinhoff, J.: AI ethics as subordinated innovation network. AI Soc. (2023). https://doi.org/10.1007/s00146-023-01658-5
7. Organization for Economic Co-operation and Development. Recommendation of the Council on Artificial Intelligence. https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449 (2019). Accessed 8 July 2023
8. United Nations Educational, Scientific and Cultural Organization. Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000381137 (2021). Accessed 28 June 2023
9. European Commission. On artificial intelligence – a European approach to excellence and trust. White paper. COM(2020) 65 final. https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en (2020). Accessed 28 June 2023
10. European Commission. The ethics guidelines for trustworthy AI. Directorate-General for Communications Networks, Content and Technology, EC Publications Office. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (2019). Accessed 10 July 2023
11. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat. Mach. Intell. **1**, 389–399 (2019). https://doi.org/10.1038/s42256-019-0088-2
12. Hagendorff, T.: The ethics of AI ethics: an evaluation of guidelines. Minds Mach. **30**, 99–120 (2020). https://doi.org/10.1007/s11023-020-09517-8
13. Su, A.: The promise and perils of international human rights law for AI governance. Law Technol. Hum. **4**(2), 166–182 (2022). https://doi.org/10.5204/lthj.2332
14. Ulnicane, I.: Emerging technology for economic competitiveness or societal challenges? Framing purpose in artificial intelligence policy. GPPG. **2**, 326–345 (2022). https://doi.org/10.1007/s43508-022-00049-8
15. Sigfrids, A., Leikas, J., Salo-Pöntinen, H., Koskimies, E.: Human-centricity in AI governance: a systemic approach. Front Artif. Intell. **6**, 976887 (2023). https://doi.org/10.3389/frai.2023.976887
16. Benkler, Y.: Don't let industry write the rules for AI. Nature **569**(7755), 161 (2019). https://doi.org/10.1038/d41586-019-01413-1
17. Phan, T., Goldenfein, J., Mann, M., Kuch, D.: Economies of virtue: the circulation of 'ethics' in Big Tech. Sci. Cult. **31**(1), 121–135 (2022). https://doi.org/10.1080/09505431.2021.1990875
18. Ochigame, R.: The invention of "ethical AI": how big tech manipulates academia to avoid regulation. Intercept. https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/ (2019). Accessed 10 July 2023
19. Ferretti, T.: An institutionalist approach to ai ethics: justifying the priority of government regulation over self-regulation. MOPP **9**(2), 239–265 (2022). https://doi.org/10.1515/mopp-2020-0056
20. van Maanen, G.: AI ethics, ethics washing, and the need to politicize data ethics. DISO **1**(9), 1–23 (2022). https://doi.org/10.1007/s44206-022-00013-3
21. Gerdes, A.: The tech industry hijacking of the AI ethics research agenda and why we should reclaim it. Discov. Artif. Intell. **2**(25), 1–8 (2022). https://doi.org/10.1007/s44163-022-00043-3
22. Amariles, D.R., Baquero, P.M.: Promises and limits of law for a human-centric artificial intelligence. Comput. Law Secur. Rev. **48**(105795), 1–10 (2023). https://doi.org/10.1016/j.clsr.2023.105795
23. Mittelstadt, B.: Principles alone cannot guarantee ethical AI. Nat. Mach. Intell. **1**(11), 501–507 (2019). https://doi.org/10.1038/s42256-019-0114-4
24. Munn, L.: The uselessness of AI ethics. AI Ethics **3**, 869–877 (2022). https://doi.org/10.1007/s43681-022-00209-w
25. Heilinger, J.C.: The ethics of AI ethics. A constructive critique. Philos. Technol. **35**(61), 1–20 (2022). https://doi.org/10.1007/s13347-022-00557-9

26 Roche, C., Wall, P.J., Lewis, D.: Ethics and diversity in artificial intelligence policies, strategies and initiatives. AI Ethics (2022). https://doi.org/10.1007/s43681-022-00218-9

27. Diercks, G., Larsen, H., Steward, F.: Transformative innovation policy: addressing variety in an emerging policy paradigm. Res. Policy **48**(4), 880–894 (2019). https://doi.org/10.1016/j.respol.2018.10.028

28. Owen, R., Pansera, M.: Responsible innovation and responsible research and innovation. In: Dagmar, S., Kuhlmann, S., Stamm, J., Canzler, W. (eds.) Handbook on Science and Public Policy, pp. 26–48. Edward Elgar, Cheltenham (2019)

29. Macq, H., Tancoigne, E., Strasser, B.J.: From deliberation to production: public participation in science and technology policies of the European Commission (1998–2019). Minerva **58**(4), 489–512 (2020). https://doi.org/10.1007/s11024-020-09405-6

30. Cath, C.: Governing artificial intelligence: ethical, legal and technical opportunities and challenges. Philos. Trans. Royal Soc. A. **376**, 20180080 (2018). https://doi.org/10.1098/rsta.2018.0080

31. Wilson, C.: The socialization of civic participation norms in government?: Assessing the effect of the Open Government Partnership on countries'e-participation. Gov. Inf. Q. **37**(4), 101476 (2020). https://doi.org/10.1016/j.giq.2020.101476

32. Hagendijk, R., Irwin, A.: Public deliberation and governance: engaging with science and technology in contemporary Europe. Minerva **44**(2), 167–184 (2006). https://doi.org/10.1007/s11024-006-0012-x

33. Weingart, P., Joubert, M., Connoway, K.: Public engagement with science - origins, motives and impact in academic literature and science policy. PLoS One **16**(7), e0254201 (2021). https://doi.org/10.1371/journal.pone.0254201

34. Wynne, B.: Public participation in science and technology: performing and obscuring a political–conceptual category mistake. East Asian Sci. **1**(1), 99–110 (2007). https://doi.org/10.1215/s12280-007-9004-7

35. Rommetveit, K., Van Dijk, N.: Privacy engineering and the techno-regulatory imaginary. Soc. Stud. Sci. **52**(6), 853–877 (2022). https://doi.org/10.1177/03063127221119424

36. Levac, D., Colquhoun, H., O'Brien, K.: Scoping studies: advancing the methodology. Implement. Sci. **5**(69), 1–9 (2010). https://doi.org/10.1186/1748-5908-5-69

37. Arksey, H., O'Malley, L.: Scoping studies: towards a methodological framework. Int. J. Soc. Res. Methodol. **8**(1), 19–32 (2005). https://doi.org/10.1080/1364557032000119616

38. Stemler, S.: An overview of content analysis. Pract. Asses. Res. Eval. **7**(17), 1–9 (2001). https://doi.org/10.7275/z6fm-2e34

39. European Commission. European Commission's ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (2021). Accessed 8 July 2023

40. Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., et al.: The moral machine experiment. Nature **563**(7729), 59–64 (2018). https://doi.org/10.1038/s41586-018-0637-6

41. Liyanage, H., Liaw, S.T., Jonnagaddala, J., Schreiber, R., Kuziemsky, C., Terry, A.L., de Lusignan, S.: Artificial intelligence in primary health care: perceptions, issues, and challenges. Yearb. Med. Inform. **28**(1), 41–46 (2019). https://doi.org/10.1055/s-0039-1677901

42. Jenkins, S., Draper, H.: Care, monitoring, and companionship: views on care robots from older people and their carers. Int. J. Soc. Robot. **7**(5), 673–683 (2015). https://doi.org/10.1007/s12369-015-0322-y

43. Tzouganatou, A.: Openness and privacy in born-digital archives: reflecting the role of AI development. AI Soc. **37**(3), 991–999 (2022). https://doi.org/10.1007/s00146-021-01361-3

44. Liljamo, T., Liimatainen, H., Pollanen, M.: Attitudes and concerns on automated vehicles. Transp. Res. Part F Traffic Psychol. Behav. **59**, 24–44 (2018). https://doi.org/10.1016/j.trf.2018.08.010

45. Couture, V., Roy, M.C., Dez, E., Laperle, S., Belisle-Pipon, J.C.: Ethical implications of artificial intelligence in population health and the public's role in its governance: perspectives from a citizen and expert panel. J. Med. Internet Res. **25**, e44357 (2023). https://doi.org/10.2196/44357

46. McCradden, M.D., Sarker, T., Paprica, P.A.: Conditionally positive: a qualitative study of public perceptions about using health data for artificial intelligence research. BMJ Open **10**(10), e039798 (2020). https://doi.org/10.1136/bmjopen-2020-039798

47. Blease, C., Kharko, A., Annoni, M., Gaab, J., Locher, C.: Machine learning in clinical psychology and psychotherapy education: a mixed methods pilot survey of postgraduate students at a Swiss University. Front. Public Health **9**(623088), 1–8 (2021). https://doi.org/10.3389/fpubh.2021.623088

48 Kieslich, K., Keller, B., Starke, C.: Artificial intelligence ethics by design. Evaluating public perception on the importance of ethical design principles of artificial intelligence. Big Data Soc. **9**(1), 1–15 (2022). https://doi.org/10.1177/20539517221092956

49. Willems, J., Schmidthuber, L., Vogel, D., Ebinger, F., Vanderelst, D.: Ethics of robotized public services: the role of robot design and its actions. Gov. Inf. Q. **39**(101683), 1–11 (2022). https://doi.org/10.1016/J.Giq.2022.101683

50. Tlili, A., Shehata, B., Adarkwah, M.A., Bozkurt, A., Hickey, D.T., Huang, R.H., Agyemang, B.: What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learn Environ. **10**(15), 1–24 (2023). https://doi.org/10.1186/S40561-023-00237-X

51. Ehret, S.: Public preferences for governing AI technology: comparative evidence. J. Eur. Public Policy **29**(11), 1779–1798 (2022). https://doi.org/10.1080/13501763.2022.2094988

52. Esmaeilzadeh, P.: Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives. BMC Med. Inform. Decis. Mak. **20**(170), 1–19 (2020). https://doi.org/10.1186/s12911-020-01191-1

53. Laïï, M.C., Brian, M., Mamzer, M.F.: Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. J. Transl. Med. **18**(14), 1–13 (2020). https://doi.org/10.1186/S12967-019-02204-Y

54. Valles-Peris, N., Barat-Auleda, O., Domenech, M.: Robots in healthcare? What patients say. Int. J. Environ. Res. Public Health **18**(9933), 1–18 (2021). https://doi.org/10.3390/ijerph18189933

55. Hallowell, N., Badger, S., Sauerbrei, A., Nellaker, C., Kerasidou, A.: "I don't think people are ready to trust these algorithms at face value": trust and the use of machine learning algorithms in the diagnosis of rare disease. BMC Med. Ethics **23**(112), 1–14 (2022). https://doi.org/10.1186/s12910-022-00842-4

56. Criado, J.I., de Zarate-Alcarazo, L.O.: Technological frames, CIOs, and artificial intelligence in public administration: a socio-cognitive exploratory study in spanish local governments. Gov. Inf. Q. **39**(3), 1–13 (2022). https://doi.org/10.1016/J.Giq.2022.101688

57. Isbanner, S., O'Shaughnessy, P.: The adoption of artificial intelligence in health care and social services in Australia: findings from a methodologically innovative national survey of values and attitudes (the AVA-AI Study). J. Med. Internet Res. **24**(8), e37611 (2022). https://doi.org/10.2196/37611

58. Kuberkar, S., Singhal, T.K., Singh, S.: Fate of AI for smart city services in India: a qualitative study. Int. J. Electron. Gov. Res. **18**(2), 1–21 (2022). https://doi.org/10.4018/Ijegr.298216

59. Kallioinen, N., Pershina, M., Zeiser, J., Nezami, F., Pipa, G., Stephan, A., Konig, P.: Moral judgements on the actions of self-driving cars and human drivers in dilemma situations from

different perspectives. Front. Psychol. **10**(2415), 1–15 (2019). https://doi.org/10.3389/fpsyg.2019.02415
60. Vrščaj, D., Nyholm, S., Verbong, G.P.J.: Is tomorrow's car appealing today? Ethical issues and user attitudes beyond automation. AI Soc. **35**(4), 1033–1046 (2020). https://doi.org/10.1007/s00146-020-00941-z
61. Bastian, M., Helberger, N., Makhortykh, M.: Safeguarding the journalistic DNA: attitudes towards the role of professional values in algorithmic news recommender designs. Digit. Journal. **9**(6), 835–863 (2021). https://doi.org/10.1080/21670811.2021.1912622
62. Kaur, K., Rampersad, G.: Trust in driverless cars: investigating key factors influencing the adoption of driverless cars. J. Eng. Technol. Manag. **48**, 87–96 (2018). https://doi.org/10.1016/j.jengtecman.2018.04.006
63 Willems, J., Schmid, M.J., Vanderelst, D., Vogel, D., Ebinger, F.: AI-driven public services and the privacy paradox: do citizens really care about their privacy? Public Manag. Rev. (2022). https://doi.org/10.1080/14719037.2022.2063934
64. Duke, S.A.: Deny, dismiss and downplay: developers' attitudes towards risk and their role in risk creation in the field of healthcare-AI. Ethics Inf. Technol. **24**(1), 1–15 (2022). https://doi.org/10.1007/s10676-022-09627-0
65 Cresswell, K., Cunningham-Burley, S., Sheikh, A.: Health care robotics: qualitative exploration of key challenges and future directions. J. Med. Internet Res. **20**(7), e10410 (2018). https://doi.org/10.2196/10410
66. Amann, J., Vayena, E., Ormond, K.E., Frey, D., Madai, V.I., Blasimme, A.: Expectations and attitudes towards medical artificial intelligence: a qualitative study in the field of stroke. PLoS One **18**(1), e0279088 (2023). https://doi.org/10.1371/journal.pone.0279088
67. Aquino, Y.S.J., Rogers, W.A., Braunack-Mayer, A., Frazer, H., Win, K.T., Houssami, N., et al.: Utopia versus dystopia: professional perspectives on the impact of healthcare artificial intelligence on clinical roles and skills. Int. J. Med. Inform. **169**(104903), 1–10 (2023). https://doi.org/10.1016/j.ijmedinf.2022.104903
68. Sartori, L., Bocca, G.: Minding the gap(s): public perceptions of AI and socio-technical imaginaries. AI Soc. **38**(2), 443–458 (2022). https://doi.org/10.1007/s00146-022-01422-1
69. Chen, Y.-N.K., Wen, C.-H.R.: Impacts of attitudes toward government and corporations on public trust in artificial intelligence. Commun. Stud. **72**(1), 115–131 (2021). https://doi.org/10.1080/10510974.2020.1807380
70. Aitken, M., Ng, M., Horsfall, D., Coopamootoo, K.P.L., van Moorsel, A., Elliott, K.: In pursuit of socially ly-minded data-intensive innovation in banking: a focus group study of public expectations of digital innovation in banking. Technol. Soc. **66**(101666), 1–10 (2021). https://doi.org/10.1016/j.techsoc.2021.101666
71. Choung, H., David, P., Ross, A.: Trust and ethics in AI. AI Soc. **38**(2), 733–745 (2023). https://doi.org/10.1007/s00146-022-01473-4
72. Hartwig, T., Ikkatai, Y., Takanashi, N., Yokoyama, H.M.: Artificial intelligence ELSI score for science and technology: a comparison between Japan and the US. AI Soc. **38**(4), 1609–1626 (2023). https://doi.org/10.1007/s00146-021-01323-9
73. Ploug, T., Sundby, A., Moeslund, T.B., Holm, S.: Population preferences for performance and explainability of artificial intelligence in health care: choice-based conjoint survey. J. Med. Internet Res. **23**(12), e26611 (2021). https://doi.org/10.2196/26611
74. Zheng, B., Wu, M.N., Zhu, S.J., Zhou, H.X., Hao, X.L., Fei, F.Q., et al.: Attitudes of medical workers in China toward artificial intelligence in ophthalmology: a comparative survey. BMC Health Serv. Res. **21**(1067), 1–13 (2021). https://doi.org/10.1186/S12913-021-07044-5
75. Ma, J., Tojib, D., Tsarenko, Y.: Sex robots: are we ready for them? An exploration of the psychological mechanisms underlying people's receptiveness of sex robots. J. Bus. Ethics **178**(4), 1091–1107 (2022). https://doi.org/10.1007/s10551-022-05059-4
76. Rhim, J., Lee, G.B., Lee, J.H.: Human moral reasoning types in autonomous vehicle moral dilemma: a cross-cultural comparison of Korea and Canada. Comput. Hum. Behav. **102**, 39–56 (2020). https://doi.org/10.1016/j.chb.2019.08.010
77. Dempsey, R.P., Brunet, J.R., Dubljevic, V.: Exploring and understanding law enforcement's relationship with technology: a qualitative interview study of police officers in North Carolina. Appl. Sci-Basel **13**(6), 1–17 (2023). https://doi.org/10.3390/App13063887
78. Lee, C.H., Gobir, N., Gurn, A., Soep, E.: In the black mirror: youth investigations into artificial intelligence. ACM Trans. Comput. Educ. **22**(3), 1–25 (2022). https://doi.org/10.1145/3484495
79. Kong, S.C., Cheung, W.M.Y., Zhang, G.: Evaluating an artificial intelligence literacy programme for developing university students? Conceptual understanding, literacy, empowerment and ethical awareness. Educ. Technol. Soc. **26**(1), 16–30 (2023). https://doi.org/10.30191/Ets.202301_26(1).0002
80. Street, J., Barrie, H., Eliott, J., Carolan, L., McCorry, F., Cebulla, A., et al.: Older adults' perspectives of smart technologies to support aging at home: insights from five world cafe forums. Int. J. Environ. Res. Public Health **19**(7817), 1–22 (2022). https://doi.org/10.3390/Ijerph19137817
81. Ikkatai, Y., Hartwig, T., Takanashi, N., Yokoyama, H.M.: Octagon measurement: public attitudes toward AI ethics. Int J Hum-Comput Int. **38**(17), 1589–1606 (2022). https://doi.org/10.1080/10447318.2021.2009669
82. Wang, S., Bolling, K., Mao, W., Reichstadt, J., Jeste, D., Kim, H.C., Nebeker, C.: Technology to support aging in place: older adults' perspectives. Healthcare (Basel) **7**(60), 1–18 (2019). https://doi.org/10.3390/healthcare7020060
83. Zhang, H., Lee, I., Ali, S., DiPaola, D., Cheng, Y.H., Breazeal, C.: Integrating ethics and career futures with technical learning to promote AI literacy for middle school students: an exploratory study. Int. J. Artif. Intell. Educ. **33**, 290–324 (2022). https://doi.org/10.1007/s40593-022-00293-3
84 Henriksen, A., Blond, L.: Executive-centered AI? Designing predictive systems for the public sector. Soc. Stud. Sci. (2023). https://doi.org/10.1177/03063127231163756
85. Nichol, A.A., Halley, M.C., Federico, C.A., Cho, M.K., Sankar, P.L.: Not in my AI: moral engagement and disengagement in health care AI development. Pac. Symp. Biocomput. **28**, 496–506 (2023)
86 Aquino, Y.S.J., Carter, S.M., Houssami, N., Braunack-Mayer, A., Win, K.T., Degeling, C., et al.: Practical, epistemic and normative implications of algorithmic bias in healthcare artificial intelligence: a qualitative study of multidisciplinary expert perspectives. J. Med. Ethics (2023). https://doi.org/10.1136/jme-2022-108850
87. Nichol, A.A., Bendavid, E., Mutenherwa, F., Patel, C., Cho, M.K.: Diverse experts' perspectives on ethical issues of using machine learning to predict HIV/AIDS risk in sub-Saharan Africa: a modified Delphi study. BMJ Open **11**(7), e052287 (2021). https://doi.org/10.1136/bmjopen-2021-052287
88. Awad, E., Levine, S., Kleiman-Weiner, M., Dsouza, S., Tenenbaum, J.B., Shariff, A., et al.: Drivers are blamed more than their automated cars when both make mistakes. Nat. Hum. Behav. **4**(2), 134–143 (2020). https://doi.org/10.1038/s41562-019-0762-8
89. Blease, C., Kaptchuk, T.J., Bernstein, M.H., Mandl, K.D., Halamka, J.D., DesRoches, C.M.: Artificial intelligence and

the future of primary care: exploratory qualitative study of UK general practitioners' views. J. Med. Internet Res. **21**(3), e12802 (2019). https://doi.org/10.2196/12802
90. Blease, C., Locher, C., Leon-Carlyle, M., Doraiswamy, M.: Artificial intelligence and the future of psychiatry: qualitative findings from a global physician survey. Digit. Health **6**, 1–18 (2020). https://doi.org/10.1177/2055207620968355
91. De Graaf, M.M.A., Hindriks, F.A., Hindriks, K.V.: Who wants to grant robots rights? Front Robot AI **8**, 781985 (2022). https://doi.org/10.3389/frobt.2021.781985
92. Guerouaou, N., Vaiva, G., Aucouturier, J.-J.: The shallow of your smile: the ethics of expressive vocal deep-fakes. Philos. Trans. R Soc. B Biol. Sci. **377**(1841), 1–11 (2022). https://doi.org/10.1098/rstb.2021.0083
93. McCradden, M.D., Baba, A., Saha, A., Ahmad, S., Boparai, K., Fadaiefard, P., Cusimano, M.D.: Ethical concerns around use of artificial intelligence in health care research from the perspective of patients with meningioma, caregivers and health care providers: a qualitative study. CMAJ Open **8**(1), E90–E95 (2020). https://doi.org/10.9778/cmajo.20190151
94. Rogers, W.A., Draper, H., Carter, S.M.: Evaluation of artificial intelligence clinical applications: Detailed case analyses show value of healthcare ethics approach in identifying patient care issues. Bioethics **36**(4), 624–633 (2021). https://doi.org/10.1111/bioe.12885
95. Tosoni, S., Voruganti, I., Lajkosz, K., Habal, F., Murphy, P., Wong, R.K.S., et al.: The use of personal health information outside the circle of care: consent preferences of patients from an academic health care institution. BMC Med. Ethics **22**(29), 1–14 (2021). https://doi.org/10.1186/S12910-021-00598-3
96. Allahabadi, H., Amann, J., Balot, I., Beretta, A., Binkley, C., Bozenhard, J., et al.: Assessing trustworthy AI in times of COVID-19: deep learning for predicting a multiregional score conveying the degree of lung compromise in COVID-19 patients. IEEE Trans. Technol. Soc. **3**(4), 272–289 (2022). https://doi.org/10.1109/TTS.2022.3195114
97. Gray, K., Slavotinek, J., Dimaguila, G.L., Choo, D.: Artificial intelligence education for the health workforce: expert survey of approaches and needs. JMIR Med. Educ. **8**(2), e35223 (2022). https://doi.org/10.2196/35223
98. Alfrink, K., Keller, I., Doorn, N., Kortuem, G.: Tensions in transparent urban AI: designing a smart electric vehicle charge point. AI Soc. **38**(3), 1049–1065 (2022). https://doi.org/10.1007/s00146-022-01436-9
99. Bourla, A., Ferreri, F., Ogorzelec, L., Peretti, C.S., Guinchard, C., Mouchabac, S.: Psychiatrists' attitudes toward disruptive new technologies: mixed-methods study. JMIR Ment. Health **5**(4), e10240 (2018). https://doi.org/10.2196/10240
100. Kopecky, R., Kosova, M.J., Novotny, D.D., Flegr, J., Cerny, D.: How virtue signalling makes us better: moral preferences with respect to autonomous vehicle type choices. AI Soc. **38**, 937–946 (2022). https://doi.org/10.1007/s00146-022-01461-8
101. Lam, K., Abramoff, M.D., Balibrea, J.M., Bishop, S.M., Brady, R.R., Callcut, R.A., et al.: A Delphi consensus statement for digital surgery. NPJ Digit. Med. **5**(100), 1–9 (2022). https://doi.org/10.1038/s41746-022-00641-6
102. Karaca, O., Çalışkan, S.A., Demir, K.: Medical artificial intelligence readiness scale for medical students (MAIRS-MS) – development, validity and reliability study. BMC Med. Educ. **21**(112), 1–9 (2021). https://doi.org/10.1186/s12909-021-02546-6
103 Papyshev, G., Yarime, M.: The limitation of ethics-based approaches to regulating artificial intelligence: regulatory gifting in the context of Russia. AI Soc. (2022). https://doi.org/10.1007/s00146-022-01611-y
104. Balaram, B., Greenham, T., Leonard, J.: Artificial intelligence: real public engagement. RSA, London. https://www.thersa.org/globalassets/pdfs/reports/rsa_artificial-intelligence---real-public-engagement.pdf (2018). Accessed 28 June 2023
105. Hagendorff, T.: A virtue-based framework to support putting AI ethics into practice. Philos Technol. **35**(55), 1–24 (2022). https://doi.org/10.1007/s13347-022-00553-z
106. Felt, U., Wynne, B., Callon, M., Gonçalves, M. E., Jasanoff, S., Jepsen, M., et al.: Taking european knowledge society seriously. Eur Comm, Brussels, 1–89 (2007). https://op.europa.eu/en/publication-detail/-/publication/5d0e77c7-2948-4ef5-aec7-bd18efe3c442/language-en
107. Michael, M.: Publics performing publics: of PiGs, PiPs and politics. Public Underst. Sci. **18**(5), 617–631 (2009). https://doi.org/10.1177/09636625080985
108 Hu, L.: Tech ethics: speaking ethics to power, or power speaking ethics? J. Soc. Comput. **2**(3), 238–248 (2021). https://doi.org/10.23919/JSC.2021.0033
109 Strasser, B., Baudry, J., Mahr, D., Sanchez, G., Tancoigne, E.: "Citizen science"? Rethinking science and public participation. Sci. Technol. Stud. **32**(2), 52–76 (2019). https://doi.org/10.23987/sts.60425
110. De Saille, S.: Innovating innovation policy: the emergence of 'Responsible Research and Innovation.' J. Responsible Innov. **2**(2), 152–168 (2015). https://doi.org/10.1080/23299460.2015.1045280
111. Schwarz-Plaschg, C.: Nanotechnology is like… The rhetorical roles of analogies in public engagement. Public Underst. Sci. **27**(2), 153–167 (2018). https://doi.org/10.1177/0963662516655686
112 Taylor, R.R., O'Dell, B., Murphy, J.W.: Human-centric AI: philosophical and community-centric considerations. AI Soc. (2023). https://doi.org/10.1007/s00146-023-01694-1
113. van Dijk, N., Tanas, A., Rommetveit, K., Raab, C.: Right engineering? The redesign of privacy and personal data protection. Int. Rev. Law Comput. Technol. **32**(2–3), 230–256 (2018). https://doi.org/10.1080/13600869.2018.1457002
114 Aradau, C., Blanke, T.: Algorithmic reason. The new government of self and others. Oxford University Press, Oxford (2022)
115. McGoey, L.: The unknowers. How strategic ignorance rules the word. Zed, London (2019)
116. Ananny, M., Crawford, K.: Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability. New Media Soc. **20**(3), 973–989 (2018). https://doi.org/10.1177/1461444816676645