A Literature Review on Ethics for AI in Biomedical Research and Biobanking

Michaela Kargl, Markus Plass, Heimo Müller
Medical University Graz, Graz, Austria

Summary

Background: Artificial Intelligence (AI) is becoming more and more important especially in datacentric fields, such as biomedical research and biobanking. However, AI does not only offer advantages and promising benefits, but brings about also ethical risks and perils. In recent years, there has been growing interest in AI ethics, as reflected by a huge number of (scientific) literature dealing with the topic of AI ethics. The main objectives of this review are: (1) to provide an overview about important (upcoming) AI ethics regulations and international recommendations as well as available AI ethics tools and frameworks relevant to biomedical research, (2) to identify what AI ethics can learn from findings in ethics of traditional biomedical research - in particular looking at ethics in the domain of biobanking, and (3) to provide an overview about the main research questions in the field of AI ethics in biomedical research.

Methods: We adopted a modified thematic review approach focused on understanding AI ethics aspects relevant to biomedical research. For this review, four scientific literature databases at the cross-section of medical, technical, and ethics science literature were queried: PubMed, BMC Medical Ethics, IEEE Xplore, and Google Scholar. In addition, a grey literature search was conducted to identify current trends in legislation and standardization.

Results: More than 2,500 potentially relevant publications were retrieved through the initial search and 57 documents were included in the final review. The review found many documents describing high-level principles of AI ethics, and some publications describing approaches for making AI ethics more actionable and bridging the principles-to-practice gap. Also, some ongoing regulatory and standardization initiatives related to AI ethics were identified. It was found that ethical aspects of AI implementation in biobanks are often like those in biomedical research, for example with regards to handling big data or tackling informed consent. The review revealed current 'hot' topics in AI ethics related to biomedical research. Furthermore, several published tools and methods aiming to support practical implementation of AI ethics, as well as tools and frameworks specifically addressing complete and transparent reporting of biomedical studies involving AI are described in the review results.

Conclusion: The review results provide a practically useful overview of research strands as well as regulations, guidelines, and tools regarding AI ethics in biomedical research. Furthermore, the review results show the need for an ethical-mindful and balanced approach to AI in biomedical research, and specifically reveal the need for AI ethics research focused on understanding and resolving practical problems arising from the use of AI in science and society.

Keywords

AI ethics; biomedical research; biobank; literature review

1 Introduction

Artificial Intelligence (AI) is becoming more and more important in all sectors [1, 2] and dominates almost all aspects of data-centric research today. It is envisaged that AI will determine the future in medicine and biomedical research [3]. However, it is widely recognized that AI brings about not only a range of advantages, opportunities, and promises, but also risks and perils. Thus, in a euphoric "gold digger" mood it is of utmost importance not to forget about the ethical aspects and consequences of AI to ensure that AI can "work for the good of humanity, individuals, societies and the environment and ecosystems" as formulated by the UNESCO recently [2].

Indeed, there is high interest in AI ethics within the scientific community, which is reflected by a huge body of (scientific) literature dealing with the topic of AI ethics. Recently, several literature reviews have been conducted to frame the topics of AI ethics in general [4, 5], or specifically AI ethics in health care [6-8].

Complementary to these existing literature reviews, the focus of this review paper is on AI ethics in biomedical research and biobanking. However, rather than shedding light on fundamental questions of ethics in this field, the main aim of this review is to address the practical needs and questions of the yearbook reader considering that medical informatics is an applied science discipline at the cross-link of computer science, information technology, and medicine.

1.1 Goals

To achieve a practical and useful overview of AI ethics in biomedical research, the objectives of this literature review are threefold:

- to provide an overview about important (upcoming) AI ethics regulations and international recommendations relevant to biomedical research;
- to identify what AI ethics can learn from findings in ethics of traditional biomedical research - in particular looking at ethics in the domain of biobanking as an inspiring field;
- to provide an overview about the main research questions in the field of AI ethics in biomedical research.

1.2 What is Out of Scope

Since AI ethics in the biomedical/medical domain is a broad topic, it would go beyond the scope of this review to give an exhaustive overview of the complete field. Therefore, this review concentrates on AI ethics in biomedical research. Topics, which are deemed out of scope for this review, include:

- general AI ethics, if the principles do not apply directly to biomedical research;
- specific legal topics, as well as national law and regulations;
- machine ethics, that is how agents/algorithms should behave when faced with ethical dilemmas;
- AI ethics in healthcare when it is about treatment of patients and non-research topics.

However, although AI ethics with respect to treatment of patients in healthcare is beyond the scope of this review, AI ethics in relation to the acquisition of medical data will be covered in this paper, since secondary use of medical data for research purposes is an issue for biomedical research. Furthermore, although specific legal topics and national law are beyond the scope of this review, broader regulatory and political initiatives will be covered in this paper.

2 Methods

Since the aim of this review was to provide useful insights to important AI ethics regulations, recommendations and current research questions relevant to biomedical research, as well as showing how the already well-established ethics of biobanking could inspire AI ethics in biomedical research, the authors adopted a modified thematic review approach focused on understanding AI ethics aspects relevant to biomedical research in a pragmatic way. The focus of the research was on the identification of a traceable research question, the development of a robust search strategy, data gathering and the synthesis of the key findings.

2.1 Research Question and Scoping Questions

The core research question of this review focuses on practically relevant aspects of AI ethics in the biomedical research field. Furthermore, the authors agreed to consider scoping questions, which should help with narrowing search terms, articulating queries and focusing on relevant points throughout retrieval and review of the literature. To help guide this review, these scoping questions included "Which of the existing AI ethics regulations, guidelines and recommendations are important for biomedical research?", "Are there any tools and methods available to support practical implementation of AI ethics in biomedical research?", and "Which aspects of the well-established ethics of biobanking could potentially inspire AI ethics in biomedical research?"

2.2 Information Sources and Search Strategy

For this review, four scientific literature databases at the cross-section of medical, technical and ethics science literature were queried: PubMed, BMC Medical Ethics, IEEE Xplore, and Google Scholar. In addition, also a grey literature search was conducted to identify current trends in the legislation and standardization area, which are not (yet) derivable from scientific literature. The initial literature sources were retrieved in December 2021, and the research was iteratively refined and completed through the middle of January 2022.

For the PubMed queries (initial query conducted on December 22, 2021) five search terms typically used in relation to AI in the research community were used: (1) 'artificial intelligence', (2) 'AI', (3) 'machine learning', (4) 'deep learning', and (5) 'big data'. These were combined with the term 'ethic*', whereby the wildcard, that is the asterisk at the end of the term 'ethic*', allowed for variations, such as 'ethics', 'ethical', or 'ethically'. The PubMed search was limited to the title and abstract fields of the database. In addition, to retrieve only the most current results, a filter for the publication years 2017-2021

was applied. Furthermore, the query results were restricted to documents in the English language where the full text was available. For querying BMC Medical Ethics (initial query conducted on December 9, 2021) there was no need to apply the search term 'ethic*' as there are only papers related to medical ethics included in this database. Thus, for querying BMC Medical Ethics we applied the search terms 'biomedical', 'research', 'artificial', and 'intelligence'. For querying IEEE Xplore (initial query on December 10, 2021) and Google Scholar (initial query on December 15, 2021) we combined the search terms 'ethic*' and 'biomedic*' with 'artificial intelligence', and restricted the results to publications since 2017. Finally, since this query strategy for Google Scholar resulted in a huge amount of results that were not relevant for our research question, we further restricted the query results from Google Scholar by excluding the terms 'healthcare', 'robot*' and 'social media'. In addition, for finding relevant grey literature, a Google search with the search terms 'standardization' & 'AI' & 'ethics' and a Google search with the search terms 'legislation' & 'AI' & 'ethics' were conducted. The first 20 results of each of these Google queries were used as a starting point for a wide and rapid search to find relevant grey literature pointing to international standards and norms as well as legislation relevant to AI ethics in biomedical research.

2.3 Paper Selection and Analysis

Through the search strategy described in the previous section, in total 2,525 scientific publications were returned, and about 20 websites/online articles pointing to potentially relevant aspects of laws and standards were found. For the first stage of document selection, these results were screened based on titles, and those documents/articles clearly not relevant to the research question were discarded. As a result of this first selection procedure and after removal of duplicates, 168 documents remained on the shortlist for review on the abstract. In a second step, the full papers of these shortlisted articles were downloaded, and two reviewers screened the documents

on abstracts against the inclusion and exclusion criteria to identify the potentially relevant papers. Consensus between the two reviewers was reached by discussion. The result of this second stage of document selection was a final list of 66 articles to be read in more detail. Resulting from this in-depth analysis, 57 documents were included as references for this review paper. The process of article selection as applied for this review is depicted in Figure 1.

3 Results

3.1 Legislation, Guidelines, and Standards

3.1.1 Relevant Regulatory and Political Initiatives

In April 2021, the European Commission proposed the Artificial Intelligence Act [9], which is worldwide the first legal framework on AI. The proposed regulation lays down harmonized rules on AI to support the objective of the European Union being a global leader in the development of secure, trustworthy and ethical artificial intelligence and to ensure the protection of ethical principles. According to this proposed regulation, 'high-risk' AI systems, which pose significant risk to the health, safety, or fundamental rights of persons, will have to comply with mandatory requirements for trustworthy AI and follow conformity assessment procedures before these systems can be placed on the European Union market. To ensure safety and protection of fundamental rights throughout the whole AI system's lifecycle, the Artificial Intelligence Act sets out clear obligations also for providers and users of these AI systems [9].

In June 2021, the World Health Organization (WHO) published a guidance document on "Ethics and Governance of Artificial Intelligence for Health" [3], and in November 2021, all 193 Member states of the UN Educational, Scientific and Cultural Organization (UNESCO) adopted an agreement that defines the common values and principles needed to
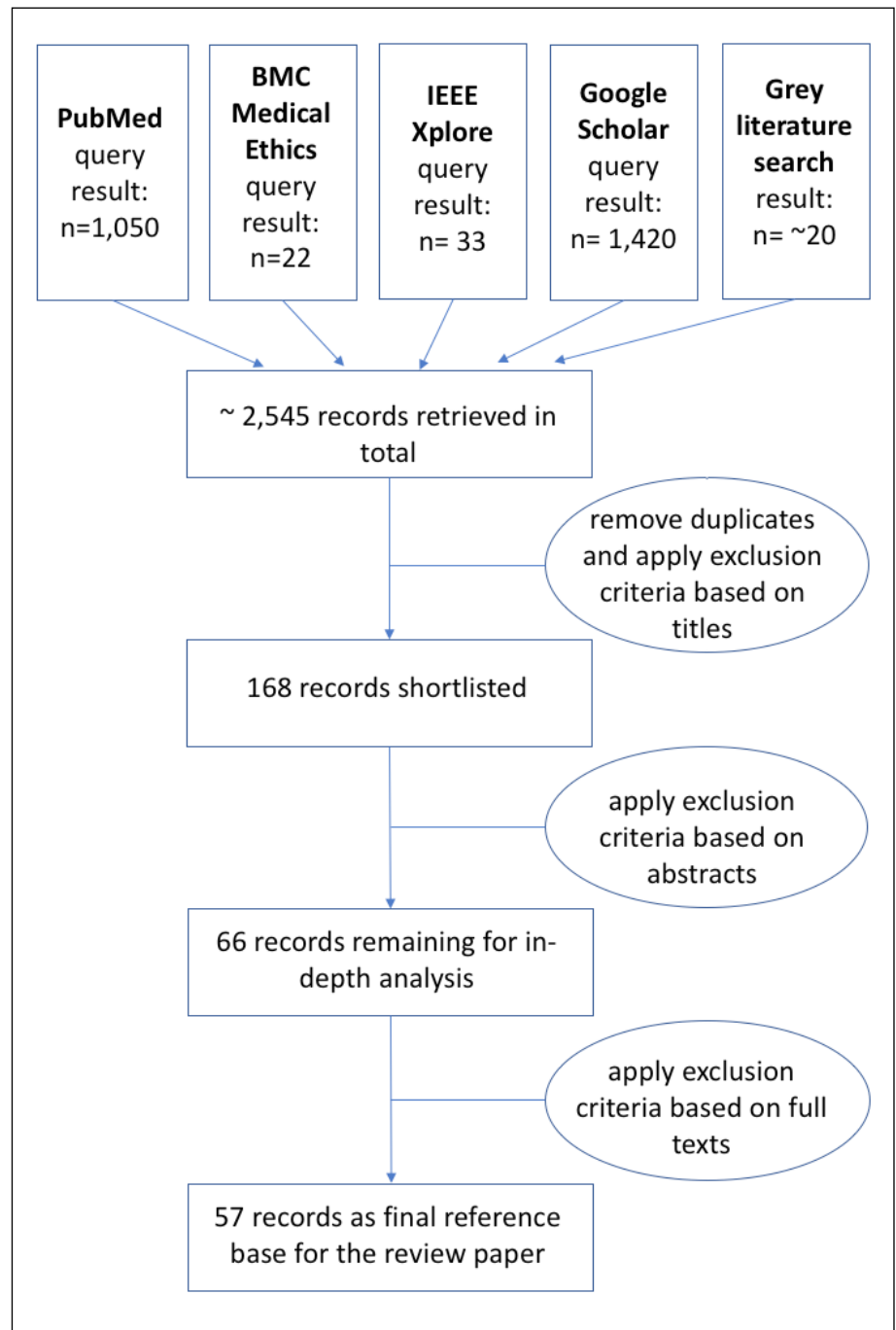

Fig. 1 Overview of the article selection process applied for this review.

ensure the healthy development of AI [2]. Furthermore, this UNESCO agreement encourages governments to set up a regulatory framework that defines a procedure, particularly for public authorities, to

carry out ethical impact assessments on AI systems to predict consequences, mitigate risks, avoid harmful consequences, facilitate citizen participation and address societal challenges [2].

3.1.2 High-level AI Ethics Principles

There is high interest in AI ethics, which is reflected in many AI ethics guidelines developed by government, science, industry and non-profit organizations in recent years. In 2019, Jobin *et al.,* identified 84 documents containing principles and guidelines for ethical AI [4]. Jobin *et al.,* found that most of these guidelines were produced by private companies (22.6%) and governmental agencies (21.4%) in more economically developed countries. By analyzing these 84 guideline documents for ethical AI, Jobin *et al.,* revealed eleven overarching ethical values and principles: transparency, justice and fairness, non-maleficence, responsibility, privacy, beneficence, freedom and autonomy, trust, dignity, sustainability, and solidarity [4]. Although none of these eleven ethical principles appeared in all of the analyzed guideline documents, the first five principles listed above were mentioned in over 50% of the guideline documents analyzed by Jobin *et al*. [4]. In line with the findings of Jobin *et al.*, also Hagendorff, who analyzed and evaluated 22 major AI ethics guidelines, found that especially the aspects of accountability, privacy and fairness appear in about 80% of these guidelines [10]. In a systematic literature study, Khan *et al.*, [5] identified 22 ethical principles relevant for AI, and found that transparency, privacy, accountability and fairness were the four most common ethical principles for AI. In a comparative analysis of 6 AI ethics guideline documents issued by high-profile initiatives established in the interest of socially beneficial AI, Floridi and Cowl also found a high degree of overlap [11]. However, Floridi and Cowl state that overlaps between different guidelines must be taken with caution since similar terms often used to mean different things [11], and also Jobin *et al.,* found significant divergences among the analyzed guidance documents regarding how ethical principles are interpreted, why they are deemed important, what domain/actors they pertain to, and how they should be implemented [4]. Also, Loi *et al.*, stated that it is difficult to compare the ethical principles in the existing AI ethics guidelines, as some guidelines cluster values that others keep separated and many definitions of values are rather vague [12].

As a result of their comparative analysis of AI ethics documents, Floridi and Cowl [11, 13] developed a framework of five overarching ethics principles for AI, composed by the four traditional bioethics principles *beneficence, nonmaleficence, autonomy, and justice*, and one additional AI-specific principle *explicability*, which incorporates both the answers to "how does it work?" and "who is responsible for the way it works?". According to Floridi and Cowl, these five principles capture all of the principles found in the analyzed documents and in addition form also the basis of the "Ethics Guidelines for Trustworthy AI" [1] and the "Recommendation of the Council of Artificial Intelligence" [14] published by the European Commission and the OECD respectively in 2019 [11].

3.1.3 Towards Actionable AI Ethics

The fundamental AI ethics principles formulated in many guidelines are rather theoretical concepts and philosophical foundations. The complexity, variability, subjectivity, and lack of standardization, including variable interpretation of the ethical principles are major challenges to practical implementation of these AI ethics principles [15]. Khan *et al.*, identified fifteen challenging factors for practical implementation of ethics in AI, whereby the lack of ethical knowledge and the vaguely formulated ethical principles were the main challenges that hinder the practical implementation of ethical principles in AI [5]. In addition, Schiff *et al.*, [16] also found socio-technical and disciplinary divides as well as functional separations within organizations as explanations for the principles-to-practices gap. To tackle these issues hindering the practical implementation of AI ethics, guidance needs to go beyond high-level principles. In the following paragraphs some approaches to make AI ethics guidelines suitable for practical use are described.

The "Ethics Guidelines for Trustworthy AI" [1] developed by the High-Level Expert Group on Artificial Intelligence (AI HLEG), which was set up by the European Commission in June 2018, aim to offer guidance for fostering and securing ethical and robust AI by providing not only high-level ethics principles but also guidance on how these principles can be implemented in practice. This guidance document identifies three high-level ethical principles, namely respect for human autonomy, prevention of harm, and fairness & explicability, which should be respected in the development, deployment, and use of AI systems [1]. To provide guidance on how these principles can be implemented, seven key requirements that AI systems should meet are listed: human agency & oversight, technical robustness & safety, privacy & data governance, transparency, diversity, non-discrimination & fairness, environmental & societal well-being, and accountability [1]. These seven key requirements are also included in the Artificial Intelligence Act proposed by the European Commission in 2021 [9].

To provide concrete practical guidance specifically for organizations developing and using AI, Ryan and Stahl [17] retrieved detailed, practically useful explanations of the normative implications of common high-level ethical principles.

Loi *et al.,* propose a framework of seven actionable principles suitable for practical use:

(1) Beneficence: do the good (promote individual and community well-being and preserve trust in trustworthy agents);

(2) Non-maleficence: avoid harm (protect security, privacy, dignity, and sustainability);

(3) Autonomy: promote the capabilities of individuals and groups (protect civic and political freedoms, privacy, and dignity);

(4) Justice: be fair, avoid discrimination, promote social justice and solidarity;

(5) Control: knowledgeably control entities, goals, process, and outcomes affecting people;

(6) Transparency: communicate your knowledge of entities, goals, process, and outcomes, in an adequate and effective way, to the relevant stakeholders;

(7) Accountability: assign moral, legal, and organizational responsibilities to the individuals who control entities, goals, process, and outcomes affecting people [12].

There are several initiatives underway to define standards to support practical implementation of AI ethics - a list of international organizations engaged in AI ethics related standardization is given in [18]. For example, the IEEE Standards Association launched the IEEE P7000® series of eleven standardization projects dedicated to societal and ethical issues associated with AI systems [19]. Also, the International Organization for Standardization's technical committee ISO/IEC JTC 1/SC 42 Artificial Intelligence is currently working on an overview of ethical and societal concerns of information technology and artificial intelligence. This forthcoming technical report ISO/IEC DTR 24368 shall provide guidance to other ISO/IEC technical committees developing standards for domain specific applications that use AI [20].

Besides these general efforts for making AI ethics principles more suitable for practical use, there are also some initiatives targeted specifically at AI ethics in the medical and biomedical domain, as described in the following paragraphs.

Müller *et al.*, [21] formulated the following ten commandments of ethical medical AI as practical guidelines for applying AI in medicine:

(1) It must be recognizable that and which part of a decision or action is taken and carried out by AI;
(2) It must be recognizable which part of the communication is performed by an AI agent;
(3) The responsibility for an AI decision, action, or communicative process must be taken by a competent physical or legal person;
(4) AI decisions, actions, and communicative processes must be transparent and explainable;
(5) An AI decision must be comprehensible and repeatable;
(6) An explanation of an AI decision must be based on state-of-the-art (scientific) theories;
(7) An AI decision, action, or communication must not be manipulative by pretending accuracy;
(8) An AI decision, action, or communication must not violate any applicable law and must not lead to human harm;
(9) An AI decision, action, or communi-

cation shall not be discriminatory. This applies to the training of algorithms;
(10) The target setting, control, and monitoring of AI decisions, actions, and communications shall not be performed by algorithms [21].

Since lack of effective interdisciplinary practices have been identified as issues hindering practical implementation of AI ethics [16], Jongsma and Bredenoord [22] suggest 'ethics parallel research', a process where ethicists are closely involved in the development of new technologies from the beginning, as a practical approach for ethical guidance of biomedical innovation.

The Primary Care Informatics Working Group of the International Medical Informatics Association (IMIA) stated 14 principles for ethical use of routinely collected health data and AI [23] and formulated the following six concrete recommendations:

(1) Ensure consent and formal process to govern access and sharing throughout the data life cycle;
(2) Sustainable data creation & collection requires trust and permission;
(3) Pay attention to Extract-Transform-Load processes as they may have unrecognized risks;
(4) Integrate data governance and data quality management to support clinical practice in integrated care systems;
(5) Recognize the need for new processes to address the ethical issues arising from AI in primary care;
(6) Apply an ethical framework mapped to the data life cycle, including an assessment of data quality to achieve effective data curation [23].

3.2 AI-Ethics in Biobanking

Biobanks are an important infrastructure in medical research and are a long established (research) field traditionally concerned with ethical issues. In the overview textbook "The ethics of research biobanking" Solbakk *et al.*, group these ethical issues related to biobanks into the four clusters: a) Issues concerning how biological materials are entered into the bank; b) Issues concerning research

biobanks as institutions; c) Issues concerning under what conditions researchers can access materials in the bank, problems concerning ownership of biological materials and of intellectual property arising from such materials; and d) Issues related to the information collected and stored, e.g., access-rights, disclosure, confidentiality, data security, and data protection [24].

In our review on AI related ethics in biobanking, we intentionally excluded the field of clinical trials, not data-centered as well as interventional studies, as all these rely on special requirements and specific legal frameworks. In this review, we focus on population-based biobanks, study-oriented biobanks and clinical biobanks aiming for secondary use of medical data e.g., in the development or validation of AI algorithms. In this case, a biobank itself has already covered a series of ethical issues during the collection of samples and data and has a clear policy for data reuse in research projects, which should cover both non-AI and AI based scenarios. How to implement AI in international biobanks covering also ethical legal and governance requirements is described by Kozlakidis [25], and future possibilities of AI in biobanking are described by Lee [26]. Data sets related to biosamples can be high-volume, high-velocity and high-variety information assets, which means we can speak of 'big-data' in biobanks. For such 'big data' AI methods as machine learning and deep learning can be applied to analyze and extract knowledge, for example to train automatic decision-making systems [27]. Whenever data from humans are used in the development of AI-based models, issues how data providers and donors are informed about their involvement arise, which are very similar in biobanking and AI development. Thus, the usage of a good, informed consent plays a central role in both biobanking and AI development. Jurate *et al.,* analyzed consent documents in terms of model of the consent, scope of future research, access to medical data, feedback to the participants, consent withdrawal, and role of ethics committees [28]. The transition of biobanks from a simple sample storage service to data banks and data curation centers, e.g., for longitudinal and population-based biobanks, brings a biobank into the role of a trusted data re-

pository and fate-keeper for secondary use of medical data. When the data objects are used in the training and validation of high-risk AI (this is the case for most medical AI solutions), biobank guidelines should follow the recommendations as laid out in article 10 (data and data governance) of the European Commission's harmonized rules in the Artificial Intelligence Act [9]. For ethical AI study design, Chauhan and Gullapalli propose an inclusive AI design and bias-covering sample choice and valuation. They also address the controversial concept of race in ethical design [29]. Besides the primary data source, they also raise the question of other stakeholders, for example how a pathologist, who worked on generating the annotation, should be compensated. In addition to clinical reporting guidelines, Baxi et al., demand for a similar approach to bias risk guidelines in data annotation, for example in digital pathology [30]. At an institutional level, Gille et al., propose mechanisms for future-proof biobank governance, which help to signal trustworthiness to stakeholders and the public. These mechanisms, which are proposed for biobank governance, can also be applied to AI ethics in biomedical research institutions [31].

3.3 Main Research Topics Regarding AI Ethics in Biomedical Research

3.3.1 General Research Topics

A scoping review of the ethics literature in the medical field conducted by Murphy et al., [8] revealed that the main research in that area was related to the common ethical themes of privacy and security, trust in AI, accountability & responsibility, and bias [8]. In a recent bibliometric analysis Saheb et al., [32] found twelve clusters of research questions in AI ethics, from which the following cover the research questions on AI ethics in biomedical research raised in the reviewed literature: *data ethics* (addressing e.g., data ownership, data sharing and usage, data privacy, data bias & skewness, and sensitivity & specificity), *algorithm and model ethics* (addressing e.g., machine decision making, algorithm selection processes, training and testing of AI models, transparency,

interpretability, explainability, replicability, algorithm bias, error risk, and transparency of data flow), *predictive analytics ethics* (addressing e.g., discriminatory decisions and contextually relevant insight), *normative ethics* (addressing e.g., discrimination by generalization of AI conclusions, justice, fairness, and inequality), *relationship ethics* (addressing e.g., user interfaces and human-computer interaction, as well as relationships between patients, physicians and other healthcare stakeholders) [32].

In addition to research in these areas identified by Saheb et al., [32], researchers in the field of AI ethics, such as for example Nebeker et al., [18] and Goodman [33], also demand for standards to support actionable ethics. Furthermore, there is still applied research needed to guide developers, users and institutions on the question how to adopt and evaluate AI ethics in health informatics and biomedical research. Blasimme and Vayena propose to structure this effort according to adaptivity, flexibility, inclusiveness, reflexivity, responsiveness and monitoring (AFIRRM) principles [34].

3.3.2 AI Ethics Tools and Methods

Morley et al., [35] conducted a review of publicly available AI ethics tools and methods, which aim to help developers, engineers, and designers of machine learning applications to translate AI ethics principles into practice. The result of their work is a typology listing for each of the five overarching AI ethics principles (beneficence, non-maleficence, autonomy, justice, and explicability) and the tools and methods available to apply that ethics principle in each stage of machine learning algorithm development. In total, 107 tools and methods are included in this typology. Morley et al., found that the availability of tools is not evenly distributed across ethical principles and across the stages of machine learning algorithm development. According to the review of Morley et al., the greatest range of tools and methods is available for the principle of explicability at the stage of testing, and these tools are primarily 'statistical' in nature such as for example LIME (Local Interpretable Model-Agnostic Explanations). Furthermore, Morley et al.,

noticed that most of the available AI ethics tools and methods lack usability and are not actionable in practice as they offer only limited documentation and little help on how to use them, and users would need a high skill-level to apply these tools in practice [35].

Checklists, Frameworks, and Processes

In 2020, the High-Level Expert Group on Artificial Intelligence, set up by the European Commission, provided "The Assessment List for Trustworthy Artificial Intelligence (ALTAI)" [36] to support practical implementation of the seven key ethical requirements listed in the "Ethics Guidelines for Trustworthy AI" [1] and referred to in the "Artificial Intelligence Act" proposed by the European Commission in 2021 [9]. "The Assessment List for Trustworthy Artificial Intelligence" is available as a pdf document as well as an interactive online version (https://altai.insight-centre.org/) containing additional explanatory notes. It is intended for self-evaluation purposes and shall support stakeholders to assess whether an AI system that is being developed, deployed, procured, or used, adheres to the seven key ethical requirements [36].

Zicari et al., describe Z-Inspection®, a process based on applied ethics to assess trustworthy AI in practice. They provide a detailed introduction to the phases of the Z-Inspection® process and accompanying material such as catalogues of questions and checklists [37].

Nebeker et al., developed a digital-health decision-making framework and an associated checklist to help researchers and other concerned stakeholders with selecting and evaluating digital technologies for use in health research and healthcare [18]. This digital-health decision-making framework comprises five domains: (1) participant privacy, (2) risks and benefits, (3) access and usability, (4) data management, and (5) ethical principles [18]. Malik et al., provide ten rules for engaging with artificial intelligence in biomedicine, where they mention liabilities of computational error, bias harmful to underrepresented groups, as well as privacy and consent challenges especially in genomics research as the main ethical implications of AI in biomedicine [38].

IBM's research group on Trusted AI provides a range of tools to support ethical principles in AI development [39]. These tools include for example "AI Fairness 360" - an open source software toolkit to detect and remove bias in machine learning models [40].

Tools for Reporting AI Transparently

IBM's research group on Trusted AI [39] provides "AI FactSheets 360" - a methodology to create complete and transparent documentation of an AI model/application [41]. Also, Mitchell *et al.*, describe a framework called "Model Cards", which supports transparent reporting and documentation of trained machine learning models including their performance evaluation and intended use context [42].

Complete and transparent reporting of clinical/biomedical study results is an essential building block of ethical research since the reporting is the precondition for reliable assessment of the validity of the study results. However, existing reporting guidelines for clinical/biomedical studies are not sufficient to address potential sources of bias specific to AI systems, such as for example the procedure for acquiring input data, data preprocessing steps, and model development choices [43, 44]. To ensure complete and transparent reporting of clinical trials of AI systems, AI-related extensions of guidelines for reporting clinical trial protocols and completed clinical trials are developed by key stakeholders in the Enhancing Quality and Transparency of Health Research (EQUATOR) network program. Shelmerdine *et al.*, [45] give a comprehensive overview of reporting guidelines for common study types in (bio)medical research involving AI. Luo *et al.*, [46] created a set of guidelines including a comprehensive list of reporting items and a practical description of the sequential steps for developing predictive models, to help the biomedical research community with correct application of machine learning models and consistent reporting of model specifications and results. To support the potential of AI in biomedical research and help to overcome the reporting deficit in biomedical AI, Matschinske *et al.*, propose the AIMe standard, a generic minimal information standard for reporting of any biomedical AI

system, and the AIMe registry, a community-driven web-based reporting platform for AI in biomedicine. This reporting platform should increase accessibility, reproducibility, and usability of biomedical AI models, and facilitate future revisions by the scientific community [47].

3.3.3 The Need for an Ethical-mindful and Balanced Approach

Although there are many AI ethics guidelines, all these guidelines do not have an actual impact on human decision-making in the field of AI and machine learning [10]. Furthermore, since the AI ethics principles do not have legally binding grounding, there is nothing to prevent any company / country from choosing to adopt a different set of ethics principles for the sake of convenience or competitiveness [35]. Jotterand and Bosco [48] argue that the ethical framework that sustains a responsible implementation of such technologies should be reconsidered and assessed in relation to anthropological implications, how the technology might disrupt or enhance the clinical encounter and how this impacts clinical judgments and the care of patients. The humanistic dimension is in the center of their chain of arguments with the dimensions of empathy, respect and emotional intelligence [48]. Buruk *et al.*, give a critical perspective on guidelines for responsible and trustworthy artificial intelligence. They analyzed three main AI ethics guidelines, finding an overlap in several principles, such as human control, autonomy, transparency, security, utility, and equality, but also a great divergence in the description of future scenarios. What is missing in their view are grounded suggestions for ethical dilemmas occurring in practical life and a strategy for reflective equilibrium between ethical principles [49]. Faes *et al.*, [50] raise the need for standardization to critically appraise machine learning studies, and promote standards as TRIPOD-ML, SPIRIT-AI, and CONSORT-AI for reporting covering also ethical issues [51-53]. Yochai Benklerer calls on society to not let industry write the rules for AI and to campaign to bend research and regulation for their benefit. He argues that organizations working to ensure that AI is fair and beneficial must be publicly funded,

subject to peer review and transparent to civil society [54]. In their critical assessment of the movement for ethical AI and machine learning, Green *et al.*, are missing shared consensus on the moral responsibility of computer engineers and data scientists towards their own inventions. They ask to generate the moral consensus (according to Popper) and warn the community that ethical design possesses some of the same elements as Value Sensitive Design but lacks their explicit focus on normative ends devoted to social justice or equitable human flourishing [55]. Brent Mittelstadt argues that principles alone cannot guarantee ethical AI. He states that we should not yet celebrate consensus around high-level principles that hide deep political and normative disagreement, as AI development lacks common aims and fiduciary duties, professional history and norms, proven methods to translate principles into practice, and robust legal and professional accountability mechanisms [56].

Whittlestone [57] highlights some of the limitations of ethics principles. In particular, she criticizes they are often too broad and high-level to guide ethics in practice. She suggests that an important next step for the field of AI ethics is to focus on exploring the tensions that inevitably arise as stakeholders try to implement principles. With the term 'tension' she refers to any conflict, whether apparent, contingent or fundamental, between important values or goals, where it appears necessary to give up one in order to realize the other. To improve the current situation, Whittlestone proposes that principles need to be formalized in standards, codes and ultimately regulation and the research topics in AI ethics to be more focused on understanding and resolving tensions as an important step towards solving practical problems arising from the use of AI in society [57].

4 Conclusions

The goal of this review was to give a practically useful overview of research strands as well as regulations, guidelines, and tools regarding AI ethics in biomedical research. To reach this goal, more than 2,500 pub-

lications were retrieved through queries of scientific databases and grey literature search, and 57 of these were analyzed in-depth. The review revealed that there is a large number of publications regarding high-level AI ethics principles, but there are only a few publications dedicated to helping practitioners with implementation of these high-level principles in practice. Furthermore, the review found that there is a large body of literature regarding AI ethics in healthcare, but comparatively fewer publications are dealing with AI ethics in (bio)medical research. Many of these analyzed publications, which are specifically dedicated to AI ethics in (bio) medical research, tackle the issue of correct, comprehensive, and transparent reporting of (bio)medical studies involving AI. From the literature, ethics in biobanking is – in contrast to AI – a long established research field covering informed consent, collection of samples, bias in population, as well as all aspects of secondary sample and data (re)-use. Ethical aspects of AI implementation in biobanking are often like those of AI in biomedical research, especially with regards to handling big data or tackling informed consent.

Overall, the review results show the need for an ethical mindful and balanced approach to AI in biomedical research, specifically the need for AI ethics research focused on understanding and resolving practical problems arising from the use of AI in science and society.

Table 1 Summary of the key findings from literature review.

Key Findings from Literature Review	Further Reading
AI ethics arrived on the political agenda and have become recognized as important topic also at political level, recently.	[1-3,9, 14]
High-level AI ethics principles are stated by many documents developed by government, science, industry, and non-profit organizations in recent years.	[4,5,10-13]
Reasons for principles-to-practice-gap are complexity, variability, and subjectivity of the AI ethics principles on the one hand, as well as lack of ethical knowledge, socio-technical divides and functional separations within organizations on the other hand.	[5,15,16]
Efforts to make AI ethics more tangible for practical implementation include practically useful formulations and explanations of AI principles (also specifically for the biomedical domain), development of concrete requirements for AI systems, and standardization initiatives.	[1,12,17-21,23]
Ethics traditions in the field of research-oriented biobanks already covered a series of ethical issues during the collection of samples and data and have a clear policy for data reuse in research projects, which should cover both non-AI and AI based scenarios.	[24-26,28,31]
Available tools and methods to help practitioners with design, development, and implementation of ethical AI, include methods addressing explicability of AI at the stage of model testing, tools to detect bias in AI algorithms, as well as checklists and processes for self-assessment of compliance with AI ethics requirements.	[18,35-40]
Available tools and methods to support ethical AI in (biomedical) research are focusing on transparent and complete reporting of AI-related studies and projects.	[39,41-47,51-53]
Current research regarding AI ethics in biomedical research covers a wide range of topics including ethics related to data, algorithms & models, predictive analytics, norms, and relationships.	[8,32]
Future research shall include applied research and standards addressing ethical dilemmas occurring in practice as well as anthropological implications of AI usage (in the biomedical domain).	[18,33,48-50,57]
Calls on society include shared consensus on the moral responsibility of computer engineers and data scientists towards their AI inventions, public funding and transparency of organizations working to ensure that AI is fair and beneficial, as well as robust legal and professional accountability mechanisms, all of which are needed to achieve ethical AI.	[54-57]

References

1. AI HLEG, European Commission. Ethics guidelines for trustworthy AI; 2019.
2. UNESCO. Recommendation on the ethics of artificial intelligence; 2021. Available from: https://unesdoc.unesco.org/ark:/48223/pf0000377897
3. WHO. Ethics and governance of artificial intelligence for health ethics and governance of artificial intelligence for health 2; 2021. Available from: http://apps.who.int/bookorders
4. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nature Machine Intelligence 2019;1:389–99.
5. Khan AA, Badshah S, Liang P, Khan B, Waseem M, Niazi M, et al. Ethics of AI: A systematic literature review of principles and challenges The International Conference on Evaluation and Assessment in Software Engineering 2022. p. 383-92.
6. Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: A mapping review. Soc Sci Med 2020 Sep;260:113172.
7. Ienca M, Ferretti A, Hurst S, Puhan M, Lovis C, Vayena E. Considerations for ethics review of big data health research: A scoping review. PLoS One 2018 Oct 11;13(10):e0204937.
8. Murphy K, Di Ruggiero E, Upshur R, Willison DJ, Malhotra N, Cai JC, et al. Artificial intelligence for good health: a scoping review of the ethics literature. BMC Med Ethics 2021 Feb 15;22(1):14.
9. European Commission. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts; 2021.
10. Hagendorff T. The ethics of AI ethics: An evaluation of guidelines. Minds and Machines 2020;30:99–120.
11. Floridi L, Cowls J. A unified framework of five principles for AI in society. Harvard Data Science Review 2019;1.

12. Loi M, Heitz C, Christen M. A comparative assessment and synthesis of twenty ethics codes on AI and big data. Institute of Electrical and Electronics Engineers Inc.; 2020. p. 41–6.

13. Floridi L, Cowls C, Beltrametti M, Chatila R, Chazerand P, Dignum V, et al. Ai4people - an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines 2018;28:689–707.

14. OECD. Recommendation of the council on artificial intelligence. oecd/legal/0449;2019. Available from: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

15. Zhou J, Chen F, Berry A, Reed M, Zhang S, Savage S. A survey on ethical principles of AI and implementations. Institute of Electrical and Electronics Engineers Inc.; 2020. p. 3010–7.

16. Schiff D, Rakova B, A. Ayesh A, Fanti A, Lennon M. Principles to practices for responsible AI: Closing the gap. arXiv preprint arXiv:2006.04707; 2020.

17. Ryan M, Stahl BC. Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. Journal of Information, Communication and Ethics in Society 2021;19:61–86.

18. Nebeker C, Torous J, Bartlett Ellis RJ. Building the case for actionable ethics in digital health research supported by artificial intelligence. BMC Med 2019 Jul 17;17(1):137.

19. IEEE, 2021, IEEE 7000™ projects | IEEE ethics in action. Available from: https://ethicsinaction.ieee.org/p7000/

20. Hjalmarson M. To ethicize or not to ethicize, in ISOfocus_137, 2019. Available from: https://www.iso.org/isofocus_137.html

21. Müller H, Mayrhofer MT, Veen EBV, Holzinger A. The ten commandments of ethical medical AI. Computer 2021;54:119–23.

22. Jongsma KR, Bredenoord AL. Ethics parallel research: an approach for (early) ethical guidance of biomedical innovation. BMC Med Ethics 2020 Sep 1;21(1):81.

23. Liaw ST, Liyanage H, Kuziemsky C, Terry AL, Schreiber R, Jonnagaddala J, et al. Ethical Use of Electronic Health Record Data and Artificial Intelligence: Recommendations of the Primary Care Informatics Working Group of the International Medical Informatics Association. Yearb Med Inform 2020 Aug;29(1):51-7.

24. Solbakk JH, Holm S, Hofmann B. The Ethics of Research Biobanking. Springer US, USA; 2009.

25. Kozlakidis Z. Biobanks and Biobank-Based Artificial Intelligence (AI) Implementation Through an International Lens. In: Artificial Intelligence and Machine Learning for Digital Pathology. Cham: Springer; 2020. p. 195-203.

26. Lee JE. Artificial intelligence in the future biobanking: Current issues in the biobank and future possibilities of artificial intelligence, Biomedical Journal of Scientific & Technical Research 2018;7(3):5937-9.

27. Kinkorová J, Topolčan O. Biobanks in the era of big data: objectives, challenges, perspectives, and innovations for predictive, preventive, and personalised medicine. EPMA J 2020;11(3):333–41.

28. Jurate S, V. Zivile V, Eugenijus G. 'Mirroring' the ethics of biobanking: What analysis of consent documents can tell us? Sci Eng Ethics 2014;20:1079–93.

29. Chauhan C, Gullapalli RR. Ethics of AI in Pathology: Current Paradigms and Emerging Issues. Am J Pathol 2021 Oct;191(10):1673-83.

30. Baxi V, Edwards R, Montalto M, Saha S. Digital pathology and artificial intelligence in translational medicine and clinical practice. Mod Pathol 2022 Jan;35(1):23-32.

31. Gille F, Vayena E, Blasimme A. Future-proofing biobanks' governance. Eur J Hum Genet 2020 Aug;28(8):989-96.

32. Saheb T, Saheb T, Carpenter DO. Mapping research strands of ethics of artificial intelligence in healthcare: A bibliometric and content analysis. Comput Biol Med 2021 Aug;135:104660.

33. Goodman KW. Ethics in Health Informatics. Yearb Med Inform 2020 Aug;29(1):26-31.

34. Blasimme A, Vayena E. The ethics of AI in biomedical research, patient care and public health. Patient Care and Public Health (April 9, 2019). Oxford Handbook of Ethics of Artificial Intelligence. Forthcoming.

35. Morley J, Floridi L, Kinsey L, Elhalal A. From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices, Sci Eng Ethics 2020;26(4):2141-68.

36. European Commission and Directorate-General for Communications Networks, Content and Technology. The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment. EC Publications Office; 2020.

37. Zicari RV, Brodersen J, Brusseau J, Dudder B, Eichhorn T, Ivanov T, et al. Z-inspection®: A process to assess trustworthy AI. IEEE Trans Technol Soc 2021;2:83–97.

38. Malik A, Patel P, Ehsan L, Guleria S, Hartka T, Adewole S, et al. Ten simple rules for engaging with artificial intelligence in biomedicine. PLoS Comput Biol 2021 Feb 11;17(2):e1008531.

39. IBM. Trusted AI | IBM research teams; 2021. Available from: https://research.ibm.com/teams/trusted-ai#tools

40. Bellamy RK, Dey K, Hind M, Hoffman SC, Houde S, Kannan K, et al. AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. IBM J Res Dev 2019;63(4/5): 4:1 - 4:15.

41. Richards J, Piorkowski D, Hind M, Houde S, Mojsilović A. A methodology for creating AI FactSheets. arXiv preprint arXiv:2006.13796;2020.

42. Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, et al. Model cards for model reporting. ACM 2019.

43. Ibrahim H, Liu X, Rivera SC, Moher D, Chan AW, Sydes MR, et al. Reporting guidelines for clinical trials of artificial intelligence interventions: the SPIRIT-AI and CONSORT-AI guidelines. Trials 2021 Jan 6;22(1):11.

44. Sounderajah V, Ashrafian H, Golub RM, Shetty S, De Fauw J, Hooft L, et al; STARD-AI Steering Committee. Developing a reporting guideline for artificial intelligence-centred diagnostic test accuracy studies: the STARD-AI protocol. BMJ Open 2021 Jun 28;11(6):e047709.

45. Shelmerdine SC, Arthurs OJ, Denniston A, Sebire NJ. Review of study reporting guidelines for clinical studies using artificial intelligence in healthcare. BMJ Health Care Inform 2021 Aug;28(1):e100385.

46. Luo W, Phung D, Tran T, Gupta S, Rana S, Karmakar C, et al. Guidelines for Developing and Reporting Machine Learning Predictive Models in Biomedical Research: A Multidisciplinary View. J Med Internet Res 2016 Dec 16;18(12):e323.

47. Matschinske J, Alcaraz N, Benis A, Golebiewski M, Grimm DG, Heumos L, et al. The AIMe registry for artificial intelligence in biomedical research. Nat Methods 2021 Oct;18(10):1128-31.

48. Jotterand F, Bosco C. Artificial Intelligence in Medicine: A Sword of Damocles? J Med Syst 2021 Dec 11;46(1):9.

49. Buruk B, Ekmekci PE, Arda B. A critical perspective on guidelines for responsible and trustworthy artificial intelligence. Med Health Care Philos 2020 Sep;23(3):387-99.

50. Faes L, Liu X, Wagner SK, Fu DJ, Balaskas K, Sim DA, et al. A Clinician's Guide to Artificial Intelligence: How to Critically Appraise Machine Learning Studies. Transl Vis Sci Technol 2020 Feb 12;9(2):7.

51. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. Lancet 2019 Apr 20;393(10181):1577-9.

52. CONSORT-AI and SPIRIT-AI Steering Group. Reporting guidelines for clinical trials evaluating artificial intelligence interventions are needed. Nat Med 2019 Oct;25(10):1467-8.

53. Liu X, Faes L, Calvert MJ, Denniston AK; CONSORT/SPIRIT-AI Extension Group. Extension of the CONSORT and SPIRIT statements. Lancet 2019 Oct 5;394(10205):1225.

54. Benkler Y. Don't let industry write the rules for AI. Nature 2019 May;569(7755):161.

55. Greene D, Hoffmann AL, Stark L. Better, nicer, clearer, fairer: A critical assessment of the movement for ethical artificial intelligence and machine learning. Proceedings of the Annual Hawaii International Conference on System Sciences; 2019. p. 2122–31.

56. Mittelstadt, B. AI Ethics–Too principled to fail. arXiv preprint arXiv:1906.06668. 2019.

57. Whittlestone J, Nyrup R, Alexandrova A, Cave S. The role and limits of principles in AI ethics: Towards a focus on tensions. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. ACM; 2019. p. 195-200. Available from: https://doi.org/10.1145/3306618.3314289

Correspondence to:
Michaela Kargl
Medical University Graz
Auenbruggerplatz 2
Graz, 8036
Austria
E-mail: michaela.kargl@medunigraz.at
www.medunigraz.at