AI Ethics Evaluation Report

1 INTRODUCTION

Artificial intelligence has become pervasive and across various domains, including finance, economy, and healthcare innovations. The increasing reliance on algorithms raises concerns about ethical issues associate with artificial intelligence and its negative consequences, garnering popular and academic attention.

1.1 Overview of ethical issues in AI generally

Advanced Artificial intelligence raises numerous ethical issues that across various aspects of its design, development, and deployment, challenging in assigning moral responsibility in AI ethical evaluation [Coeckelbergh 2019]. Broadly, Public privacy violations and data protection are the most paramount issues as vast amounts of data are necessary in the process of AI design, raising great concerns about data collection, processing and sharing as well as the lack of awareness, consent of collecting and utilization of personal information [Onik et al. 2019].

The lack of transparency, explainability, responsibility and accountability, especially black box operation system in machine learning algorithms, leading to difficulties in understanding the decision-making process in Artificial intelligences technologies and infrastructures, resulting in unfairness, and amplifying social inequities [Kim et al. 2020]. Bias, opacity, and unfairness, such as gender and race bias, led to unjust treatment and discrimination in AI applications, resulting in blatant or potential discriminatory [Nelson 2019]. Notley, the erosion of human autonomy, freedom and human right is another considerable ethical issue that artificial intelligence may cause as AI systems is increasing becoming more and more advanced [Calvo et al. 2020]. Additionally, Vulnerable cybersecurity of AI algorithms systems is inevitable to be attacked, leading to sensitive information leakage and even harm to individuals and society [Kuzlu et al. 2021]. The impact of artificial intelligence on potential employment displacement exacerbates economic disparities, reinforce existing power structures and existing social inequities [Abuselidze and Mamaladze 2021].

Minimizing negative ethical impacts in AI requires interdisciplinary collaboration among ethicists, policymakers, and other responsible parties for algorithmic decisions and the emphasizing of importance of transparence in AI algorithmic decisionmaking processes [Mahmud et al. 2022].

1.2 AI ethical issues in healthcare

Long history of AI applications and technologies in healthcare and medicine has proved that AI has the capacity to revolutionize various aspect of healthcare [Johnson et al. 2020], ultimately leading to better health outcomes for human populations and society safety [Bur et al. 2019] by augmenting human capabilities improving and enhancing overall healthcare delivery, for example, its significant impact of AI advancements on diagnostic methods and treatments, potential to adapt to development in image diagnosis and robotic surgery [Singh 2018]. While AI provides such significant beneficial potential, it also raises ethical and practical issues, proactively address ethical issues arising in applications of AI in healthcare need efforts from global evaluations and policy coordination [Santosh and Gaur 2021].

1.2.1 Key AI ethical concerns in healthcare

Unequal access to AI-enabled healthcare is the one of the most paramount ethical concerns when employing AI technologies in healthcare, patricianly geographic disparities, and socioeconomic disparities, exacerbating existing healthcare disparities [Curtis and Brolan 2022]. In another words, the gap in healthcare, opportunities, and outcomes will be widened because certain populations from rural or remote area where patients have limited chances to access to AI-enabled healthcare and advanced healthcare infrastructure [Kuiler and McNeely 2023], while patients in urban areas greatly benefit from timely and accurate diagnoses facilitated by AI [Goralski and Tan 2023]. Similarly, Patients from lower socioeconomic backgrounds, especially for those who from lower-income communities and do no enrolled in public health insurance plans, face financial constraints or not able to afford it due to costly infrastructure, equipment, and trained personnel of implementation of AIdriven healthcare detection [Zhou 2020].

Privacy concerns about storage and transmission of sensitive patients' information when implementing artificial intelligence in healthcare receives public attention lately [Murdoch 2021]. AI-assisted clinical trials that handle large amounts data collecting form patient health records, including medical history, test results, and genetic information [Reddy et al. 2019], although the notable advantage is to produce accurate outcomes and improve patient care, there is high risk of unauthorized access to such sensitive information. If patients' data is not adequately protected or personally identifiable information is inadvertently revealed during the data processing and analyzing, personal information could be misused, potentially violate and compromise patient privacy and breach confidentiality [Zarifis et al. 2020].

Patient eligibility for treatment based on AI-oriented recommendations in healthcare and patient understating of AI rationale behind the recommendation is another ethical issue that cannot be unnoticed [Liu et al. 2021]. Informed consent, and patient autonomy ensure patient clearly understand the risks, benefits, and alternative of receiving medical intervention from AI algorithms clinical devices [Young 2010], but in practice, patients rarely comprehend the underlying technology, such as medical imaging, and healthcare operations and resource management due to the complexity and opacity of algorithms, resulting in difficulties to retain control and agency in their healthcare choices [Välimäki et al. 2000].

Bias in decision-making and algorithms and lack of evidence and transparency is a severe ethical issue when applying artificial intelligence in healthcare domain, especially in the utilization of predictive algorithms in determining patient treatment plans [Parikh et al. 2019]. Patients' historical data, including demographics, medical history, and treatment outcome have been trained on AI algorithms, However, a specific demographic group, such as white guys, were primarily represents, while darker females' group may inadvertently be ignored by flawed algorithms, resulting in underrepresented groups may receive suboptimal care [Celi 2022].

1.2.2 Risks associate with AI ethical issues in healthcare

Specific risk associated with AI ethical issues in healthcare manifest misdiagnosis and treatment errors caused by overreliance on AI algorithms and lack of human oversight and responsibility [Angioi et al. 2014].

AI-powered diagnosing machine was tasked to identify patient's condition based on various previous symptoms and historical test results; however, the AI algorithm may not adequately account for complex nature of healthcare, rare and atypical case, leading to misdiagnosis and inappropriate or incorrect treatment if healthcare professionals solely rely on AI [Harada et al. 2020].

Another strong example of risks associate with AI ethical issues in healthcare is regulatory challenges and negligence legal liability [Miller and Miller 2007]. Misleading diagnoses for a significant number of patients lead to unnecessary procedures or even harm to patients, in such cases, compare to traditional legal frameworks, how to determine responsibility and liability becomes complex due to involvement of AI. Additionally, rapid advancements in AI Technology and continuously evolve poses challenges in terms of establishing guidelines and standard for AI deployment in healthcare [Guerra et al. 2021].

2 REVIEW CURRENT APPROACHES

2.1 Current Approaches to ensure AI ethics

2.1.1 Ethic Guidelines for trustworthy artificial intelligence

Guidelines and principles outline the core principles such as fairness, privacy, transparency, and accountability that should be upheld to ensure AI ethics. For instance, The EU approach to Ethic Guidelines for trustworthy artificial intelligence [Smuha 2019], The Ethical Guidelines of the Japanese Society for Artificial Intelligence (JSAI), Korea Institute for Robot Industry Advancement (KIRIA) published a Robots Ethics Charter, which was further revised in 2016 [Gal 2019], which are designed to be adaptable and compatible to different contexts and across various domains, establishing an ethical foundation in AI deployment, raising awareness about potential ethical risk associated with AI, and fostering better understating toward AI ethical considerations [Nguyen 2022].

The guidelines and principles are highly relevant to healthcare as it provides a compressive framework to ensure AI systems could be used in a ethical and patient-centric manner.

However, lack of concrete instructions, specific methodologies [Klontzas 2023], and robust measures can be challenging to be interpreted compliance, making it difficult to insure consistent between ideal principles and practices [Ponte 2021].

2.1.2 Data-centric technical approaches to AI Ethics by Design

Technical approaches aim to embed values in AI systems, anticipate ethical and epistemic issues early in the development stages, and necessitates careful curation of training datasets to ensure AI algorithmic ethics [D'Aquin et al. 2018]. For instance, AI Value Sensitive Design (AVD) [Gerdes 2021], a participatory data-centric driven approach to AI ethics as a manageable remedy to mitigate ethical issues in AI system, providing targeted solutions, tangible implementation, concrete method, practical techniques, and frameworks that can be implemented directly by AI developers in the process of development and enable a measurable assessment of AI system to achieve specific ethical goals [Wong et al. 2020].

Healthcare domain thoroughly evaluate and validate AI to ensure accurate and consistent performance in clinical setting by technical approaches, particularly important in treatment planning and patient monitoring.

However, ethical dilemmas in AI based on purely technical approaches typically are context-driven and normally involve inherent varying values, requiring the reestablishment of interdisciplinary effort from diverse perspectives to design broader AI system [Nurock et al. 2021].

2.1.3Regulatory and policy measure and Stakeholders Engagement

The legal and government framework aiming to address ethical issues in AI, implemented by governments and organizations, such, a lawful processing, GDPR Principles, outlining the principles around transparency, data minimisation, accuracy, storage limitation, security, and accountability, ensuing responsible in AI applications [Milossi et al. 2021].There two are critical components of effective healthcare governance and work together to contribute to patient-cantered healthcare services.

However, emerging ethical challenges posed by advancing AI technologies can be changeling due to the rapid pace of technological advancements that outpaces the establishment of ethical regulations related to AI. Additionally, Regulatory standard across different state and counties, make it hard to remain consistent on a global scale and international cooperation level [Ray 2023].

2.2   Key projects that sought to address AI ethical issues in healthcare
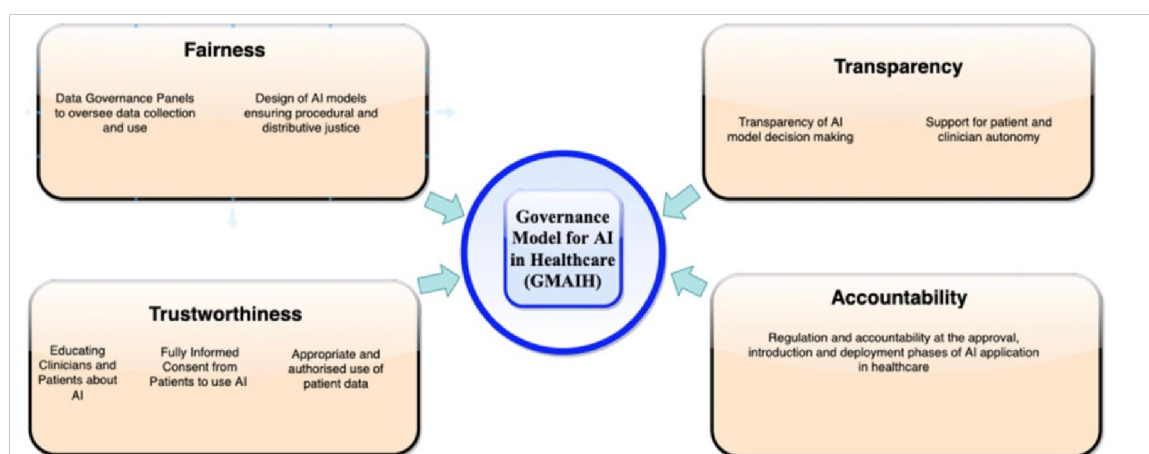
2.2.1Governance Model for AI in Healthcare (GMAIH)

Figure 1: Outline of the Governance Model for Artificial Intelligence in Health Care

A key project that presents a responsible governance model aiming to address the ethical and regulatory challenges for the use of Artificial intelligence in AI in healthcare was proposed to Sandeep Reddy and Sonia Allan in 2020.It provides guidelines and standards for AI deployment and monitoring by taking into account data privacy, data protection, algorithmic transparency, accountability and patient safety. Additionally, it highlights the importance of inveigling participation from diverse stakeholder, including healthcare professionals, AI developers, patient representative and policymakers when model is applied to

governance AI ethical issues. Ongoing monitoring and continuous evaluations of AI systems in healthcare setting was emphasized in order to assess effectives, safety and adherence to ethical principles during the decision-making processes, ultimately striking a balance between applying innovation and ensuring ethical deployment of AI in healthcare applications and maximizing the benefits from AI technologies [Reddy et al. 2019b].

2.2.2 Ethical AI principle in healthcare setting

Table 1: Ethical principles for Artificial Intelligence in Health Care

Human value [22]	Ethical principle [14]
Security (safety, harmony, and stability of society, of relationships, and of self)	Non-maleficence (protection from harm, precaution, prevention, non-subversion)
Self-direction (independence in thought and action; creating, exploring, being curious)	Freedom and autonomy (consent, choice, self-determination, liberty, empowerment)
	Dignity
	Privacy (protection of personal or private information)
Benevolence (preserving and enhancing others' welfare, voluntary concern for others' welfare)	Beneficence (benefits, well-being, peace, social good, common good)
	Responsibility (accountability, liability, acting with integrity)
	Trust
	Transparency (explainability, understandability, interpretability, acts of communication and disclosure)
	Solidarity (social security and cohesion)
Universalism (understanding, appreciation, tolerance, and protection for the welfare of all people and for nature)	Justice and fairness (consistency, inclusion, equality, equity, non-discrimination, respect for diversity, plurality, accessibility, redress)
	Sustainability (conserving environment and natural resources)

Key human values play a crucial role in guiding AI ethical considerations, particularly in healthcare setting. Corresponding ethical principles based on human values was discussed by Pravik Solank, John Grundy, Waqar Hussian in 2022.For instance, security focuses on the safety, harmony, the stability of society and relationship, and corresponding principles involve minimizing harm and avoid bias and discrimination, requiring AI developers to reduce potential harms caused by technical failures and risk of violation. Similarly, Benevolence refers to enhance and preserve the welfare of oneself and others, the ethical AI principle is tasked to emphasize the solidarity, beneficence, and responsibility, ensuring the well-being of individual and mitigate disparities in AI application related to healthcare domain [Solanki et al. 2022]

Universalism encompasses appreciation for the welfare of all planet and people, ensuring to represent diverse populations and minimize the ecological footprint of AI initiatives. By considering these human values, developers, and users of AI in heartcare can promote inclusivity when safeguarding patient well-being.

3 REVIEW APPROACHES TO ETHICAL EVALUATIONS SYSTEMS

3.1 Approaches to ethical evaluations of systems

3.1.1 Auditing algorithmic decision-making system (ADS)

The purpose of ethical matrix for auditing algorithmic decision-making systems (ADS) is to assess the underlying assumptions that are based on pseudoscientific level of human nature and ensure AI ethical evaluations systems. The SocioTechnical Matrix examines the goals, data, function, assumptions of Auditing ADS .By containing the matrix of different stage of the hiring funnel, such as " heretual " for matching candidates' profiles can be matched to job fit , "codility "for assessing a predictor of job skills, "pymetrics" for analysing gameplay scores, it serves as a tool to be used by auditors, evaluaters to assess the transparency and accountability in using ADS in hiring domain [Waltl and Vogl 2018].

Applying Auditing algorithmic decision-making systems (ADS) used in healthcare becomes possible. For instance, assess AI algorithmic ethics performance included in ADS could be used to examine ethical evaluations systems' performance across

different patient populations, ensuring fairness and avoid disparate impacts. The requirement of evaluating transparency and explainability of the ADS could be interpreted by healthcare professionals and patients [Koreff et al. 2021].

However, this socio-technical Matrix primarily was considered an incomplete assessment due to its focus on ethical consideration in hiring domain and a wider range of ethical dimensions and refinement is required to design a more comprehensive assessment framework [Matias et al. 2022].

3.1.2 ALTAI Audit

Assessment List for Trustworthy Artificial Intelligence was designed to evaluate the compliance of the trustworthy AI system with 7 requirements, including human agency and oversight, technical robustness and safety, privacy, and data governance, transparency, diversity, non-discrimination and fairness, environmental and societal well-being, and accountability. It has considered to be the structured, flexible, sustainable, and systematic approach by High-Level Expert Group on AI, providing a comprehensive set of requirements that covers diverse dimension of trustworthy AI [Ammanath 2022].

ALTAI framework relates to healthcare domain manifesting in several ways. In healthcare domain, it is crucial to ensure that AI systems are developed to assistant human healthcare professionals instead of replacing them, which is compactable with the requirement of human agency and oversight [Rajamaki et al. 2022].

However, The ALTAI lack specificity and fail to focus on context-specific consideration making it challenging to deployment in the different domains. Additionally, applying ALTAI audit requires intensive resources and expertise, led to difficult to conduct for some smaller companies with limited resources [Ala-Pietilä et al. 2020].

3.1.3 Clinical audit ---COIIN

The COIN project focused on developing evidence-based guidelines on clinical core data and creating computer workstations tailored for efficient data capture. By setting standards and guidelines that can be adopted by the National Health Service (NHS), enabling healthcare professionals to assess technical practices against benchmarks. Although it is may not perfectly applied in nowadays artificial intelligence, advocating collaboration with specialty working groups and assessing innovative technologies in healthcare initiative can be drawn on [Karp et al. 1995].

3.1.4 Deloitte AI ethical audit

The audit of the trustworthy AI framework was developed by Deloitte to guide organizing in designing and deploying AI systems in a responsibly an ethically level, providing actionable recommendations to mitigate ethical risk, such as fair and impartial use checks to circumvent bias, implementing transparency and explainable AI to ensure justification for AI-driven decision and AI algorithms, safeguarding privacy, monitoring for reliability, and putting proper security in place.

3.2   Clear need for standard ethical audit in healthcare

There is a certainly clear need for standard ethical audit of AI in healthcare, as ethical audits play a significant role in evaluating AI systems and corresponding impact on healthcare providers, patients, and whole society, upholding professional and ethical standards, specifically for healthcare professionals have ethical obligation to offer high-standard service to patients. By conducting standard ethical audit of AI systems in healthcare, healthcare regulator can provide comprehensive framework to evaluate ethical implication of AI systems in healthcare domain and promote societal values. [Scott et al. 2020]. **4 DESGIN AN ETHICAL AUDIT**

4.1   Key considerations to guide design process

Careful consideration of design an AI ethics include defining the purpose and scope to clarify the specific objectives and scope of AI ethics, identifying primary focus areas and domains, composition, and expertise to ensure diverse representation, define the roles and responsibilities, establish a clear governance structure, foster transparency, and promote continuous

learning, establish a periodic review process, recognize the AI systems adaptability and flexibility. Additionally, specific design varies based on the organization's size and industry's unique needs [Fagerberg et al. 2014].

4.2   Ethical Audit for healthcare AI systems

According to the AI system used by the organization shown in Figure 2, The ethical audit that tailored to healthcare based on the seven categories of the ALTAI framework has been design in a high-level principal manner.
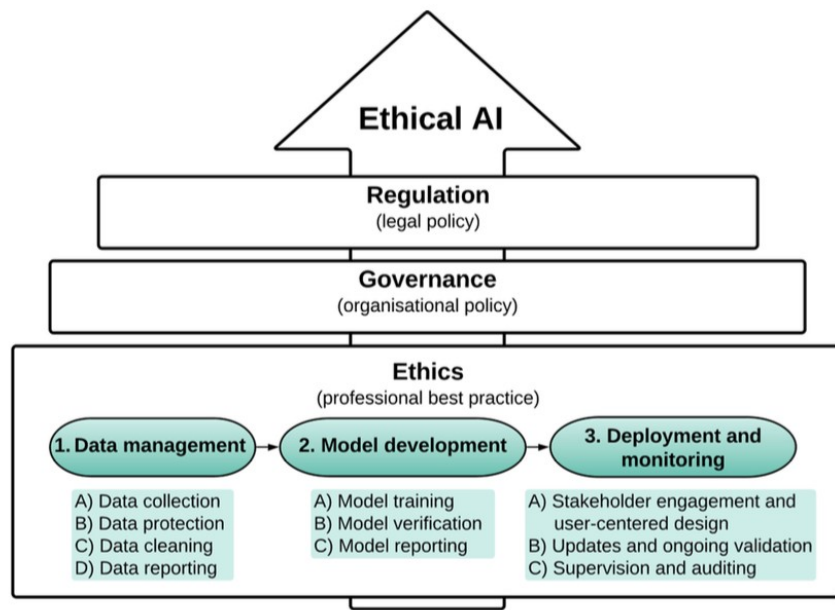

Figure 2: Artificial Intelligence Systems in Health Care

4.2.1Human Agency and oversight in healthcare

Persevering human control and understanding in healthcare decision-making process contains whether healthcare professionals and patients could actively engage with AI systems and determine human agency is preserved in AI-decision making systems [Floridi et al. 2018]. Clearly communication to healthcare professionals and patients to ensure voice of stakeholders are considered, helping to weight and balance the benefits and risks of AI systems and prioritizing the retention of autonomy [Butz 2013].

4.2.2Technical Robustness and safety in healthcare

Assessing technical robustness and reliability including evaluate the AI system's capabilities to handle complex healthcare scenarios and ensure reliable, accurate and dependable results, safeguarding the integrity of patient healthcare data [Cook 2000]. Rigorous testing to verify and validate the systems outputs against the benchmarks. Monitoring and maintaining AI system in healthcare can track system performance and safety over time [Terry 2019].

4.2.3Privacy and data governance in healthcare

Conduct governance policies to ensure healthcare data used in AI system is handled in compliance with the general data protection regulation in the European union, obtaining informed consent from patients, proper anonymization, and encryption to protect patient's data confidentiality, reduce the risk of personally identifiable information during transmission, adhering to legal requirements and mitigate unauthorized access [Omar et al. 2019].

4.2.4 Transparency in healthcare AI systems

A clear explanations of AI algorithm and decision-making process, documenting and disclosing information about the model architecture and user-friendly interfaces and visualization that help patients and healthcare providers to interpret the output offered by AI system should be considered to ensure transparency [Ehsan et al. 2021]. By workshops, feedback sessions, interactive dashboard to present an intuitive manner result [Kiseleva et al. 2022].

4.2.5 Diversity, Non-discrimination, and Fairness in healthcare AI systems

Considering the needs preferences, and characteristics of diverse patient populations, identifying bias that may perpetuate existing disparities or disadvantage certain individual, providing equitable healthcare outcomes for all individuals, including for someone with disability or have limitation in accessing technology [Cobianchi et al. 2022].

4.2.6 Environmental and societal wee-being in healthcare AI systems

Energy consumption, resource usage associated with AI infrastructure should be considered to ensure sustainability of healthcare practices. The implications of AI systems on patient-provider relationships to ensure healthcare equity to enhance patient experience. Tackle public health issues and optimization of healthcare operations also should be considered to promote the positive impact of AI on public health [Matos 2021].

4.2.7 Accountability in healthcare AI systems

Accountability mechanisms and culture of accountability involve clearly allocating responsibilities and reporting structures to different stakeholders, regularly conducting evaluation independent audits to assess the healthcare setting outcomes and patient safety, ensuring avenues for recourse in case of negative impacts posed by AI systems in healthcare [Habli et al. 2020].

Conclusion

AI ethical issues in general and particularly in healthcare setting, various approaches proposed to reduce ethical bias, ethical evaluation of AI system, the design of an ethical audit specifically for healthcare AI domain have been discussed in this report, enable healthcare organizations to assess the ethical implication of AI system comprehensively.

In conclusion, multifaced approaches involving collaboration between healthcare professionals and policymakers are necessarily required to address ethical issues. Only through such effort can we strive toward a future where AI systems serve the best of society at large.

REFERENCES

[1]   Matsuzaki, T. (2018). Ethical issues of artificial intelligence in medicine. Cal. WL Rev., 55, 255.

[2]   Coeckelbergh, M. 2019. Artificial Intelligence: Some ethical issues and regulatory challenges. *DOAJ (DOAJ: Directory of Open Access Journals)*.

[3]   Onik, M.H., Kim, C.S., and Yang, J. 2019. Personal Data Privacy Challenges of the Fourth Industrial Revolution.

[4]   Kim, B., Park, J.-S., and Suh, J. 2020. Transparency and accountability in AI decision support: Explaining and visualizing convolutional neural networks for text information. *Decision Support Systems 134*, 113302.

[5]   Nelson, G.A. 2019. Bias in Artificial Intelligence. *North Carolina Medical Journal 80*, 4, 220–222.

[6] Calvo, R.A., Peters, D., Vold, K., and Ryan, R.M. 2020. Supporting Human Autonomy in AI Systems: A Framework for Ethical Enquiry. In: *Philosophical studies series*. Springer International Publishing, 31–54.

[7] Kuzlu, M., Fair, C., and Guler, O. 2021. Role of Artificial Intelligence in the Internet of Things (IoT) cybersecurity. *Discover Internet of Things 1*, 1.

[8] Abuselidze, G. and Mamaladze, L. 2021. The impact of artificial intelligence on employment before and during pandemic: A comparative analysis. *Journal of Physics 1840*, 1, 012040.

[9] Mahmud, H., Islam, A.K.M.N., Ahmed, S.H., and Smolander, K. 2022. What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change 175*, 121390.

[10] Johnson, K.P., Wei, W.I., Weeraratne, D., et al. 2020. Precision Medicine, AI, and the Future of Personalized Health Care. *Clinical and Translational Science 14*, 1, 86–93.

[11] Bur, A.M., Shew, M.L., and New, J. 2019. Artificial Intelligence for the Otolaryngologist: A State-of-the-Art Review. *Otolaryngology-Head and Neck Surgery 160*, 4, 603–611.

[12] Singh, M. 2018. Revolutionizing Healthcare through Artificial Intelligence and Virtual Reality. *International Journal of Education and Management Studies 8*, 4, 406.

[13] Santosh, K.C. and Gaur, L. 2021. Introduction to AI in Public Health. In: *Springer eBooks*. 1–10.

[14] Curtis, C. and Brolan, C.E. 2022. Health care in the metaverse. *The Medical Journal of Australia 218*, 1, 46.

[15] Kuiler, E.W. and McNeely, C.L. 2023. Panopticon implications of ethical AI: equity, disparity, and inequality in healthcare. In: *Elsevier eBooks*. 429–451.

[16] Nizam, V. and Aslekar, A. 2021. Challenges of Applying AI in Healthcare in India. *Pharmaceutical Research International*, 203–209.

[17] Goralski, M.A. and Tan, T.K. 2023. Artificial Intelligence: Poverty Alleviation, Healthcare, Education, and Reduced Inequalities in a Post-COVID World. In: *Springer eBooks*. 97–113.

[18] Zhou, X.-Y., Snoswell, C.L., Harding, L., et al. 2020. The Role of Telehealth in Reducing the Mental Health Burden from COVID-19. *Telemedicine Journal and E-health 26*, 4, 377–379.

[19] Liu, R., Rizzo, S., Whipple, S., et al. 2021. Evaluating eligibility criteria of oncology trials using real-world data and AI. *Nature 592*, 7855, 629–633.

[20] Young, R.H. 2010. Informed Consent and Patient Autonomy. In: *Wiley-Blackwell eBooks*. 530–540.

[21] Välimäki, M., Dassen, T., and Arndt, M. 2000. *Patient's Autonomy, Privacy and Informed Consent*. .

[22] Murdoch, B. 2021. Privacy and artificial intelligence: challenges for protecting health information in a new era. *BMC Medical Ethics 22*, 1.

[23] Reddy, S.K., Allan, S., Coghlan, S., and Cooper, P.R. 2019. A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association 27*, 3, 491–497.

[24] Zarifis, A., Kawalek, P., and Azadegan, A. 2020. Evaluating If Trust and Personal Information Privacy Concerns Are Barriers to Using Health Insurance That Explicitly Utilizes AI. *Journal of Internet Commerce 20*, 1, 66–83.

[25] Parikh, R.B., Teeple, S., and Navathe, A.S. 2019. Addressing Bias in Artificial Intelligence in Health Care. *JAMA 322*, 24, 2377.

[26] Celi, L.A., Cellini, J., Charpignon, M., et al. 2022. Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review. *PLOS Digital Health 1*, 3, e0000022.

[27] Angioi, K., Terrada, C., Locatelli, A., Hoang, P.L., Béné, M.C., and Bodaghi, B. 2014. Ocular Manifestations of X-linked Chronic Granulomatous Disease: About Two Atypical Case Reports. *Ocular Immunology and Inflammation*.

[28] Harada, T., Shimizu, T., Kaji, Y., et al. 2020. A Perspective from a Case Conference on Comparing the Diagnostic Process: Human Diagnostic Thinking vs. Artificial Intelligence (AI) Decision Support Tools. *International Journal of Environmental Research and Public Health 17*, 17, 6110.

[29] Miller, R.A. and Miller, S. 2007. Legal and regulatory issues related to the use of clinical software in health care delivery. In: *Elsevier eBooks*. 423–444.

[30] Guerra, A., Parisi, F., and Daniel, P. 2021. Liability for robots I: legal challenges. *Journal of Institutional Economics*, 1–13.

[31] Gal, D. 2019. Perspectives and Approaches in AI Ethics: East Asia. *Social Science Research Network*.

[32] Smuha, N.A. 2019. The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence. *Computer Law Review International 20*, 4, 97–106.

[33] Nguyen, A., Ngo, H.N., Hong, Y., Dang, B., and Nguyen, B.-P.T. 2022. Ethical principles for artificial intelligence in education. *Education and Information Technologies 28*, 4, 4221–4241.

[34] Klontzas, M.E., Gatti, A.A., Tejani, A.S., and Kahn, C.E. 2023. AI Reporting Guidelines: How to Select the Best One for Your Research. *Radiology 5*, 3.

[35] Ponte, E.B., Lavorato, D., Lamboglia, R., and Mancini, D. 2021. Artificial intelligence activities and ethical approaches in leading listed companies in the European Union. *International Journal of Accounting Information Systems 43*, 100535.

[36] D'Aquin, M., Troullinou, P., O'Connor, N.E., Cullen, A., Faller, G., and Holden, L. 2018. *Towards an "Ethics by Design" Methodology for AI Research Projects*.

[37] Gerdes, A. 2021. A participatory data-centric approach to AI Ethics by Design. *Applied Artificial Intelligence 36*, 1.

[38] Wong, R.Y., Boyd, K.S., Metcalf, J., and Shilton, K. 2020. *Beyond Checklist Approaches to Ethics in Design*. .

[39] Nurock, V., Chatila, R., and Parizeau, M.-H. 2021. What Does "Ethical by Design" Mean? In: *Lecture Notes in Computer Science*. 171–190.

[40] Milossi, M., Alexandropoulou-Egyptiadou, E., and Psannis, K.E. 2021. AI Ethics: Algorithmic Determinism or Self-Determination? The GPDR Approach. *IEEE Access 9*, 58455–58466.

[41] Ray, P.P. 2023. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-physical Systems 3*, 121–154.

[42] Reddy, S.K., Allan, S., Coghlan, S., and Cooper, P.R. 2019b. A governance model for the application of AI in health care. *Journal of the American Medical Informatics Association 27*, 3, 491–497.

[43] Solanki, P., Grundy, J., and Hussain, W. 2022. Operationalising ethics in artificial intelligence for healthcare: a framework for AI developers. *AI And Ethics 3*, 1, 223–240.

[44] Koreff, J., Weisner, M.M., and Sutton, S.G. 2021. Data analytics (ab) use in healthcare fraud audits. *International Journal of Accounting Information Systems 42*, 100523.

[45] Matias, J.N., Hounsel, A., and Feamster, N. 2022. Software-Supported Audits of Decision-Making Systems: Testing Google and Facebook's Political Advertising Policies. *Proceedings of the ACM on Human-computer Interaction 6*, CSCW1, 1–19.

[46] Waltl, B. and Vogl, R. 2018. Increasing Transparency in Algorithmic- Decision-Making with Explainable AI. *Datenschutz Und Datensicherheit - Dud 42*, 10, 613–617.

[47] Deloitte Introduces Trustworthy AI Framework to Guide Organizations in Ethical Application of Technology in the Age of With. *Deloitte United States*. https://www2.deloitte.com/us/en/pages/about-deloitte/articles/press-releases/deloitte-introduces-trustworthy-ai-framework.html.

[48] Karp, S.J., Squire, C.J., and Sizer, B. 1995. 166 National comparative audit in clinical oncology: The royal college of radiologists' coin project. *European Journal of Cancer*.

[49] Rajamaki, J., Rocha, P.A.L., Perenius, M., and Gioulekas, F. 2022. *SHAPES Project Pilots' Self-assessment for Trustworthy AI.* .

[50] Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. 2020. *Shaping Europe's Digital Future*. https://digitalstrategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment.

[51] Ala-Pietilä, P., Bonnet, Y., Bergmann, U., et al. 2020. *The Assessment List for Trustworthy Artificial Intelligence (ALTAI)*. European Commission.

[52] Ammanath, B. 2022. Trustworthy AI: A Business Guide for Navigating Trust and Ethics in AI. John Wiley & Sons.

[53] Fagerberg, L., Hallström, B.M., Oksvold, P., et al. 2014. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & Cellular Proteomics 13*, 2, 397–406.

[54] Scott, A.M., Kolstoe, S., Ploem, M.C., Hammatt, Z., and Glasziou, P. 2020. Exempting low-risk health and medical research from ethics reviews: comparing Australia, the United Kingdom, the United States and the Netherlands. *Health Research Policy and Systems 18*, 1.

[55] Smuha, N.A. 2019b. The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence. *Computer Law Review International 20*, 4, 97–106.

[56] Floridi, L., Cowls, J., Beltrametti, M., et al. 2018. AI4People—An Ethical Framework for a Good AI Society: Opportunities, risks, principles, and recommendations. *Minds and Machines 28*, 4, 689–707.

[57] Butz, S. 2013. Health care comes home: the human factors. *International Journal of Integrated Care 13*, 4.

[58] Cook, R.I. 2000. Gaps in the continuity of care and progress on patient safety. *BMJ 320*, 7237, 791–794.

[59] Terry, N.P. 2019. Of regulating healthcare AI and robots. *Social Science Research Network*.

[60] Omar, A.A., Bhuiyan, Z.A., Basu, A., Kiyomoto, S., and Rahman, M.M. 2019. Privacy-friendly platform for healthcare data in cloud based on blockchain environment. *Future Generation Computer Systems 95*, 511–521.

[61] Kiseleva, A., Kotzinos, D., and De Hert, P. 2022. Transparency of AI in healthcare as a multilayered system of accountabilities: Between legal requirements and technical limitations. *Frontiers in Artificial Intelligence 5*.

[62] Ehsan, U., Liao, Q.V., Muller, M., Riedl, M.O., and Weisz, J.D. 2021. *Expanding Explainability: Towards Social Transparency in AI systems*. .

[63] Cobianchi, L., Verde, J.M., Loftus, T.J., et al. 2022. Artificial intelligence and Surgery: ethical dilemmas and open issues. *Journal of the American College of Surgeons 235*, 2, 268–275.

[64] Matos, F. 2021. ECIAIR 2021 3rd European Conference on the Impact of Artificial Intelligence and Robotics. Academic Conferences and publishing limited.

[65] Habli, I., Lawton, T., and Porter, Z. 2020. Artificial intelligence in health care: accountability and safety. *Bulletin of the World Health 98*, 4, 251–256.

APPENDICES

Assessment List for Trustworthy Artificial Intelligence

Category		Recommendations
Human agency and oversight	1.	To avoid end-users' full trust in AI systems.
	2.	To avoid that, the system inadvertently affects human autonomy.
	3.	To provide training to exercise oversight (in-the-loop, on-the-loop, in-command).
	4.	To clarify all potential negative consequences for end-users or targets (e.g., develop attachments).
	5.	To provide means for end-users to have control of the interactions and preserve autonomy.
	6.	To have means to reduce the risk of manipulation (clear information about ownership and aims).
	7.	To establish detection and response mechanisms in case of undesirable effects for the end-users.
	8.	To establish control measures that reflect the self-learning/autonomous nature of the system.
	9.	To involve experts from other disciplines, such as psychology and social work.
Technical robustness and safety	1.	To assess risks of attacks to which the AI system could be vulnerable.
	2.	To assess AI system threats and their consequences (design, technical, environmental, human).
	3.	To assess the risk of possible malicious use, misuse, or inappropriate use of the AI system.
	4.	To assess the dependency of the critical system's decisions on its stable and reliable behavior.
	5.	To control the AI system changes and its technical robustness and safety permanently.
Privacy and data governance	1.	To adopt mechanisms that flag privacy and data protection issues.
	2.	To implement the rights to withdraw consent, object, and be forgotten in the AI systems.
	3.	To protect privacy and personal data during the lifecycle of an AI system (data processing).
	4.	To protect non-personal data during the lifecycle of an AI system (data processing).
	5.	To align the AI system with widely accepted standards (e.g., ISO) and protocols.
Transparency	1.	To continuously survey users about their decisions and understanding of AI systems.
	2.	To continuously assess the quality of the input data to the AI system.
	3.	To explain to the end-users the AI system decisions or suggestions (answers).
	4.	To explain to the end-users that AI system is an interactive machine (that he/she communicates with).
Diversity, non-discrimination, and fairness	1.	To teach/educate the AI system developers about potential system bias.
	2.	To implement fair AI systems and be sensitive to the variety of preferences/abilities in society.
	3.	To build accessible AI systems and interfaces for all people (Universal Design principles).
	4.	To assess the AI systems' disproportional impacts considering individuals and groups.
	5.	To assess the AI systems' bias related to algorithm design (data inputs) permanently.
	6.	To build algorithms that include diversity and representativeness of individuals and groups.
	7.	To assess permanently the AI systems' bias related to discrimination (e.g., race, gender, age).
	8.	To adopt mechanisms to identify subjects (in) directly affected by the AI system.
	9.	To adopt mechanisms that flag diversity, non-discrimination, and fairness issues.
	10.	To adopt mechanisms to continuously measure the risk of bias.
	11.	To provide AI systems with widely accepted definitions, concepts, and frameworks.
	12.	To involve or consult the end-users in all phases of AI system development.
	13.	To provide publicly available educational materials based on research and state of the art.
	14.	To assess "Conflicts of Interest" of the team/individuals involved in building the AI system.
Societal and environmental well-being	1.	To adopt mechanisms to identify AI systems' positive/negative impacts on the environment.
	2.	To define measures to reduce the environmental impact of AI system's lifecycle.
	3.	To involve the AI systems to tackle societal, environmental, and well-being problems.
	4.	To reduce the AI systems' negative impacts to the work and workers.
	5.	To provide people with re-skill educational tools to counteract de-skilling based on AI systems.
	6.	To ensure people understand the AI systems' positive/negative impacts very well.
Accountability	1.	To ensure AI systems' auditability, modularity, and traceability (also by third parties).
	2.	To fit the best practices and industry standards available and acknowledged.
	3.	To ensure that all conflicts of values or tradeoffs be well-documented and explained.
	4.	To include a non-technical method to assess the trustworthiness of AI (e.g., "ethical review board").
	5.	To consistently provide multisectoral and multidisciplinary auditing or guidance.
	6.	To have and update the legal framework considering a wide range of impacts.
	7.	To assess vulnerabilities and risks to identify and mitigate potential pitfalls.