The Double-Edged Sword: Ethical Implications of AI in Healthcare Decision-Making

Sanket Bhujbal
sanketb@vt.edu
Virginia Tech
Blacksburg, VA, USA

Sindhura Kommu
sindhura@vt.edu
Virginia Tech
Blacksburg, VA, USA

Robin Lu
robinlu@vt.edu
Virginia Tech
Blacksburg, VA, USA

Ritish Shailly
rshailly@vt.edu
Virginia Tech
Blacksburg, VA, USA

Alok Mehandale
alokm@vt.edu
Virginia Tech
Blacksburg, VA, USA

ABSTRACT

Artificial intelligence (AI), particularly in the form of Biomedical Large Language Models (LLMs) [13, 19], has revolutionized healthcare and bio-medicine, offering immense potential for improved diagnostics, personalized treatment plans, drug discovery and streamlined medical processes. However, along with these benefits come significant ethical challenges that demand careful consideration. The urgency to tackle these ethical challenges is underscored by AI's rapid integration into healthcare, with profound implications for societal norms and individual rights. This report aims to spotlight these issues, advocating for immediate action to formulate ethical guidelines and safeguards. This report explores the "double-edged sword" nature of AI in healthcare decision-making, highlighting both its potential benefits and inherent ethical dilemmas.

KEYWORDS

Artificial intelligence (AI); Large Language Models (LLMs); Healthcare decision-making; Informed consent; Equitable healthcare; Biomedical AI.

ACM Reference Format:
Sanket Bhujbal, Sindhura Kommu, Robin Lu, Ritish Shailly, and Alok Mehandale. 2024. The Double-Edged Sword: Ethical Implications of AI in Healthcare Decision-Making. In *Proceedings of Blacksburg, VA (CS 5024, Spring 2024).* ACM, New York, NY, USA, 7 pages. https://doi.org/XXXXXXX.XXXXXXX

CS 5024, Spring 2024,

,

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

As AI becomes more deeply integrated into various sectors, including healthcare and biomedicine [18], understanding its ethical implications becomes increasingly crucial. While society appears to embrace this evolution, it is essential to critically examine its ethical implications for both our current generation and future ones, as it could significantly impact the direction of research and influence public perception.

AI models such as Large language models (LLMs) have swiftly become integral to our daily lives. They can process enormous medical literature, suggest diagnoses, identify therapeutic targets in drug discovery and aid decision-making. However, trusting their results is a major ethical concern due to Factual Errors and Misinformation, Lack of Transparency (Black Box Problem) and Potential for Bias among other things.

Nevertheless, there is a clear movement towards implementing biomedical AI models in healthcare and biomedicine for purposes like patient care and predicting disease progression. If this trend continues, future societies might exist in a world where AI is present in many aspects of life, leading to intricate ethical challenges. These challenges include issues of transparency, accountability, fairness, and how AI affects human autonomy and decision-making. It will be crucial for policymakers, researchers, and industry leaders to address these concerns to ensure that AI development and use bring positive outcomes to society.

We explore important ethical considerations in using biomedical AI, including legal aspects, biases and disparities in care, privacy risks, explainability, challenges in securing informed consent, and the risk of excessive reliance on AI recommendations. Through analyzing these issues and their potential impacts, we aim to highlight the urgent need for strong

ethical frameworks and responsible implementation strategies. We stress the importance of ongoing discussion and collaboration among healthcare professionals, AI developers, policymakers, and the public to ensure that AI is a positive force, promoting fair and ethical healthcare for everyone. Our work emphasizes the necessity of trustworthy and responsible use of biomedical AI models in healthcare and bio-medicine decision-making.

2 LITERATURE REVIEW

2.1 Legal Considerations

AI has quickly become a disruptive force in healthcare, presenting previously unseen possibilities to improve patient outcomes, optimize clinical procedures, and progress medical research. But integrating AI into biological and healthcare systems also raises some difficult legal issues that need to be carefully considered and supervised by regulators.

Regulatory Work

The legislative frameworks controlling data privacy, security, and medical device approval are among the most important legal factors in the application of AI in healthcare. The Food and Drug Administration (FDA) is one of the regulatory bodies in the United States for medical devices, which includes software used in healthcare settings. Before medical products, including AI algorithms, are offered to consumers and healthcare professionals, the FDA's premarket approval procedure makes sure they meet safety guidelines. Similarly, some strict guidelines for the gathering, storing, and processing of patient data are imposed by data privacy laws like the Health Insurance Portability and Accountability Act (HIPAA) [4] in the US and the General Data Protection Regulation (GDPR) [15] in the EU. These rules apply to AI-driven healthcare apps in order to protect patient confidentiality and privacy.

The ever-changing nature of AI presents obstacles for regulatory bodies charged with assessing them. Novel healthcare AI solutions may experience some delays in approval and market access due to traditional regulatory procedures' potential incompatibility with the fast pace of AI innovation. Regulatory agencies must use flexible strategies that are suited to the particular features of AI technology in order to strike the right balance between encouraging innovation and guaranteeing patient safety.

Liability

Liability concerns arise when AI is included into important decision-making procedures and leads to unfavorable results or mistakes. In contrast to normal medical procedures, in which healthcare personnel are accountable for patient care, AI systems add some levels of complexity. It is important here to carefully consider a number of different elements when identifying who is responsible for AI-generated errors, including the AI system's design and implementation, the caliber of the input data, and the decisions made by human operators in response to the system's recommendations.

Legal liability may extend to various stakeholders [6] involved in the development, deployment, and use of AI in healthcare, including:

- Manufacturers and Developers: Organizations and scholars in charge of creating AI algorithms may be held liable for mistakes or negative outcomes that their systems generate.
- Healthcare providers: Physicians who make decisions about patient treatment based on AI recommendations could be held responsible for the results, especially if they don't take the necessary steps to confirm the outputs.
- Regulatory bodies: If the government organizations in charge of regulating AI-based medical devices don't discover and reduce the risks connected with these technologies, there may be grounds for concern.

In order to navigate liability issues, all stakeholders should clearly define their roles and responsibilities, communicate AI's capabilities and limitations in a transparent manner, and have strong processes in place for tracking safety concerns through the product lifecycle.

Intellectual Property

Complex IP issues relating to algorithms, data ownership, and licensing are brought up by the development of AI algorithms in biomedical and healthcare settings. AI algorithms are valuable assets that, depending on their uniqueness, usefulness, and financial relevance, may be protected by a patent, trade secret, or copyright law [17].

Because the discoveries are abstract and algorithmic in nature, patent protection for AI algorithms poses special difficulties. When determining whether AI inventions can be patentable, courts and patent offices have to consider issues of enablement, non-obviousness, and patent eligibility and because AI research and development is collaborative in nature, it usually incorporates contributions from a number of people and organizations, which raises questions about joint ownership rights.

Apart from the AI algorithms themselves, data produced by AI systems could also be protected by IP rights. This is especially true when private datasets are utilised to train machine learning models. In AI research collaborations and commercial partnerships, the rights and duties of the parties involved are determined in large part by data ownership and access agreements.

The use of AI algorithms and associated IP are governed by technology licensing agreements. Enforcable and transparent licensing agreements safeguard the rights of licensees and IP owners while promoting cooperation and innovation. [8]

A clear understanding of legal requirements, liability issues, and intellectual property rights is necessary to successfully navigate the legal terrain of AI in the biomedical and healthcare fields. Policymakers, business stakeholders, and regulatory agencies must work together to create flexible legal standards and regulatory frameworks.

2.2 Privacy and Confidentiality

Privacy and confidentiality are foundational to the ethical use of AI in healthcare, where sensitive medical data is used for training and application of large language models (LLMs). Ensuring privacy means protecting patient data from unauthorized access, while confidentiality relates to the obligations of healthcare providers to secure patient information from disclosure.

- **Data Anonymization Techniques:** The use of advanced anonymization techniques to safeguard patient data is a primary method for ensuring privacy. Researchers like Zeng and Pathak [21] discuss the effectiveness of differential privacy techniques in LLMs, which add randomness to the training data, thus preventing individual data re-identification without significantly compromising the utility of the models.
- **Regulatory Compliance:** Compliance with legal frameworks such as GDPR in Europe and HIPAA in the USA is crucial. Mikk et al [14] provides a comprehensive review of how AI applications can align with these regulations by implementing strict data governance and security measures.
- **Privacy-Preserving Technologies:** The integration of technologies such as federated learning, where AI models are trained across multiple decentralized devices holding local data samples, is highlighted by Kumar et al [11]. This approach minimizes the risk of data exposure by not centralizing patient data.

2.3 Bias and Fairness

AI has great potential to improve healthcare outcomes and delivery, there are also concerns about bias, data discrimination, and treating patients unfairly. When biased presumptions in training data or algorithm design are encoded into decision-making processes, it can lead to systemic errors or unfairness in AI systems. [16] In order to provide fair access to high-quality care and minimize potential risks to patient populations who are already vulnerable, it is critical to address prejudice in AI-driven healthcare.

Data Bias

Data bias occurs when training datasets for AI algorithms are not representative of the varied populations they are meant to serve. Predictions can become skewed or wrong due to biases present in training data, such as the overrepresentation of privileged populations or the underrepresentation of specific demographic groups. This can increase health disparities and sustain structural inequities [17]

For instance, biases in diagnosis rates, treatment recommendations, or access to care based on racial, ethnic, gender, or socioeconomic status may reflect in AI algorithms trained on historical healthcare data. AI systems run the potential of creating discrepancies in healthcare delivery if intentional steps are not taken to eliminate data bias.

Detailed approaches for varied data collection, preparation, and validation are necessary to mitigate data bias. The gathering of representative datasets that represent the whole range of patient demographics, clinical situations, and healthcare settings must be given top priority by healthcare institutions. Also, biases in training data and algorithmic decision-making processes can be found and decreased with data augmentation, bias detection algorithms, and fairness-aware machine learning.

Algorithmic Bias

When some prejudices are embedded into AI algorithms throughout their design, development, or optimization process, it can lead to unfair or discriminating outcomes for particular groups of people. This phenomenon is known as algorithmic bias. There are some quantitative measures for evaluating algorithmic fairness and equity like demographic parity, equalized probabilities, and disproportionate impact analysis. [5]

More transparent AI models make it possible for all parties involved—including regulators, physicians, and patients—to understand the variables that influence algorithmic judgments and spot any bias sources.

Addressing prejudice in AI-driven healthcare shows us some ethical problems regarding justice, fairness, and patient autonomy. AI systems undermine patient confidence in medical facilities and technology. The application of ethical frameworks in healthcare, such as beneficence, nonmaleficence, autonomy, and justice, offer guiding principles for responsible AI. These ethical frameworks should be a core part of the development and deployment of AI in healthcare.

Overall, tackling bias in AI-driven healthcare shows that a multidisciplinary strategy incorporating social, ethical, and technical factors is required.

2.4 Transparency and Explainability

The Need for AI's Transparency Taxonomy:

Reseachers Anastasiya Kiseleva, Dimitris Kotzinos, Paul De Hert [10] address the lack of clarity in defining terms related to transparency in AI, such as interpretability and explainability in their article. It highlights the importance

of establishing a common taxonomy to facilitate communication and knowledge transfer within the field. While some scholars use these terms interchangeably, others propose distinguishing between them, with interpretability being seen as a broader concept encompassing explainability measures. While some researchers view transparency as a component of interpretability or explainability, others argue for a broader understanding that encompasses the entire process of AI development and use. Transparency is depicted as a fundamental value that permeates the entire AI lifecycle, influencing its development, use, and governance.

External transparency:

The paper acknowledges the challenge of providing explanations for opaque AI systems, particularly when even their developers may struggle to understand them fully. It argues that while efforts should be made to offer the best possible explanations, some level of opacity may be deemed acceptable, following a risk-benefit approach. This perspective suggests that the inherent opacity of certain algorithms should be assessed alongside their performance and other benefits when verifying AI-based devices. Additionally, in the context of transparency toward patients, the paper emphasizes the importance of tailoring explanations to meet individual needs and preferences. It highlights that overly complex explanations may hinder rather than facilitate decision-making, and compares the process to how physicians distill complex medical information for patients. [10]

Internal transparency:

Internal transparency is deemed crucial for holding healthcare professionals accountable to patients and aiding patients in making informed treatment decisions. The Medical Devices Framework (MDF) serves as the cornerstone for ensuring internal transparency in AI, outlining the obligations of AI developers (device manufacturers) in supplying relevant information to users (healthcare professionals). The framework emphasizes the provision of comprehensive information tailored to the technical knowledge and experience of users, along with detailing clinical benefits, performance characteristics, risks, and limitations of AI devices. [10]

Black box problem:

To address the black-box issue in AI, several considerations are proposed [10]. Firstly, it's argued that AI technologies with inherent opacity should not be banned, even in critical domains like healthcare, as evidenced by the European Commission's inclusive approach in its AI regulation proposal. Secondly, AI providers are urged to implement the best possible explainability measures, incentivizing progress in research and technical capabilities for AI transparency. Thirdly, while complete technical resolution of algorithmic opacity may be challenging, it could be deemed an acceptable risk given careful evaluation and demonstration of the benefits of AI use outweighing the opacity. This aligns with the inherent uncertainty in healthcare where some level of opacity is unavoidable, making state-of-the-art explanations crucial for justifying the acceptance of opacity. Finally, adopting a multilayered transparency system can balance algorithmic opacity by strengthening other transparency measures from AI providers and other stakeholders, offering a pragmatic approach to implementing transparency within accountability frameworks. [10]

Explainability:

In the article on Explainability for AI in healthcare by Amann et. al., from a technological perspective, explainability in healthcare AI involves addressing two key issues. Firstly, explainability methods can either be inherent to the algorithm or approximated through other techniques, with inherent explainability generally being more accurate but often associated with traditional methods like linear regression, contrasting with the higher performance of modern methods like artificial neural networks (ANNs). Secondly, explainability is crucial for developers to ensure AI models are not making predictions based on irrelevant factors, such as metadata or hardware-related information, rather than clinically relevant data. Explainability methods allow developers to identify and rectify such errors before clinical validation, saving time and development costs. [1]

From a medical perspective, they think explainability in AI-based clinical decision support systems (CDSS) is crucial for understanding how the system arrives at conclusions and for ensuring transparency in the clinical setting. Two levels of explainability are distinguished: first-level explainability helps understand the system's general conclusions, while second-level explainability identifies the important features for individual predictions. Clinical validation, while important, is often focused on prediction accuracy, but explainability becomes instrumental in resolving disagreements between AI systems and human experts, particularly in cases of systematic errors or bias. Explainability assists clinicians in evaluating recommendations, strengthening trust in the system, and identifying instances of poor performance for quality improvement. [1]

From the patient perspective, the issue of explainability in AI-powered decision aids is examined for its compatibility with patient-centered care principles, which prioritize patients' values, needs, and active participation in medical decisions. Shared decision-making relies on open conversations between patients and clinicians. However, the emergence of "black-box medicine," where clinicians cannot fully understand or explain the inner workings of decision aids, poses a challenge to this approach. Explainability in AI decision aids becomes essential as it allows clinicians to provide personalized explanations to patients, fostering informed decision-making and boosting patients' confidence in their treatment choices. By visualizing how various factors contribute to risk

assessment, explainable AI systems can enhance patients' understanding, encourage engagement in shared decision-making, and promote more accurate risk perceptions, ultimately aligning with the principles of patient-centered care. [1] They conclude that omitting explainability in clinical decision support systems poses a threat to core ethical values in medicine and may have detrimental consequences for individual and public health. [1]

2.5 Informed Consent

Informed consent in AI-driven healthcare involves ensuring that patients understand how their medical data will be used, particularly in training and applying AI models like LLMs.

- **Dynamic Consent Models:** Traditional consent processes may not suffice for the complexities involved in AI applications. Researchers have proposed dynamic consent models, where patients can adjust their consent preferences in real-time as their data is used in new ways. Johnson et al [7] explore this model, highlighting its potential to enhance patient autonomy and trust.
- **Transparency in Consent Processes:** Ensuring that consent forms and processes are transparent about the use of AI is essential. This involves clearly explaining the purposes of data use, the potential risks, and the benefits. In [12], the authors discuss the development of AI-specific consent protocols that are designed to be understandable to non-experts .
- **Patient Education:** With the complexity of AI technologies, educating patients on what AI in healthcare means is crucial for informed consent. Thomson and Borenstein suggest structured patient education programs that can demystify AI processes and encourage informed decision-making [20].

2.6 Sustainability

One issue that is hotly debated in the field of AI at large is the sustainability of it, currently being presented as a factor against the complete adoption of the technology. There is a significant amount of research already conducted on this topic, in one existing systematic review conducted on 287 papers using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). Based on literature, it was found as a common theme that although we are moving toward sustainable development, the integration of AI still presents various issues and challenges. The most forefront topics however, can be grouped into three categories, environmental, social, and technological. That is not to say there are not benefits to the integration of this technology. Using AI, we can accurately predict weather changes and potential catastrophic conditions, which allows us to prepare ahead of time to divert resources, and evacuate to minimize losses.

The application of AI in the healthcare sector has vast potentials. AI is able to help healthcare providers better understand the day-to-day needs and changes in them, in order to provide a better support system that can efficiently help people stay healthy, providing long term benefits. One of the most discussed benefits of AI is its ability to predict disease with high accuracy, allowing people to proactively incorporate healthier changes to their environment and their behaviors, in order to create a healthier lifestyle. AI additionally offers the capability of helping healthcare professional with better feedback and guidance in real time, in the effort of keeping patients healthy. The predictive models of AI can provide accurate and reliable results about environment factors such as air quality, information that can be used by city designers in making informed choices about emission and regulation to protect the population's overall health and life quality [9].

2.7 Trustworthiness and Reliability

Addressing the reliability of AI/ML in healthcare involves a multifaceted approach spanning product development, selection, validation, performance calibration, implementation, evaluation, and oversight. Developers must prioritize transparency in algorithm design and training processes to ensure systems are auditable and explainable, mitigating the risks associated with black box models. Interdisciplinary teams should assess the clinical relevance and outcomes of AI/ML adoption, considering factors such as patient safety, efficiency, and equity. Rigorous validation processes, including biological validation where applicable, are necessary to verify claims about completeness, data quality, and effectiveness. Calibration ensures the AI/ML output matches actual diagnoses, while ongoing pilot testing in real-world clinical settings identifies and addresses potential errors and improves user acceptance. Implementation and oversight involve education initiatives, ongoing monitoring, and quality assurance measures to track performance, correct for drift, and ensure dynamic safety assurances throughout the AI/ML lifecycle [3].

Ensuring the reliability of AI in healthcare necessitates addressing human factors and system properties that influence trust in AI systems. Key factors influencing trust include user education, past experiences, biases, and properties of the AI system such as controllability, transparency, and complexity. Reliability, particularly concerning the predictability and consistency of AI performance, is paramount in healthcare due to the potential impact on patient outcomes. However, concerns about biased or overfitted outcomes generated by AI, especially in the presence of new data, hinder

trust and acceptance of AI systems. Achieving an optimal level of trust requires balancing skepticism between humans and AI to mitigate the risks associated with overly high trust, especially in life-critical applications. Incorporating fairness, transparency, and robustness into AI development can help establish and maintain trust in human-AI collaboration, thereby enhancing the reliability of AI systems in healthcare [2].

3 CONCLUSIONS AND FUTURE WORKS

The ethical implications surrounding the use of AI models such as biomedical large language models (Bio-LLMs), in healthcare decision-making are profound and multifaceted. This report has explored the complex ethical implications of AI models in healthcare decision-making. We have highlighted key concerns around data bias, privacy, informed consent, transparency, and the shift in autonomy that AI can bring. Addressing these ethical challenges will be paramount for policymakers, researchers, and industry leaders to ensure that AI development and deployment benefit society. Robust ethical frameworks and responsible implementation strategies must be developed to guide the trustworthy and responsible use of Biomedical AI in healthcare decision-making and bio-medicine. Additionally, further research is needed to address the specific technical and regulatory challenges posed by AI in healthcare, such as liability, intellectual property, and the evolving nature of these technologies. Flexible regulatory approaches that balance innovation and patient safety will be crucial.

Addressing these ethical challenges will require coordinated efforts from interdisciplinary stakeholders - healthcare providers, AI developers, policymakers, ethicists, and the public. Robust governance frameworks, stringent data practices, fairness-aware algorithm development, AI literacy initiatives, and ongoing monitoring are essential to mitigate risks and uphold core ethical principles. Future work should focus on:

- Developing comprehensive ethical guidelines and certifications specific to AI use in healthcare contexts.
- Researching advanced privacy-preserving techniques like federated learning and differential privacy for biomedical data.
- Furthering explainable AI methods to enhance transparency and facilitate trust between clinicians, patients, and AI systems.
- Implementing dynamic informed consent models that empower patients with Agency over how their data is utilized.
- Establishing interdisciplinary collaborations to continually assess AI reliability, validate performance, and correct for embedded biases or drift over time.
- Examining the societal impact of increased AI integration on human decision-making autonomy and healthcare workforce dynamics.

Proactive efforts to navigate these ethical terrains will shape a future where the transformative power of AI is effectively harnessed to elevate healthcare quality, equity, and trust for all members of society. Responsible AI development and adoption today paves the way for equitable, ethical, and trustworthy AI-assisted healthcare tomorrow.

REFERENCES

[1] Julia Amann, Alessandro Blasimme, Effy Vayena, Dietmar Frey, Vince I Madai, and Precise4Q Consortium. 2020. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC medical informatics and decision making* 20 (2020), 1–9.

[2] Onur Asan, Alparslan Emrah Bayrak, and Avishek Choudhury. 2020. Artificial intelligence and human trust in healthcare: focus on clinicians. *Journal of medical Internet research* 22, 6 (2020), e15154.

[3] Katherine Drabiak. 2022. Leveraging law and ethics to promote safe and reliable AI/ML in healthcare. *Frontiers in Nuclear Medicine* 2 (2022), 983340.

[4] Barbara J Evans. 2023. The HIPAA privacy rule at age 25: privacy for equitable AI. *Florida State University Law Review, Forthcoming* (2023).

[5] Benedetta Giovanola and Simona Tiribelli. 2023. Beyond bias and discrimination: redefining the AI ethics principle of fairness in healthcare machine-learning algorithms. *AI & society* 38, 2 (2023), 549–563.

[6] Frank Griffin. 2021. Artificial intelligence and liability in health care. *Health Matrix* 31 (2021), 65.

[7] Amy Johnson, Barbara Smith, and Xiaoyu Liu. 2024. Towards Dynamic Consent in Healthcare AI. *Ethics and Information Technology* 26, 1 (2024), 11–29.

[8] Sarah Kamensky. 2019. Artificial intelligence and technology in health care: overview and possible legal implications. *DePaul J. Health Care L.* 21 (2019), 1.

[9] Arpan Kumar Kar, Shweta Kumari Choudhary, and Vinay Kumar Singh. 2022. How can artificial intelligence impact sustainability: A systematic literature review. *Journal of Cleaner Production* 376 (2022), 134120. https://doi.org/10.1016/j.jclepro.2022.134120

[10] Anastasiya Kiseleva, Dimitris Kotzinos, and Paul De Hert. 2022. Transparency of AI in healthcare as a multilayered system of accountabilities: between legal requirements and technical limitations. *Frontiers in Artificial Intelligence* 5 (2022), 879603.

[11] Praveen Kumar, Greg Cohen, and Timothy Bickmore. 2023. Federated Learning for Privacy-Preserving AI in Healthcare. *Health Informatics Journal* 29, 1 (2023), 45–56.

[12] Hongwei Li, Qiang Meng, and Miaomiao Zhou. 2023. Enhancing Transparency in AI-driven Healthcare. *Journal of Medical Ethics* 49, 1 (2023), 77–83.

[13] Renqian Luo, Liai Sun, Yingce Xia, Tao Qin, Sheng Zhang, Hoifung Poon, and Tie-Yan Liu. 2022. BioGPT: generative pre-trained transformer for biomedical text generation and mining. *Briefings in Bioinformatics* 23, 6 (Sept. 2022). https://doi.org/10.1093/bib/bbac409

[14] Raivo Mikk, Greg Cohen, and Timothy Bickmore. 2022. GDPR and HIPAA Compliance for Healthcare AI Systems. *Health Policy and Technology* 11, 1 (2022), 134–142.

[15] Lilian Mitrou. 2018. Data protection, artificial intelligence and cognitive services: is the general data protection regulation (GDPR)'artificial intelligence-proof'? *Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR)'Artificial Intelligence-Proof*

(2018).

[16] Lama H Nazer, Razan Zatarah, Shai Waldrip, Janny Xue Chen Ke, Mira Moukheiber, Ashish K Khanna, Rachel S Hicklen, Lama Moukheiber, Dana Moukheiber, Haobo Ma, et al. 2023. Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS digital health* 2, 6 (2023), e0000278.

[17] Claudio Novelli, Federico Casolari, Philipp Hacker, Giorgio Spedicato, and Luciano Floridi. 2024. Generative AI in EU Law: Liability, Privacy, Intellectual Property, and Cybersecurity. *arXiv preprint arXiv:2401.07348* (2024).

[18] Pranav Rajpurkar, Emma Chen, Oishi Banerjee, and Eric J. Topol. 2022. AI in health and medicine. *Nature Medicine* 28, 1 (01 Jan 2022), 31–38. https://doi.org/10.1038/s41591-021-01614-0

[19] Karan Singhal, Shekoofeh Azizi, Tao Tu, S. Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, Perry Payne, Martin Seneviratne, Paul Gamble, Chris Kelly, Abubakr Babiker, Nathanael Schärli, Aakanksha Chowdhery, Philip Mansfield, Dina Demner-Fushman, Blaise Agüera y Arcas, Dale Webster, Greg S. Corrado, Yossi Matias, Katherine Chou, Juraj Gottweis, Nenad Tomasev, Yun Liu, Alvin Rajkomar, Joelle Barral, Christopher Semturs, Alan Karthikesalingam, and Vivek Natarajan. 2023. Large language models encode clinical knowledge. *Nature* 620, 7972 (01 Aug 2023), 172–180. https://doi.org/10.1038/s41586-023-06291-2

[20] Lisa Thomson and Jason Borenstein. 2023. Educating Patients About AI in Healthcare: A Necessity for Informed Consent. *AI & Society* 38, 1 (2023), 207–215.

[21] Xiang Zeng and Jyotishman Pathak. 2023. Differential Privacy in Healthcare: A Robust Method for Protecting Patient Data. *Journal of Medical Internet Research* 25, 2 (2023), e24105.