



Algor-ethics: charting the ethical path for AI in critical care

Jonathan Montomoli^{1,2} · Maria Maddalena Bitondo¹ · Marco Cascella³ · Emanuele Rezoagli^{4,5} · Luca Romeo⁶ ·
Valentina Bellini⁷ · Federico Semeraro⁸ · Emiliano Gamberini¹ · Emanuele Frontoni⁹ · Vanni Agnoletti¹⁰ ·
Mattia Altini¹¹ · Paolo Benanti¹² · Elena Giovanna Bignami⁷

Received: 18 March 2023 / Accepted: 22 March 2024
© The Author(s) 2024

Abstract

The integration of Clinical Decision Support Systems (CDSS) based on artificial intelligence (AI) in healthcare is groundbreaking evolution with enormous potential, but its development and ethical implementation, presents unique challenges, particularly in critical care, where physicians often deal with life-threatening conditions requiring rapid actions and patients unable to participate in the decisional process. Moreover, development of AI-based CDSS is complex and should address different sources of bias, including data acquisition, health disparities, domain shifts during clinical use, and cognitive biases in decision-making. In this scenario algor-ethics is mandatory and emphasizes the integration of ‘Human-in-the-Loop’ and ‘Algorithmic Stewardship’ principles, and the benefits of advanced data engineering. The establishment of Clinical AI Departments (CAID) is necessary to lead AI innovation in healthcare, ensuring ethical integrity and human-centered development in this rapidly evolving field.

Keywords Algor-ethics · Artificial intelligence · Machine learning · Ethics · Data engineering

✉ Jonathan Montomoli
jonathan.montomoli@gmail.com

¹ Department of Anesthesia and Intensive Care, Infermi Hospital, Romagna Local Health Authority, Viale Settembrini 2, Rimini 47923, Italy

² Health Services Research, Evaluation and Policy Unit, Romagna Local Health Authority, Viale Settembrini 2, Rimini 47923, Italy

³ Unit of Anesthesia and Pain Medicine, Department of Medicine, Surgery and Dentistry “Scuola Medica Salernitana,” University of Salerno, Baronissi, Salerno, Italy

⁴ School of Medicine and Surgery, University of Milano-Bicocca, Via Cadore, 48, Monza 20900, Italy

⁵ Dipartimento di Emergenza e Urgenza, Terapia intensiva e Semintensiva adulti e pediatrica, Fondazione IRCCS San Gerardo dei Tintori, Via Pergolesi, 33, Monza 20900, Italy

⁶ Department of Economics and Law, University of Macerata, Macerata 62100, Italy

⁷ Anesthesiology, Critical Care and Pain Medicine Division, Department of Medicine and Surgery, University of Parma, Via Gramsci 14, Parma 43125, Italy

⁸ Department of Anesthesia, Intensive Care and Prehospital Emergency, Ospedale Maggiore Carlo Alberto Pizzardi, Largo Bartolo Nigrisoli, 2, Bologna 40133, Italy

⁹ Department of Political Sciences, Communication and International Relations, University of Macerata, Macerata 62100, Italy

¹⁰ Department of Surgery and Trauma, Anesthesia and Intensive Care Unit, Maurizio Bufalini Hospital, Romagna Local Health Authority, Viale Giovanni Ghirelli, 286, Cesena 47521, Italy

¹¹ Hospital Care Sector, Emilia-Romagna Region, Via Aldo Moro, 21, Bologna 40127, Italy

¹² Pontifical Gregorian University, Piazza della Pilotta 4, Roma 00187, Italy

1 Introduction

Artificial intelligence (AI) algorithms have steadily gained prominence in healthcare and are expecting to revolutionize the landscape of clinical decision support systems (CDSSs) [1–3]. However, in the field of critical care, despite the increasing number of developed AI-based algorithms, the vast majority remains within the testing and prototyping phase, lacking external validation and certification [4]. CDSSs harness AI to enhance patient outcomes and assist healthcare professionals in making informed diagnostic and therapeutic choices. Theoretically, the amalgamation of AI's advanced capabilities with the acumen and skill of clinicians represents a transformative force in healthcare. However, an intrinsic complexity of critical care intensifies the challenges associated with implementing AI-based methods. Therefore, ensuring fairness and responsibility in AI implementation is crucial. Relying solely on “algorithmic fairness audits” may not guarantee the equitable development and deployment of such technology [5, 6]. Instead, they should be integrated within a comprehensive and structured approach to AI implementation in critical care settings [7].

In the rapidly evolving landscape of healthcare, the concept of Algorithmic Stewardship emphasizes the judicious management and integration of these technologies, advocating for a robust framework that addresses potential biases and ensures clinicians' readiness [8]. This framework underscores the importance of a synergistic interplay between humans and algorithms throughout the machine learning

process grouped under the umbrella term of Human-in-the-Loop (HITL) [9].

Questions regarding ethical use and implications of AI in healthcare become increasingly critical and an ‘algor-ethics’ emerges at the forefront. This novel ethical paradigm addresses the dynamic nature of AI encompassing the entire process of its development, from its nascent stages to its real-world deployment. By setting the ‘rules of the game’, algor-ethics establishes the bedrock upon which Algorithmic Stewardship and HITL are built, aligning AI systems with human values and societal norms, and fostering accountability, fairness, and transparency. In this manuscript, we aim to present our perspective about the principles at the foundation of the AI development in critical care, starting by identifying the peculiarities in this field, addressing the most relevant sources of bias, and suggesting a structured process.

1.1 Peculiarities to account of AI deployment in critical care

Human beings have always actively promoted technological advancement as part of the human evolution. In the past, the rate of innovation was often limited by technology itself in comparison with human adaptability. However, recent significant advancements in the computational power, have exponentially accelerate technological growth to the point it has reach, or even outpaced, our ability to adapt and control such development (Fig. 1). This concept was firstly introduced by Thomas Friedman emphasizing the need for lifelong learning as a pathway to harmonizing

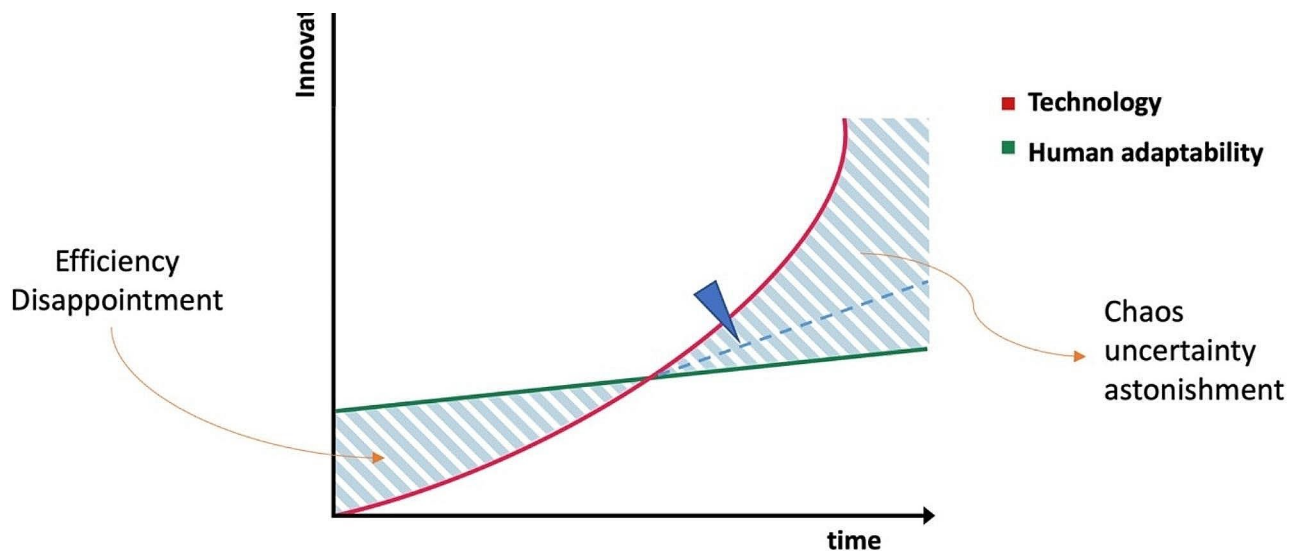


Fig. 1 Graphical depiction illustrating the progression of technological innovation in relation to human adaptability. The evident divergence emphasizes the accelerated rate of technological advancements compared to human capacity for adaptation, leading to potential areas of

disparity and unpredictability. Such observations necessitate a considered approach to technological developments to ensure alignment with human evolutionary pathways

our relationship with advancing technology [10]. Within healthcare, critical care heavily utilizes technology, ranging from continuous invasive and noninvasive monitoring [11], imaging tools [12, 13], and organ support devices such as ventilators [14, 15], and hemodialysis [16]. Consequently, a massive amount of data is continuously stored in electronic healthcare records (EHRs). This data could potentially serve to assist physicians in understanding the complexity and heterogeneity of patient conditions. However, data stored in EHRs are not designed for rapid assessment or to support timely decision-making. The fact that intensivist usually coordinates a team while being required to make quick decisions within limited information adds layers of complexity to decision-making. Unlike other specialties, the intensivist's role often demands rapid decisions without direct patient input such as in the decision to withdraw life-sustaining treatment for futility or not to resuscitate [17]. These decisions are often precipitated by acute, unexpected events that preclude the possibility of discussing choices not only with the patient but sometimes even with their closest relatives. Therefore, the intensivist may face the profound responsibility of acting in the patient's best interests based on clinical judgment and existing ethical frameworks, without the patient's articulated consent or advanced directives. As we explore potentiality of AI in critical care, it is crucial to establish a deployment foundation that upholds the best interests of all individuals involved—patients, healthcare practitioners, and when possible, their families—while also confronting the inherent challenges this alignment presents.

1.2 Bias in data acquisition for AI in healthcare

The accuracy and reliability of AI in healthcare are challenged by varied institutions' practices. This variability affects AI performance across different settings. While data standardization is important, its broad application is impractical due to different clinical protocols and environments. Instead, the focus should shift towards the development of advanced data engineering such as Extract, Transform, Load (ETL) pipelines [18, 19]. These pipelines play a critical role in effectively harmonizing and integrating data from disparate sources and formats enabling the collection and processing of data in a manner that acknowledges the complexity inherent in critical care and respects the unique operational workflows of each healthcare institution [20, 21]. This approach allows the development of more robust and adaptable AI algorithms [22].

1.3 Sources of bias in algorithm development for AI in healthcare

Although data engineering may markedly help with the collection and harmonization of data from different sources, health disparities can inadvertently be exacerbated by AI algorithms if not duly considered during the development and deployment stages [23]. "Algorithmic fairness audit" is a useful tool to prevent AI algorithms from unintentionally promoting health disparities that may stem from residual and unmeasured confounding, which may be associated with patients' characteristics such as gender [24], comorbidities [25], or ethnicity [5]. Additionally, although the prognostic impact of comorbidities may be associated to ethnicity other associated conditions such as socioeconomic status and educational levels may partly explain the association and such information are not routinely recorded or available to researchers [26]. In such scenario, audits should evaluate the performance of the AI-based algorithms in population subgroups by different metrics and lead to an appropriate calibration and discrimination of the models. Nevertheless, health disparities are not solely a product of unmeasured or residual confounding. Historical biases present in training data, algorithmic biases favoring certain demographics, limited access to advanced healthcare technology, and cultural differences or language barriers, can all contribute to disparities [27]. Despite also other sources of bias have been reported in the phases of algorithm development such as the managing of missing data or outliers, features selection, or modeling of underrepresented groups [23, 28], data sharing restrictions and data privacy represent a fundamental issue that should be considered. Data protection and patient protection are both critically important in the context of healthcare. However, when implementing AI algorithms, it is necessary to question the priority between the two as they may, at times, conflict with each other or require different approaches for optimization. For instance, strict data protection may sometimes limit the availability of data needed for the development of AI algorithms that can potentially save lives or significantly improve patient outcomes. The dilemma is highlighted by Floridi, who has questioned the priority to data protection over patient protection [29]. While data protection is important, it should not be regarded as an absolute value, but rather as an instrumental value that serves broader ethical principles, such as the protection of human dignity and well-being. In some cases, the protection of patients may necessitate the use of data in a manner that can conflict with stringent data protection norms. If different approaches have been proposed to limit methodological sources of bias, in regard to data protection, the values and potential outcomes should be carefully weighted to arrive at decisions that uphold the core

ethical principles of promoting well-being and protecting human dignity.

1.4 Issues to take into consideration in the use of AI-based CDSS

In the AI deployment in healthcare it is vital to ensure that the advancements benefit all patient demographics equitably. Although the process of re-calibration and external validation may improve the generalizability of the algorithms, it poses the problem of the “domain shift” referring to the phenomenon wherein the statistical properties or distribution of the data that an algorithm is applied to during real-world usage differ from the properties or distribution of the data that the algorithm was trained on [30]. This discrepancy can be due to various factors including changes in population characteristics, technological advancements, evolving clinical practices, or even subtle differences in data collection methods between different settings. In the context of critical care, domain shift may significantly impact AI algorithm. For example, a predictive AI algorithm to diagnose acute kidney injury based on data where prevail data on prerenal causes might fail to accurately perform on patients developing intrarenal acute kidney injury. Domain adaptation techniques are often employed to tackle domain shift by adjusting the algorithm to perform better on the new data distribution without the need for extensive retraining (i.e. transfer learning, feature-level domain adaptation, instance-weighted domain adaptation) [31, 32]. In recent years, there have been also advancements in machine learning that specifically address fairness such as adversarial de-biasing, where a model is trained to make predictions that are statistically independent of certain sensitive attributes, such as gender or ethnicity [33]. Moreover, the introduction of AI-based CDSS is not merely an addition to the existing clinical practice but a transformative influence that reshapes it. As these systems are integrated into the workflow of critical care, where novel monitoring systems and experimental diagnostic tools are frequently adopted, the practice itself evolves. This evolution can alter the predictive accuracy of AI-based CDSS, necessitating their continuous adaptation to the shifting clinical landscape. Particularly in critical care, the rapid pace of technological innovation and the trial of emerging treatments demand that AI systems are not static but are capable of learning and adapting in tandem with the field’s advancement.

To sustain the relevance and accuracy of these systems, an ongoing process of audits and performance supervision is indispensable. This continuous oversight ensures that AI-based CDSS remains attuned to the ever-changing reality of patient care, capable of identifying domain shifts and deviations from established data patterns. By institutionalizing a

regimen of rigorous, ongoing evaluation, healthcare providers can maintain the integrity and utility of AI applications, ensuring these tools continue to reflect and respond to the latest clinical evidence and practice standards.

1.5 Human factors contributing to bias in the use of AI in healthcare

Human reasoning is not inherently structured for optimal decision-making but rather for maximizing decision-making efficiency. This understanding is crucial in healthcare, where the decisional process is usually complex and clinical judgment should follow a structured pathway and should not merely follow the intuitive approach [34]. The dual-process theory differentiates between ‘System 1’ and ‘System 2’ thinking: System 1 operates quickly and intuitively, often relying on heuristics, while System 2 is slower, more deliberate, and analytical [35]. Clinicians frequently rely on System 1 for immediate, routine decisions, but complex, high-stakes scenarios necessitate the engaged, analytical reasoning of System 2. Understanding this dichotomy is pivotal in grasping how healthcare professionals interact with AI-based systems and the potential biases that can arise from these interactions. In critical care, the urgency of decision-making often leads clinicians to predominantly utilize System 1 thinking. This reliance on intuition and heuristics, while efficient, can be prone to biases and errors. Recognizing, evaluating, and addressing these cognitive dynamics is crucial when considering the use of IA-based CDSS. Gaube et al. highlight a critical concern where physicians might unconsciously adopt biases present in AI systems [36]. Diagnostic accuracy was significantly worse when participants received inaccurate advice. Notably, the ability to detect the low quality of the advice from the AI system correlated with the degree of the physician task-expertise. This phenomenon underscores the risk of a ‘feedback loop’ where AI becomes a source of reinforced bias. Over-reliance on AI might lead clinicians to overlook its limitations, leading to potential errors in judgment when AI advice is flawed [37]. Nonetheless, the bias carried by an AI system affects clinicians’ decisions also when AI is no longer making recommendation suggesting human inheritance bias of AI [37]. The evidence collectively calls for a balanced approach to AI implementation in healthcare. It necessitates not only technical precision but also an awareness of human cognitive biases and the development of a digital culture among all medical stakeholders to critically engage with AI-based CDSS [38]. This approach is especially pivotal in critical care, where the ramifications of decision-making are immediate and profound.

1.6 Algor-ethics: ensuring ethical integrity in AI development and implementation in healthcare

Prioritizing ethical considerations sets the stage for responsible and beneficial AI integration, reinforcing the commitment to patient welfare and equitable healthcare practices. Notably, the leadership of this process should originate from the identification of specific healthcare needs, aims, and priorities, ensuring that technology serves these predefined objectives. This human-led direction ensures that technological development is a response to actual clinical demands, rather than developing algorithms first and subsequently seeking applications for them. This approach helps to avoid the pitfall of technology-driven solutions in search of problems, focusing instead on problem-driven innovations that are more likely to benefit patient care and healthcare outcomes. To this end, it is crucial to devise an ethical framework that can address the complexity and value of humans - their unique rational and emotional functionalities, and the intrinsic relationship between humans and artifacts that underlies our existence (techno-human condition). The unique challenges in critical care further underscore the need for an ethically guided, human-centric approach to AI implementation. The ethical framework previously advocated should, therefore, not only addresses the technical and clinical aspects of patient care but also deeply considers patient autonomy and dignity. This framework is “algor-ethics” [39] and must be thought of in a cooperative manner. In other words, AI is not an evolutionary adversary but a tool that must be thought of as cooperative with the person. Indeed, AI must be created to augment the cognitive capability that is a unique and peculiar prerogative of the human being, and they must never replace it. The ultimate objective, then, is to enhance human cognition, not to turn cognition into an algorithmic function separated from human beings. In this regard, the definition of the “wayfinding” approach for the purpose of AI is particularly helpful [40]. A shift towards a “generalized wayfinding” approach both in the diagnostic and therapeutic process may significantly improve the entire care process by helping with the understanding of the pathophysiological mechanisms underlying the development of clinical complications [41]. Moreover, such an approach promotes a culture of continuous learning with the aim of improving the path to the correct diagnosis, synthesizing complex patient data, and determining the best next steps, rather than predicting a predefined outcome. AI’s growing autonomy and decision-making capability necessitate control mechanisms to ensure safety, ethical alignment, and accountability. Furthermore, the implementation of AI technologies carries a displacement of “power” that could impact who benefits from and who is adversely affected by these technologies. Without an ethical framework, the use

of AI could lead to inequitable outcomes, potentially benefiting some while inadvertently harming others [28, 42]. The role of algor-ethics can be elucidated through an analogy between AI-based algorithms and cars. The manufacturing process of cars, akin to the development framework, ensures that cars are equipped with essential safety features like brakes and steering. However, the mere presence of safety features does not guarantee accident-free roads. It is imperative for constructors and drivers, akin to developers and users of AI algorithms, to adhere to a set of rules and behaviors to ensure safety. This analogy underscores the importance of not only having safety features (ethical considerations in development) embedded in AI algorithms but also ensuring that they are implemented and employed within a responsible and regulated behavioral framework. Algor-ethics also guides the technical and pathway aspects of algorithmic development. To achieve this comprehensive scope, algor-ethics embraces two pivotal concepts: ‘Human-in-the-Loop’ (HITL) and ‘Algorithmic Stewardship’ (Fig. 2).

The ‘Human-in-the-Loop’ (HITL) concept is critical to question the role of human expertise and judgment within the AI workflow and outlines four main domains that encapsulate the interplay between humans and algorithms [9]:

- “Learning with Humans”: It delineates who controls the learning process, distinguishing between active learning (algorithm-controlled), machine teaching (human-controlled), and interactive machine learning (a dynamic and cooperative interaction);
- “Curriculum Learning”: This refers to the algorithm’s iterative learning process, whereby simpler tasks precede more complex ones, mimicking the learning trajectory of a human learner;
- “Explainable AI”: Given that AI systems often operate as a ‘black box’, transparency is key. Explainable AI ensures that AI model decisions are interpretable and transparent, promoting trust among clinicians.
- “Beyond Learning – Useful and Usable AI”: AI models need to be both useful (improving outcomes or efficiency) and usable (seamlessly integrating into existing workflows) to successfully enhance clinical practice.

“Algorithmic stewardship” has been recently advocated and may help with defining a framework capable of addressing these four domains and creating an AI that can be effectively integrated and improve clinical practice [8]. Algorithmic Stewardship should provide guidelines on AI accountability, transparency, and fairness, enabling a balance between human judgement and automated decision-making. It oversees the entire lifecycle of AI systems: their training, validation, implementation, and ongoing evaluation. Crucially, it

ensures that AI applications enhance patient care without compromising ethical standards. To be effective, the process to define the “algorithmic stewardship” should have a collaborative governance promoting the collaboration between various stakeholders, including technologists, ethicists, regulators, and the healthcare professionals. Furthermore, it should be locally adapted to the characteristics of each institutions in term of the maturity of the “three pillars” of digital transformational: data quality and quantity, technological infrastructures, and digital culture (Fig. 2) [43].

Although we commonly think that “two heads are better than one”, it is crucial to acknowledge that in human-AI interactions, this synergy is likely missing. Indeed, the effectiveness of such collaboration depends significantly on the task’s nature and uniformity of the knowledge between human and AI [44]. Given that AI-based models are trained on extensive data and undergo intricate data processing not easily interpretable, the interaction is prone to knowledge asymmetry between the parties [44]. In setting requiring rapid interventions such as critical care, this imbalance increases further worsening the interaction [45]. One potential consequence is facing the situation where technology overcome human expertise, transforming physicians in passive executors. Although consequences of the imbalance are difficult to predict there is almost no doubt that such scenario is likely to impair clinical reasoning and motivation while increasing dependence on technology [46]. Algorithmics, emphasizing a human-centered approach, works to ensure a meaningful engagement of physicians and other operators with tasks and environments.

In conclusion, the main requirements of algor-ethics may be summarized as following (Fig. 2):

- *Human-centered*: AI should not replace, but rather augment human decision-making. AI should preserve uncertainty in its output, providing information about the accuracy and precision of its estimates. This allows for the preservation of human intervention in the decision-making process, rather than ceding complete control to AI. Similarly, the scope of the algorithm should be calibrated to serve the patient, addressing specific and meaningful problems maximizing the use of limited resources. This approach ensures that AI not only supports medical professionals but also contributes efficiently to patient-centered care.
- *Traceability*: every step of the algorithm’s journey, from creation to validation, must be transparent and documented. An AI’s applicability, its outputs, and its developmental steps should be disclosed in a standardized datasheet.
- *Customization*: instead of a one-size-fits-all approach, AI algorithms must be capable of adapting and interacting

with each individuality. This adaptability should be evident from both the perspective of the users (healthcare professionals) and the patients themselves.

- *Adequation*: The ultimate aim of AI in healthcare should be to align with and serve the patient’s best interest. The priority of the patient should become the priority of the algorithm, not vice versa.

1.7 The clinical AI department: the place for a human-oriented AI development

The development and exploitation of AI-based algorithms involves multifaceted considerations, from data acquisition and quality to awareness of algorithmic biases and training of healthcare professionals. These advancements have the potential to enable physicians to make more informed diagnostic and therapeutic decisions, emphasizing the necessity for a seamless and effective digital technology integration in healthcare. The Clinical AI Department (CAID) is envisioned as a pivotal entity for addressing the complexities surrounding AI-based CDSS in healthcare [47]. As a global virtual institution comprised of numerous local CAIDs, it embodies the principle of “think globally, act locally”. This initiative not only bolsters data-driven decision-making but also triggers a virtuous cycle of perpetual enhancements in medical diagnostics and therapeutic methodologies (Fig. 2). The CAID should lead the AI innovation in healthcare, serving as a collaborative hub where stakeholders converge to identify problems, define strategies, assess infrastructure and feasibility, recognize limitations, develop algorithms, and evaluate the use and performance of AI as CDSS. Furthermore, it actively fosters an AI-informed culture within healthcare. Local CAIDs are crucial in promoting an ETL framework and contributing to national and international projects, ensuring a cohesive and comprehensive approach to AI in healthcare. This collaborative approach not only enhances the algorithm’s reliability and accuracy but also ensures that the AI tools developed are equitable, contextually relevant, and capable of addressing the diverse needs of patient populations.

2 Conclusions

Although the integration of AI into clinical practice may significantly enhance patient outcomes and support clinician in the decisional process, it comes with distinct challenges. There is a need of a strong leadership defining the aims and the pathway to follow for the AI revolution in healthcare. The CAID may carry this challenge by promoting a collaborative and multidisciplinary environment

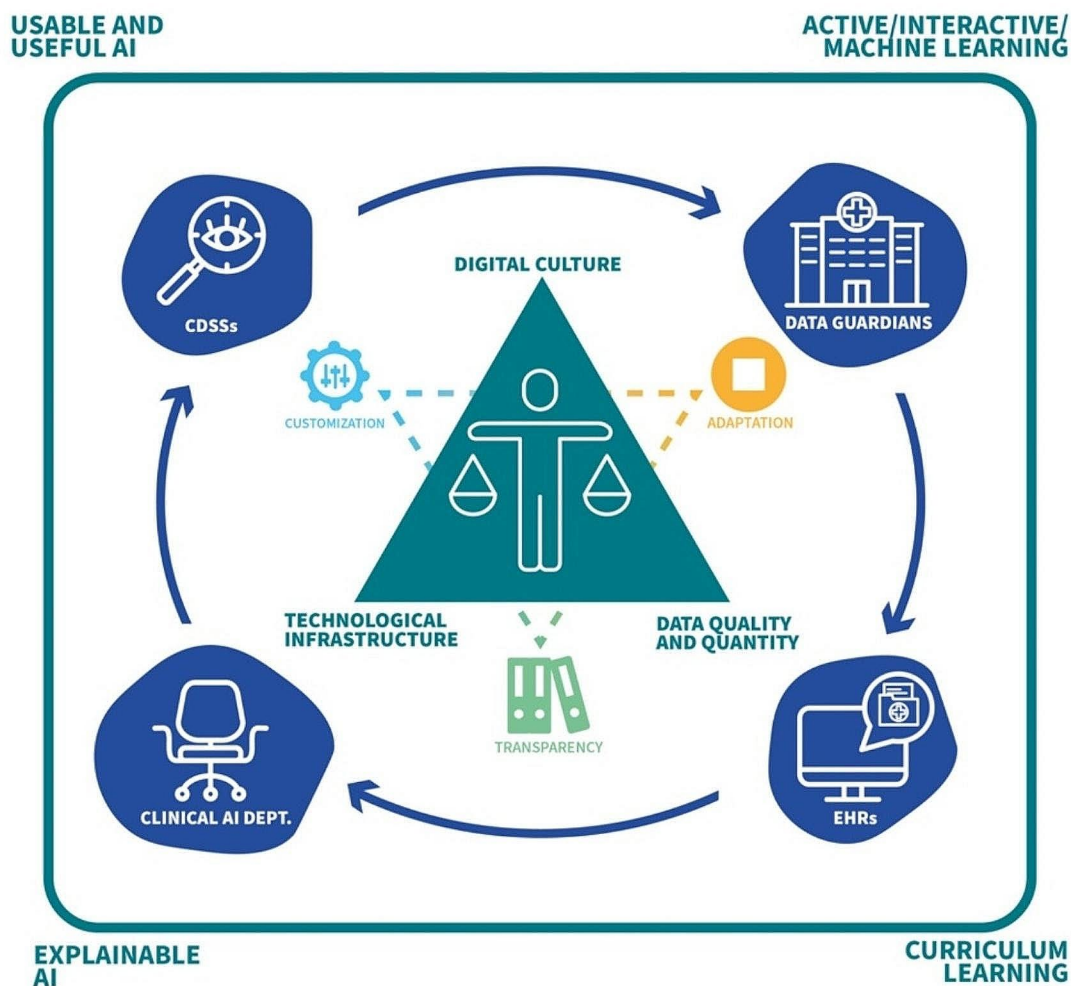


Fig. 2 A Mind Map of AI Application in Healthcare. The looping structure illustrates the various steps in the development and deployment of AI models, starting with data acquisition from hospitals or institutions (data guardians), to the storage of this information in electronic health records (EHRs), to the prototyping and validation phase of AI algorithms within a clinical AI department (CAID). The process culminates in the implementation of clinical decision support systems (CDSSs) that can alter clinical practice. This cycle is built upon the three pillars of digital transformation: data quality and quantity, a sound technological infrastructure, and a nurturing digital culture. Algorithmic

stewardship guides and oversees the various domains: learning with humans, curriculum learning, explainable AI, and Beyond learning – Useful and Usable AI. This entire process aligns with the principles of algor-ethics, putting humans at the center of the process and adhering to the principles of traceability, customization, and adequation. This underscores the human-centered approach of algor-ethics, with AI designed to augment human decision-making and adapt to the unique needs and interests of each individual. It serves as an ethical layer that embraces and interacts with all aspects of the AI development and deployment process

while safeguarding against biases and ensuring equitable patient care This work underscores the importance of algor-ethics as the foundation for AI development in healthcare, balancing technological advancement with the preservation of human values, autonomy, and the complexities of clinical decision-making. The future of healthcare AI lies not just in technological innovation but -further - in its harmonious

integration with human expertise and vision, ensuring that AI serves as a tool for timely enhancement rather than replacement, ultimately contributing to improved patient outcomes and healthcare delivery.

Acknowledgements The authors acknowledge Matilde Pretolesi for drawing the figures of the paper.

Author contributions J.M. has written the first draft of the manuscript. J.M., P.B. and E.G.B. have given substantial contributions to study conception. M.M.B., M.C., E.R., L.R., V.B., F.S., E.G., E.F., V.A. and M.A. revised and provided critical intellectual content to the manuscript. All authors read and approved the final version of the manuscript to be published.

Funding Not applicable.

Data availability Not applicable.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests JM is a co-founder and shareholder of Callisia srl, a University Spin-off at Università Politecnica delle Marche developing a smart bracelet collecting patient data intelligently for real-time visualization and data analysis. The rest of the authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Moazemi S, Vahdati S, Li J, Kalkhoff S, Castano LJV, Dewitz B, et al. Artificial intelligence for clinical decision support for monitoring patients in cardiovascular ICUs: a systematic review. *Front Med.* 2023;10:1109411.
- Benjamens S, Dhunoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *Npj Digit Med.* 2020;3:118.
- Muehlematter UJ, Bluethgen C, Vokinger KN. FDA-cleared artificial intelligence and machine learning-based medical devices and their 510(k) predicate networks. *Lancet Digit Health.* 2023;5:e618–26.
- van de Sande D, van Genderen ME, Huiskens J, Gommers D, van Bommel J. Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. *Intensive Care Med.* 2021;47:750–60.
- van de Sande D, van Bommel J, Fung Fen Chung E, Gommers D, van Genderen ME. Algorithmic fairness audits in intensive care medicine: artificial intelligence for all? *Crit Care.* 2022;26:315.
- McCadden MD, Joshi S, Mazwi M, Anderson JA. Ethical limitations of algorithmic fairness solutions in health care machine learning. *Lancet Digit Health.* 2020;2:e221–3.
- Grote T, Keeling G. Enabling Fairness in Healthcare through Machine Learning. *Ethics Inf Technol.* 2022;24:39.
- Eaneff S, Obermeyer Z, Butte AJ. The case for Algorithmic Stewardship for Artificial Intelligence and Machine Learning Technologies. *JAMA.* 2020;324:1397.
- Mosqueira-Rey E, Hernández-Pereira E, Alonso-Ríos D, Bobes-Bascarán J, Fernández-Leal Á. Human-in-the-loop machine learning: a state of the art. *Artif Intell Rev.* 2023;56:3005–54.
- Friedman TL. Thank you for being late: an Optimist's guide to thriving in the age of accelerations. London, UK: Allen Lane; 2016.
- Nakanishi R, Okubo R, Sobue Y, Kaneko U, Sato H, Fujimoto S, Nozaki Y, Kajiya T, Miyoshi T, Ichikawa K, Abe M, Kitagawa T, Ikenaga H, Osawa K, Saji M, Iguchi N, Nakazawa G, Takahashi K, Ijich T, Mikamo H, Kurata A, Moroi M, Iijima R, Malkasian S, Crabtree T, Chamie D, Alexandra LJ, Min JK, Earls JP, Matsuo H. Rationale and design of the INVICTUS Registry: (Multicenter Registry of Invasive and Non-Invasive imaging modalities to compare Coronary Computed Tomography Angiography, Intravascular Ultrasound and Optical Coherence Tomography for the determination of Severity, Volume and Type of coronary atherosclerosis). *J Cardiovasc Comput Tomogr.* 2023 Sep 5:S1934-5925(23)00427-6.
- Gerard SE, Herrmann J, Xin Y, Martin KT, Rezoagli E, Ippolito D, Bellani G, Cereda M, Guo J, Hoffman EA, Kaczka DW, Reinhardt JM. CT image segmentation for inflamed and fibrotic lungs using a multi-resolution convolutional neural network. *Sci Rep.* 2021;11(1):1455.
- Connell M, Xin Y, Gerard SE, Herrmann J, Shah PK, Martin KT, Rezoagli E, Ippolito D, Rajaei J, Baron R, Delvecchio P, Humayun S, Rizi RR, Bellani G, Cereda M. Unsupervised segmentation and quantification of COVID-19 lesions on computed tomography scans using CycleGAN. *Methods.* 2022;205:200–9.
- Maddali MV, Churpek M, Pham T, Rezoagli E, Zhuo H, Zhao W, He J, Delucchi KL, Wang C, Wickersham N, McNeil JB, Jauregui A, Ke S, Vessel K, Gomez A, Hendrickson CM, Kangelaris KN, Sarma A, Leligdowicz A, Liu KD, Matthay MA, Ware LB, Laffey JG, Bellani G, Calfee CS, Sinha P. LUNG SAFE investigators and the ESICM Trials Group. Validation and utility of ARDS subphenotypes identified by machine-learning models using clinical data: an observational, multicohort, retrospective analysis. *Lancet Respir Med.* 2022;10(4):367–77.
- Stephens AF, Šeman M, Diehl A, Pilcher D, Barbaro RP, Brodie D, Pellegrino V, Kaye DM, Gregory SD, Hodgson C. Extracorporeal Life Support Organization Member centres. ECMO PAL: using deep neural networks for survival prediction in venoarterial extracorporeal membrane oxygenation. *Intensive Care Med.* 2023;49(9):1090–9.
- Chen YY, Liu CF, Shen YT, Kuo YT, Ko CC, Chen TY, Wu TC, Shih YJ. Development of real-time individualized risk prediction models for contrast associated acute kidney injury and 30-day dialysis after contrast enhanced computed tomography. *Eur J Radiol.* 2023;167:111034.
- Avidan A, Sprung CL, Schefold JC, Ricou B, Hartog CS, Nates JL, et al. Variations in end-of-life practices in intensive care units worldwide (Ethicus-2): a prospective observational study. *Lancet Respiratory Med.* 2021;9:1101–10.
- Denney MJ, Long DM, Armistead MG, Anderson JL, Conway BN. Validating the extract, transform, load process used to populate a large clinical research database. *Int J Med Inf.* 2016;94:271–4.
- Quiroz JC, Chard T, Sa Z, Ritchie A, Jorm L, Gallego B. Extract, transform, load framework for the conversion of health databases to OMOP. Deserno TM, editor. *PLoS ONE.* 2022;17:e0266911.
- Henke E, Peng Y, Reinecke I, Zoch M, Sedlmayr M, Bathelt F. An extract-transform-load process design for the Incremental Loading of German Real-World Data based on FHIR and OMOP

- CDM: Algorithm Development and Validation. *JMIR Med Inf*. 2023;11:e47310.
21. Unberath P, Prokosch HU, Gründner J, Erpenbeck M, Maier C, Christoph J. EHR-Independent Predictive decision Support Architecture based on OMOP. *Appl Clin Inf*. 2020;11:399–404.
 22. Fleuren LM, Dam TA, Tonutti M, De Bruin DP, Lalisang RCA, Gommers D, et al. The Dutch Data Warehouse, a multicenter and full-admission electronic health records database for critically ill COVID-19 patients. *Crit Care*. 2021;25:304.
 23. Nazer LH, Zatarah R, Waldrip S, Ke JXC, Moukheiber M, Khanna AK et al. Bias in artificial intelligence algorithms and recommendations for mitigation. Kalla M, editor. *PLOS Digit Health*. 2023;2:e0000278.
 24. McNicholas BA, Madotto F, Pham T, Rezoagli E, Masterson CH, Laffey JG, LUNG SAFE Investigators, the ESICM Trials Group. Demographics, management and outcome of females and males with acute respiratory distress syndrome in the LUNG SAFE prospective cohort study. *Eur Respir J*. 2019;54(4):1900609.
 25. Rezoagli E, McNicholas BA, Madotto F, Pham T, Bellani G, Laffey JG, LUNG SAFE Investigators, the ESICM Trials Group. Presence of comorbidities alters management and worsens outcome of patients with acute respiratory distress syndrome: insights from the LUNG SAFE study. *Ann Intensive Care*. 2022;12(1):42.
 26. Majid Z, Welch C, Davies J, Jackson T. Global frailty: the role of ethnicity, migration and socioeconomic factors. *Maturitas*. 2020;139:33–41.
 27. Bellini V, Montomoli J, Bignami E. Poor quality data, privacy, lack of certifications: the lethal triad of new technologies in intensive care. *Intensive Care Med* 2021;47:1052–3.
 28. Agarwal R, Bjarnadottir M, Rhue L, Dugas M, Crowley K, Clark J, et al. Addressing algorithmic bias and the perpetuation of health inequities: an AI bias aware framework. *Health Policy Technol*. 2023;12:100702.
 29. Floridi L. On human dignity as a Foundation for the right to privacy. *Philos Technol*. 2016;29:307–12.
 30. Pooch EHP, Ballester P, Barros RC et al. Can We Trust Deep Learning Based Diagnosis? The Impact of Domain Shift in Chest Radiograph Classification. In: Petersen J, San José Estépar R, Schmidt-Richberg A, Gerard S, Lassen-Schmidt B, Jacobs C, editors. *Thoracic Image Analysis* [Internet]. Cham: Springer International Publishing; 2020 [cited 2023 Jun 16]. pp. 74–83. https://link.springer.com/https://doi.org/10.1007/978-3-030-62469-9_7.
 31. Zhang A, Xing L, Zou J, Wu JC. Shifting machine learning for healthcare from development to deployment and from models to data. *Nat Biomed Eng*. 2022;6:1330–45.
 32. Farahani A, Voghoei S, Rasheed K, Arabnia HR. A Brief Review of Domain Adaptation. In: Stahlbock R, Weiss GM, Abou-Nasr M, Yang C-Y, Arabnia HR, Deligiannidis L, editors. *Advances in Data Science and Information Engineering* [Internet]. Cham: Springer International Publishing; 2021 [cited 2023 Jun 16]. pp. 877–94. https://link.springer.com/https://doi.org/10.1007/978-3-030-71704-9_65.
 33. Nelson GS. Bias in Artificial Intelligence. *N C Med J*. 2019;80:220–2.
 34. Croskerry P. A Universal Model of Diagnostic reasoning. *Acad Med*. 2009;84:1022–8.
 35. Kahneman D. *Thinking, fast and slow*. London: Penguin Books; 2012.
 36. Gaube S, Suresh H, Raue M, Merritt A, Berkowitz SJ, Lermer E, et al. Do as AI say: susceptibility in deployment of clinical decision-aids. *Npj Digit Med*. 2021;4:31.
 37. Vicente L, Matute H. Humans inherit artificial intelligence biases. *Sci Rep*. 2023;13:15737.
 38. Sujan M, Furniss D, Grundy H, Nelson D, Elliott M, et al. Human factors challenges for the safe use of artificial intelligence in patient care. *BMJ Health Care Inf*. 2019;26:e100081.
 39. Benanti P. *Homo Faber: The Techno-Human condition* [Internet]. EDB - Edizioni Dehoniane Bologna; 2018. <https://books.google.it/books?id=7-wCEAAQBAJ>.
 40. Adler-Milstein J, Chen JH, Dhaliwal G. Next-generation Artificial Intelligence for diagnosis: from Predicting Diagnostic labels to Wayfinding. *JAMA*. 2021;326:2467.
 41. Montomoli J, Rezoagli E, Bellini V, Finazzi S, Bignami EG. A generalized wayfinding paradigm for improving AKI understanding and classification: insights from the Dutch registries. *Minerva Anesthesiol*. 2023.
 42. Celi LA, Cellini J, Charpignon M-L, Dee EC, Dernoncourt F, Eber R et al. Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review. Fraser HS, editor. *PLOS Digit Health*. 2022;1:e0000022.
 43. Montomoli J, Hilty MP, Ince C. Artificial intelligence in intensive care: moving towards clinical decision support systems. *Minerva Anesthesiol*. 2022.
 44. Blanchard MD, Kleitman S, Aidman E. Are two naïve and distributed heads better than one? Factors influencing the performance of teams in a challenging real-time task. *Front Psychol*. 2023;14:1042710.
 45. Koriati A. When two heads are better than one and when they can be worse: the amplification hypothesis. *J Exp Psychol Gen*. 2015;144:934–50.
 46. Aporta C, Higgs E. *Satellite Culture: Global Positioning Systems, Inuit Wayfinding, and the need for a New Account of Technology*. *Curr Anthropol*. 2005;46:729–53.
 47. Cosgriff CV, Stone DJ, Weissman G, Pirracchio R, Celi LA. The clinical artificial intelligence department: a prerequisite for success. *BMJ Health Care Inf*. 2020;27:e100183.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.