Cureus
Part of SPRINGER NATURE

Open Access Technical Report

Bioethics Artificial Intelligence Advisory (BAIA): An Agentic Artificial Intelligence (AI) Framework for Bioethical Clinical Decision Support

Taposh P. Dutta Roy [1, 2]

1. Responsible AI, Kaiser Permanente, Oakland, USA 2. Bioethics, Harvard Medical School, Boston, USA

Corresponding author: Taposh P. Dutta Roy, taposh.dr@gmail.com

Review began 02/24/2025
Review ended 03/10/2025
Published 03/12/2025

DOI: 10.7759/cureus.80494

Abstract

Healthcare professionals face complex ethical dilemmas in clinical settings in cases involving end-of-life care, informed consent, and surrogate decision-making. These nuanced situations often lead to moral distress among care providers. This paper introduces the Bioethics Artificial Intelligence Advisory (BAIA) framework, a novel and innovative approach that leverages artificial intelligence (AI) to support clinical ethical decision-making. The BAIA framework integrates multiple bioethical approaches, including principlism, casuistry, and narrative ethics, with advanced AI capabilities to provide comprehensive decision support. The framework employs a structured methodology that includes data collection, paradigmatic case review, analysis through "mattering maps," and scenario-based decision reasoning. A detailed analysis of two challenging cases, an end-of-life care decision and a complex conjoined twins case, demonstrates BAIA's potential to harmonize diverse ethical perspectives while reducing the moral burden on healthcare providers. The framework's agentic architecture additionally allows integration with any new and existing ethical AI systems like METHAD, Delphi, and EAIFT, enabling multiframework collaboration. This work also acknowledges limitations related to data quality, bias, and complexity of ethical decisions and proposes mitigation strategies, including standardized databases, fairness algorithms, and maintaining human oversight. Thus, this work represents a significant step toward combining technological advancement in agentic AI with established bioethical principles to improve the quality and consistency of clinical ethical decision-making, thus reducing moral distress for clinicians.

Categories: Neurology, Pediatrics, Healthcare Technology
Keywords: agentic ai, ai, ai bioethics, bioethics framework, bioethics recommendations

Introduction

The integration of analytics in healthcare traces back to 1854 when Dr. John Snow [1] first illustrated the use of systematic data analysis to mark the end of cholera in London. In the following 170 years, significant advances have emerged in medicine and computer science. While technology advances, clinical decision-making remains particularly challenging when healthcare teams face ambiguous, emotional, and complex decisions involving end-of-life care, informed consent [2], surrogate decision-making [3], genetics [4], futility [5], harm principle [6], and others. These decisions impact the care team and lead to moral distress [7], residue [8], and injury. The potential of artificial intelligence (AI) to serve as an advisor to support decision-making [9] and reduce moral impact can significantly benefit the team, patient, and family. This essay proposes an innovative Bioethics Artificial Intelligence Advisory (BAIA) framework to augment human reasoning in clinical decision-making. BAIA complements healthcare teams in navigating complex ethical dilemmas by integrating bioethical approaches, including principlism, casuistry, narrative ethics, and agentic AI capabilities. Through the analysis of two challenging cases, an end-of-life care decision and a complex conjoined twin's case, this framework demonstrates its potential to harmonize diverse ethical perspectives, reduce moral distress and moral burden on the care providers, and enhance the quality and consistency of decisions in highly complex and emotional clinical environments.

Technical Report

Agentic AI system

An AI system is trained on a large amount of data and learns statistical patterns to predict the next word in a sequence [10]. When enhanced with the capability of invoking other programs, these are called "Agents". Chawla et al. [11] define "Agentic AI" as a framework in which large language models enable workflows, supporting four capabilities: tool usage for accuracy enhancement using external sources, self-correction, structured task breakdown, and multi-model collaboration. As of this writing, there are three distinct AI systems for ethical decision-making: DELPHI [12], Medical Ethics Advisor (METHAD) [13], and Ethical Artificial Intelligence Framework Theory (EAIFT) [14]. EAIFT embeds ethical reasoning within AI systems to guarantee their ethical operation. DELPHI is a framework for moral reasoning, leveraging AI to determine ethically acceptable actions. METHAD focuses on clinical ethics dilemmas and models Beauchamp and Childress's (B&C) [15] autonomy, beneficence, and nonmaleficence utilizing fuzzy cognitive maps (FCMs)

[16], a method for modeling cause-and-effect relationships and interconnected concepts. However, METHAD misses the concrete case-specific details and narrative approaches [17], which add context to the family and patient's perspective. Each of the above systems uses a different methodology, with its strengths and weaknesses, and investigates different sides of ethical AI (Figure *1*).

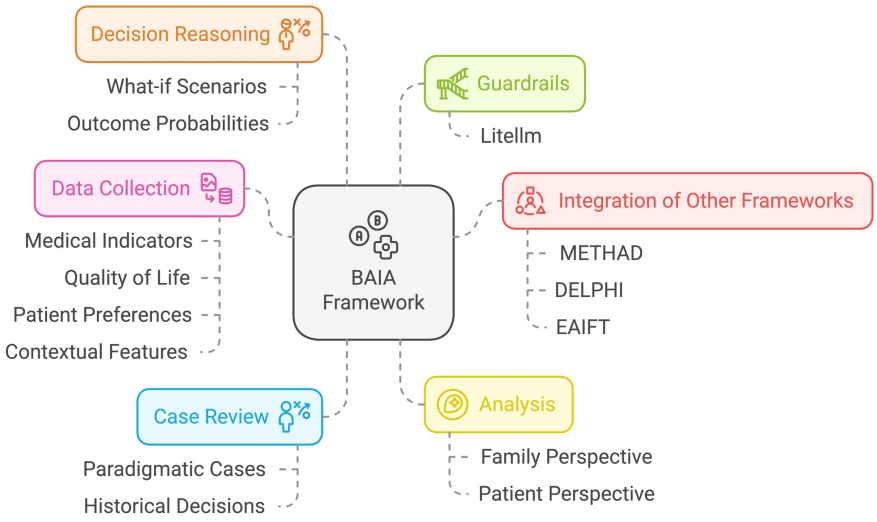

FIGURE 1: Components and capabilities of the BAIA framework

BAIA: Bioethics Artificial Intelligence Advisory

Figure credits: Taposh Dutta Roy, image created using napkin.ai

BAIA framework

Today, clinical ethics teams use principlism as outlined by B&C to support complex, time-sensitive, and strenuous healthcare decisions. B&C's principlism [15] has stood the test of time and provides a robust yet abstract approach to ethical decision-making. Other moral theories, such as casuistry [18] and narrative ethics [17], provide case-level details and storytelling to make the decision-making approach concrete. Current frameworks such as METHAD follow principlism, DELPHI leverages AI for moral reasoning, and EAIFT embeds ethical decision-making in AI. This work proposes BAIA, a novel framework developed in response to the limitations of existing AI-driven ethical decision-making tools. BAIA uses a scaleable agentic AI strategy that incorporates B&C's principlism [15], casuistry [18], and narrative ethics [17]. BAIA expands casuistry's first step, "topics or case container," [18] to collect data for medical indicators, quality of life, patient preferences, and contextual features by adding features from narrative ethics such as storytelling and extracting data on voice, character, plot, and resolution [17]. Next, we review the paradigmatic [18] cases so we can learn from past decisions. A paradigmatic case review is part of casuistry, where one reviews a past case similar to the case in hand to get a historical perspective. The third step is "analysis," developing "mattering maps [19]," a narrative ethics concept used for the representation of the family and patient's perspective of what is most important in their life and how they got to this point. It also weighs B&C's principles based on the data available from prior steps. The fourth step is decision reasoning, where, based on the information, the system develops "what-if" capability for the scenarios and their probability of outcomes. Additional methods and theories, such as deontology, utilitarianism, etc., can be added to the final step to incorporate different viewpoints. The BAIA becomes one agent in our agentic strategy, while METHAD, DELPHI, and EAIFT become other decision-making agents. As more frameworks evolve, our agentic system can easily be expanded to incorporate newer details. Additionally, we define guardrails such as bias and drift detection, human-in-the-loop oversight, and explainability mechanisms for the agentic framework utilizing the open-source LiteLLM [20]. With BAIA's multimodel agentic capability, we can review clinical cases and provide advisory data points (Figure *2*).

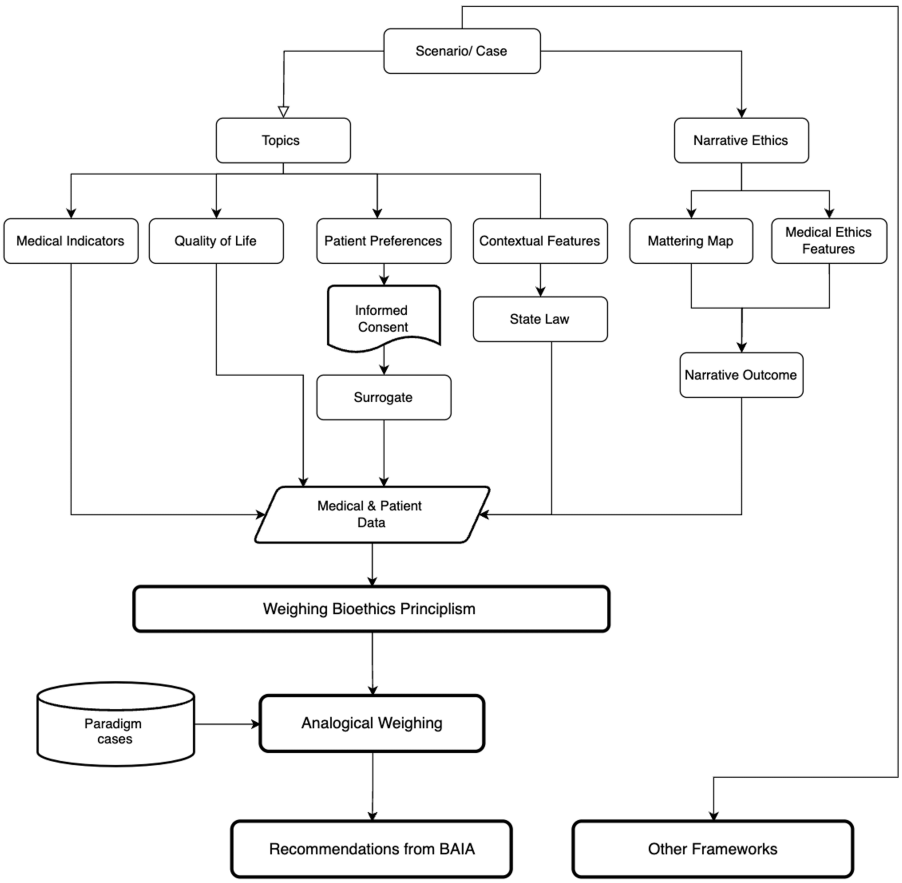

FIGURE 2: BAIA framework details

BAIA: Bioethics Artificial Intelligence Advisory

Figure credits: Taposh Dutta Roy, image created using the flowcharting tool draw.io

Discussion

Case analysis

Case 1: End-of-Life Care

Consider the case of a 68-year-old male patient [21] with severe impairments, myocardial infarction, stroke, hemiplegia, and multiple organ failure. His family insisted on “full code,” including aggressive life-prolonging interventions such as cardiopulmonary resuscitation (CPR) to save his life, despite the physician’s view that these may be futile. The hospital requested the Court of Protection to withhold CPR, invasive hemodynamic support, and renal replacement therapy in the event of future degradation, which was rejected. Applying the proposed BAIA framework in this situation, the ethics team gathers topical data, such as medical indicators, quality of life, patient preferences, and contextual features, and conducts narrative interviews with family members, physicians, and nursing leaders to understand the voice, character, plot, and resolution. The BAIA framework reveals the family's emotional motivations and cultural beliefs through these interviews. Next, the framework will look for a similar case from the past; if one is found, it will become the "paradigmatic" case utilized in this context. The system analyzes two key points. First, it creates “mattering maps” that highlight the moral weight of prolonging life versus alleviating suffering from the perspectives of both patient and family. Second, it evaluates the principles of beneficence and nonmaleficence to develop a balanced scorecard, and finally, it formulates a “what-if” analysis, which is a simulation capability that provides outcomes and explanations through scenario modeling. For example, one scenario could involve discontinuing futile interventions and transitioning to palliative care, while another might consider continuing with “full code” treatment. The patient's family insisted on doing everything possible to save him. This situation falls under “positive rights,” where patients have the right to receive medical care but do not have the right to interventions that exceed appropriate medical care. The BAIA framework will take into account the family’s perspective along with all case details, such as topical data and paradigm cases. This structured process ensures that the family's voice is acknowledged while adhering to the ethical principles of beneficence and nonmaleficence. Additionally, the BAIA framework will

also seek guidance from other systems such as METHAD and DELPHI. The BAIA framework synthesizes all information to provide recommendations and the potential to run additional scenarios and consult other frameworks or approaches. Utilizing this AI framework would reduce moral distress, enhance quality, and bring consistency to decision-making for the patient.

Case 2: Conjoined Twins

Cummings et al. published a case report about 22-month-old conjoined twins (“Twin A” and “Twin B”) [22], highlighting the tension between medical possibility and ethical boundaries. The twins were born in East Africa and arrived at Massachusetts General Hospital for evaluation of separation. They shared a single liver, an abdominal cavity, and a portion of their gastrointestinal tract. Twin B was larger and healthier, while Twin A had complex congenital heart disease and relied on her sibling’s circulation for support. Unfortunately, Twin A’s condition worsened, which required the twins to be admitted to the pediatric intensive care unit for stabilization and treatment. Applying our proposed BAIA framework to this case, specific medical, quality of life, and contextual information such as Islamic religion and advice from their local Imam are obtained. Since the patients are pediatric, parental consent was a necessity for any intervention. The ethics team conducts narrative interviews with parents and other care providers. Given the rarity of conjoined twins, we may not find a good paradigmatic case. The system will develop from the parent’s perspective a “mattering map” and analyze the case considering various concepts such as beneficence, nonmaleficence, the doctrine of double effect [23,24], pediatric informed consent, self-driven car facing a choice between hitting someone on a crosswalk or killing themselves, etc. The decision-reasoning step will provide a recommendation and the ability to do a scenario analysis considering various possibilities. The BAIA framework will evaluate various perspectives [22], including each twin’s likelihood of survival, the parents' religious beliefs, and refusal of surgery. Additionally, it will take into account the doctrine of double effect, which recognizes the intention to act in the best interest while acknowledging that “Twin A” may not survive, effectively designating her as a “marked for death” patient [22]. BAIA will provide an advisory recommendation and explanation that respects the family's values, thus reducing the moral distress and providing consistent decisions for the case.

BAIA strengths

Using the two cases, we show that the proposed BAIA framework provides a comprehensive and structured approach to making complex treatment decisions. It utilizes data from the case, narrative stories, and a principled approach. The framework’s reliance on data collection ensures that all pertinent information, such as medical information, quality of life, patient preferences, contextual information, and narrative stories, is collected. Incorporating a paradigmatic case ensures that we draw insights from similar scenarios in the past. In the analysis phase, we develop “mattering maps” [17], a patient perspective on “how they got here” and what their wishes are, adding depth and developing a human context. Further, the ability to do scenario analysis provides ways to plan the situation and weigh the pros and cons of each. Compared to existing frameworks METHAD [13] and DELPHI [12], BAIA provides concrete case-specific depth, reasoning, and a data-driven approach. Finally, using an agentic technology makes the BAIA framework expandable and additive to any new approach.

BAIA opportunities

Despite its comprehensive approach, BAIA has several limitations. First, its analysis relies on high-quality, unbiased, and comprehensive datasets, which can be challenging due to access issues or incomplete data capture. Second, the outcomes of the BAIA algorithm must be appropriate, fair, and unbiased. Validating these outcomes is increasingly important for BAIA, as it can be complex to determine the correct answer. Third, ethical decisions are multifaceted and nuanced, which AI systems might oversimplify. We should set up the following tools and strategies to mitigate these limitations. First, develop a standardized database with diverse anonymized cases. These cases should be revisited for validation and appropriately tagged if they contribute to any decision-making. Second, fairness and bias [25] detection algorithms should be established to validate the outcomes. The validation strategy should include model outcome explanation methods such as SHAPley Additive exPlanations (SHAP) [26], causality [27,28], and counterfactual [29] analysis. Furthermore, every outcome report should contain a probability of consideration and a reasoning-based chain of thought [30] informing the decision recommendation. Additionally, a “human in the loop [31]” approach will ensure that care professionals remain central to the decision-making process. Finally, the time and resources required to use the framework could limit its feasibility in time-sensitive situations. Addressing these limitations, including a standardized database of anonymized cases, data fairness, bias detection, explainable outcomes, and “humans in the loop,” will enhance BAIA’s ability to support complex decisions while upholding human values.

Conclusions

Safeguarding patient well-being and preserving human values are at the heart of healthcare. This theoretical approach utilizes the latest technological advancements, such as large language models and agentic AI, to develop a solution for nuanced real-world problems. It builds on the work done by prior scholars and develops a comprehensive system that looks at abstract bioethical principles and case-specific details to provide advisory support. Further, the ability to extend the framework to existing developed methods makes

it flexible to adjust and scale. In this report, we analyze two real cases, one in pediatrics and one for end-of-life. We showcase how the BAIA framework can reduce moral distress on the care providers, harmonize differing perspectives, and enhance the quality and consistency of decisions. In highly emotional and critical scenarios, this advice from BAIA might bring a rational angle to advising surrogates and their families. Our next step is to apply this framework in real time to actual cases, validate outcomes, and establish baseline measures to assess its impact on moral distress and ethical residue.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Taposh P. Dutta Roy

Acquisition, analysis, or interpretation of data: Taposh P. Dutta Roy

Drafting of the manuscript: Taposh P. Dutta Roy

Critical review of the manuscript for important intellectual content: Taposh P. Dutta Roy

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

I am grateful to Leanne Homan, RN, BSN, MBE, Associate Director of Clinical Ethics at Harvard Medical School Center for Bioethics, for her invaluable review and guidance throughout the publication process, which significantly improved the quality of this work. I also extend my sincere appreciation to Dr. Anthony Breu, MD, Assistant Professor in Medicine, Harvard Medical School and Dr. Brian M. Cummings, MD, Assistant Professor of Pediatrics, Massachusetts General Hospital, for their instruction on clinical ethics at Harvard Medical School, which provided both the foundational knowledge and the motivation to pursue this research. Their insights and encouragement have been instrumental in shaping the development of this paper.

References

1. Tulchinsky TH: John Snow, cholera, the broad street pump; waterborne diseases then and now . Case Studies in Public Health. Elsevier, Amsterdam; 2018. 77-99. 10.1016/B978-0-12-804571-8.00017-2
2. Katz AL, Webb SA: Informed consent in decision-making in pediatric practice . Pediatrics. 2016, 138:e20161485. 10.1542/peds.2016-1485
3. Berger JT, DeRenzo EG, Schwartz J: Surrogate decision making: reconciling ethical theory and clinical practice. Ann Intern Med. 2008, 149:48-53. 10.7326/0003-4819-149-1-200807010-00010
4. Garrison NA: Genomic justice for Native Americans: impact of the Havasupai case on genetic research . Sci Technol Human Values. 2013, 38:201-23. 10.1177/0162243912470009
5. Misak CJ, White DB, Truog RD: Medical futility: a new look at an old problem . Chest. 2014, 146:1667-72. 10.1378/chest.14-0513
6. Diekema DS: Parental refusals of medical treatment: the harm principle as threshold for state intervention . Theor Med Bioeth. 2004, 25:243-64. 10.1007/s11017-004-3146-6
7. Morley G, Ives J, Bradbury-Jones C, Irvine F: What is 'moral distress'? A narrative synthesis of the literature . Nurs Ethics. 2019, 26:646-62. 10.1177/0969733017724354
8. Booth Adam T., Christian, Becky J. : Surgical intensive care unit nurses' coping with moral distress and moral residue: a descriptive qualitative approach. Dimens Crit Care Nurs. 2024, 43:298-305. 10.1097/DCC.0000000000000665
9. Grady D: Artificial intelligence as surrogate decision-maker. JAMA Intern Med. 2024, 184:1007. 10.1001/jamainternmed.2024.2679
10. Telenti A, Auli M, Hie BL, Maher C, Saria S, Ioannidis JP: Large language models for science and medicine . Eur J Clin Invest. 2024, 54:e14183. 10.1111/eci.14183
11. Chawla C, Chatterjee S, Gadadinni SS, Verma P, Banerjee S: Agentic AI: the building blocks of sophisticated AI business applications. AIRWA. 2024, 3:196-210. 10.69554/XEHZ1946
12. Can machines learn morality? The Delphi experiment . (2021). https://www.semanticscholar.org/paper/CAN-MACHINES-LEARN-MORALITY-THE-DELPHI-EXPERIMENT-Jiang-Bhagavatula/6ef1edae425....

13. Meier LJ, Hein A, Diepold K, Buyx A: Algorithms for ethical decision-making in the clinic: a proof of concept. Am J Bioeth. 2022, 22:4-20. 10.1080/15265161.2022.2040647
14. Ejjami R: Ethical artificial intelligence framework theory (EAIFT): a new paradigm for embedding ethical reasoning in AI systems. Int J Multidiscip Res. 2024, 6:10.36948/ijfmr.2024.v06i05.28231
15. Beauchamp TL, Childress JF: Principles of Biomedical Ethics . Oxford University Press, Oxford (UK); 2019.
16. Hein A, Meier LJ, Buyx AM, Diepold K: A fuzzy-cognitive-maps approach to decision-making in medical ethics. IEEE Xplore. 2022, 1-8. 10.1109/FUZZ-IEEE55066.2022.9882615
17. Montello M: Narrative ethics. Hastings Cent Rep. 2014, 44:S2-6. 10.1002/hast.260
18. Tomlinson T: Chapter 7, casuistry & clinical ethics . United Kingdom Methods in Medical Ethics: Critical Perspectives. Oxford University Press , Oxford (UK); 2012.
19. McCarthy J: Principlism or narrative ethics: must we choose between them? . Med Humanit. 2003, 29:65-71. 10.1136/mh.29.2.65
20. LiteLLM Guardrails. (2024). Accessed: February 20, 2025: https://docs.litellm.ai/docs/proxy/guardrails.
21. Szawarski P: Classic cases revisited: Mr David James, futile interventions and conflict in the ICU . J Intensive Care Soc. 2016, 17:244-51. 10.1177/1751143716628885
22. Cummings BM, Gee MS, Benavidez OJ, Shank ES, Bojovic B, Raskin KA, Goldstein AM: Case 33-2017. 22-month-old conjoined twins. N Engl J Med. 2017, 377:1667-77. 10.1056/NEJMcpc1706105
23. Regnard C, George R, Grogan E, et al.: So, farewell then, doctrine of double effect . BMJ. 2011, 343:d4512. 10.1136/bmj.d4512
24. Boyle JM: Toward understanding the principle of double effect . Ethics. 1980, 90:527-38. 10.1086/292183
25. Carey S, Pang A, Kamps M: Fairness in AI for healthcare . Future Healthc J. 2024, 11:100177. 10.1016/j.fhj.2024.100177
26. Edin J, Maistro M, Maaløe L, Borgholt L, Havtorn JD, Ruotsalo T: An unsupervised approach to achieve supervised-level explainability in healthcare records. Future Healthc J. 2024, 10.48550/arXiv.2406.08958
27. Sontag D, Johansson F: AI for health needs causality. Broad Institute. 2018, Accessed: February 20, 2025: https://www.youtube.com/watch?v=MBz9hVFYDl8.
28. Sanchez P, Voisey JP, Xia T, Watson HI, O'Neil AQ, Tsaftaris SA: Causal machine learning for healthcare and precision medicine. R Soc Open Sci. 2022, 9:220638. 10.1098/rsos.220638
29. Pfohl S, Duan T, Ding DY, Shah NH: Counterfactual reasoning for fair clinical risk prediction . Proc Mach Learn Res. 2019, 106:325-58. 10.48550/arXiv.1907.06260
30. Jason Wei, Xuezhi Wang, Dale Schuurmans et.al: Chain-of-thought prompting elicits reasoning in large language models. NeurIPS. 2022, 1800:24824-37.
31. Mosqueira-Rey E, Hernández-Pereira E, Alonso-Ríos D, , Bobes-Bascarán J, Fernández-Leal A: Human-in-the-loop machine learning: a state of the art. Artif Intell Rev. 2023, 56:3005-54. 10.1007/s10462-022-10246-w