Special Issue on Human-Centered Artificial Intelligence for One Health

Paolo Buono[a*], Nadia Berthouze[b], Maria Francesca Costabile[a], Adela Grando[c], Andreas Holzinger[d]

a. *University of Bari, Bari, Italy*
b. *University College London, London, UK*
c. *College of Health Solutions, Arizona State University, (AZ, USA)*
d. *University of Natural Resources and Life Sciences Vienna, Austria*

- *corresponding author*

Foreword

There is a lot of talk about Artificial Intelligence (AI) these days, in different contexts: scientific, economic, social, etc. Indeed, AI technologies can significantly increase the efficiency of processes in many areas and help people with everyday problems. However, there is still a lack of clarity as to how the benefits of AI can be effectively exploited and, at the same time, provide reliability, security, traceability, explainability, and thus the trustworthiness of the system for humans. AI is usually viewed from the perspective of autonomy: from autonomous behavior to autonomous decision-making. In various situations, a high degree of autonomy can be useful (although not without risks).

The new discipline “Human-Centered Artificial Intelligence” (HCAI) aims to employ AI as a means to support and facilitate humans by complementing (and valuing) human cognitive abilities rather than replacing them. Thus, not only are humans the focus of the design of innovative systems, but they also represent the ultimate goal of each innovative solution. The HCAI systems should complement, empower, and in no way replace human capabilities; in particular, they should be more reliable, safe, and trustworthy for humans. HCAI acknowledges that AI systems should allow a high level of automation while still guaranteeing appropriate levels of human control [1]. To ensure that such properties are maintained in the long run, strategies must be provided to detect and correct errors and common AI system issues (e.g., concept drifting). To this end, interactive machine learning introduces a feedback system that allows humans to annotate data points on the fly, improving the overall performance of the system [2]. However, especially in the context of medicine, explainability is of crucial importance (see, e.g., [3, 4]). Explainable Interactive Machine Learning, by introducing explainability within the loops of interactive machine learning, allows establishing a bi-directional communication between AI and humans (through explanations provided by the AI system and new data provided by humans, respectively [5]) and a consequent collaboration, which allows not only to improve the AI outcomes through manual intervention, but also to improve humans’ knowledge through AI explanations [6]. This collaboration is thus essential to achieve Human-Machine Symbiosis, as defined some years ago [7]. New human-machine interfaces play a significant role to ensure that humans can understand a particular machine explanation [8], thus enabling a symbiotic human-AI collaboration.

Symbiotic AI is very relevant for the new HCAI: it requires a progressive and deeper integration between human and AI, i.e., a symbiosis of such two forms of intelligence, where both humans and AI augment each other's capabilities thanks to a collaboration that balances each other's strengths and weaknesses. Symbiotic AI systems are AI-enabled systems that improve themselves through users, while also providing a way for users to improve themselves. However, to create systems that are ethical and sustainable, this symbiosis requires users to be in control. Only in this way, the objective of creating systems that ensure reliability, safety, and trustworthiness, supporting humans rather than replacing them, will be reached.

The aim of this Special Issue was to provide an opportunity to demonstrate the progress made in the HCAI methods and tools and their application to the maintenance of One Health. The World Health Organization (WHO) definition of "One Health" calls for the collaborative efforts of multiple disciplines, working locally, nationally, and globally, to attain optimal health for humans and the whole ecological environment. However, most of the 33 submitted articles focused on medical systems rather than on well-being. They were peer-reviewed by at least two reviewers and mostly in rounds of two to comply with the high-quality standards of the journal. At the end, 10 articles were selected and published in this special issue. They present analyses of users' requirements and new work practices emerging from the use of AI in medical and healthcare domains, as well as lessons learned from experimental studies of HCAI systems.

The 10 articles are briefly described here, organized into two groups according to their main topics and goals:

1) *Methods and tools supporting humans in decision-making and in better understanding decisions of AI-driven applications* [9, 12, 13, 14, 15, 19];
2) *Methods and algorithms enabling a better human-AI collaboration* [20, 21, 22, 23].

Since the 6 articles in group 1) primarily focus on supporting decision-making, they discuss different ways of providing explanations about system behaviors. Several of the 10 articles also report about user research, performed to understand the way clinicians want to interact with AI and what kind of support they expect from an AI application in their medical activities.

Göndöcs and Dörfler remark that AI has long been regarded as a panacea for decision-making, i.e. something that helps humans get rid of their shortcomings [9]. Instead, there is increasing consensus that AI can provide useful support to decision-makers, but in no way it should replace them. Indeed, decision-making is not solely algorithmic analysis, but it involves other very human factors, such as creativity, intuition, emotions, feelings, and value judgments. Such findings come out from semi-structured open-ended interviews conducted with 17 dermatologists about their expectations and experience involving AI in the diagnostic process of melanoma. The results of this user study are presented organized around four aggregated dimensions: role of AI, responsibility, explainability, and the mindset needed to work with AI. The interviewees expect support from AI but do not believe they might be replaced; as in several other contexts (see, e.g., [10]), humans dislike automation and favor augmentation in many situations, they want to take responsibility for their decisions. Explainability usually refers to the possibility to understand how a particular decision has been reached. It involves transparency and traceability of black-box ML/DL models [11]. Their study suggests that physicians are also interested to another level of AI explainability: they want to understand the AI that they use, what it is like and how it generally works. Finally, the need for a new mindset emerges from the collected data and a new way of diagnosing, which benefits from the physicians' knowledge as well as the AI processing power; this invokes the symbiotic AI to maximize the joint performance of human and AI.

Jin et al. conducted a nation-wide clinical study in Canada on the glioma grading task to assess the utility of AI assistance and its explanations [12]. More specifically, the authors aimed to investigate if and how AI predictions, obtained with state-of-the-art AI methods, may assist physicians and improve their task performance, and also to assess the utility of the explanations provided by the AI on the glioma grading tasks. The study involved 35 neurosurgeons, each one was asked to read a set of 25 brain MRIs of patients with glioma and to give her/his judgment without and with the assistance of AI prediction and explanation. Both quantitative and qualitative data were collected. Interesting results are related to physicians' task performance: they showed that physicians' task performance increased with statistical significance when assisted by AI prediction but remained at almost the same level with the additional assistance of AI explanations. Thus, the study indicated the clinical utility of AI to assist physicians on the glioma grading task but also showed the limitations of current methods to provide explanations. In fact, the change in physicians' performance when assisted by the AI explanations was insignificant because such explanations did not provide explicit reasons, contexts, or descriptions of clinical features to help doctors discern potentially incorrect AI predictions. Such drawbacks provide useful insights for improving current explanation methods in medicine.

The article by Cabitza et al. discusses a different type of explanations [13], called pro-hoc explanations, which provide support to the users by offering alternative explanations than post-hoc explanations: instead of providing the user with an explanation of the machine's answer after a machine advice (post-hoc explanation), this solution provides a pro-hoc explanation (pro-hoc, from Latin, 'instead of that'), which substitutes the machine advice. In particular, the authors focus on *explanation by example*, i.e. the system presents cases similar to the current one. The objective of the research is to provide decision support via pro-hoc explanations; indeed, the authors investigate the impact of presenting physicians with similar cases as the sole form of explanation. The article presents an exploratory study involving 16 orthopedists, requesting to analyze X-rays; users are provided with similar cases depending on their prior judgment. While the application of this kind of support improved accuracy only marginally, its utility was greatly appreciated by the involved medical practitioners, particularly those with less experience. Participants also reported increased confidence in their final decisions. The main limitation of this study is the small scale. However, the main interest of the study is not in statistical significance, rather its aim is to provide a foundational exploratory analysis that can set the stage for more extensive research.

Lombardi et al. present a Machine Learning (ML) system interpretable by human experts, whose aim is to support the preoperative differential diagnosis of leiomyosarcomas and leiomyomas [14]. The work is based on both clinical data and gynecological ultrasound assessment of 68 patients. To make explanations human-interpretable, natural language explanations are proposed, so that specialists may better understand both the decision-making mechanism of the ML algorithms and the impact of the features on each automatic decision. The ML pipeline was designed according to a human-centered approach, in order to deeply understand physicians' work practice and their needs and demands. Some physicians were also involved in the evaluation of the interactive system, but the authors are aware that a deeper evaluation with a much larger number of domain specialists should be performed.

Explainability in clinical decision-making is again the main topic addressed by Kobayashi et al. [15]. The authors remark that there exists a significant gap between physicians' expectations for model explainability and the actual behavior of these models. They explored the capabilities of attention maps but their experiment did not confirm the anticipated benefit of the explanation in enhancing model reliability. Thus, they warn that superficial explanations could do more harm than good by misleading physicians into relying on uncertain predictions, suggesting that the current state of attention mechanisms should not be overestimated in the context of model explainability. The presented case study was instrumental to discover that explanations, even with higher agreement with physicians do not necessarily lead to better uncertainty estimation. The article confirms the need for improved explainability techniques in clinical decision-making. Actually, several researchers have also remarked on the numerous problems with many post-hoc approaches (see, e.g. [16, 17, 18]), emphasizing the importance of ante-hoc approaches that incorporate mechanisms of *explainability by design.*

The article by Ashraf et al. [19] presents a novel work investigating whether Code-Free Deep Learning (CDSL) is a viable approach for developing pill recognition models. The fundamental idea behind CFDL is to allow more people, especially non-AI experts, to develop models and implement them in tools intended to aid decision-making in clinical settings. The authors highlight the challenges and present viable solutions to achieve optimal performance and real-world applicability of pill detection models. Mobile approaches on the smartphone are pursued. The article shows that using a code-free DL approach is a feasible and cost-effective method for developing AI-based pill recognition tools. The work is a good example of interdisciplinary collaboration that ultimately leads to human-centered AI models with improved practicality; it shows how AI can add value for people and also such models require a human-centered approach to development.

Of the 10 published articles, the remaining 4, discussed in the following, primarily address *methods and algorithms enabling a better human-AI collaboration*. A solution that enables the evaluation of visual-spatial neglect is presented by De Boi et al. [20]. The authors have developed an active learning method (interactive machine learning with the human in the loop) based on Gaussian processes regression to model data gathered from patients. It is an example of human-AI collaboration as both the patient and the algorithm interact directly and the system continuously adapts to the data entered by the patient. During the assessment, the active learning algorithm chooses a new location in the patient's field of vision to perform a measurement, i.e. it chooses a location where a new stimulus should be placed. This location is based on the patient's behavior. After each individual stimulus, the model is retrained with new information. The article describes how the proposed model can be utilized in treatments by using a virtual reality application. The assessment module has been evaluated with clinical trials involving patients in a real-world setting, using a virtual reality application. The results obtained using the described AI-based assessment have been compared with the widely used conventional visuospatial neglect tests employed in the clinical practice. The VR application proved to be more sensitive, confirming its potential as a valuable tool for diagnosing and monitoring visuospatial neglect.

Turchi et al. present a human-centered methodology for the development of an AI-as-a-service platform to facilitate access to personalized healthcare [21]. Indeed, the digital revolution in the healthcare sector has the potential to enable person-centric healthcare models, which consider individual needs of the patients who shift from passive recipients to active participants. The central research question of this article is: "How can HCAI principles be considered in the development of an AI-as-a-Service platform that democratizes access to personalized healthcare?". The approach presented includes a design fiction method that involves clinicians from different fields. The aim is to gather diverse views on the collaborations among medical professionals, patients and AI as well as to envision potential future scenarios and address possible ethical and social challenges. The authors incorporate meta-design principles and explore the possibilities for users to modify the AI system - or rather the AI components - based on their experiences, thus promoting a platform that evolves with the user and considers many different perspectives. The authors point out several limitations. First, the feasibility of the proposed approaches under existing clinical workflow pressures and within limited resources has to be better explored; indeed, the intense workloads and time constraints faced by physicians may hinder the proposed solutions in their daily routines. Moreover, a detailed discussion on the data quality and availability issues, which are pivotal for the effective functioning of AI systems in healthcare, is lacking. Another issue is interoperability, since integration with existing hospital systems as well as the continuous learning and adaptation of AI systems to new medical knowledge and practices remain underexplored.

Wang et al. illustrate an innovative approach that integrates segmentation models and the physicians' expertise [22]; the approach fosters human-AI collaboration through a query strategy, involving physicians in the training process. The article demonstrates the efficacy of the proposed human-AI collaboration, which results in reduced physician annotation time and model training time that lead to efficiency gains in medical image analysis. The experimental results showcase the effectiveness of the proposed human-centric AI approach. The utilization of a query strategy to identify uncertain unlabeled data for expert correction demonstrates a tangible improvement in the system-generated tracheal central lines. Interactive machine learning techniques are used to improve (semi)automated segmentation in medical imaging. Given the complexity of pulmonary tracheal segmentation and the limitations in this area, such as the lack of large datasets, interactive machine learning appears a viable way to utilize AI systems.

The article by Desolda et al. remarks that a comprehensive conceptual framework able to guide designers of HCAI systems is still lacking [23]. Indeed, there is the need of design models and methods that promote explainability techniques adequate for domain experts rather than AI experts, as well as the need of interaction paradigms fostering collaboration of humans and AI. Decision-making is a negotiation process that empowers humans not only to understand the reasons of a specific system behaviour, but also to be in control, providing proper interaction mechanisms that allow humans to intervene and even to modify system behaviour through interactive reconfiguration. The authors present a novel interaction paradigm that foster human-AI collaboration, emphasizing explainability and user control of AI systems. It is based on three interaction strategies: Clarification, Reconfiguration and Iterative exploration. Clarification means explaining the behavior or decision of the AI model/algorithm. Reconfiguration refers to enabling users to initiate the retraining of AI models easily. Iterative exploration means that the system outcome is reached through a sequence of iterative steps driven by the interaction between the user and the AI system. The interaction paradigm has been implemented in the HCAI-based redesign of an existing AI tool for microscopic analysis of the nasal mucosa, which has been evaluated in the medical domain with rhinocytologists. The results reveal the value of the interaction strategies. Finally, the article provides interesting lessons learned, which contribute to the understanding and application of HCAI principles for AI in medicine.

We like to conclude by saying that another commonality of the articles in this special issue is that they all remark that the role of AI in the future is not to replace the users but to empower them. This opinion is shared by leading industries (see, for example, Apple's announcement of its version of AI called Apple Intelligence). Physicians worldwide state that they do not want AI tools that replace doctors, but tools that collaborate well with doctors (see, e.g. [24]). Almost all the 10 articles performed studies to gather the perspectives of all involved stakeholders; this is instrumental to envision solutions that are not only technically sound but also align with the practical needs and experiences of everyone in the healthcare community. Finally, we want to remark that ethical and legal aspects of AI in healthcare, including patient privacy, consent, decision-making biases, and liability concerns, is mentioned, often as one of the limitations of the proposed approaches, but not discussed in detail. Future work should also emphasize the importance of robust clinical validation of AI models and of the proposed methodologies and techniques for designing HCAI systems and explore strategies to enhance clinician trust and acceptance of such systems.

References:

1. Shneiderman, Ben. "Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy." International Journal of Human–Computer Interaction, March 23, 2020. https://doi.org/10.1080/10447318.2020.1741118.
2. Fails, Jerry Alan, and Dan R. Olsen. "Interactive Machine Learning." In Proceedings of the 8th International Conference on Intelligent User Interfaces, 39–45. Miami Florida USA: ACM, 2003. https://doi.org/10.1145/604045.604056.
3. Rudin, Cynthia. "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead." Nature Machine Intelligence 1, no. 5 (May 1, 2019): 206–15. https://doi.org/10.1038/s42256-019-0048-x.
4. Combi, Carlo, Beatrice Amico, Riccardo Bellazzi, Andreas Holzinger, Jason H. Moore, Marinka Zitnik, and John H. Holmes. "A Manifesto on Explainability for Artificial Intelligence in Medicine." Artificial Intelligence in Medicine, no. 133 (2022): 102423. https://doi.org/10.1016/j.artmed.2022.102423.
5. Pfeuffer, Nicolas, Lorenz Baum, Wolfgang Stammer, Benjamin M. Abdel-Karim, Patrick Schramowski, Andreas M. Bucher, Christian Hügel, Gernot Rohde, Kristian Kersting, and Oliver Hinz. "Explanatory Interactive Machine Learning: Establishing an Action Design Research Process for Machine Learning Projects." Business & Information Systems Engineering 65, no. 6 (December 2023): 677–701. https://doi.org/10.1007/s12599-023-00806-x.
6. Selvaraju, Ramprasaath R., Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization." International Journal of Computer Vision 128, no. 2 (February 2020): 336–59. https://doi.org/10.1007/s11263-019-01228-7.

7. Grigsby, Scott S. “Artificial Intelligence for Advanced Human-Machine Symbiosis.” In Augmented Cognition: Intelligent Technologies, edited by Dylan D. Schmorrow and Cali M. Fidopiastis, 10915:255–66. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-91470-1_22.
8. Holzinger, A., and Mueller, H.: “Toward Human-AI Interfaces to Support Explainability and Causability in Medical AI”, IEEE COMPUTER, 2021, 54, (10), 78-86, https://doi.org/10.1109/MC.2021.3092610
9. Dóra Göndöcs, Viktor Dörfler, AI in medical diagnosis: AI prediction & human judgment, Artificial Intelligence in Medicine, Volume 149, 2024, 102769, ISSN 0933-3657, https://doi.org/10.1016/j.artmed.2024.102769.
10. Andrea Esposito, Giuseppe Desolda, and Rosa Lanzilotti. 2024. The Fine Line between Automation and Augmentation in Website Usability Evaluation. Scientific Reports 14, 1 (May 2024), 10129. DOI:10.1038/s41598-024-59616-0
11. Holzinger A, Langs G, Denk H, Zatloukal K, Müller H. Causability and explainability of artificial intelligence in medicine. Wiley Interdiscip Rev: Data Min Knowl Discov 2019;9(4):e1312. https://doi.org/10.1002/widm.1312.
12. Weina Jin, Mostafa Fatehi, Ru Guo, Ghassan Hamarneh, Evaluating the clinical utility of artificial intelligence assistance and its explanation on the glioma grading task, Artificial Intelligence in Medicine, Volume 148, 2024, 102751, ISSN 0933-3657, https://doi.org/10.1016/j.artmed.2023.102751.
13. Federico Cabitza, Chiara Natali, Lorenzo Famiglini, Andrea Campagner, Valerio Caccavella, Enrico Gallazzi, Never tell me the odds: Investigating pro-hoc explanations in medical decision making, Artificial Intelligence in Medicine, Volume 150, 2024, 102819, ISSN 0933-3657, https://doi.org/10.1016/j.artmed.2024.102819.
14. Angela Lombardi, Francesca Arezzo, Eugenio Di Sciascio, Carmelo Ardito, Michele Mongelli, Nicola Di Lillo, Fabiana Divina Fascilla, Erica Silvestris, Anila Kardhashi, Carmela Putino, Ambrogio Cazzolla, Vera Loizzi, Gerardo Cazzato, Gennaro Cormio, Tommaso Di Noia, A human-interpretable machine learning pipeline based on ultrasound to support leiomyosarcoma diagnosis, Artificial Intelligence in Medicine, Volume 146, 2023, 102697, ISSN 0933-3657, https://doi.org/10.1016/j.artmed.2023.102697.
15. Kazuma Kobayashi, Yasuyuki Takamizawa, Mototaka Miyake, Sono Ito, Lin Gu, Tatsuya Nakatsuka, Yu Akagi, Tatsuya Harada, Yukihide Kanemitsu, Ryuji Hamamoto, Can physician judgment enhance model trustworthiness? A case study on predicting pathological lymph nodes in rectal cancer, Artificial Intelligence in Medicine, Volume 154, 2024, 102929, ISSN 0933-3657, https://doi.org/10.1016/j.artmed.2024.102929.
16. Kundu S. AI in medicine must be explainable. Nature Med 2021;27:1328, URL: https://doi.org/10.1038/s41591-021-01461-z.
17. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health 2021;3:e74550. http://dx.doi.org/10.1016/S2589-7500(21)00208-9, URL https://www.sciencedirect.com/science/article/pii/S2589750021002089.
18. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019;1:206–15, URL https://doi.org/10.1038/s42256-019-0048-x.
19. Amir Reza Ashraf, Anna Somogyi-Végh, Sára Merczel, Nóra Gyimesi, András Fittler, Leveraging code-free deep learning for pill recognition in clinical settings: A

multicenter, real-world study of performance across multiple platforms, Artificial Intelligence in Medicine, Volume 150, 2024, 102844, ISSN 0933-3657, https://doi.org/10.1016/j.artmed.2024.102844.
20. Ivan De Boi, Elissa Embrechts, Quirine Schatteman, Rudi Penne, Steven Truijen, Wim Saeys, Assessment and treatment of visuospatial neglect using active learning with Gaussian processes regression, Artificial Intelligence in Medicine, Volume 149, 2024, 102770, ISSN 0933-3657, https://doi.org/10.1016/j.artmed.2024.102770.
21. Tommaso Turchi, Giuseppe Prencipe, Alessio Malizia, Silvia Filogna, Francesco Latrofa, Giuseppina Sgandurra, Pathways to democratized healthcare: Envisioning human-centered AI-as-a-service for customized diagnosis and rehabilitation, Artificial Intelligence in Medicine, Volume 151, 2024, 102850, ISSN 0933-3657, https://doi.org/10.1016/j.artmed.2024.102850.
22. Shiyi Wang, Yang Nan, Sheng Zhang, Federico Felder, Xiaodan Xing, Yingying Fang, Javier Del Ser, Simon L.F. Walsh, Guang Yang, Probing perfection: The relentless art of meddling for pulmonary airway segmentation from HRCT via a human-AI collaboration based active learning method, Artificial Intelligence in Medicine, 2024, 102930, ISSN 0933-3657, https://doi.org/10.1016/j.artmed.2024.102930.
23. Giuseppe Desolda, Giovanni Dimauro, Andrea Esposito, Rosa Lanzilotti, Maristella Matera, Massimo Zancanaro, A human-AI interaction paradigm and its application to rhinocytology, Artificial Intelligence in Medicine, 2024, 102933, ISSN 0933-3657, https://doi.org/10.1016/j.artmed.2024.102933.
24. Huang, Z., Yang, E., Shen, J. et al. A pathologist–AI collaboration framework for enhancing diagnostic accuracies and efficiencies. Nat. Biomed. Eng (2024). https://doi.org/10.1038/s41551-024-01223-5

Acknowledgments
The activities of Paolo Buono and Maria Francesca Costabile about this special issue have been partially supported by the co-funding of the European Union - Next Generation EU: NRRP Initiative, Mission 4, Component 2, Investment 1.3 - Partnerships extended to universities, research centers, companies, and research D.D. MUR n. 341 del 15.03.2022 - Next Generation EU (PE0000013 - "Future Artificial Intelligence Research - FAIR" - CUP: H97G22000210007); the activities are part of Work package 6.1 Design of Symbiotic AI systems.