FrameSAI: a Three-Layer Framework to Create Symbiotic AI Systems

Miriana Calvano[1], Antonio Curci[1,2], Rosa Lanzilotti[1] and Antonio Piccinno[1]

[1]Department of Computer Science, University of Bari, Via Edoardo Orabona 4 70125, Bari, Italy
[2]Department of Computer Science, University of Pisa, Largo B. Pontecorvo 4 50756, Pisa, Italy

Abstract

As Artificial Intelligence (AI) becomes increasingly widespread in numerous domains of everyday life, humans must maintain their agency, valuing their expertise and judgment and preserving their privacy and safety. The research presented in this article aims at laying the groundwork for a standardized approach for creating systems, that allow a symbiotic relationship between the human and the machine. Specifically, we present a framework, called FRAMework to creatE SAI (FrameSAI), which guides practitioners in developing high-quality SAI systems while complying with law. It is composed of three layers (i.e., Principles, Guidelines, and Success Criteria), which become increasingly operational indications as one moves from one level to the next. A preliminary version of FrameSAI is presented, illustrating a first set of guidelines and success criteria mapped to the four levels of risks of the AI Act, the European Union (EU)'s regulation for the design, development, and deployment of AI systems.

Keywords

Human-AI Symbiosis, Framework, Design, Evaluation, AI Act

1. Introduction

The rapid advancement of AI is influencing the way humans carry out activities in many contexts, becoming more and more pervasive. Thus, it is important to allow humans to properly communicate with it through appropriate interaction mechanisms. For this reason, shifting the focus to creating Human-Centered Artificial Intelligence (HCAI) systems becomes necessary to align with users' needs, preferences, and expertise [1]. As the strong and close relationship between humans and AI raises crucial challenges to address, a novel branch of HCAI is emerging: Symbiotic Artificial Intelligence. Symbiotic Artificial Intelligence (SAI) is encompasses all the AI systems that support rather than replace humans, fostering a bidirectional relationship that lays the groundwork for mutual and continuous exchange to improve both parties over time [2, 3]. To design high-quality SAI systems, a human-centered approach must be undertaken, following the Human-Centered Design (HCD) and considering users during the whole creation process [4].

To foster human-AI symbiosis, users must be enabled to understand the system's decisions by comprehending the processes and motivations that lie behind them. Despite being a complex task to fulfill, this aspect is crucial in establishing a trustworthy relationship [5], in which an AI system deserves human trust [1]. In this new and ever-evolving scenario, designers and developers must create AI systems that allow humans to make informed decisions, solving problems to the best of their abilities, being supported by them. Ensuring that humans fully understand system's output plays a particularly important role because it can influence individuals' behaviors and decisions that can consequently have a huge impact on their or others' lives and/or environment. New frontiers are opening with respect the human-centeredness of AI systems, being considered also by legal and governmental bodies, who are underlining how humans must be safeguarded in all of their dimensions, preserving the well-being of individuals, the society they live in and their environment. The EU's AI Act represents a huge leap

Ital-IA 2025: 5th National Conference on Artificial Intelligence, organized by CINI, June 23-24, 2025, Trieste, Italy
miriana.calvano@uniba.it (M. Calvano); antonio.curci@uniba.it (A. Curci); rosa.lanzilotti@uniba.it (R. Lanzilotti); antonio.piccinno@uniba.it (A. Piccinno)
0000-0002-9507-9940 (M. Calvano); 0000-0001-6863-872X (A. Curci); 0000-0002-0877-7063 (R. Lanzilotti); 0000-0002-2039-8162 (A. Piccinno)

forward in this direction, providing the foundations to build human-centric AI-based systems. It adopts a risk-based approach classifying AI systems according to their level of risk: (i) high, (ii) limited, (iii) low, and (iv) minimal risk [6].

As the requirements and standards evolve, designers and developers must adapt to them. Thus, the objective of the research presented in this article is to propose a three-layer framework, called FRAMework to creatE Symbiotic Artificial Intelligence that undertakes a top-down approach to guide practitioners in the creation of SAI systems that comply with the current regulations. The three-layer structure of the framework proposes indications in term of principles, guidelines, and success criteria that become increasingly operational indications as one moves from one level to the next.

The article is structured as follows: Section 2 reports the main challenges of SAI, presenting its main characteristics with respect to the multi-disciplinary and risk-based approaches; Section 3 presents FRAMework to creatE Symbiotic Artificial Intelligence (FrameSAI); Section 4 draws the conclusions and explores the future work of this research.

2. Background

Traditionally, the development of systems, products, or services was predominantly system-centric [7, 1], a mindset that has also reflected in the process of building AI systems, focusing on functionalities and high-performing models [8]. However,there is a growing need to shift from a mere automation to a more appropriate augmentation [9], where both AI and humans learn from each other to improve their capabilities and collaborating effectively. It is therefore a necessity to create systems that put humans in control and that enable them to communicate with them appropriately. Preserving human rights and complying with the law (i.e. AI Act) can foster trust in AI performance and suggestions [2]. In order to meet these requirements and to match these objectives, two approaches are being undertaken: *multi-disciplinary* and *risk-based*, each described in the following sections and merged in FrameSAI.

2.1. Multidisciplinary Approach

Since the interaction between humans and AI systems is particularly complex and multifaceted, to create high-quality SAI systems a multidisciplinary approach must be adopted, encompassing the following main research areas: Human-Computer Interaction (HCI), Law & Ethics, Software Engineering (SE), and AI [10]; each of these disciplines contributes equally, but in different ways, to the creation of these systems. They are intrinsically intertwined because HCI, which can be considered the bridge between the technical side of Computer Science and the human studies of psychology [11], serves to create interactive solutions that are intuitive, accessible, and usable. Developing usable AI systems is key to establishing a symbiotic relationship because they must allow users to reach their goals by providing them useful feedback while being affordable fostering a fruitful interaction process [12]. Beyond usability, as mentioned earlier, legal and ethical considerations are crucial to ensure the creation of systems that safeguard humans regardless from system functioning [13]. Thus, they must be integrated into the processes of the Agile development, following the best practices and methodologies that belong to Software Engineering (SE). In the end, to ensure a fruitful symbiotic relationship, accurate and appropriate AI models must be built, allowing humans to be supported in their activities [14].

2.2. AI Act Risk-based Approach

In the field HCAI, researchers face many challenges, such as providing users with low-level explainability, data privacy, biased data and ethical issues that represent a huge risk to humans. For this reason, the ethical challenges associated with AI must be addressed to reduce biases and discriminatory behaviors as much as possible [15].

To fill this gap, the EU AI Act must be considered when creating AI systems that are legally compliant while fostering a human-AI symbiosis. It is a regulatory framework enacted by the EU to govern the usage of AI, classifying systems in four levels of risk: *Unacceptable Risk* refers to AI systems that pose

risks to human rights and societal well-being, such as those employing manipulative techniques to distort behavior, exploiting vulnerable individuals, or enabling discriminatory behaviors, and encompasses prohibited applications that must be banned to safeguard humans; *High Risk* includes AI systems integrated as safety components in products or subject to conformity assessments due to potential risks to health, safety, or fundamental rights. It encompasses applications that can profile individuals through automated analysis of personal data evaluating aspects of their lives, including job performance, economic status, or health; *Limited Risk* encompasses AI systems that must comply with specific transparency obligations to warn humans that they are interacting with AI to make them aware about the fact that they are subject to emotional recognition or biometric categorization (e.g., biometric recognition systems, emotion recognition systems); *Low or Minimal Risk* refers to systems that feature AI without requiring law conformity checks since they do not require interaction with humans avoiding an impact on their lives [6].

Considering these two approaches, a framework that encompasses principles, guidelines, and specific success criteria is presented in the following section to guide designers and developers in the process of creating AI systems that comply with regulations and human needs.

3. Structure and Contents of FrameSAI

This section explores the structure of FrameSAI, illustrated in Figure 1, providing insights concerning the motivations and the choices behind it. The elements of each component are still being developed and refined, as they derive from Systematic Literature Reviews (SLRs) that are being carried out and will be refined through the generalization of case studies' results.

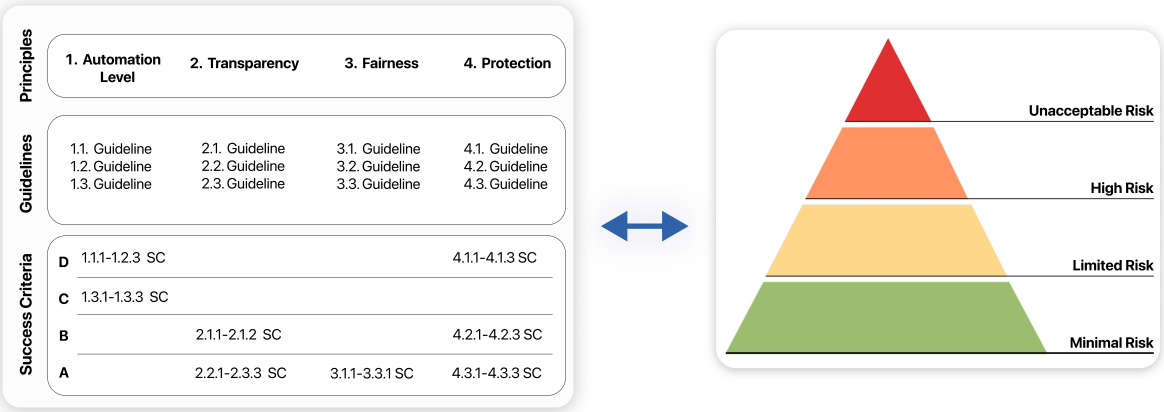

Figure 1: The FrameSAI framework structure mapped with the risk levels of the AI Act

The SLR that laid the groundwork for the definition of the framework had the objective of defining the principles that can guide the creation of SAI systems, while promoting a symbiotic human-AIrelationship, and adhering to the European AI Act [6]. This research provides the insights necessary to investigate how the AI and Human Computer Interaction (HCI) communities are absorbing and adopting the regulation while considering the symbiotic relationship between humans and AI, with respect to the new standards it is setting for these systems. The principles are defined based on the results of this SLR, but they can be refined on the basis of the results of ongoing work.

With the objective of provide support to designers and developers to create SAI systems by guiding them with different levels abstraction, the FrameSAI consists of three main components: *Principles*, *Guidelines*, and *Success Criteria*. Undertaking a top-down approach and considering the human-centeredness of SAI systems, the structure of the framework is reported in the following paragraphs.

Principles Currently, the principles are the core of the framework, representing the guiding pillars that characterize the symbiosis between the human and AI. They are: *Automation Level*, *Transparency*, *Fairness*, and *Protection*. The principles dictate the main requirements for SAI systems, which must guarantee an appropriate level of automation to humans, be properly transparent and explain the motivations behind their outputs while exhibiting a fair and non-discriminatory behavior, protecting humans privacy, safety, and security [16, 17, 18].

Guidelines This component encompasses a set of guidelines for each principle that provide the foundations for SAI. The guidelines are expressed in terms of what features the AI system must provide and the behavior it is expected to exhibit in order to foster effective collaboration with humans. Some guidelines reflect concepts that belong to more than one principle, thus overlapping and highlighting the multidisciplinary nature of symbiosis. The guidelines integrated in the framework appear in the following form: <Subject> must <obligation> <action> [<motivation>]. In addition, the guidelines should be defined to provide an appropriate level of coercivity and maintain an high level of generality to be applied to any domain [19]. Examples of guidelines related to the Automation Level principle are:

(GL 1.1) The human must be able to check, monitor, and supervise the AI system's behavior, taking into account how it functions and the decisions it makes.

(GL 1.2) The human must be enabled to assign proper meaning to the AI model's outputs, processes, and structure to eventually modify its behavior.

(GL 1.3) The human must be guided towards ethical behaviors when reconfiguring the AI model to avoid harms to both parties.

Success Criteria Deriving from each guideline and respective principle(s), they consist of verifiable instructions that make it possible to assess whether the guideline to which they refer is complied with. The success criteria will be statements with a lower level of generality with respect to the guidelines, detailing instructions depending on the technologies, contexts of use, and requirements based on the AI system's level of risk.

When it comes to AI systems, symbiosis cannot be a "black and white" concept. Thus, we identified four categories of success criteria. In order to adhere with the current legal scenario concerning AI systems, the compliance with the European Union's AI Act will be considered by mapping the success criteria to four levels of risk of the regulation: A (minimal risk), B (Limited risk), C (High risk), and D (Unacceptable risk). An example of some success criteria for the guideline (GL 1.2) of the Automation Level principle are:

(SC 1.2.1) The system shows human-understandable explanations representing the system's underlying functioning and reasoning depending on the type of data that the model processes (e.g., activation maps for images, shap for structured tabular data).

(SC 1.2.2) The system allows humans to input corrections and/or meaningful feedback that result in verifiable changes in future interactions with respect to the behavior of the model (e.g., via fine-tuning or rule updates).

(SC 1.2.3) The system allows humans to revise its behavior, modifying the decision (e.g., via rules, weights, thresholds) through suitable User Interface (UI) elements (e.g. checkbox, buttons, text fields, etc.).

These SCs must be verified in each high-risk system. At the same time, in case of minimal-risk, (SC 1.2.1) remains necessary for the establishment of symbiosis but can also be not considered as particularly relevant from a legal point of view. Contrarily, in the case of a medical high-risk system, (SC 1.2.1) must be verified since a suggestion can have a severe impact on health.

Ensuring that the characteristics, indicated by the success criteria, are included into AI systems aims at increasing their trustworthiness, establishing interaction mechanisms that put humans in-the-loop and avoiding them from blindly relying on technology [20, 21].

4. Conclusions and Future Works

Designing, developing, and deploying AI systems in this era of modern society brings multiple challenges that practitioners must face. This article has proposed a framework, called FrameSAI, aiming at systematizing the process of creating AI systems that support humans rather than replacing them. The framework is structured in three layers: *Principles*, *Guidelines*, and *Success Criteria*. Although the definition of the inner workings of the framework is still on-going, we present examples of guidelines and success criteria that it includes, making considerations concerning their mapping with the risk levels of the AI Act.

Future work of this research will focus on the refinement of the current set of guidelines and success criteria through their application to real case studies. Currently, they are being applied to two contexts of use in the medical field, but we intend to expand this investigation to other areas that bring diverse challenges. In addition, an extensive study is being planned to evaluate the comprehensibility, relevancy, and feasibility of the framework as a whole with designers and developers.

Acknowledgments

The research of Miriana Calvano, Antonio Curci, and Rosa Lanzilotti is 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).

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] B. Shneiderman, Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy, International Journal of Human–Computer Interaction 36 (2020) 495–504. URL: https://www.tandfonline.com/doi/full/10.1080/10447318.2020.1741118. doi:10.1080/10447318.2020.1741118.

[2] G. Desolda, A. Esposito, R. Lanzilotti, A. Piccinno, M. F. Costabile, From human-centered to symbiotic artificial intelligence: a focus on medical applications, Multimedia Tools and Applications (2024). URL: https://link.springer.com/10.1007/s11042-024-20414-5. doi:10.1007/s11042-024-20414-5.

[3] S. S. Grigsby, Artificial Intelligence for Advanced Human-Machine Symbiosis, in: D. D. Schmorrow, C. M. Fidopiastis (Eds.), Augmented Cognition: Intelligent Technologies, volume 10915 of *Lecture Notes in Computer Science*, Springer International Publishing, Cham, 2018, pp. 255–266. URL: https://link.springer.com/10.1007/978-3-319-91470-1_22. doi:10.1007/978-3-319-91470-1_22.

[4] I. O. for Standardization, Iso 9241:210 - ergonomics of human-system interaction: Human-centred design for interactive systems, 2019. URL: https://www.iso.org/standard/77520.html.

[5] G. Stettinger, P. Weissensteiner, S. Khastgir, Trustworthiness Assurance Assessment for High-Risk AI-Based Systems, IEEE Access 12 (2024) 22718–22745. doi:10.1109/ACCESS.2024.3364387.

[6] European Parliament, Council of the European Union, Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations

(EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), 2024.

[7] R. Polillo, Facile da Usare, volume 1, 1 ed., Apogeo, 2010. URL: http://www.rpolillo.it/facile-da-usare/.

[8] T. Jiang, Z. Sun, S. Fu, Y. Lv, Human-AI interaction research agenda: A user-centered perspective, Data and Information Management (2024) 100078. URL: https://linkinghub.elsevier.com/retrieve/pii/S2543925124000147. doi:10.1016/j.dim.2024.100078.

[9] A. Esposito, G. Desolda, R. Lanzilotti, The fine line between automation and augmentation in website usability evaluation, Scientific Reports 14 (2024) 10129. URL: https://www.nature.com/articles/s41598-024-59616-0. doi:10.1038/s41598-024-59616-0.

[10] A. Curci, A Comprehensive Framework Proposal to Design Symbiotic AI Systems, in: Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering, ACM, Salerno Italy, 2024, pp. 460–465. URL: https://dl.acm.org/doi/10.1145/3661167.3661219. doi:10.1145/3661167.3661219.

[11] G. M. Olson, J. S. Olson, Human-Computer Interaction: Psychological Aspects of the Human Use of Computing, Annual Review of Psychology 54 (2003) 491–516. URL: https://www.annualreviews.org/doi/10.1146/annurev.psych.54.101601.145044. doi:10.1146/annurev.psych.54.101601.145044.

[12] H. Sharp, J. Preece, Y. Rogers, Interaction Design: beyond human-computer interaction, 5 ed., John Wiley & Sons, Inc., 2019.

[13] L. D. Urquhart, G. McGarry, A. Crabtree, Legal Provocations for HCI in the Design and Development of Trustworthy Autonomous Systems, in: Nordic Human-Computer Interaction Conference, ACM, Aarhus Denmark, 2022, pp. 1–12. URL: https://dl.acm.org/doi/10.1145/3546155.3546690. doi:10.1145/3546155.3546690.

[14] N. Anamaria, M. M. Monica Mihaela, M. Cristina, Artificial Intelligence: Friend or Foe? Experts' Concerns on European AI Act, Economic Computation and Economic Cybernetics Studies and Research 57 (2023) 5–22. doi:10.24818/18423264/57.3.23.01.

[15] B. Shneiderman, Human-Centered AI, 1 ed., Oxford University Press, Oxford, 2022.

[16] G. Pavlidis, Unlocking the black box: Analysing the EU artificial intelligence act's framework for explainability in AI, Law, Innovation and Technology 16 (2024) 293–308. doi:10.1080/17579961.2024.2313795.

[17] High Level Expert Group on Artificial Intelligence, Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment, 2020.

[18] L. Enqvist, 'Human oversight' in the EU artificial intelligence act: What, when and by whom?, Law, Innovation and Technology 15 (2023) 508–535. doi:10.1080/17579961.2023.2245683.

[19] A. Dix (Ed.), Human-computer interaction, 3rd ed ed., Pearson/Prentice-Hall, Harlow, England ; New York, 2004.

[20] C. Keding, P. Meissner, Managerial overreliance on AI-augmented decision-making processes: How the use of AI-based advisory systems shapes choice behavior in R&D investment decisions, Technological Forecasting and Social Change 171 (2021) 120970. URL: https://www.sciencedirect.com/science/article/pii/S0040162521004029. doi:https://doi.org/10.1016/j.techfore.2021.120970.

[21] A. Esposito, M. Calvano, A. Curci, F. Greco, R. Lanzilotti, A. Piccinno, Explanation-Driven Interventions for Artificial Intelligence Model Customization: Empowering End-Users to Tailor Black-Box AI in Rhinocytology, 2025. URL: https://arxiv.org/abs/2504.04833. doi:10.48550/ARXIV.2504.04833, version Number: 2.