The Human-Centered Approach to Design and Evaluate Symbiotic AI Systems

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

[1]University of Bari Aldo Moro, Via Edoardo Orabona 4, 70125, Bari, Italy
[2]University of Pisa, Largo B. Pontecorvo 3, 56127, Pisa, Italy

Abstract

Artificial Intelligence (AI) is spreading in many domains, revolutionizing the way individuals conceive their working and private life; it enhances many tasks by automating decision-making and augmenting human capabilities. It is necessary to design high-quality AI systems that focus on the users' priorities and avoid potential unethical and unexpected behaviours. The widespread adoption of AI solutions faces challenges related to their transparency, since humans must be enabled to fully understand the outputs of such systems in order to make informed decisions. To address these concerns, a shift toward a human-centered approach is emerging when it comes to interact with AI systems. In this new scenario, Human-Computer Interaction (HCI) plays a pivotal role and contaminates AI to reach the human-AI symbiosis. Designers and developers should gravitate towards Symbiotic AI (SAI), which has the goal to support humans without replacing them and establish a symbiotic relationship with users, adapting to their cognitive models.
This contribution aims to present a proposal of a framework to design high-quality SAI systems and metrics that can be employed to appropriately evaluate them. Opportunities and challenges that characterize this new research context are also presented and discussed.

Keywords

Symbiotic AI, Human-Centered Design, Design, Evaluation

1. Introduction

The technological advancement of modern society has introduced Artificial Intelligence (AI) in multiple fields (e.g., medicine, transportation, education), bringing innovation with new services and products that can boost productivity and reduce the demand for repetitive tasks in terms of time and resources [1, 2]. However, the current scenario concerning AI is raising questions about important ethical and legal factors because it must be used responsibly, avoiding potential misuse, biases, and infringement of human rights. In this regard, new guidelines and regulations are emerging to ensure a responsible and correct design, development, and use of systems that feature AI, prioritizing human well-being and societal values.

Professionals in diverse domains broadly use AI to make decisions that often lead to irreversible consequences [3]. Such issues must be addressed when creating AI systems; they lie in the fact that end users are not computer scientists and cannot fully comprehend the processes that lead to outputs from a technical point of view. Designers and developers must consider users' cognitive models, skills, and needs to create AI systems that establish a symbiotic relationship between humans and machines. This concept plays a pivotal role in the introduction of Symbiotic AI (SAI). This expression refers to AI systems that enhance and support humans in performing their activities without replacing them, adapting to their mental and physical models, allowing them to make informed decisions and avoid negative feelings [4, 5]. For example, a physician can rely on an AI-powered system that classifies patients as ill or healthy based on MRI scans; at the same time, they must be enabled to correctly interpret the system's output and to comprehend the motivations that generated the response to make informed

Proceedings of the 1st International Workshop on Designing and Building Hybrid Human–AI Systems (SYNERGY 2024), Arenzano (Genoa), Italy, June 03, 2024.
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); 0000-0002-2039-8162 (R. Lanzilotti); 0000-0003-1561-7073 (A. Piccinno)

decisions [6]. The human-AI interaction must be enhanced and improved by recognising humans' responses and emotions, embracing their needs and adapting to their behaviors [4]. Therefore, humans should play a central role throughout the design of SAI systems, employing the Human-Centered Design (HCD) methodology while following the processes and techniques that belong to the agile development of Software Engineering (SE) [7]. This approach aims at merging and linking disciplines, such as Law, Human-Computer Interaction (HCI) and Software Engineering (SE), boosting the productivity of designers and developers while improving users' experience when interacting with such systems.

This contribution presents a proposal of a novel approach to design and evaluate SAI systems; it illustrates a conceptual multidisciplinary framework and preliminary considerations about metrics to assess the desired properties of such systems (e.g., trustworthiness, safety, reliability, etc.).

This paper is structured as follows: Section 2 describes the motivations of this research work; Section 3 presents its main goals and research questions; Section 4 illustrates the proposal of a comprehensive conceptual framework; Section 5 presents preliminary considerations about metrics for the evaluation of SAI systems and Section 6 concludes and details the future steps of this research.

2. Research Motivations

Understanding AI models is a complex and challenging task for humans because of their *black-box* nature, such as deep learning models which are characterized by complex mathematical operations involving millions or billions of parameters. They can recognize intricate and nonlinear patterns, which are often highly difficult for humans to interpret how specific inputs lead to particular outputs [8]. This issue feeds the need for *Transparency*, in which *Explainability* and *Interpretability* are reinforced: in transparent AI models, adequate explanations are provided to users about the processes that were employed to produce specific outputs, making them interpretable by humans, which implies that they can map the abstract concepts to something that they can make sense of [9]. Understanding the motivations behind specific outputs is crucial to guarantee humans the right level of control while balancing automation because, even if it can increase efficiency, users should be allowed to intervene and control the system's performance when appropriate [10, 11, 12].

This research work emphasizes that creating AI systems not only involves mere technicalities but also ethical, legal, and anthropological dimensions. For instance, Kieseberg et al. state that "AI possess three key characteristics throughout its entire life-cycle: Lawfulness, adherence to ethical principles and technological, as well as, social robustness" [6]. In this regard, the main topics that are worth researching for this study are represented in the diagram shown in Figure 1. Specifically, *Technicalities* ensure that through the human-AI symbiosis, the user's cognitive abilities are enhanced while guaranteeing the right level of control of the system; *Ethical Aspects* are focused on ethical concerns in SAI systems; *Human Factors* highlight the need for a usable and explainable AI to allow the user to make informed decisions and to easily comprehend the system's performance [13]. It is important to underline that these three dimensions are also related to each other due to the interdisciplinary nature of the field. Thus, it would be possible to design artefacts that, being governance-compliant and fair, can guarantee human-driven decision-making.

In this regard, the European Union (EU) has formalized the factors concerning AI by releasing the Artificial Intelligence Act (AIA), which delineates a regulation with respect to the design, development, and employment of AI through a risk-based approach [14]. Similarly to the General Data Protection Regulation (GDPR) for privacy and security, the soft law is changing the future of AI in the Union and in the rest of the world, highlighting that users must be always preserved and protected. The GDPR governs how data in the EU is stored, processed, and transferred, implying that designers and developers must necessarily comply with this regulation to create systems that can be appropriately used by end users [15]. It is important to consider the users' characteristics since integrating information about their mental model from the training phase of AI models can lead to the creation of non-biased and ethically compliant SAI systems.

Establishing best practices and guidelines for designers and developers to create AI systems that

foster symbiosis must start with defining principles that act as the leitmotif of the research in this context.

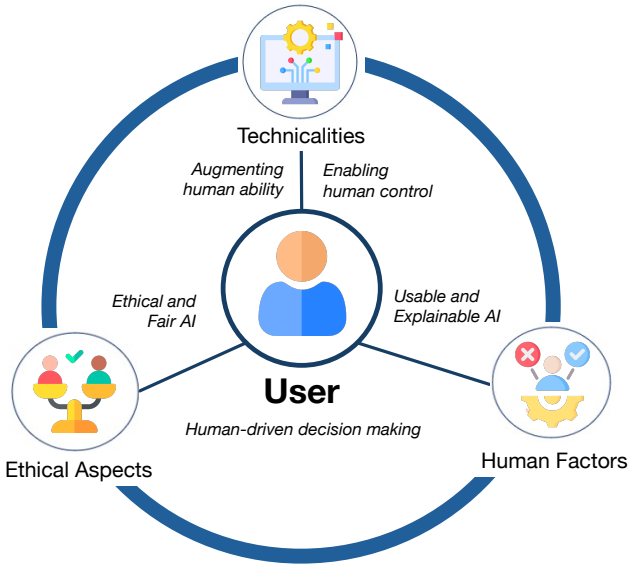

Figure 1: The three main dimensions of SAI: Technicalities, Ethical Aspects and Human Factors (Adapted from [13]).

3. Objectives and Research Questions

Based on the topics explored in Section 1 and 2, the overall objectives of this research are listed below:

- Create a comprehensive framework tailored to the design of SAI systems that encompasses AI, HCI, SE, Law, and Ethics.
- Define new interaction paradigms and new user interfaces that align with SAI requirements.
- Design transparent models that allow users to understand the behavior and make informed decisions.
- Delineate an evaluation framework and metrics to assess the human-AI symbiosis considering its desired properties (e.g., trustworthiness, safety, and reliability). This framework will provide an instrument to assess the effectiveness of the human-AI relationship in SAI systems.

This contribution represents a starting point to achieve the research goal, which is guided by the following research questions:

(RQ1) How can the methodologies of HCI be integrated into the processes that belong to SE to develop SAI systems?

(RQ2) How can the legal and ethical requirements concerning AI be integrated into a framework for the development of compliant SAI systems?

(RQ3) How the current challenges in conventional metrics for evaluating SAI systems can be faced to assess the human-AI symbiosis?

4. Comprehensive Framework

Creating SAI systems can be a challenging objective and must be addressed through a multidisciplinary approach, encompassing diverse domains that range from Computer Science to Law. Figure 2 presents a conceptual version of the comprehensive framework that embraces four main research areas: Human-Computer Interaction (HCI), Law & Ethics, Software Engineering (SE), and Artificial Intelligence. Although these disciplines are characterized by their own principles, guidelines, and techniques, this framework aims to define the connections and influences among them and find the links that reinforce human-machine symbiosis. It is underlined that the domains involved in this research are all relevant and they equally contribute to the achievement of the goal.

The following sections describe each component of the framework, illustrating its role in the SAI scenario.

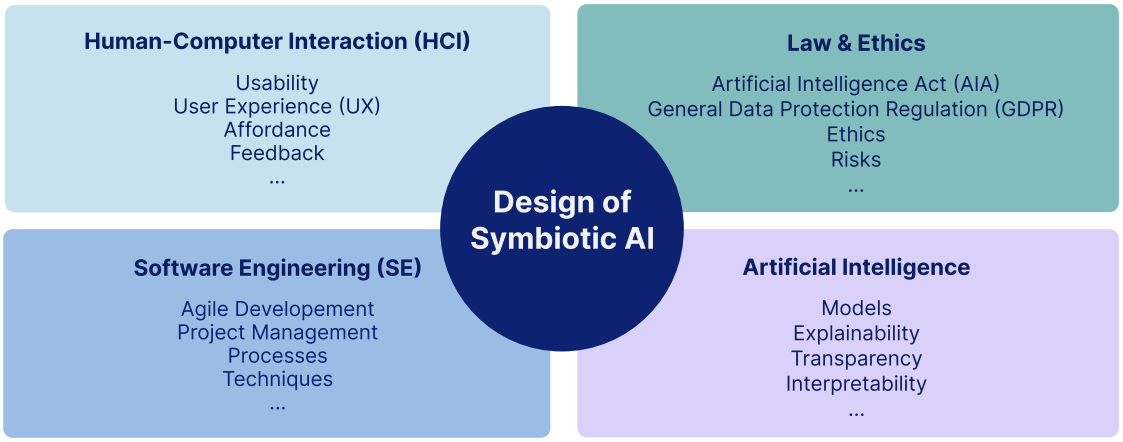

Figure 2: Conceptual Framework for the Design of SAI systems

Human-Computer Interaction (HCI) It is the bridge that stands between the technical side of Computer Science and the human studies of Psychology. It is the discipline that stresses how crucial end users are, defining methodologies and techniques that must be employed when designing any kind of product. Developing usable AI systems is key to establishing a symbiotic relationship because they must allow users to reach their goals with effectiveness, efficiency, and satisfaction. Thus, they need to provide useful feedback, they must be affordable, and deliver gratification to users during their interaction with the system [16]. It emerges that the Human-Centered Design (HCD) is the core of HCI and addresses the complete and continuous involvement of humans by employing different techniques, such as interviews, questionnaires, field studies, and focus groups, providing quantitative and qualitative data to obtain rich insights about the users' needs, preferences, behaviors, and cognitive models [16, 7].

Law and Ethics They represent one of the four pillars of this framework, encompassing the factors that influence the creation of AI systems from a regulatory, philosophical, and ethical standpoint. Designers and developers must be aware of these regulations in order to create compliant products that preserve users' social, working, and personal well-being. From a legal standpoint, the main elements to consider are the AIA and the GDPR. These regulations define the ethical principles that any kind of system should possess in order to be available to society [14, 15].

Software Engineering (SE) It defines how software is created through standardized methodologies [17]. This framework aims at guiding design and developers in the creation of SAI, ensuring that they operate following a human-centered approach, while complying with the legal requirements

and implementing high-performing AI systems [18]. Therefore, the objective is to integrate the Agile principles and the processes of the Agile Development Lifecycle with those belonging to the SAI design, creating a mapping that does not exclude any discipline.

Artificial Intelligence This dimension is related to technical elements of the design of SAI systems by suggesting techniques and standards in relation to the actual implementation of AI models. The models can be applied in various real-world domains - e.g., business, finance, healthcare, agriculture, smart cities, and cybersecurity. Depending on the activities that have to be performed, different models can be employed, along with different AI tasks, such classification, prediction, description [19]. Other than the mere metrics for the evaluation of AI models, this framework revolves around *Transparency*; it embraces other techniques, such as *Explainability* and *Interpretability* and has the goal of providing insight into how models work, why specific decisions are made, and what data is used to reach those decisions.

These disciplines are intrinsically intertwined because HCI serves to create interactive solutions that are intuitive, accessible, and usable. Beyond usability, considerations of legality and ethics are crucial to guide the development process to ensure compliance with regulations and adherence to ethical principles, safeguarding humans. The process has to be carried out following standardized SE practices to build robust architecture and ensure a reliable implementation of systems powered by AI. Such systems not only enhance human capabilities but also foster symbiotic relationships between humans and technology, redefining how we interact with AI.

5. Metrics for Evaluating SAI

User behavior and AI system performances have been considered unrelated aspects to analyse and evaluate [20]. This point of view led to the definition of metrics that separately evaluate user-side and AI-side (i.e. User Experience (UX) and AI metrics).

In this new scenario, where there is a strict correlation between human and AI performances, it is strictly necessary to define novel metrics that are able to evaluate both user and AI performance and, consequently, the human-AI symbiotic relationship [21, 22]. These metrics should revolve around the principle of trustworthiness which ensures that SAI systems can be trusted, operate safely, and exhibit reliable behavior [10].

The starting point for the definition of these metrics is considering the gaps in the traditional approach in which users and AI systems are evaluated separately. As shown in Figure 3, these two dimensions must contaminate in order to properly assess the symbiotic relationship. Considering both UX metrics to evaluate human behavior and AI metrics to measure the AI model performance, it is possible to obtain an initial definition of metrics able to assess the human-AI symbiosis. For instance, in this scenario, it is important to consider not only information about the dataset but also about the user's mental model and cognitive abilities [23].

Preliminary considerations about novel metrics are presented below based on known research available in the literature.

- *Trustworthiness* can be employed to evaluate how much through the human-AI symbiotic relationship, it is possible to enhance user trust. In particular, this can measure the level of user trust level in terms of prevention from undesired system behaviors and the correctness of decisions taken. Trustworthiness can be achieved if an AI system possesses other several properties, for example, safety, fairness, sustainability, etc [1].
- *Interaction Enhancement (IE)* is a novel metric that can be defined to evaluate to which extent user cognitive abilities can be enhanced through SAI systems. In this way, whether the human-AI strict collaboration brings benefit can be assessed considering the effort employed by the user during the interaction process.

[1] https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html

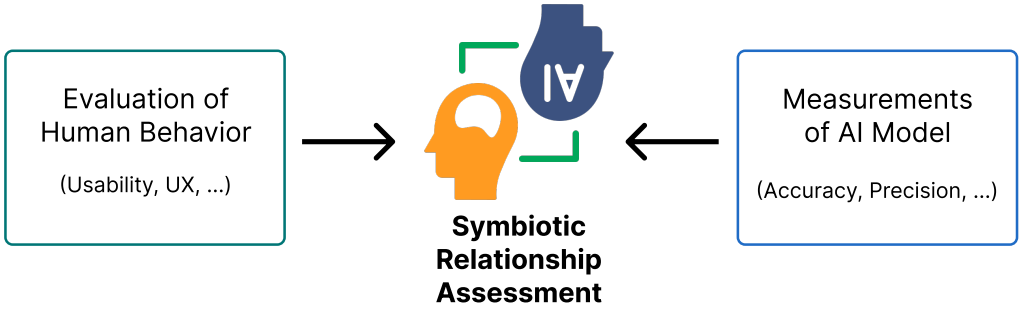

Figure 3: UX and AI metrics' role for the definition of metrics to assess the human-AI symbiotic relationship.

In the end, to understand if the proposed solutions are valid and how they can be improved, they will be assessed through user studies to understand if they can be correctly applied in real-world contexts. They will be used to obtain insights into the effectiveness of proposed metrics and the SAI system's user interface.

6. Conclusions and Future Work

This contribution presents the current challenges and opportunities that characterize this new scenario concerning the interaction between humans and AI.

SAI systems have the objective to enhance the user's cognitive abilities and to guarantee the right balance between human control and AI system automation. Being a new field of research with specific demands and delicate requirements, it is important to determine principles, guidelines, and practices to guide designers and developers in the process of creating AI systems that comply with regulations and human needs. Undertaking a methodological and empirical approach in this context also involves the proper evaluation of these systems [24].

This paper presents a proposal for a framework to lead computer scientists in the creation of SAI systems and preliminary considerations about an assessment strategy based on novel metrics that measure the extent to which such systems are trustworthy and compliant with the defined guidelines. In this regard, preliminary ideas about how to address the existing challenges are proposed which will be further revised and validated.

The future work of this research concerns the formal definition of the framework and metrics in question and their validity assessment. In this process, end-users will be involved to have direct feedback and suggestions. Real-case scenarios and user studies will be used to evaluate the entire methodology, examining how the framework performs when used by designers and developers and determining the effectiveness of metrics when employed for the evaluation of existing SAI systems.

Acknowledgments

The research of Miriana Calvano and Antonio Curci 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).

References

[1] M. Holt, Artificial intelligence in modern society, 2018. URL: https://api.semanticscholar.org/CorpusID:115683678.

[2] I. M. Cockburn, R. Henderson, S. Stern, The impact of artificial intelligence on innovation, IRPN: Innovation & Cyberlaw & Policy (Topic) (2018). URL: https://api.semanticscholar.org/CorpusID:91169962.

[3] S. Andolina, J. A. Konstan, Introduction to the special issue on ai, decision-making, and the impact on humans, International Journal of Human–Computer Interaction 39 (2023) 1367 – 1370. URL: https://api.semanticscholar.org/CorpusID:256892177.

[4] S. S. Grigsby, Artificial intelligence for advanced human-machine symbiosis, in: D. D. Schmorrow, C. M. Fidopiastis (Eds.), Augmented Cognition: Intelligent Technologies, Springer International Publishing, Cham, 2018, pp. 255–266.

[5] M. Melles, A. Albayrak, R. Goossens, Innovating health care: key characteristics of human-centered design, International Journal for Quality in Health Care 33 (2020) 37–44. URL: https://doi.org/10.1093/intqhc/mzaa127. doi:10.1093/intqhc/mzaa127, tex.eprint: https://academic.oup.com/intqhc/article-pdf/33/Supplement_1/37/36350849/mzaa127.pdf.

[6] P. Kieseberg, E. Weippl, A. M. Tjoa, F. Cabitza, A. Campagner, A. Holzinger, Controllable ai - an alternative to trustworthiness in complex ai systems?, in: A. Holzinger, P. Kieseberg, F. Cabitza, A. Campagner, A. M. Tjoa, E. Weippl (Eds.), Machine Learning and Knowledge Extraction, Springer Nature Switzerland, Cham, 2023, pp. 1–12.

[7] 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.

[8] R. Iyer, Y. Li, H. Li, M. Lewis, R. Sundar, K. Sycara, Transparency and explanation in deep reinforcement learning neural networks, in: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, AIES '18, Association for Computing Machinery, New York, NY, USA, 2018, p. 144–150. URL: https://doi.org/10.1145/3278721.3278776. doi:10.1145/3278721.3278776.

[9] G. Montavon, W. Samek, K.-R. Müller, Methods for interpreting and understanding deep neural networks, Digital Signal Processing 73 (2018) 1–15. URL: https://linkinghub.elsevier.com/retrieve/pii/S1051200417302385. doi:10.1016/j.dsp.2017.10.011.

[10] B. Shneiderman, Human-Centered AI, 1 ed., Oxford University PressOxford, 2022. URL: https://academic.oup.com/book/41126. doi:10.1093/oso/9780192845290.001.0001.

[11] D. Gunning, D. W. Aha, Darpa's explainable artificial intelligence program, AI Magazine 40 (2019) 44–58. URL: https://onlinelibrary.wiley.com/doi/abs/10.1609/aimag.v40i2.2850. doi:https://doi.org/10.1609/aimag.v40i2.2850. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1609/aimag.v40i2.2850.

[12] L. Enqvist, 'Human oversight' in the EU artificial intelligence act: what, when and by whom?, Law, Innovation and Technology 15 (2023) 508–535. URL: https://www.tandfonline.com/doi/full/10.1080/17579961.2023.2245683. doi:10.1080/17579961.2023.2245683.

[13] L. G. Wei Xu, Marvin J. Dainoff, Z. Gao, Transitioning to human interaction with ai systems: New challenges and opportunities for hci professionals to enable human-centered ai, International Journal of Human–Computer Interaction 39 (2023) 494–518. doi:10.1080/10447318.2022.2041900.

[14] T. E. Commission, Proposal for a regulation of the european parliament and of the council laying down harmonised rules on asrtificial intelligence (artificial intelligence act) and amending certain union legislative acts, 2024. URL: http://thomas.loc.gov/cgi-bin/query/z?c102:H.CON.RES.1.IH.

[15] Gazzetta Ufficiale dell'Unione Europea, General Data Protection Regulation (GDPR): Regulation (EU) 2016/679, 2018.

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

[17] P. Bourque, R. E. Fairley (Eds.), SWEBOK: guide to the software engineering body of knowledge, version 3.0 ed., IEEE Computer Society, Los Alamitos, CA, 2014. OCLC: 880350861.

[18] D. Salah, R. F. Paige, P. Cairns, A systematic literature review for agile development processes and user centred design integration, in: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, ACM, London England United Kingdom, 2014, pp. 1–10. URL: https://dl.acm.org/doi/10.1145/2601248.2601276. doi:10.1145/2601248.2601276.

[19] I. Sarker, Ai-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems, 2022. doi:10.20944/preprints202202.0001.v1.

[20] V. S. Barletta, F. Cassano, A. Pagano, A. Piccinno, New perspectives for cyber security in software development: when end-user development meets artificial intelligence, 2022, p. 531 – 534. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85146434281&doi=10.1109%2f3ICT56508.2022.9990622&partnerID=40&md5=8f8c310fdc8f750cd071562bef2f23be. doi:10.1109/3ICT56508.2022.9990622.

[21] S. Ruhela, Thematic correlation of human cognition and artificial intelligence, 2019 Amity International Conference on Artificial Intelligence (AICAI) (2019) 367–370. URL: https://api.semanticscholar.org/CorpusID:139165389.

[22] V. S. Barletta, F. Cassano, A. Pagano, A. Piccinno, A collaborative ai dataset creation for speech therapies, volume 3136, 2022, p. 81 – 85. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85130701005&partnerID=40&md5=e5349d9fedc09b849dc4fc1c8debf3be, cited by: 7.

[23] J. E. Block, E. D. Ragan, Micro-entries: Encouraging deeper evaluation of mental models over time for interactive data systems, 2020 IEEE Workshop on Evaluation and Beyond - Methodological Approaches to Visualization (BELIV) (2020) 38–47. URL: https://api.semanticscholar.org/CorpusID:221677042.

[24] R. Conradi, A. I. Wang, ESERNET (Eds.), Empirical methods and studies in software engineering: experiences from ESERNET, number 2765 in Lecture notes in computer science, 1 ed., Springer, Berlin; New York, 2003. URL: https://doi.org/10.1007/b11962.