Multimedia Tools and Applications (2025) 84:32109–32150
https://doi.org/10.1007/s11042-024-20414-5

From human-centered to symbiotic artificial intelligence: a focus on medical applications

Giuseppe Desolda[1] · **Andrea Esposito**[1] · **Rosa Lanzilotti**[1] · **Antonio Piccinno**[1] · **Maria F. Costabile**[1]

Received: 2 September 2024 / Revised: 10 October 2024 / Accepted: 21 October 2024 /
Published online: 28 November 2024

Abstract
The rapid growth in interest in Artificial Intelligence (AI) has been a significant driver of research and business activities in recent years. This raises new critical issues, particularly concerning interaction with AI systems. This article first presents a survey that identifies the primary issues addressed in Human-Centered AI (HCAI), focusing on the interaction with AI systems. The survey outcomes permit to clarify disciplines, concepts, and terms around HCAI, solutions to design and evaluate HCAI systems, and the emerging challenges; these are all discussed with the aim of supporting researchers in identifying more pertinent approaches to create HCAI systems. Another main finding emerging from the survey is the need to create Symbiotic AI (SAI) systems. Definitions of both HCAI systems and SAI systems are provided. To illustrate and frame SAI more clearly, we focus on medical applications, discussing two case studies of SAI systems.

Keywords Human-centered design · Human-AI collaboration · Human-AI Symbiosis · Survey.

1 Introduction and background

The rapid growth of Artificial Intelligence (AI) has significantly driven research and business activities in recent years. AI applications range from creative fields, like image generation [109], to critical areas, like medical decision support systems [24, 31]. This interest in AI technologies is exacerbating some latent concerns, of which the interaction with AI systems is one of the most critical. Indeed, while AI models promise impressive performance, their potential is often constrained by how users interact with AI. For example, in the field of medicine, an AI-based diagnosis system is typically able to determine whether a patient has a tumor based on an MRI scan. However, while these AI-based systems are highly precise, they often remain confined to research settings because, in a real-world

✉ Andrea Esposito
andrea.esposito@uniba.it

[1] Computer Science Department, University of Bari Aldo, Via E. Orabona 4, 70125 Bari, Italy

context, a doctor does not want to and cannot rely on a black-box system that, for instance, does not provide insights into why an MRI led the AI to recognize a tumor.

In response to these challenges, a new field of study has emerged in recent years at the intersection of Human-Computer Interaction (HCI) and AI, namely Human-Centered AI (HCAI): it proposes a new perspective on the interaction with AI, aiming at *augmenting* rather than *replacing* humans and their expertise [121]. HCAI envisions AI systems that are *ethically aligned*, *reflecting human intelligence*, and are *designed considering human factors* [143]. HCAI also promotes the design of AI systems that are *reliable*, *safe*, and *trustworthy*, addressing some of the latent concerns [122]. Recent research in HCAI suggests the need for a truly human-centered process in designing AI systems [47]. In fact, studies have shown how there is no real "one size fits all" approach when designing AI systems, as different user goals bring the need for a different level of automation and control in a system [47].

While HCAI aims to perfect the interaction with AI systems, it brings problems peculiar to newly emerging disciplines. One example concerns the need for a comprehensive approach to develop systems that considers interaction design methods and the integration of AI models that support such interaction. This is because, during user requirements analysis, the need to design a user interface that exposes functions that cannot be implemented might emerge; for example, users might require a certain type of explanation that cannot be implemented due to the use of a black-box model [66]. Another problem relates to the need for different scientific communities to collaborate. Multidisciplinary HCAI design teams should include experts in HCI, AI, software engineering, ethics, and experts in the specific application domain (e.g., medicine, automotive). This multidisciplinary nature exacerbates the terminology inconsistencies (e.g., synonyms used to refer to the same concept, same words used to refer to different concepts) and the lack of awareness of existing solutions in related disciplines, since the different research communities have not yet found a common ground (e.g., an HCI expert knows how to design a user interface, but probably might not know in detail the AI models needed). Such issues are not novel: this is a rather well-known problem in HCI (multidisciplinary itself) and, in general, in every multidisciplinary field. However, the collaboration with experts from other disciplines could mitigate the need, of a specific expert, for multidisciplinary knowledge. Still, it becomes essential to share a common vocabulary and be aware of the advantages and benefits of existing solutions in related disciplines.

Survey studies are instrumental to shed light on critical emerging issues of new research fields (see, e.g [20, 66]). With a similar goal, this article first reports a survey of the scientific literature on HCAI. The following 4 Research Questions drove this survey: (1) *What disciplines are involved in HCAI? (2) What are the concepts underlying HCAI and the terminology inconsistencies among them? (3) What are the solutions available in the related disciplines that might support the creation of HCAI systems? (4) What are the emerging challenges?* The process of surveying the state of the art started from a seed of 7 articles reporting definitions of HCAI [20, 64, 117, 121, 122, 124, 143].

Based on this deep survey of the literature, one of the contributions of this article is that it clarifies the disciplines, concepts, and terms around HCAI, solutions to design and evaluate HCAI systems, as well as the emerging challenges. For each definition of HCAI provided in the 7 articles of the initial seed, we identified the main concepts on which it is based. These concepts were then extended by searching for articles that discuss them. This process led to the identification of 7 scientific disciplines that revolve around HCAI, 15 fundamental concepts, 25 solutions for designing HCAI systems, 23 solutions for evaluating HCAI systems, and 11 challenges.

Another valuable contribution provided by this article, which results from the lessons learned from the survey, is the need for *Symbiotic AI* systems as a specialization of HCAI systems. Symbiotic AI extends the old and well-known concept of Man-Computer Symbiosis [90] and requires a continuing and deeper integration between humans and AI, i.e., a symbiosis of human intelligence and artificial intelligence, where both humans and AI augment each other's capabilities thanks to a collaboration that balances each other's strengths and weaknesses [17]. In other words, Symbiotic AI leads to the creation of AI systems functioning as cognitive orthotic systems that collaborate with humans rather than cognitive prostheses that replace human abilities [64]. SAI systems are of particular interest in medicine since, as reported in [17, 31] physicians do not want AI systems that replace doctors but tools that collaborate well with doctors. To illustrate and frame symbiotic AI more clearly, two case studies on medical applications are also presented.

The article is, therefore, structured as follows. Sections 2, 3, 4 and 5 discuss the findings related to the above 4 Research Questions, respectively. Section 6 clarifies the meaning of Symbiotic AI, while Sect. 7 discusses its practical implications by analyzing two case studies on medical applications. Finally, Sect. 8 concludes the article.

2 Disciplines involved in HCAI

Before analyzing basic concepts and terms used in the HCAI and illustrating what solutions exist for creating HCAI-based systems, it is crucial to understand what disciplines are converging in this new research area. This section discusses the findings related to Research Question 1. Besides the two pillars of HCI and AI, other disciplines converge into HCAI. Table 1 summarizes these disciplines, their key contributions to HCAI, and the main interrelationships among them. It should be noted that in this manuscript we refer to "disciplines" even when referring to subfields. This because all fields and subfields are relevant enough to HCAI that it is not useful to classify them differently.

The beating heart of HCAI is undoubtedly **Human-Computer Interaction** (HCI). It concerns the design, implementation, and evaluation of interactive computing systems for human use and the study of major phenomena surrounding them [70]. Leveraging guidelines and knowledge from social sciences (including psychology and cognitive science), HCI aims to improve the quality of the dialog between users and computers by developing and evaluating usable software, i.e., software that can be used efficiently, efficaciously, and satisfactorily by well-specified groups of users in specified contexts [74]. In HCAI, the methodologies proposed in HCI aim to ensure that AI systems are user-friendly, accessible, and tailored to the human context of use. Furthermore, HCI approaches help create AI systems aligned with human values and needs, allowing the creation of systems that augment and complement human capabilities rather than replace them.

The other pillar of HCAI is **Artificial Intelligence** (AI). In this survey, AI broadly refers to all systems with some form of "intelligence", regardless of their implementation [116]. AI focuses on creating systems that perform tasks that usually require human intelligence traits, such as perception, reasoning, learning, and decision-making. In the context of HCAI, AI provides the technical foundation of systems and ways to build intelligent machines and evaluate their performance for the tasks at hand.

A discipline related to AI that is fundamental to HCAI is **Machine Learning** (ML). It involves the development of algorithms that enable computers to learn from data and make predictions or decisions based on it [116]. Most recent AI models are based on neural

Table 1 Disciplines converging in HCAI, their key contributions, and the main interrelationships among them

Discipline	Key contributions to HCAI	Interrelationships
Human-Computer Interaction (HCI)	To take care of the human-centered aspects, addressing usability and user experience	HCI principles and design methodologies inform the design and evaluation of the interfaces of HCAI systems
Artificial Intelligence (AI)	To provide computational models and algorithms	AI solutions are influenced by Ethics and HCI principles
Machine Learning (ML)	To develop adaptive models that improve with data	ML models are the core of AI systems
Explainable AI (XAI)	To provide transparency in AI decision-making processes	HCI principles influence XAI to improve trust and confidence in AI systems
Software Engineering (SE)	To develop and maintain HCAI systems	SE provides quality properties and methodologies for integrating AI in broader software systems, ensuring scalability and reliability
End-User Development (EUD)	To empower users to customize HCAI systems	EUD provides some indications to democratize AI, making it more accessible and adaptable by lay users.
Ethics	To guide development ensuring fairness, transparency, and accountability	Ethics provides principles for AI design that prevent biases and discrimination

networks, which are a specific type of ML models that have proven to be very versatile and powerful and are the main players in deep learning, a subset of ML focusing on complex neural networks. Differently from ML, AI also includes expert rule-based systems, first-order logic systems, symbolic reasoning systems, etc. In a nutshell, AI is the overarching concept of machines being able to perform tasks in a way that we consider "intelligent". ML is a subset of AI that involves learning from data and improving over time without being explicitly programmed for specific tasks.

EXplainable Artificial Intelligence (XAI) has risen very quickly in recent years. It has often become an integral part of HCAI studies [20]. XAI focuses on creating AI (or ML) models whose decision processes are interpretable and understandable by humans [66]. Therefore, XAI aims to provide transparency in AI decision-making processes, to allow users to comprehend how and why decisions are made. In HCAI, XAI is crucial as it aids in building trust and confidence in AI systems and ensures that users can understand, verify, and challenge AI outcomes. Furthermore, XAI can help in guaranteeing accountability in case of incidents.

While HCI is concerned with the interaction with user interfaces and AI is concerned with creating models that impart intelligence to the application, the creation of systems that can be used outside of experimental settings falls into the area of **Software Engineering** (SE). In the context of HCAI, SE enables the creation of quality systems, ensuring that AI systems are designed, developed, and maintained efficiently and effectively [14]. SE practices are also crucial for integrating AI systems into larger software systems, ensuring reliability, scalability, and maintainability while guaranteeing human-centered design principles.

A subfield of HCI that overlaps with SE is **End-User Development** (EUD). It aims to empower users to modify, or even to create, software artifacts without requiring deep (if any) programming knowledge [91]. In the context of HCAI, EUD involves designing systems that allow users to customize and extend AI functionalities to meet their needs. This approach enables the democratization of AI technology, making it more accessible and adaptable by a broader range of users, including lay users [46]. Furthermore, EUD proves helpful when dealing with situational exceptions or errors, allowing for reconfiguring existing systems to better fit users' needs [118]. The goals of EUD are to enable HCAI systems to be: (i) created from scratch by the same end-users who could use AI to satisfy their needs, (ii) adapted to fulfill the users' goals better, or (iii) corrected in case of errors or exceptions, either instantaneously or by changing the AI behavior. Research on the first two goals of EUD for HCAI is still in its early stages [46], even though *interactive machine learning* aims at allowing domain experts to change the dataset to train AI interactively [2, 51]. Regarding the third goal, *intervention user interfaces* empower users with mechanisms to deal with situational exceptions or AI reconfiguration [118]. An example of an intervention user interface for AI is discussed in Sect. 7.1.

While the former disciplines involved in HCAI are concerned with technical and technological aspects, **Ethics** is a pure humanistic discipline that strongly intervenes in HCAI. Ethics is generally concerned with the study of morality, i.e., the principles of "right" and "wrong" behavior [13]. In the context of HCAI, ethics proves helpful in guiding the development and deployment of AI systems that are fair, transparent, accountable, and respectful of human rights [20]. Thus, the ethical considerations that characterize HCAI literature help prevent bias, discrimination, and other negative impacts on individual users and society. The discussion of ethical AI (or machine ethics) mainly focuses on three principles: *fairness*, *transparency*, and *accountability* of AI systems. *Fairness* is defined as the lack of any prejudice or favoritism towards an individual or a group based on intrinsic or acquired properties in the context of the decision-making process [96]. In other words, *fairness* is the lack of *bias* and *discrimination* [96]. *Transparency* is a complex concept that, according to [97], can be given two different meanings: the ability of the algorithm to (i) explain itself and its inner workings or (ii) make itself seamless and make objective outcomes of the users more apparent. Transparency is, therefore, a requirement of *accountability*, which is related to responsibility. Accountability refers to the role that should be considered "responsible" for the AI system's behavior [97]. Accountability is an important principle, especially in unexpected or harmful behaviors. The discussion on machine ethics in the context of HCAI underlines the importance of *human inclusion*. Including humans in the loop of AI system design may, in fact, mitigate some of the ethical issues discussed in the previous paragraphs [20].

3 HCAI concepts and possible inconsistencies

Due to its intrinsic multidisciplinary nature and the great interest in this novel discipline, HCAI still lacks a commonly accepted definition. Two main trends have emerged in the scientific literature on AI and on HCI. The AI community primarily indicates as "human-centered" those AI systems that adopt a formal description of the final users in their decision-making process [15, 147]. On the other side, HCI identifies as "human-centered" those systems that are designed and evaluated by involving users in the iterative process defined by ISO 9241 – 210 [75]. More specifically, the HCI community identifies HCAI

systems as AI-based systems that are *useful* and *usable* for their users [143] while being *reliable*, *safe*, and *trustworthy* [122]. In HCI, the overall process to create systems requires accurate user research to deeply analyze real users and to specify their requirements and context of use, using the collected information to define design solutions and to evaluate them. Often, real users are also directly involved in design and evaluation. Therefore, leveraging the formal description of users employed by the AI community and the perspective of HCI, we define HCAI systems as follows.

Definition 1. Human-Centered AI Systems

Human-Centered AI systems are AI systems that are designed, developed, and evaluated by involving users in the process, with the goal of increasing performances and satisfaction of humans in specified tasks. HCAI systems aim to be *useful* and *usable* for specified users, who might be described through a formal model, to reach their specified goals in their context of use, while being *reliable*, *safe* to use, and *trustworthy*.

Different basic concepts characterize HCAI. In some cases, the same terms identifying certain concepts are used by different disciplines with varying meanings; in other cases, different terms refer to the same concept. These terminological inconsistencies are typical in multidisciplinary and new fields such as HCAI. In this section, referring to Research Question 2, we present and analyze our survey findings about fundamental concepts that revolve around HCAI, identifying possible synonyms and homonyms.

3.1 Interpretability vs. Explainability vs. Transparency

The terms "interpretability" and "explainability" are frequently used interchangeably. However, although similar, slight differences can be recognized between the two terms. In the context of AI, explainability is the property of an AI model that allows for generating human-understandable explanations that provide insights into the model's decision-making process [1, 39, 98]. *Interpretability* is the ability of a system's decisions to be interpreted [66]. Thus, interpretability inherits its meaning from logic, where an *interpretation* is an assignment of meaning to the symbols of a formal language [71]. Another meaning of interpretability is the clarity of the model's inner workings and mechanisms [111]. Therefore, one can conclude that, in AI, a model is interpretable if its decision-making process is transparent and understandable by humans [39]. Although this definition implies that a system is interpretable only if its decision-making process is comprehensible without the need for explicit explanations, most of the literature uses *interpretability* as a generalization of *explainability* [65, 118]. As previously stated, *transparency* is instead a complex concept that can be given two different meanings: the ability of the algorithm to (i) explain itself and its inner workings or (ii) make itself seamless and make objective outcomes of the users more apparent [97]. Therefore, *explainability* is related to *transparency* through its first meaning, while *interpretability* is related to its second meaning.

In the context of XAI, *explainability* is usually subdivided into *ante-hoc* and *post-hoc* [111]. Post-hoc explainability uses methods that, given the model itself and its output, provide insight into how the model arrived at its output [111]. There are several methods for post-hoc explainability [66], and some of the most adopted are SHAP [92], LIME [112], Grad-CAM [119]. On the contrary, ante-hoc explainability, also known as "intrinsic explainability" requires a "white-box" model that is explainable by default without needing external methods [111]. Therefore, as defined before, one can conclude that ante-hoc explainability is synonymous with interpretability.

In the context of HCAI, sometimes, post-hoc explainability is frowned upon, since it has been argued that one cannot be sure that the explanation provided by a post-hoc model is faithful to the actual decision-making process employed by the AI model [115]. However, recent research has shown that some end-users may not require full explainability [31], suggesting that AI's explanations should be designed through user studies and by factoring in the level of risk that AI decisions pose to its end-users [47].

3.2 Automation vs. Autonomy

Although strongly related, *automation* and *autonomy* are slightly different concepts. In HCAI, automation refers to employing technology to perform tasks without human intervention (if not for dealing with errors or misbehaviors) [131]. Automation does not necessarily require AI, but simple algorithms or rules may suffice to automate more straightforward tasks [131]. As per its definition, automation reduces the amount of required human interaction: depending on the "level of automation" (usually measured on a 10-point scale), the human may have complete control or may not have any way to intervene in the system decisions [105]. Examples of automated systems are manufacturing robots. However, due to the limited possibility of intervention, automated systems are frowned upon in higher-risk and safety-critical scenarios.

Autonomy goes slightly beyond the typical concept of automation. Autonomous systems are not only capable of performing tasks without the need for human intervention, but they also can adapt to new scenarios [138]. For this reason, autonomous systems usually require some AI model to power their decisions. A critical aspect of an autonomous system to adapt to new scenarios is the ability to continuously learn through interaction with humans: this instantiates a symbiotic relationship between the human (who benefits from the automation of the task) and the system (which benefits from the continuous stream of new data) [64]. Examples of autonomous systems are autonomous vehicles and personal assistants.

3.3 Trust vs. Trustworthiness

In general, an AI system can be considered *trustworthy* if it is deemed worthy of being trusted by its end-users based on specific verifiable requirements [80]. However, *trust* is an abstract concept that makes defining it highly complex. Recent research suggests that providing context-dependent definitions of trust benefits one's understanding of the concept [7]. A first definition expresses trust as *the willingness of a party (the trustor) to be vulnerable to the actions of another party based on the expectation that the other (the trustee) will perform a particular action important to the trustor*, *irrespective of the ability to monitor or control that other party* [94]. Therefore, this definition sees trust as a property of the dialog between two agents, the AI and the user. Other definitions see trust as *the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability* [86] or as *the belief on whether the application could fulfil a task as expected (the trustworthiness of mobile applications relates to their dependability*, *security*, *and usability)* [144]. Thus, trust is also subordinate to the users' goals and safety, underlining the relevance of a human-centered approach to designing AI systems to meet users' goals and expectations. Regarding trustworthiness, it is an interesting metric for the HCAI system, which is on par with the usability of the classical system. However,

most of the literature adopts custom-made questionnaires to evaluate users' trust, as well as qualitative methods [7], suggesting the need for more research on objectively evaluating trust. Examples of adopted questionnaires are the "Trust in Automated System Test (TOAST)" [141], the "Multidimensional Measure of Trust" [133], and the "Trust in Automation (TiA)" questionnaire [77].

3.4 Usability vs. User eXperience (UX)

Usability and *User eXperience* (UX) are among the main concerns of HCI. The HCI community has been discussing the usability of interactive systems since the 1980s. More recently, HCI has become increasingly concerned with UX [114]. It is now acknowledged that designing for experience includes but it is much more than designing for efficiency and other traditional attributes of usability. While efficiency is focused on attributes such as fast, easy, functional, and error-free, UX involves feelings and thus focuses on beautiful (harmonious, clear), emotional (affectionate, lovable), stimulating (intellectual, motivational), and on tactile (smooth, soft), acoustic (rhythmic, melodious) in case of multimodal interfaces [69].

Usability is defined in the ISO standard 9241 as *the extent to which a system*, *product*, *or service can be used by specified users to achieve specified goals with effectiveness*, *efficiency*, *and satisfaction in a specified context of use* [74]. The same ISO standard defines UX as *the whole set of a person's perceptions and responses resulting from the use and/or anticipated use of a product*, *system*, *or service* [74]. Therefore, UX is a slightly broader concept: a good level of usability is required to avoid hampering UX, but UX also includes hedonistic aspects not involved in usability quality [69].

Current approaches for designing AI systems are mainly data-based, rather than user-based. AI models serve a single purpose (classification, prediction, or generation once provided with a specified input) that highly depends on data and on the designer's decisions. However, recent research shows that the end-users, their goals, and context of use highly impact how AI should be designed, influencing the overall UX [47]. Thus, an AI system capable of evoking a good UX requires a purely human-centered approach to its design [143].

3.5 Human-in-the-Loop (HITL) vs. Human-on-the-Loop (HOTL)

A cornerstone of HCAI is the inclusion of humans (the users) in the design of AI systems. However, user involvement is crucial not only in the initial stages of the life of an AI system but also throughout its use. More precisely, two types of user involvement are considered, depending on the role that humans have: human-in-the-loop (HITL) and human-on-the-loop (HOTL) [54, 142].

A HITL approach involves users in the decision-making process during the system usage [142]. The AI system is not created as an "oracle" that processes users' inputs to provide an outcome. The interaction is thus bi-directional: users actively contribute to AI decisions by providing expertise and experience; the system employs its inner AI algorithm to support users.

On the contrary, a HOTL approach does not involve end-users in decision-making. Still, it involves users in detecting and fixing errors or misbehaviors of the system [54]. HOTL is beneficial when high levels of automation or autonomy are possible, but risks may arise in the case of wrong decisions. Thus, the capability to let users *intervene* (for example,

through intervention user interfaces [118]) proves extremely useful in dealing with exceptions or errors. Although systems created through an HOTL approach are not entirely symbiotic in their relationship with humans, they are still human-centered if they are designed according to HCD. They should not be discarded in favor of HITL systems: recent studies show that different approaches may be beneficial to reach different goals, thus leaving the final choice to the end-users for which the system is being designed [47].

The two approaches imply the design of different interaction mechanisms and different implementations of AI models. For example, the interaction with HITL systems should be designed to encourage users to provide continuous input to the AI system. At the same time, the AI model should be able to learn from these inputs and improve its performance over time. Of course, final decisions should never be completely delegated to the AI. On the contrary, the design of interactions with HOTL systems should prioritize autonomous behavior of the AI models but must include interaction mechanisms that allow users to act as supervisors, monitoring and intervening when necessary to either stop or control the behavior of the AI model. It should be noted that the two approaches are not mutually exclusive, and users may sometimes benefit from choosing between the two approaches, thus fostering a symbiotic human-AI relationship [31].

3.6 Fairness vs. Bias

One of the cornerstones of HCAI is *fairness*. In a survey on bias and fairness [96], fairness is defined as the lack of bias and discrimination. Different biases can affect the model's inner decision-making process when dealing with AI systems, hampering fairness. AI bias can be classified into three groups, depending on where it has the most effect: data to algorithm, algorithm to user, and user to data [96]. Bias from data to algorithm only derives from the dataset itself. Examples of such bias are measurement bias (which arises from how a particular feature is measured), sampling bias (which arises from how data is sampled, especially in underrepresented classes), and representation bias (which arises from how a feature is represented). Bias from algorithm to user, instead, derives from the algorithm itself. Examples of such bias are algorithmic bias (which arises from the algorithm itself and its parameters, for example, due to limitations in the functions used) and presentation bias (which arises from how a particular outcome is presented). Finally, bias from user to data derives from how end-users behave, affecting how new data points are generated before being used to improve an existing AI model. Examples of such biases are historical bias (which arises from socio-technical issues that seep into how users behave) and population bias (due to differences in the population for which the AI was designed and the actual end-user population).

A truly HCAI system must present strategies to avoid or, at least, mitigate the different types of biases. Various strategies can be considered. For example, techniques for unbiasing data can be adopted [96]. Similarly, governance structures can be instated to ensure control over the fairness of the used dataset and the final AI decision-making process [123].

3.7 Ethical AI vs. Responsible AI

When discussing HCAI, the AI system must follow norms and regulations. Such norms and regulations should not be intended as merely legal; it is also fundamental that AI systems adhere to moral guidelines and norms. Therefore, one may discuss "ethical AI" and "responsible AI". Ethical AI is a term used to describe an AI system that follows moral

norms and principles [126]. Their primary aim is to avoid any harm to its users and society at large. On the other hand, responsible AI is a term used to describe designing, developing, and deploying AI in a way that considers the legal implications of the technology, ensuring its accountability [107].

In the context of HCAI, AI systems should be developed to guarantee adherence to moral guidelines and respect legal requirements [122]. Thus, a truly HCAI system is an example of ethical and responsible AI. To aid in creating an HCAI system, sets of AI-specific guidelines, which encompass ethical and legal requirements and governance structure to ensure adherence to such guidelines, will be useful [123]. The European AI Act is an example of such regulation that may guide the design and development of both Ethical and Responsible AI systems [49]. Sadly, although ethical and legal aspects of HCAI are often mentioned in the surveyed articles, they are still not discussed in detail.

4 Solutions supporting the creation of HCAI systems

To successfully create HCAI systems, it is useful to have in-depth knowledge of the design and evaluation solutions that exist in the related disciplines and emerged in this survey. By design we refer to both design and development. An overview of such solutions is provided, so that experts can engage in a holistic approach to the creation of HCAI systems. Design is addressed in Sect. 4.1, while evaluation in Sect. 4.2.

4.1 Designing HCAI systems

Each discipline identified by the survey proposes solutions that can be useful in the design of HCAI systems, as reported in the following. Figure 1 summarizes the main solutions.

4.1.1 HCI design solutions

HCAI emphasizes the idea of creating systems that are both performant from the technological point of view and valuable from the human values and user requirements perspective. Applying HCI principles and methods to the design process of HCAI systems provides a strong foundation for realizing this balance. This section explores a range of established HCI solutions for integrating user perspectives into the design of HCAI systems.

Human-Centered Design (HCD), originally called *User-Centered Design* (UCD) [114] is the general model adopted in HCI for the design of systems that satisfy users' needs and expectations; it prescribes that users are involved from the very beginning of the planning stage, and identifying user requirements becomes a crucial phase [114]. Various design processes based on HCD are now available, like UCD Sprint [84]. HCD requires understanding who will use the system, where, how, and to do what. Then, the system is designed by iterating a design-evaluation cycle. Being the design based on empirical knowledge of user behavior, needs, and expectations, it is possible to avoid serious mistakes and to save re-implementation time to solve such mistakes. This model is reported in the standard ISO 9241 – 210 [75]. In HCAI, adopting HCD is a must when a direct interaction between users and system is expected (e.g., in consumer applications, educational software, or health-related support). UCD may be less important when users do not interact frequently with the system, or an AI system is developed to run independently without users' frequent interference, such as backend AI.

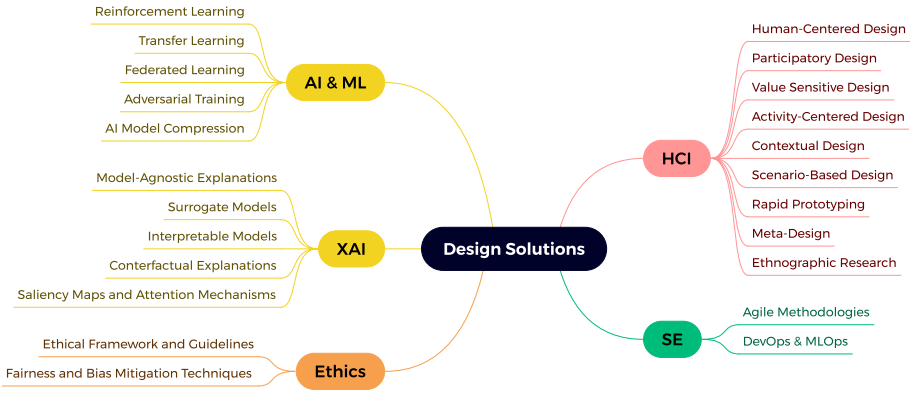

Fig. 1 Overview of the design solutions proposed in the disciplines converging in HCAI

Several important design solutions that are based on HCD are available. One is *Participatory design* (PD), a broader version of UCD, which difference is that users are not only involved in the design process as subjects of study but as participants in the design team, involved in making choices about the system being designed. This approach is useful when it comes to creating HCAI systems for specific subgroups of people, such as accessible technologies and culturally suitable AI products. However, users' intervention may be limited by the amount of time available in a project and the resources at a project's disposal and, hence, not very efficient for projects requiring quick turnaround time or projects whose user base is relatively more homogenized.

More specific HCD solutions can be also adopted to design HCAI systems. *Value Sensitive Design* (VSD) is an approach that suggests design choices that reflect and consider ethical considerations and human values [57]. VSD recognizes that AI systems are not isolated entities but operate in social, cultural, and moral environments. VSD aims to develop appropriate systems that meet privacy, equity, and transparency requirements. Even though VSD is valuable for handling ethical issues, it may not be as useful in cases when the AI system acts within tightly framed technical fields with relatively low ethical characteristics.

Activity-centered design (ACD) is a design process that starts with the users and the activity they carry out, with the idea of better designing a system that aligns with specific tasks [59]. This design approach is particularly adequate in professional and enterprise settings where comprehensiveness of work accomplishment is critical. ACD is appropriate for creating HCAI systems for productivity improvement, including smart tools for engineers or business intelligence specialists. As ACD focuses on specific activities, creating systems that can fit the work schemes and enhance user productivity is possible. Nevertheless, ACD may not be practical for some other applications where user experience, in terms of their emotions, is more valuable than efficiency, for example, in entertainment or leisure-based AI systems.

Contextual Design (CD) is an approach that entails observational techniques to understand the environmental and social aspects relevant to users' behaviors in technology use [72]. CD is especially useful when designing HCAI systems either for particular contexts or that are employed in flexible contexts, such as emergency services. Applying CD may be impractical in situations where the context of use is either strictly defined or has a negligible impact on the system's performance, such as in static or virtual settings.

A similar approach is represented by *Scenario-Based Design* (SBD), which includes the specification of interaction between the user and the system in the form of narrative scenarios [21]. These scenarios, which can be based on finely developed user research contexts, illustrate to the designers how users will use the system in real life and how the system should respond to such usage. SBD is especially helpful in designing a strategy where the behavior of users may differ from one condition or task to the other. Considering different scenarios in the design process might help designers identify potential challenges and opportunities, allowing for a more robust and flexible system design. However, SBD could be less effective in scenarios where user interactions are quite standardized and do not differ from user to user.

The HCD design model includes prototyping as a crucial phase to test different design ideas. However, the term *rapid prototyping* is also used to indicate the creation of simple and preliminary prototypes of user interfaces that can be iteratively evaluated in a few cycles [68]. Rapid prototyping is particularly useful while designing HCAI systems because it enables designers to model and experiment with a system before it is fully developed. It is especially beneficial for systems where the interface is experimental or the interaction is unclear, and users' feedback is crucial.

As stated above, Participatory Design considers the participation of end-users in the design process. The rationale is that end-users are experts of the work domain so a system can be effective only if these experts are allowed to participate in its design, indicating their needs and expectations. End-users are increasingly willing to shape the software they use to tailor it to their own needs [10]. As said in Sect. 2, End-User Development (EUD) is a sub-field of HCI that promotes a more active involvement of end-users in the overall software design, use, and evolution processes; end-users thus become co-designers of their tools and products [5, 55], enabled to shape software artifacts at run time. One of the most influential frameworks for supporting EUD is *Meta-design* [53]. It considers a two-phase design process: first, the meta-design phase consists of designing software environments where different stakeholders (domain experts, end-users, etc.) can work and are also allowed to create various versions of the final system by tailoring it to their needs (design-phase). Thus, Meta-design promotes the design of open systems that various people, acting as end-user developers, can modify and evolve at use time. The advantages of EUD solutions for HCAI systems are already mentioned in Sect. 2.

All the above HCI design solutions suggest considering end-users in the system's lifecycle at different phases and with different levels of involvement. Thus, it is instrumental to apply HCI techniques that allow designers to gather data from the end-users correctly. In HCI, a valuable solution for gathering user requirements in the best way is *Ethnographic Research*, which consists of a long-term observation of the users in their natural environments to understand their behaviors, needs, and challenges [29]. It also enables the researcher to understand the cultural, social and context in which the users will engage with the technology. When it comes to HCAI, ethnographic studies can be useful in designing equipment for cultural or community usage by making the technology applicable to the culture of a particular community. It is especially helpful in designing a system for special or marginalized users as, in most cases, their experience cannot be identified by conventional user research methods. Truly ethnographic research is costly and may take a long time to complete, thus it is often replaced by faster ethnography-based techniques like shorter *field observations* and *contextual interviews* [106].

4.1.2 AI and ML design solutions

AI and ML algorithms are fundamental to designing HCAI systems. They provide a vast variety of approaches, which are beneficial to solving the different issues occurring during HCAI system development, each with its advantages and disadvantages.

Reinforcement learning is an area of ML that includes models such as Q-learning, deep Q-learning, deep Q-network, and policy gradient methods, which are used to help the AI agent choose the best action to engage with the environment [78, 89]. These models are especially useful in developing HCAI systems that necessitate flexibility in hardly predictable contexts. For instance, in self-driving cars, RL using deep Q-network or policy gradients can make decisions instantaneously in real-world complexities, including the ability to learn new tactics from its trial-and-error experiences. However, reinforcement learning models could be computationally heavy and might need a lot of training data, making their use sometimes impractical, especially when it is needed to quickly deploy or where resources that could be harnessed are limited. In addition, the exploration required by reinforcement learning models can lead to ethically questionable actions during the learning phase. This makes these models less adequate for applications where mistakes could result in harm, as in the case of medical applications. Therefore, reinforcement learning models are helpful when continuous adaptation and optimization are crucial, but they should be used cautiously in high-stakes or resource-constrained settings.

When creating HCAI systems across different domains or in cases with little labeled data, *transfer learning* techniques should be considered [132]. This class of solutions can be applied to architectures like convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and transformer models such as BERT or GPT. For example, a CNN trained on a large image database may be used in a medical application by subsequent training with a small set of labeled images, which is likely to increase diagnosis accuracy. Likewise, large language models such as BERT trained on vast text collections can be fine-tuned for more specific tasks such as sentiment analysis, or question answering in different languages. A particular example of transfer learning follows the *student-teacher* model [137]. However, transfer learning is more effective when the source and target tasks are similar, as differences may lead to poor transfer of performance or learning or even introduce potentially biased learning. In addition, the biases of these models may be inherited, which may raise ethical questions in such human-oriented fields. Transfer learning models are thus useful for broadening the domains to which an AI system can be transferred, if the results are generalizable. However, it needs to be done properly to avoid degrading the transferred model's relevance and fairness.

Other approaches beneficial for HCAI systems where data privacy and security are critical (e.g., in healthcare or finance) come from the area of *federated learning* [148]. This can be exploited with classifiers such as neural networks (e.g., CNNs or LSTMs), gradient-boosted trees, and support vector machines (SVMs), when they are trained across multiple devices or servers while keeping the data localized. For instance, using CNNs in a federated learning setup allows the development of a global model for medical image analysis while ensuring that patient data never leaves the local device, thereby complying with privacy regulations. However, federated learning can introduce challenges related to communication overhead, model synchronization, and data heterogeneity across devices, which can affect the overall performance and fairness of the system. Additionally, the decentralized nature of federated learning might make it difficult to enforce uniform training standards, potentially leading to biased models if the data distributions are not representative.

Federated learning is most effective in scenarios where data privacy is essential, but it may not be ideal in cases where centralized models offer superior accuracy and consistency.

Other approaches to be considered when developing HCAI systems requiring AI decisions to be secure (e.g., cyberspace, self-driving cars, or stock exchanges) come from the area of *adversarial training* [81]. These techniques are often applied together with deep neural networks, convolutional neural networks and recurrent neural networks to improve models' robustness against adversarial perturbations. For instance, enhancing CNN with adversarial examples allows the network to deal with adversarial manipulations, making it more secure in image classification tasks. Nevertheless, adversarial training can be computationally heavy, and it often leads to a trade-off in the model performance and robustness when the adversarial instances employed for training are insufficient in capturing all the possible attacks. Besides, in situations where the adversarial attack is unlikely, the cost of the adversarial training could be higher than the advantages gained. Adversarial training is, therefore, effective for increasing the resilience of HCAI systems against adversarial threats. However, it should be performed when the adversarial risk in the given domain has been assessed.

HCAI systems may be designed for contexts with limited resources, such as smartphones, IoT devices, or edge computing nodes. *AI model compression* techniques (like pruning, quantization, and knowledge distillation) are used on models (including DNNs, CNNs, and transformer models) to make them smaller, thus requiring less computational power [26]. For instance, a DNN can be optimized by eliminating unnecessary links, which means cutting off unimportant connections. This can lead to a sharp decrease in the model size and, at the same time, an increase in productivity, though the performance will not be influenced dramatically. Equally, quantization can convert model weights from floating-point to lower precision to make the inference process lighter. Knowledge distillation is the process of letting a small model (the so-called student) learn from a great, complex model (the teacher), producing a compact model that does not lose much of the performance. However, these compression techniques sometimes cause a reduction in the accuracy of the model or the removal of some features, which can affect the system's performance in handling difficult problems. In some cases, where computational resources are not a limiting factor and accuracy is crucial (e.g., in cloud machine learning services), the potential drawbacks of model compression might offset the advantages. Thus, AI model compression is relevant for HCAI systems that need to be efficient, but it should be considered more thoroughly for systems that may benefit from increased accuracy.

4.1.3 XAI design solutions

For HCAI systems, explaining model decisions and interpreting AI models becomes crucial to establishing trust, reducing unethical practices, and increasing the overall transparency of the AI systems. By understanding the strengths and weaknesses of XAI methods, designers and developers can select the most appropriate tools to enhance transparency and trust in AI systems, thereby creating more effective and ethically sound HCAI solutions.

HCAI systems are often based on "black box" models, such as deep neural networks, where understanding individual evaluation is critical to fostering user trust and fulfilling accountability. In this situation, *model-agnostic explanations* prove to be quite versatile to explain predictions made by different ML models, regardless of their application domain [66]. Examples of techniques that provide such explanations are LIME and SHAP. LIME

(Local Interpretable Model-Agnostic Explanations) offers interpretable approximations of the model [112], while SHAP (Shapley Additive Explanations) provides the *feature importance*, i.e., it provides explanations by analyzing the relevance of all features [92]. However, both LIME and SHAP can be quite costly depending on the amount of data, or the complexity of the model used. In addition, the approximations that LIME gives may sometimes be completely wrong, while both methods might not offer enough detail about specific and limited ways of model behavior.

Explanations can also be provided by implementing *global surrogate models*, which are models that estimate the original (black-box) model and hence provide a broad-spectrum view that can be easily understood [66]. This is especially useful in settings where various stakeholders require an overall perspective of the decisions that are being made in the system, hence promoting accountability and compliance with set rules and regulations. The main problem is that global surrogate models are less accurate than the original model, which in turn can produce less accurate explanations of the AI system's work in HCAI systems. Moreover, it may not adequately capture the essence of the decision or decision-making process and may result in misinterpretations due to simplification involved in the mathematically structured models, particularly in a system that needs instance-by-instance routing.

HCAI systems could also be based on *interpretable models*. Thanks to their nature, interpretable models such as decision trees and rule-based models are easily understandable [22]. These models present decision macros that are quite rational, unambiguous and therefore easy for end-users to comprehend and act upon. This is especially desirable in areas such as medical triage or legal analysis, where the user has to understand how the AI decision was made. The main disadvantage of using such models is that they may be less suitable in cases where the relationship between the features or the interactions between them are highly nonlinear, so they may be outperformed by other methods, including deep learning techniques.

Other HCAI systems could need to provide users with explanations about how slight input variations could affect results, and then users can make the right decision with recommendations from the AI system. In this case, AI solutions for c*ounterfactual explanations* must be considered [87, 98]. These solutions are especially useful in analytical areas such as healthcare and prescription, where the users may wish to learn how to manipulate the system to their advantage. However, counterfactual explanations may not always be possible or efficient in advanced HCAI systems where the model is very complex and nonlinear. Thus, it is difficult to come up with valid counterfactual instances. Moreover, when users can or should not change the input data, such as in the diagnosis of diseases, the counterfactual explanation can be of little help.

The last type of explanation regards HCAI systems that work on vision or sequence data, such as MRI and NLP systems. In this case, s*aliency maps and attention mechanisms* might be useful [119, 134]. These techniques emphasize certain aspects of the input for which the model comes to a conclusion, and the information can be easily described visually to the users. This is especially important to improve user confidence in the systems that have to work with information critical for making decisions, such as from medical images for disease diagnosis or legal documents for contract understanding. Nonetheless, the application of these approaches might be disappointing since they contribute to the generation of noise and, hence, hinder decision-making by complicating the process. Also, they are not as useful in natural language explication of tabular data, non-graphics HCAI systems, or other forms of explanation.

4.1.4 Software Engineering design solutions

Software Engineering offers principles, methods, and techniques useful in developing robust, scalable, and ethical HCAI systems.

Considering the need to include the users in the software development lifecycle to create HCAI systems, *agile methodologies* can be considered to get the end-users' feedback on the development process at all stages [146]. Agile methodologies are based on iterative development, feedback, and cross-functional teams [37]. Its emphasis on changes to requirements and users' needs is another advantage of Scrum and Kanban models, which allow the incorporation of Agile into HCAI systems [19]. This iterative process helps ensuring that the system is changing in accordance with actual practice and ethical questions, which is critical to keep the system loyal to people. Moreover, the essential feature of agile practices is flexibility, which allows for a quick reaction to new ethical standards or shifts in user demands to solve unanticipated issues quickly. However, there is a weakness in Agile's approach: it is based on iteration and unsuitable for long-term planning or stability. The continuous cycle of feedback and adjustments might blur the project's goals. Besides, it should be noted that there might be an issue of lesser suitability of agile practices when imposed on organizations heavily bound by compliance standards and protocols since the Agile approach runs counter to the prescriptive approach typical of compliance-intensive environments.

The development of the HCAI system could sometimes require continuous integration, continuous delivery, and automated testing. *DevOps* should be considered in this situation since it joins two aspects of a software engineering process, including software development (Dev) and IT operation (Ops) [42]. DevOps, more precisely a specialization of DevOps for ML called *MLOps* [82], is useful for HCAI systems since this approach allows the release of updates as often as possible – given the constant shifts in users' demands and ethical issues. With integration and testing procedures in place, a developer can guarantee that new code changes, for example, will not contain mistakes, or at least mistakes that are not ethical, to keep the system's integrity. In addition, DevOps is about collaboration between developers and operators, which might ensure that all ethical issues are likely to be discussed and considered by the various stakeholders. The first challenge of DevOps, when applied in developing HCAI systems, is that there is always the risk of overdoing automation. Although the system becomes automated, it also cuts the possibilities of human intervention in the process, which is highly important for determining whether all ethical aspects are covered. Further, since speed is one of the driving forces in DevOps, there might be a lack of reflection on the general ethical component of the system, not taking into consideration that IT solutions, being created for actualizing various objectives, also contain some ethical (sometimes even potential negative) features, if the number one priority of the work is technical results and delivery velocity.

4.1.5 Ethics design solutions

Integrating ethical principles into HCAI systems is often required to ensure that the resulting systems align with societal values and expectations. In this direction, *ethical frameworks and guidelines* are helpful to integrate ethical principles into HCAI systems; this means that, when developing any HCAI system, there are necessarily sound ethical foundations in place. Following these guidelines means that many of the ethical issues can be flagged at an early stage in the design when it could be easier and cheaper to fix rather

than at a later stage of development. To ensure that HCAI systems are widely accepted, developers must follow existing ethical standards, therefore gaining the trust of users and other stakeholders. For example, in the case of HCAI systems for medical applications, this means that ethical rules can guide the process of diagnosing diseases and, at the same time, respect the privacy and/or the autonomy of both patients and physicians. One potential limitation is that while quite useful, ethical guidelines can sometimes be very broad and may not be easily translated into successful heuristics. Strict adherence to those principles might hamper design or generate conflicts with other more technical specifications. However, ethical standards may not reflect cultural diversity or societal changes, leading to some dilemmas when applied globally.

Another important aspect for several HCAI systems regards fairness; specifically, HCAI systems must avoid the amplification of societal bias. In this context, *fairness and bias mitigation techniques* can enhance the reliability of AI systems by reducing the effects of bias, especially in fields with serious consequences, e.g., credit ratings or criminal charges [9, 96, 97]. Through these strategies, the techniques of avoiding bias assist the system in making a fair and just decision to serve the community to build public trust and avoid liability. The decision fairness and bias mitigation techniques come with the problem that they can deprive of further optimization of other measures, for instance, accuracy. For example, techniques that balance the measures across various groups of individuals will decrease the general levels of prediction accuracy, which is deleterious in such areas as the diagnosis of diseases. Moreover, most methods aimed at reducing bias use a set of predefined criteria of fairness, which sometimes do not reflect the concept of fairness in practice. There is also an added capability to create new biases if the techniques are instituted inappropriately.

4.2 Evaluating HCAI systems

System evaluation is a crucial activity in the software lifecycle since it permits to identify problems that may occur during the usage of the system. The survey indicated several methods proposed by the different disciplines and adopted to evaluate HCAI systems (Fig. 2); they are discussed in this section.

4.2.1 HCI evaluation solutions

HCI offers several methods to evaluate systems, focusing primarily on qualities that are of interest for the users, i.e., usability and UX. It would take too long to describe the different methods. Moreover, various methods are slight variations of a more popular method.

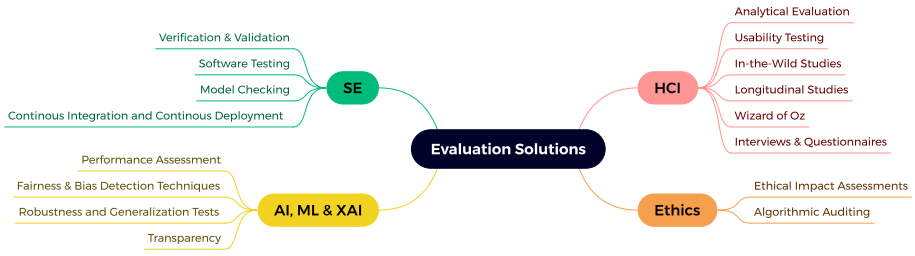

Fig. 2 Overview of the evaluation solutions proposed in the disciplines converging in HCAI

Sometimes practitioners complain that evaluation is difficult to perform since it requires too many resources. Indeed, involving users in the evaluation may be difficult and costly. To overcome these drawbacks, HCI researchers have proposed methods that do not involve users and require much less resources, still being capable of providing valuable indications about system usability. Rogers et al. cluster the evaluation methods in three main approaches: *analytical evaluation* (users are not involved), *usability testing*, and *in-the-wild studies* [114].

The analytical approach includes *inspections* and the application of *formal models* to predict users' performance. Formal methods are not suitable for providing indications about UX thus they are not very used anymore. Inspections are typically conducted by usability experts who systematically inspect the interface for compliance with usability principles, checklists, or standards. The main advantage is related to cost-saving: they ''save" users and do not require special equipment or lab facilities [6]. In addition, experts can detect a wide range of problems of complex systems in a limited amount of time and the inspection can be performed at any stage of the system lifecycle. A common inspection method is *heuristic evaluation* [103]: experts inspect the system interface and evaluate it against a list of usability principles, i.e., the heuristics. It is most effective when implemented before the advanced stages of systems development since it allows the designers to correct any usability issue before investing a lot of time and resources in conducting user testing. Heuristic evaluation may not effectively identify the problems of new interfaces, or any system based on a fairly new interaction modality that may defy conventional usability principles. Thus, in the context of HCAI systems, it is very important to refer to ad-hoc heuristics, as reported in other contexts [83, 99].

Usability testing and in-the-wild studies are empirical evaluation methods; they are the most reliable since by observing actual users interacting with a system, they can reveal how the system performs in practice, uncovering usability issues and user preferences that other methods might miss. *Usability testing* concerns the analysis of users' performance and behavior on the tasks for which the system is designed [114]. Real users are asked to perform some tasks interacting with the system, to identify usability problems, gain user feedback, and evaluate the overall users' satisfaction. This method is essential in the evaluation of HCAI systems, as it provides an important opportunity to gain first-hand information about the UX. It is well suited for the prototyping and final testing phases, especially for consumer applications where the consumer experience plays a direct and critical role in the success of the system. However, reproducing realistic situations of usage in a laboratory is difficult, e.g., selecting a representative sample of users and tasks, training users to master advanced features of the system in a limited time, or weighting the effect of important contextual factors on their performance [129]. The cost and time needed to set up usability testing may also be considerable. A frequently used technique is the *user test with thinking aloud protocol*, where the user is invited to express his/her perceptions, emotions and actions while using the prototypes or the final systems. Evaluators detect problems by observing the behavior of the users and listening to their thoughts, so that they can follow users' reasoning and understand the difficulty they experience in the usage of the product, thus resulting more effective than a simple observation. For HCAI systems where users might be involved in decision-making processes and directly interacting with the system, this protocol can be of great help in understanding how the user makes sense and responds to the new AI-generated outputs. This test is more informal, and the user sample is smaller, so it is easier and cheaper to perform, but it is still capable of providing valuable indications for improvements.

In-the-wild studies differ from the other evaluation approaches because they are conducted in natural settings, in which the system will be used, to gain better information about how the users will utilize the real environment. They are very useful for assessing HCAI systems employed in specific or dynamic environments, including newly developed mobile applications for usage in other environments aside from the comfort of our homes or AI tools in industrial settings. In contrast, field studies may be less necessary for systems that will be used in controlled or static environments where contextual variability is minimal.

Evaluation can also span over time, monitoring and assessing a system's user experience to understand any changes in the user behavior/attitude and satisfaction levels. This is the case of *longitudinal studies* [85]. This method is useful, especially when evaluating HCAI systems projected to be a long-term service or a system, e.g., enterprise or healthcare-related systems. It is also useful in finding out the sustainability of user engagement, the evolution of user needs, and the long-term impact of the system on user performance. Nevertheless, this method may be less suitable for short-term applications or those that are targeted at interactively used software where top users' engagement is only for a short period.

The development of the AI component of an HCAI system is very resource-demanding. Thus, it is useful to test and refine a prototype of the user interface before the AI component is fully developed. The *Wizard of Oz* (WoZ) technique can be adopted in this case: a simulation of the final system is quickly built to be managed by the wizard (an evaluator), and end-users have the feeling they are interacting with the actual system [4]. WoZ enables designers to experiment with how the AI can be used, what the users expect, and how they respond to the system.

Data gathering techniques, such as *interviews* and *questionnaires*, can be also used to collect information from the users on system usability and UX. Questionnaires are more used since they permit to gather both quantitative and qualitative data. These instruments can try to measure, for example, satisfaction, ease of use perceived value, and emotions evoked by the system. An example of a questionnaire used to evaluate UX is AttrakDiff [69]. For HCAI systems, questionnaires can help to understand how the users perceive AI behavior, the system's general utility, and vice versa. This method is particularly helpful for collecting data from a massive number of users or for understanding shifts in attitude during some time. However, questionnaire effectiveness depends on the design and the willingness of users to provide honest and thoughtful responses.

Evaluation methods can be conducted in person or remotely [33]. When it is conducted remotely, the evaluator can follow the study synchronously or asynchronously and is mediated by ad-hoc tools [32, 43, 52, 102]. The remote usability testing is most relevant for HCAI systems designed for large-scale or cross-cultural use where users might be placed worldwide. Finally, it is worth mentioning that there are tools that partially automate usability testing [18, 47, 48, 102]. For example, Fabo and Durikovic present a tool to evaluate usability using eye tracking data [50]. However, automated tools are not so reliable in evaluating qualitative characteristics, such as emotional responses or satisfaction, which typically require direct user feedback or other measurements.

4.2.2 AI, ML and XAI evaluation solutions

An HCAI system must be evaluated aby considering also effectiveness, fairness, transparency, robustness, and compliance with ethics. The most common evaluation of AI models is based on *assessing the performance of AI and ML models*, which is typically carried out by computing and analyzing performance metrics. Among the most common, there are

accuracy, precision, recall, F1 score, and area under the receiver-operator curve (AUC); they provide a measure as to how well a model predicts or classifies data [116]. Especially in predictive applications, such as in health diagnosis or stock predicting, the accuracy measures are significant. However, such measures do not consider ethical factors and the corresponding issues of fairness, transparency, and stability intrinsic to HCAI systems. For example, a model with good accuracy may contain prejudice in the results regarding demography. Thus, these metrics do not help assess HCAI systems regarding ethical implications.

Fairness and bias detection techniques must be considered to detect bias in AI and ML models. Solutions like disparate impact analysis, fairness constraints, and adversarial debiasing are applied to test how fair the predictions of a specific model are to members of the different demographic groups [149]. Such techniques are useful for HCAI systems where a concern of fairness and non-discrimination is important, e.g., in case of hiring algorithms, loan approval systems, and even the criminal justice systems. While these metrics are essential, they often involve trade-offs with other performance metrics. Thus, it is suggested that enforcing strict fairness constraints could potentially create a lower mean accuracy across targeting models. Also, the definitions of fairness may change with respect to a certain context, making it difficult to determine what is fair without considering various factors. Therefore, all these techniques must be applied with respect to the HCAI system's ethical goals. Additional assistance to evaluate AI systems from a human-centered perspective is the adoption of human-centered guidelines. Examples of such guidelines are the ones presented by Amershi et al. [3] and by the Google PAIR team [63].

Another important aspect, already discussed in the design of AI and ML solutions, regards the ability of a model to perform across different scenarios (e.g., when the data are outside of the training distribution, or when adversarial data is used). Thus, *robustness and generalization tests* must be performed [56]. Adversarial accuracy, perturbation resilience, and generalization error are used to measure the model's robustness level. These tests are important in conditions where HCAI systems have to function under uncertainty, which is the case with self-driving vehicles or diagnostic equipment. Making systems resilient is useful in discovering areas that, if exploited by malicious actors, lead to calamities or unethical circumstances in real-world environments. However, the effort and resources needed to conduct and improve robustness testing can be significant and therefore less practical for constraint projects in terms of budget or time. Therefore, robustness testing should be exercised in conjunction with one or the other evaluation techniques depending on the requirement of the specific HCAI application.

Regarding the interpretability of AI or ML models, *transparency* is one of the two main pillars of eXplainable AI (XAI). As stated in Sect. 4.1.3, this might be achieved by implementing tools such as LIME or SHAP that provide details on how the decisions are being made by the AI models or by implementing models explainable by design, such as decision trees. Explainability is extremely helpful in HCAI systems as users have to rely on the AI's judgments and comprehend them, for instance, in healthcare, the legal sphere, or customer support services. Transparent models allow for increased confidence of the users and for easier detection of any such bias or mistakes in judgment. Also, for XAI, metrics are essential to measure the quality of the model's interpretability or explanation. A comprehensive list of metrics for interpretability in ML has been reported in [22], while solutions for deep learning models have been reported in [25]. Examples of metrics for interpretability are *fidelity*, which evaluates how well the explanation reflects the model's actual behavior; *sparsity*, which measures the number of features used in an explanation; and *consistency*, which assesses whether similar inputs yield similar explanations across different

instances or models. Metrics to quantify the degree of which AI model predictions can be easily explainable by its features has been recently reported in [100]. These metrics summarize different aspects of explainability into scalars, providing a more comprehensive understanding of model predictions and facilitating communication between decision-makers and stakeholders, thereby increasing the overall transparency and accountability of AI systems. However, although transparency is often a desirable property, there could be a trade-off between model complexity and model performance. Indeed, achieving high levels of transparency, for example by implementing decision trees, may reduce classification performance. On the contrary, black-box models such as neural networks may be more accurate than highly interpretable models; moreover, over-simplification of complex black-box models in a bid to increase their explainability may lead to a reduction of important information. Also, reaching interpretability may lead to neglecting other important evaluation dimensions, such as robustness or fairness. Thus, depending on the needs of the HCAI system and user community, certain XAI techniques are more appropriate than others.

4.2.3 Software engineering evaluation solutions

Software engineering offers different evaluation techniques for assessing HCAI systems, ranging from technical performance to ethical and user-centered attributes. This section presents the existing solutions in software engineering for evaluating HCAI systems, providing a guide on the pros and cons of these solutions.

Verification and Validation (V&V) are first-order processes in software engineering that aim at ascertaining whether a system being developed will indeed deliver the expected results [128]. Verification concerns itself with the accuracy of the implementation of the system (for instance, to ensure that the system has been developed correctly), while validation considers whether the developed system meets the intended usage (for instance, to ensure that the right system has been developed). V&V for HCAI can also incorporate the assessment of algorithms, inclusion of linkages, systems and conformity to user desires. It's crucial to note that both V&V processes are critical assets of HCAI systems, especially in safety-critical domains such as healthcare, self-driving cars, and electronics' finance. However, V&V is important to establish the technical rigor, it may not properly address the ethical issues such as fairness, transparency, or user's self-determination. These processes are also CPU and time-consuming and depend on the professional level of the specialists who are to understand the software system and its environment. Thus, as the HCAI systems are complex, it is suggested to expand V&V by ethical assessments and user-centered evaluations.

Another evaluation technique is *software testing*, which involves the process of evaluation of software based on a comparison of the actual result and expected result during the processes of the test [128]. Some of the common testing approaches embraced by organizations include unit testing, integration testing, system testing, and acceptance testing. Various testing techniques are involved in testing HCAI, including system testing where AI models are run through different situations to test their reliability and scenario-based testing in which AI models are tested on sample scenarios. However, software testing should be used in conjunction with other forms of evaluations like ethical impact assessments and user centered evaluations to gain a comprehensive understanding of the system's performance.

Another automatic evaluation useful for evaluating HCAI systems is represented by *model checking*. It is a method of proving the correctness of finite-state systems concerning

certain specifications, which are normally expressed in temporal logic [79]. This checks whether the features in question are true for the system in all the transitions that can occur systematically go through all the possible states of the system. In the case of HCAI systems, model checking can be applied to ensure that the AI systems implement ethical standards, such as non-discriminate or non-user profiling. Model checking is especially beneficial in HCAI systems where it is important to validate that the system's behavior is appropriate as per the prevailing or explicitly specified ethics and laws, including finance or legal systems. However, in the case of large or complex systems, such as deep learning models, the computational cost of model checking could be high and practically unmanageable. Further, model checking involves the definition of the property that one wants to verify, and this can be hard to do in complex ethical terms. Hence, using model checking is useful in figuring out specific behavioral patterns; however, this form of functional verification should be used in combination with other verification techniques for broader ethical issues.

The last type of evaluation is related to the DevOps practices mentioned in Sect. 4.1.5. In this context, *continuous integration and continuous deployment* (CI/CD) are the evaluation techniques to be considered [38, 58]. They consist of the frequent integration of code changes, and this integration and deployment are automated, making sure that the software that is deployed all the time is ready to be deployed. In the case of HCAI systems, the CI/CD pipeline may also have integrated self-testing as well as self-scanning tools to check whether the system is functioning, ethical and compliant with the set regulations at all times. CI/CD solutions are useful for HCAI systems, whereby the frequent rollout and updates act as a strength since such systems are usually in environments that demand constant tweaking and deployment of new services or applications. Even though CI/CD practices make the process of deploying the software more efficient and effective, the latter does not always address the ethical issues of HCAI systems. The emphasis on speed may cause such that not enough time is given to conduct ethical analysis and there is a possibility of deploying systems that have not been through ethical scrutiny. It is for this reason that CI/CD pipelines should be implemented with consideration of ethical evaluation steps to ensure that, while delivering high velocity, developers do not compromise on ethical standards.

4.2.4 Ethics evaluation solutions

A solution to conduct ethical evaluation of HCAI systems is *Ethical Impact Assessments* (EIA), which is a framework to assess ethical implications of systems before, during, and after deployment [93]. EIAs mainly consist of a systematic examination of how an HCAI system might affect its stakeholders. Some aspects to be considered in this analysis are privacy, fairness, people's self-determination, and prevention of the occurrence of harm. They are especially useful when the HCAI system that is being appraised has a far-reaching impact on society (e.g., health care and law enforcement). Through EIAs, the developers and stakeholders can review the ethical risks that are likely to occur and avoid or address them appropriately, hence making the system work ethically. Although EIAs are comprehensive, they can be labor and cost-intensive and depend on expertise in ethics, law, and technology. In addition, the results of EIA depend on the best ethical standards and values that the evaluators incorporate, and it should note that the acceptable ethical standards may vary. For this reason, EIA should be applied when there is a need to solve possible ethical issues, and results should be considered in the context of societal values.

Another approach is represented by *algorithmic auditing*, which entails the process by which an organization investigates all the algorithms used in an AI system for possible

ethical issues, including bias, discrimination, and lack of transparency [108]. Part of this process entails comparing the inputs, decision mechanism and outputs of the system to certain ethical benchmarks. Algorithmic audits may be performed by a third party to make the findings more impartial, suggesting the need to set up governance structures [123]. Algorithmic auditing is useful to HCAI systems, which necessitate the disclosure of the decision-making process and fairness, e.g., in credit scoring systems, job recruitment, or criminal courts. Algorithmic auditing can only be effective if detailed data is available and the underlying algorithms are well understood, highlighting the importance of transparency. On certain occasions, due to patents or the black-box nature of AI systems, the auditing might remain restricted to a certain level, limiting the ethical review. Furthermore, auditing does not provide recommendations on how to solve recognized issues.

5 Emerging challenges in HCAI

HCAI seeks to develop AI systems designed, developed, and deployed by taking into account human needs, values, and contexts. While the promise of HCAI is significant, the discipline revolving around it faces several critical challenges, which span technical, ethical, social, and interdisciplinary domains. Thus, realizing HCAI systems becomes a complex endeavor. Our survey identified 11 important challenges that might drive the research in HCAI in the next year; they are discussed in this section.

5.1 Terminological inconsistencies

This survey has shown that many terminological inconsistencies pervade the development of HCAI. Considering the wide range of disciplines involved, each with its concepts and vocabulary that have already been established. For instance, the terms "explainable AI," "interpretable AI," and "transparent AI" are often employed as if they were perfect *synonyms* when, in fact, they may represent some slight nuance of difference depending upon context or field of study.

Polysemous terms complicate interdisciplinary work in HCAI: a term like "model" can carry different meanings depending on the disciplinary perspective with which it is interpreted. In ML, a "model" refers to a learning algorithm, created to perform some tasks like classification or prediction [62]. In HCI, a "model" may also refer to a conceptual or theoretical framework that seeks to explain cognitive processes like perception, memory, or decision-making [130]. These different interpretations often need clarification in projects that require collaboration, as the same word may be used to mean different things by team members trained in different disciplines. This challenge can only be tackled by developing a shared vocabulary and a conceptual framework across these disciplines to ensure more effective communication and collaboration.

5.2 Interdisciplinary integration

Another important aspect of this survey is the need for interdisciplinary integration for the success of HCAI because of the methodologies, epistemologies, and research priorities that characterize different disciplines. In practice, it is the exception that expertise in one of

these disciplines will be strongly informed by deep knowledge of the foundational theories and methods in other disciplines. A technically sophisticated model developed by an AI researcher may make powerful predictions yet remain without insights into human cognitive limitations, thus leading to systems that fail to align with human needs [88]. On the other hand, a detailed account of AI's technical capabilities and limitations may escape the notice of an ethicist, thus undermining realistic expectations and posing technical problems. Thus, efforts should be made to develop cross-disciplinary education and research initiatives between the disciplines by establishing multidisciplinary teams to share and synthesize knowledge.

Moreover, each discipline within HCAI contributes different methodologies and solutions. For example, AI research often focuses on the quantitative metrics associated with measures like accuracy or efficiency through large-scale data analysis and empirical testing [110]. Contrastingly, disciplines like HCI or ethics might focus on qualitative insights from user studies, interviews, or observational research [85]. Method differences can sometimes lead to conflicts or misunderstandings in integrating findings across disciplines, hence developing a coherent HCAI approach. In light of the aforementioned challenges, it becomes evident that an alternative interdisciplinary research approach is necessary. This approach should demonstrate due respect for the strengths inherent to each tradition, while simultaneously seeking common ground where methodologies of a mixed kind might prove effective in HCAI research.

Therefore, this stresses the collaborative frameworks that could address the issues generated from the knowledge gaps and differences in methodologies but that invoke effective work in interdisciplinarity. These kinds of frameworks could integrate other points of view in the design and evaluation process when developing an HCAI system. It is one instance of how useful it would be to involve stakeholders from more than one discipline as active participants of a series of participatory design methodologies. Such frameworks would entail conflict resolution mechanisms to balance ethical considerations with technical feasibility so that technical goals do not overshadow a human-centered perspective. This means that developing such frameworks will have to be critical in ensuring the HCAI systems are technically robust, ethically sound, and human centered.

5.3 Ethical considerations and bias mitigation

Due to the potentially important impact that AI systems may have on individuals and society, challenges in HCAI raise ethical considerations and bias mitigation at the forefront.

As already stated in the previous section, bias in AI systems can arise from multiple sources, including biased training data, biased algorithmic design, or biases introduced during human-AI interactions. They are not very easy to detect or mitigate. This includes both technical solutions—e.g., algorithmic fairness techniques that change models to have less bias [9, 67, 96]—and ethical oversight and ongoing monitoring. Mitigation for bias also should consider the context in which the AI systems will be deployed, as biases can manifest very differently with regard to domain, population, or application. For example, an AI system used in hiring might inadvertently favor certain demographics if the training data reflects historical inequalities in employment [16]. All of these require continuous research on developing fairness-aware algorithms, diversifying and making data collection more representative, and establishing ethical guidelines and regulations that make bias audits and accountability mechanisms a requirement.

Proper transparency and explainability are crucial to making AI systems work in practice and be trusted by their users. However, achieving transparency and explainability in models with acceptable performance, especially in very complex models such as deep neural networks, remains a challenge. In the end, contemporary approaches to explainability range from model-agnostic techniques—such as LIME or SHAP that give local approximations of model behavior—to inherently interpretable models designed to be simple and transparent. However, such explanations must make sense for the concrete type of users: AI experts, domain professionals, or lay users, and should facilitate appropriate decision-making and oversight. The challenge is to balance the need for detailed and accurate explanations without increasing the cognitive load for users and to ensure explanations are informative and useful.

5.4 Privacy and data protection

HCAI systems increasingly rely on personal data on a very wide scale to offer fully personalized experiences, which is an important issue. AI systems should be developed with the observance of privacy regulations, like the GDPR of the EU [135] and other emerging regulatory frameworks in different parts of the world. Therefore, for the safety of user data, it seems to be the right one to develop AI using such a privacy-preserving technique as differential privacy [40], and federated learning [95], in which to continue research. All these are challenges in implementing these at a large scale without reducing the utility of AI systems. Furthermore, privacy issues are above and beyond technical solutions; they are linked with the issues of consent, data ownership, and the right to be forgotten.

5.5 Evaluation metrics and benchmarks

Defining appropriate assessment metrics for HCAI systems is a critically important but challenging issue, primarily because a proper performance evaluation should be done along two dimensions: technical performance and human-centered outcomes.

On one side, traditional AI assessment metrics, such as accuracy, precision, and recall, put a singular focus on algorithmic performance. However, those metrics often ignore a broader context of human-AI interaction, which includes user satisfaction, trust, and the effectiveness of AI in supporting human decision-making. Holistic performance measures must combine these human-centered dimensions—possibly combining qualitative feedback from users, metrics related to cognitive load, and measures of user engagement and understanding [39]. The issue is, therefore, to develop sound and integrated evaluation frameworks that can reconcile the requirement for technical rigor with a sensitivity to the emergent fine details of human experience, enabling such systems to be both effective and user-friendly.

The effectiveness of HCAI systems can vary significantly depending on the context in which they are used, including the specific task, the characteristics of the user population, and the cultural or organizational setting. Developing evaluation frameworks that account for such variations is critical so that HCAI systems are robust and adaptable in different contexts. This calls for developing context-sensitive evaluation methods to capture the unique problems and opportunities the environment presents. For instance, an HCAI system developed for clinical applications may require an evaluation not only on diagnostic performance but also on the assessment of usability by clinicians, its integration into existing workflows, and its acceptance by patients [27]. Contextual evaluation is a multidimensional challenge because it must consider both technical and human aspects.

5.6 Long-term impact

The long-term impact on human capabilities, the decision-making processes, and society at large of the adoption of HCAI systems is rather challenging and less explored. Understanding how sustained interaction with AI/ML systems shapes users longitudinally with respect to potential changes in trust, reliance, and decision-making behavior is necessary [125]. A second point involves the general societal impacts of the wide-scale adoption of AI—employment effects, social equity, and public trust in technology—which also need to be sensitively monitored and assessed. Long-term assessment is, therefore, a challenge where interdisciplinary work is to conceive and develop new testing tools and some kind of 'ethics protocols' for the responsible deployment of HCAI systems. In this regard, longitudinal studies can be powerful instruments to assess how HCAI systems affect humans and society at large. Sadly, throughout this literature review we were not able to retrieve tools, protocols and studies with this goal. Indeed, it is important to note that longitudinal studies require time, and the novelty of the HCAI field may limit the availability of published studies.

5.7 Generalizability across domains

Ensuring that HCAI principles and solutions can be scaled and generalized across different domains and user populations is crucial for HCAI's widespread adoption and impact.

For example, adapting HCAI systems to new domains while maintaining their human-centered properties is a significant challenge. Domain adaptation involves transferring knowledge and skills from one domain to another, often with limited data or domain-specific expertise available in the new context [104]. For example, an AI system trained on medical data from one region may need to be adapted to work effectively in another region with different patient demographics, disease prevalence, or healthcare practices. Ensuring that the system remains effective, fair, and interpretable across domains requires robust transfer learning techniques and domain-specific adaptations that account for the unique challenges of each new environment.

Personalization is a key aspect of HCAI, as AI systems need to be tailored to individual user needs and preferences. However, achieving personalization at scale presents a significant challenge, as it requires balancing the customization of AI experiences with the scalability demands of large-scale systems. Techniques such as collaborative filtering, content-based filtering, and reinforcement learning can be used to personalize AI systems, but these approaches must be carefully managed to avoid issues such as filter bubbles, overfitting, and unintended biases [96, 113]. Moreover, personalization efforts must be transparent and give users control over how their data is used and how their AI experiences are shaped.

5.8 Human-AI trust dynamics

Building and maintaining trust between humans and AI systems is a critical factor in the success of HCAI, as trust influences how users interact with, rely on, and benefit from AI technologies.

Trust calibration involves ensuring that users have an appropriate level of trust in AI systems, neither over-relying on them nor under-utilizing them. Over-reliance on AI can occur when users place too much trust in the system, potentially leading to complacency,

errors, or failures in oversight [41]. On the other hand, under-utilization can result from a lack of trust, where users ignore or reject valuable AI insights, diminishing the potential benefits of the technology [125]. Achieving proper trust calibration requires designing AI systems that are accurate and reliable but also transparent and explainable, helping users understand when and how to trust the system. This involves providing users with clear feedback about the AI's confidence levels, limitations, and the rationale behind its decisions and training users to interact effectively with AI systems.

Moreover, when AI systems make mistakes or fail, it can lead to a loss of user trust, which can be difficult to regain. Trust recovery is a critical challenge in HCAI, requiring the development of mechanisms that allow AI systems to recover from errors and rebuild user confidence. This might involve transparent communication about the nature of the error, steps taken to correct it, and assurances that similar mistakes will be avoided in the future [12]. Trust recovery also involves providing users with tools to intervene, correct, or override the AI's decisions, empowering them to maintain control and confidence in the system. Developing effective trust recovery strategies is essential for ensuring long-term user engagement and the successful integration of AI into human workflows.

Finally, anthropomorphism, or the attribution of human-like characteristics to AI systems, can significantly influence user trust and expectations. While anthropomorphic design can make AI systems more relatable and easier to interact with, it can also lead to unrealistic expectations about the system's capabilities and reliability [44]. Managing the effects of anthropomorphism is a delicate balancing act, requiring designers to carefully consider how human-like features, such as voice, appearance, or behavior, are implemented in AI systems. The goal should be to enhance usability and trust without misleading users or creating dependencies that could undermine human agency or responsibility.

5.9 Cognitive load and human factors

Optimizing the cognitive aspects of human-AI interaction is crucial for ensuring that AI systems are both effective and user-friendly, minimizing the cognitive burden on users.

AI systems often generate large amounts of data and insights, which can overwhelm users and lead to information overload. Balancing the provision of AI-generated insights with human cognitive limitations is a significant challenge, as too much information can reduce decision-making quality and increase user frustration [45, 76]. Effective HCAI design must prioritize the clarity, relevance, and timeliness of the information presented to users, using techniques such as information filtering, summarization, and adaptive interfaces that adjust the level of detail based on the user's context and preferences. The goal is to provide users with the right amount of information at the right time, supporting informed decision-making without overwhelming them.

Capturing and maintaining user attention is another essential aspect for ensuring that AI systems are used effectively. However, it must be done in a way that avoids distraction or cognitive fatigue. AI systems should be designed to deliver notifications, alerts, and other forms of feedback that align with the user's current task and attention level, avoiding interruptions that could disrupt workflow or concentration [8]. Attention management strategies might include prioritizing notifications based on urgency, using non-intrusive cues, and allowing users to customize their interaction settings. The challenge is to balance keeping users informed and allowing them to focus on their primary tasks without unnecessary distractions.

Therefore, ensuring that users develop accurate mental models of AI capabilities and limitations is crucial for effective human-AI interaction. Misaligned mental models can lead to inappropriate use of the system, either through over-reliance or under-utilization. HCAI systems must be designed to align mental models with clear, consistent, and intuitive interfaces that help users understand the system's functions, limitations, and decision-making processes. This might involve visualizations, tutorials, and interactive explanations that guide users in forming accurate mental models. Additionally, ongoing user training and education can help to refine these models over time, ensuring that they remain aligned with the evolving capabilities of the AI system.

5.10 Adaptability and learning

Creating AI systems that adapt to changing user needs and environments is a critical challenge in HCAI, requiring advanced learning mechanisms and continuous user interaction.

AI systems that continuously learn and improve from ongoing user interactions offer the potential for more personalized and effective human-AI collaboration. However, continuous learning also introduces risks, such as the potential for overfitting recent data or introducing biases as the system adapts to new information [34]. Adaptivity and adaptability are solutions for enabling system modification when new needs arise during its use. Adaptivity is automatically determined by the system that tailors itself on the basis of repeated user interactions and habits; adaptability is determined by users, who are provided with tools to manually perform actions to tailor the system [53]. Balancing the need for adaptability with the stability and reliability of the system is a crucial challenge. Techniques such as incremental learning, transfer learning, and online learning can enable continuous adaptation while mitigating risks. Additionally, involving users in the learning process is essential, allowing them to provide feedback, correct errors, and guide the system's learning trajectory.

Moreover, enabling AI systems to transfer knowledge and skills learned in one context to new, related contexts is essential for their scalability and versatility. Transfer learning techniques allow AI systems to leverage existing knowledge to perform well in new domains with limited data [104]. For example, an AI system trained to recognize medical conditions in one population might need to transfer its knowledge to another population with different characteristics. The challenge lies in ensuring that the transferred knowledge remains accurate and relevant, avoiding negative transfer where performance degrades in the new context. Effective transfer learning requires robust models that can generalize well across different contexts and mechanisms to adapt and fine-tune the system to the specific needs of the new environment.

In this context, another important concept is co-evolution, which in HCAI refers to the mutual adaptation and growth of human users and AI systems over time. As AI systems learn from human interactions, they should improve their performance and enhance the user's capabilities, such as decision-making skills or domain knowledge [55, 64]. Similarly, users should become more proficient in using AI tools, developing a deeper understanding of leveraging AI effectively in their tasks. Designing for co-evolution requires creating AI systems that are adaptable and responsive to user feedback and capable of fostering user learning and development. This might involve interactive learning environments, personalized feedback, and the gradual introduction of more advanced AI features as the user's proficiency increases.

5.11 Regulatory and legal frameworks

Navigating the complex legal and regulatory landscape surrounding HCAI ensures that AI systems are developed and deployed responsibly, with appropriate safeguards and accountability.

Determining accountability in cases where AI systems contribute to decision-making errors or harmful outcomes is a significant legal and ethical challenge. As AI systems become more autonomous and integrated into critical decision-making processes, questions of liability become increasingly complex [61]. For example, suppose an AI-driven medical diagnosis system provides incorrect advice that leads to patient harm. In that case, whether the responsibility lies with the AI developer, the healthcare provider, or the user who relied on the AI's advice may be unclear. Addressing this challenge requires the development of clear legal frameworks that assign responsibility and liability in a way that is fair, transparent, and consistent with ethical principles.

Moreover, AI regulations are rapidly evolving, with different jurisdictions implementing various rules and guidelines to govern the development and use of AI technologies. Ensuring that HCAI systems comply with these regulations is a significant challenge, particularly for companies that must adhere to diverse regulatory environments [136]. Compliance requires ongoing monitoring of regulatory developments, adapting AI systems to meet new legal requirements, and ensuring that AI practices align with local laws and international standards. This challenge is further complicated by the need to balance regulatory compliance with innovation, as overly restrictive regulations could stifle the development of new HCAI technologies.

Finally, intellectual property issues related to AI are complex and evolving, particularly regarding the ownership and protection of AI-generated content and innovations. Questions arise about who owns the rights to AI-generated works, such as art, music, or software, and how traditional intellectual property laws should be applied to AI-driven creativity [127]. Additionally, concerns about protecting AI algorithms and data and the potential for patenting AI techniques could hinder innovation and competition. Addressing these intellectual property challenges requires updating existing legal frameworks to account for the unique aspects of AI, promoting fair and open access to AI technologies, and fostering collaboration between legal experts, technologists, and policymakers.

6 Need for symbiotic AI systems

Previous sections have detailed critical aspects of HCAI, emphasizing the importance of designing AI systems that complement and enhance human decision-making processes. An explicit need for this arises in complex, high-stakes environments like medical applications, where the implications of decisions made by AI may have consequences for patients' life. Indeed, traditional AI has been well integrated into the fields of medical imaging and diagnosis and, to some extent, in treatment planning. However, these capabilities significantly fall short of supporting integration with clinical workflows because of interpretability issues, lack of trust, and ethical concerns.

These challenges are highlighted in the vision of "symbiotic AI," which has received much attention over the past few years as a vehicle to bridge the gap between AI and humans [64, 101]. Symbiotic AI (SAI) requires a progressive and deeper relationship between humans and AI, specifically, the symbiosis of their two forms of intelligence,

where both humans and AI augment each other's knowledge and capabilities thanks to a collaboration that balances each other's strengths and weaknesses [17]. Therefore, SAI systems improve themselves through users while also providing a way for users to improve themselves. However, to create ethical and sustainable SAI systems, this symbiosis requires users to be in control, to reach the goal of creating systems that ensure reliability, safety, and trustworthiness, supporting humans rather than replacing them. There is a growing body of research indicating that the best AI systems are augmentations of human expertise, not replacements [17, 124]. For instance, medical imaging studies have shown that diagnostic accuracy improves significantly when AI systems assist human radiologists, instead fan working independently by themselves [145]. This symbiotic relation allows the strength of AI—its capability to deal with large volume data and recognize complex patterns—to be coupled with human intellect, thereby enabling proper understanding of context, managing uncertainty, and making ethical judgments. Based on the above, we provide the following definition of SAI systems.

Definition 2. Symbiotic AI Systems

Symbiotic AI (SAI) systems are HCAI systems powered by a continuing and deeper collaboration between humans and AI, i.e., a symbiosis of human intelligence and artificial intelligence. Humans and AI mutually augment their capabilities, balancing each other's strengths and weaknesses without hampering neither the autonomy of humans nor the performances of AI, leaving humans in control of the system's decisions. In an SAI system, AI benefits from a continuous stream of new user-provided data to refine itself, while humans benefit from AI's improved performances and knowledge.

AI, HCAI, and SAI systems are therefore specializations of each other (Fig. 3). For example, let's consider a simple AI system for cell classification. Physicians may require a better understanding of its decision-making process, asking for explanations, different ways of providing outputs, etc. Thus, the AI system can be specialized into an HCAI system by using HCI methods, which involve users in the system design and evaluation. Finally, the HCAI system can be further specialized into an SAI system by adopting ways of improving the system and the human knowledge incrementally and interactively through Human-AI collaboration. Better examples of incremental improvements are provided in Sect. 6.

In light of these considerations, the transition to SAI systems represents not just a technical shift but also a paradigm shift in how we conceive the role of AI. Such systems are designed to function as partners in decision-making processes, enabling a more holistic approach that best integrates human and machine intelligence.

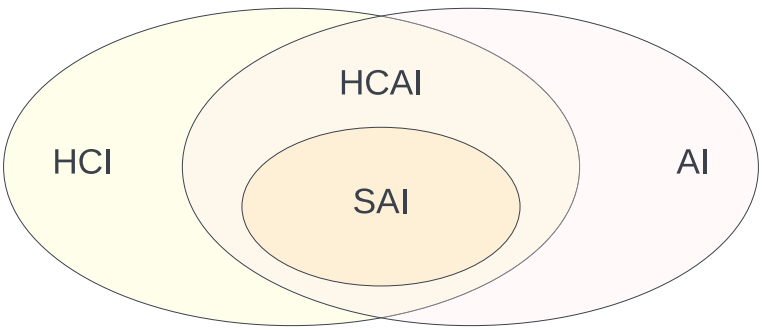

Fig. 3 SAI is a specialization of HCAI, which results from the adoption of HCI methods in AI

7 Two case studies on medical applications

This section elaborates on SAI by providing two case studies in the medical field. These are specific, real-life examples of how SAI might be implemented and illustrate how human-AI symbiosis can be effectively achieved for sound, reliable, and ethically correct AI systems in medicine, leading to better patient care and positive outcomes. More precisely, Sect. 7.1 reports the example of an AI system that has been successfully redesigned as a SAI system and which implementation is currently almost completed. On the other hand, Sect. 7.2 reports the example of a more embryonic project, which provides a new testing ground in a higher-risk scenario.

7.1 Low-risk scenarios: Rhinocytology

Rhinocytology is a specialization of medical cytology that focuses on studying the cells of nasal mucosa [60]. Unlike other similar fields (for example, hematology), nasal cytology does not yet benefit from a network of laboratories that are able to carry out the analysis. Therefore, the diagnostic process is slow and costly, as it requires direct observation under the microscope, requiring prolonged effort by rhinocytologists [60]. In this context, Rhino-Cyt, an AI system, has been developed aiming to assist physicians' activities through the automatic count of cells [36]. The system employs a CNN to analyze the digital image of a nasal cytological preparation to identify (i.e., segment) and classify cells, recognizing nine different cytotypes: (i) ciliated, (ii) muciparous, (iii) basal cells, (iv) striated cells, (v) neutrophils, (vi) eosinophils, (vii) mast cells, (viii) lymphocytes, and (ix) metaplastic cells [35]. Finally, Rhino-Cyt produces the cell count for each cytotypes, supporting the cytological examination and, by merging the result with other medical procedures and exams, the final diagnosis [35, 60]. It should be noted that the general permissiveness of the rhinocytological exam (reflected in its wide ranges [60]), the generally acceptable consequences of a wrong diagnosis, and the support provided by additional exams that do not use medical imaging techniques lower the overall risk of the use of AI with respect to other fields.

In a recent redesign, Rhino-Cyt's interface has been heavily expanded to support the requirements of a HCAI system [28, 31]. Namely, the system now provides explanations for its classifications, allowing end-users to intervene in both the classification itself and the explanation, thus enabling the customization of the classifier [31]. The result of the redesign follows a HOTL approach, but it moves towards a HITL approach by allowing customizations to reduce the number of interventions in the long run. However, the current state of Rhino-Cyt does not yet fully represent an example of SAI system. In fact, although its design followed a human-centered approach, the system is still used as an "oracle" and does not learn continuously from the users' behaviors. To improve this aspect, additional governance structures may be put in place to evaluate the decisions of physicians (to check that the data is unbiased and untainted, ensuring correctness and fairness) and to add new data points back into the dataset, transforming the basic AI model into an *interactive machine learning* model. The benefits of a symbiotic relationship between the classifier and the physicians are multiple:

i. *Human benefits*

a. Through the support of an AI-enabled system, still using a direct manipulation graphical interface, physicians can more easily (and potentially more objectively) recognize and classify cells, leading to a faster and better diagnosis.
a Through interventions mechanisms, the classification can be double-checked ensuring better performances.
c. Through explanations, physicians can improve their knowledge by recognizing and classifying cells that would leave them uncertain due to uncommon features (providing an example of machine teaching [119]).

ii. *AI system benefits*

a. Through interactive machine learning, the AI model can continuously learn and improve its performance thanks to improved and larger datasets.
b. Through interactive machine learning, the AI model can benefit from always up-to-date data, mitigating the risk of concept drifting [139] in the long run and ensuring a correct classification throughout the whole lifecycle of the AI system.

7.2 High-risk scenarios: tumor detection

One of the successful applications of medical imaging techniques has the human brain as a subject of study. In fact, Magnetic Resonance Imaging (MRI) and Positron-Emission Tomography (PET) scans of the brain have been successfully used to aid in the diagnosis of various diseases. Examples of possible applications are tumor detection [30, 73] and to aid the diagnosis of Alzheimer's disease [11, 23]. Focusing on the former application, the importance of an accurate and early recognition of tumors is crucial, as it improves the odds of surgical removal, potentially saving lives. Therefore, also considering the invasiveness of a wrong surgical procedure in case of wrong recognition, this particular application of AI is an extremely higher risk scenario with respect to the previous one. Still, AI use is beneficial as it may improve physicians' performance and objectivity, as per the previous case study.

Although the employment of medical imaging techniques with AI is not limited to brain tumor detection (e.g., MRIs are also employed for breast cancer [120]), we focus on the simple case study presented in [30]. A CNN has been trained on a publicly available dataset to detect brain tumors accurately [30]. However, the AI model itself does not prove to be helpful for physicians' actual use. For example, a user-centered design approach would allow us to provide answers to various relevant questions, such as: *Are physicians interested only in the location of the tumor or in other variables? Are physicians interested in knowing the confidence of the AI detection? Are physicians interested in knowing the parameters that brought to the recognition of a tumor? Are they interested in only knowing that "something" is there, or should AI disambiguate between tumors and (as an example) cysts?*

Such questions are extremely relevant as they immediately impact the way the AI system is to be engineered, and some answers may require some modifications to the architecture of the ML model that powers the recognition. Although some of the answers are potentially found in the literature (e.g., the variables to which physicians are interested), user studies are still crucial: recent research shows that differences in demographic details of the end-users' sample heavily affect the perceived usefulness of explanations and other

variables [47]. Similarly to the previous case study, exacerbated by the lack of an actual diagnosis-support system and related user studies, such ML models do not represent examples of SAI systems. However, we are confident that, by implementing a collaboration between the AI system and the end-users, thus investing into fostering a symbiotic relationship between the two, several benefits may be obtained:

i. *Human benefit*.

   a Through the support of an AI-enabled system, physicians can more easily (and potentially more objectively) recognize tumors, also in early stages where humans may have troubles, leading to a faster and better diagnosis [140, 145].
   b Through interventions mechanisms, the classification can be double-checked ensuring better performances.

ii. *AI system benefits*.

   a Through interactive machine learning, the AI model can continuously learn and improve its performance thanks to improved and larger datasets.
   b Through interactive machine learning, the AI model can benefit from always up-to-date data, mitigating the risk of *concept drifting* [138] in the long run and ensuring a correct classification throughout the whole lifecycle of the AI system.

8 Conclusions

This article provided a survey of articles related to the emerging field of HCAI, highlighting its interdisciplinary nature and the complexities involved in creating HCAI systems. The diverse range of disciplines, concepts, and existing solutions that play a crucial role in designing and evaluating HCAI systems are discussed. The findings emphasize the importance of the collaboration among experts from various disciplines (including HCI, AI, SE, ethics) as well as experts from specific application domains to tackle the challenges, also discussed in the article, of creating AI systems that are truly human-centered.

Definitions of HCAI systems and of SAI systems are provided. SAI builds on the principles of HCAI, but requires a deeper, more integrated collaboration between humans and AI. The case studies on medical applications presented in this article demonstrate the potential of SAI.

Overall, this work contributes to providing valuable resources for researchers and practitioners, offering insights and recommendations for creating both HCAI and SAI systems.

Acknowledgements We are grateful to the anonymous reviewers for their useful comments, which helped improve the quality of this article.

This research is 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 Italian Ministry of University and Research (MUR) under grant PRIN 2022 "DevProDev: Profiling Software Developers for Developer-Centered Recommender Systems" — CUP: H53D23003620006;

• the Italian Ministry of University and Research (MUR) and by the European Union - NextGenerationEU, under grant PRIN 2022 PNRR "PROTECT: imPROving ciTizEn inClusiveness Through Conversational AI" (Grant P2022JJPBY) — CUP: H53D23008150001;

• the Italian Ministry of University and Research (MUR) and by the European Union - NextGenerationEU, under grant PRIN 2022 PNRR "DAMOCLES: Detection And Mitigation Of Cyber attacks that exploit human vuLnerabilitiES" (Grant P2022FXP5B) — CUP: H53D23008140001.

The research of Andrea Esposito is funded by a Ph.D. fellowship within the framework of the Italian "D.M. n. 352, April 9, 2022" - under the National Recovery and Resilience Plan, Mission 4, Component 2, Investment 3.3 – Ph.D. Project "Human-Centered Artificial Intelligence (HCAI) techniques for supporting end users interacting with AI systems," co-supported by "Eusoft S.r.l." (CUP H91I22000410007).

Author contributions Giuseppe Desolda: Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. Andrea Esposito: Conceptualization, Data curation, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. Rosa Lanzilotti: Conceptualization, Funding acquisition, Methodology, Supervision, Validation, Visualization, Writing – review & editing. Antonio Piccinno: Conceptualization, Methodology, Supervision, Validation, Visualization, Writing – review & editing. Maria F. Costabile: Conceptualization, Methodology, Project administration, Validation, Writing – review & editing.

Data Availability No new data were created or analyzed during this study. Data sharing is not applicable to this article. All relevant literature records used within this survey are cited within the body of the manuscript and can be found in the References section.

Declarations

Competing Interests The authors declare no competing interests.

References

1. Adadi A, Berrada M (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE Access 6:52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052
2. Amershi S, Cakmak M, Knox WB, Kulesza T (2014) Power to the people: the role of humans in interactive machine learning. AI Mag 35:105–120. https://doi.org/10.1609/aimag.v35i4.2513
3. Amershi S, Weld D, Vorvoreanu M, Fourney A, Nushi B, Collisson P, Suh J, Iqbal S, Bennett PN, Inkpen K, Teevan J, Kikin-Gil R, Horvitz E (2019) Guidelines for human-AI interaction. In: Proceedings of the 2019 CHI conference on human factors in computing systems. Association for Computing Machinery, New York, NY, pp 1–13. https://doi.org/10.1145/3290605.3300233
4. Ardito C, Buono P, Costabile MF, Lanzilotti R, Piccinno A (2009) A tool for wizard of Oz studies of multimodal mobile systems. In: 2009 2nd conference on human system interactions. IEEE, Catania, pp 344–347. https://doi.org/10.1109/HSI.2009.5091003
5. Ardito C, Buono P, Costabile MF, Lanzilotti R, Piccinno A (2012) End users as co-designers of their own tools and products. J Vis Lang Comput 23:78–90. https://doi.org/10.1016/j.jvlc.2011.11.005
6. Ardito C, Costabile MF, De Angeli A, Lanzilotti R (2006) Systematic evaluation of e-learning systems: an experimental validation. In: Proceedings of the 4th Nordic conference on human-computer interaction: changing roles. ACM, Oslo, pp 195–202. https://doi.org/10.1145/1182475.1182496
7. Bach TA, Khan A, Hallock H, Beltrão G, Sousa S (2024) A systematic literature review of user trust in AI-enabled systems: an HCI perspective. Int J Human–Computer Interact 40:1251–1266. https://doi.org/10.1080/10447318.2022.2138826

8. Bailey BP, Konstan JA (2006) On the need for attention-aware systems: measuring effects of interruption on task performance, error rate, and affective state. Comput Hum Behav 22:685–708. https://doi.org/10.1016/j.chb.2005.12.009
9. Barocas S, Hardt M, Narayanan A (2023) Fairness and machine learning: limitations and opportunities. The MIT Press, Cambridge, Massachusetts
10. Barricelli BR, Cassano F, Fogli D, Piccinno A (2019) End-user development, end-user programming and end-user software engineering: a systematic mapping study. J Syst Softw 149:101–137. https://doi.org/10.1016/j.jss.2018.11.041
11. Ben Ahmed O, Benois-Pineau J, Ben Amar C, Allard M, Catheline G (2013) Early Alzheimer disease detection with bag-of-visual-words and hybrid fusion on structural MRI. In: Proceedings of the 11th international workshop on content-based multimedia indexing (CBMI). IEEE, Veszprém, pp 79–83. https://doi.org/10.1109/CBMI.2013.6576557
12. Binns R, Van Kleek M, Veale M, Lyngs U, Zhao J, Shadbolt N (2018) "It's reducing a human being to a percentage": perceptions of justice in algorithmic decisions. In: Proceedings of the 2018 CHI conference on human factors in computing systems. ACM, Montreal QC, pp 1–14. https://doi.org/10.1145/3173574.3173951
13. Blackburn S (2003) Ethics: a very short introduction. Oxford University Press, Oxford
14. Boehm BW (1976) Software engineering. IEEE Trans Comput C–25:1226–1241. https://doi.org/10.1109/TC.1976.1674590
15. Bonhard P, Sasse MA (2006) 'Knowing me, knowing you' — using profiles and social networking to improve recommender systems. BT Technol J 24:84–98. https://doi.org/10.1007/s10550-006-0080-3
16. Buolamwini J, Gebru T (2018) Gender shades: intersectional accuracy disparities in commercial gender classification. In: Proceedings of the 1st conference on fairness, Accountability and transparency. PMLR, New York, pp 77–91
17. Buono P, Berthouze N, Costabile MF, Grando A, Holzinger A (2024) Special issue on human-centered artificial intelligence for one health. Artif Intell Med 156:102946. https://doi.org/10.1016/j.artmed.2024.102946
18. Buono P, Caivano D, Costabile MF, Desolda G, Lanzilotti R (2020) Towards the detection of UX smells: the support of visualizations. IEEE Access 8:6901–6914. https://doi.org/10.1109/ACCESS.2019.2961768
19. Caldwell S, Sweetser P, O'Donnell N, Knight MJ, Aitchison M, Gedeon T, Johnson D, Brereton M, Gallagher M, Conroy D (2022) An agile new research framework for hybrid human-AI teaming: trust, transparency, and transferability. ACM Trans Interact Intell Syst 12. https://doi.org/10.1145/3514257
20. Capel T, Brereton M (2023) What is human-centered about human-centered Ai? A map of the research landscape. In: Proceedings of the 2023 CHI conference on human factors in computing systems. ACM, Hamburg, pp 1–23. https://doi.org/10.1145/3544548.3580959
21. Carroll JM (1995) Scenario-based design: envisioning work and technology in systems development, 1st edn. Wiley, New York, NY
22. Carvalho DV, Pereira EM, Cardoso JS (2019) Machine learning interpretability: a survey on methods and metrics. Electronics 8:832. https://doi.org/10.3390/electronics8080832
23. Castellano G, Esposito A, Lella E, Montanaro G, Vessio G (2024) Automated detection of Alzheimer's disease: a multi-modal approach with 3D MRI and amyloid PET. Sci Rep 14:5210. https://doi.org/10.1038/s41598-024-56001-9
24. Cerri S, Puonti O, Meier DS, Wuerfel J, Mühlau M, Siebner HR, Leemput KV (2020) A contrast-adaptive method for simultaneous whole-brain and lesion segmentation in multiple sclerosis. Neuroimage 225:117471. https://doi.org/10.1016/j.neuroimage.2020.117471
25. Chakraborty S, Tomsett R, Raghavendra R, Harborne D, Alzantot M, Cerutti F, Srivastava M, Preece A, Julier S, Rao RM, Kelley TD, Braines D, Sensoy M, Willis CJ, Gurram P (2017) Interpretability of deep learning models: a survey of results. In: Proceedings of the 2017 IEEE SmartWorld, ubiquitous intelligence & computing, advanced & trusted computed, scalable computing & communications, cloud & big data computing, internet of people and smart city innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). IEEE, San Francisco, CA, pp 1–6. https://doi.org/10.1109/UIC-ATC.2017.8397411
26. Choudhary T, Mishra V, Goswami A, Sarangapani J (2020) A comprehensive survey on model compression and acceleration. Artif Intell Rev 53:5113–5155. https://doi.org/10.1007/s10462-020-09816-7
27. Combi C, Amico B, Bellazzi R, Holzinger A, Moore JH, Zitnik M, Holmes JH (2022) A manifesto on explainability for artificial intelligence in medicine. Artif Intell Med 102423. https://doi.org/10.1016/j.artmed.2022.102423

28. Costabile MF, Desolda G, Dimauro G, Lanzilotti R, Loiacono D, Matera M, Zancanaro M (2022) A human-centric Ai-driven framework for exploring large and complex datasets. In: Barricelli BR, Fischer G, Fogli D, Mørch A, Piccinno A, Valtolina S (eds) Proceedings of the 6th international workshop on cultures of participation in the digital age: AI for humans or humans for AI? CEUR-WS, Aachen, pp 9–13
29. Crabtree A (2003) Designing collaborative systems: a practical guide to ethnography. Springer, London, Berlin, Heidelberg
30. Curci A, Esposito A (2024) Detecting brain tumors through multimodal neural networks. In: Proceedings of the 13th international conference on pattern recognition applications and methods. SCITEPRESS - Science and Technology Publications, Rome, pp 995–1000. https://doi.org/10.5220/0012608600003654
31. Desolda G, Dimauro G, Esposito A, Lanzilotti R, Matera M, Zancanaro M (2024) A human–AI interaction paradigm and its application to rhinocytology. Artif Intell Med 155:102933. https://doi.org/10.1016/j.artmed.2024.102933
32. Desolda G, Gaudino G, Lanzilotti R, Federici S, Cocco A (2017) UTAssistant: a web platform supporting usability testing in Italian public administrations. Proc Dr Consort Posters Demos CHItaly 1910:138–142
33. Desolda G, Lanzilotti R, Caivano D, Costabile MF, Buono P (2023) Asynchronous remote usability tests using web-based tools versus laboratory usability tests: an experimental study. IEEE Trans Hum-Mach Syst 53:731–742. https://doi.org/10.1109/THMS.2023.3282225
34. Dietterich TG (2017) Steps toward robust artificial intelligence. AI Mag 38:3–24. https://doi.org/10.1609/aimag.v38i3.2756
35. Dimauro G, Ciprandi G, Deperte F, Girardi F, Ladisa E, Latrofa S, Gelardi M (2019) Nasal cytology with deep learning techniques. Int J Med Inform 122:13–19. https://doi.org/10.1016/j.ijmedinf.2018.11.010
36. Dimauro G, Girardi F, Gelardi M, Bevilacqua V, Caivano D (2018) Rhino-Cyt: a system for supporting the Rhinologist in the analysis of nasal cytology. In: Huang D-S, Jo K-H, Zhang X-L (eds) Intelligent computing theories and application. Springer International Publishing, Cham, pp 619–630. https://doi.org/10.1007/978-3-319-95933-7_71
37. Dingsøyr T, Nerur S, Balijepally V, Moe NB (2012) A decade of agile methodologies: towards explaining agile software development. J Syst Softw 85:1213–1221. https://doi.org/10.1016/j.jss.2012.02.033
38. Donca I-C, Stan OP, Misaros M, Gota D, Miclea L (2022) Method for continuous integration and deployment using a pipeline generator for agile software projects. Sensors 22:4637. https://doi.org/10.3390/s22124637
39. Doshi-Velez F, Kim B (2017) Towards a rigorous science of interpretable machine learning. https://doi.org/10.48550/arXiv.1702.08608
40. Dwork C, Roth A (2013) The algorithmic foundations of differential privacy. Found trends®. Theor Comput Sci 9:211–407. https://doi.org/10.1561/0400000042
41. Dzindolet MT, Peterson SA, Pomranky RA, Pierce LG, Beck HP (2003) The role of trust in automation reliance. Int J Hum-Comput Stud 58:697–718. https://doi.org/10.1016/S1071-5819(03)00038-7
42. Ebert C, Gallardo G, Hernantes J, Serrano N (2016) DevOps. IEEE Softw 33:94–100. https://doi.org/10.1109/MS.2016.68
43. Edmonds A (2003) Uzilla: a new tool for web usability testing. Behav Res Methods Instrum Comput 35:194–201. https://doi.org/10.3758/BF03202542
44. Epley N, Waytz A, Cacioppo JT (2007) On seeing human: a three-factor theory of anthropomorphism. Psychol Rev 114:864–886. https://doi.org/10.1037/0033-295X.114.4.864
45. Eppler MJ, Mengis J (2004) The concept of information overload: a review of literature from organization science, accounting, marketing, MIS, and related disciplines. Inf Soc 20:325–344. https://doi.org/10.1080/01972240490507974
46. Esposito A, Calvano M, Curci A, Desolda G, Lanzilotti R, Lorusso C, Piccinno A (2023) End-user development for artificial intelligence: a systematic literature review. In: Spano LD, Schmidt A, Santoro C, Stumpf S (eds) End-user development. Springer Nature, Switzerland, pp 19–34. https://doi.org/10.1007/978-3-031-34433-6_2
47. Esposito A, Desolda G, Lanzilotti R (2024) The fine line between automation and augmentation in website usability evaluation. Sci Rep 14:10129. https://doi.org/10.1038/s41598-024-59616-0
48. Esposito A, Desolda G, Lanzilotti R, Costabile MF (2022) SERENE: a web platform for the UX semi-automatic evaluation of website. In: Proceedings of the 2022 international conference on advanced visual interfaces. Association for Computing Machinery, Frascati, pp 1–3. https://doi.org/10.1145/3531073.3534464

49. European Parliament, Council of the European Union (2024) 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)
50. Fabo P, Durikovic R (2012) Automated usability measurement of arbitrary desktop application with eyetracking. In: 2012 16th international conference on information visualisation. IEEE, Montpellier, pp 625–629. https://doi.org/10.1109/IV.2012.105
51. Fails JA, Olsen DR (2003) Interactive machine learning. In: Proceedings of the 8th international conference on intelligent user interfaces. ACM, Miami, pp 39–45. https://doi.org/10.1145/604045.604056
52. Federici S, Mele ML, Lanzilotti R, Desolda G, Bracalenti M, Buttafuoco A, Gaudino G, Cocco A, Amendola M, Simonetti E (2019) Heuristic evaluation of eGLU-box: a semi-automatic usability evaluation tool for public administrations. In: Kurosu M (ed) Human-computer interaction. Perspectives on design. Springer International Publishing, Cham, pp 75–86. https://doi.org/10.1007/978-3-030-22646-6_6
53. Fischer G, Fogli D, Piccinno A (2017) Revisiting and broadening the meta-design framework for end-user development. In: Paternò F, Wulf V (eds) New perspectives in end-user development. Springer International Publishing, Cham, pp 61–97. https://doi.org/10.1007/978-3-319-60291-2_4
54. Fischer JE, Greenhalgh C, Jiang W, Ramchurn SD, Wu F, Rodden T (2021) In-the-loop or on-the-loop? Interactional arrangements to support team coordination with a planning agent. Concurr Comput Pract Exp 33:e4082. https://doi.org/10.1002/cpe.4082
55. Fogli D, Piccinno A (2013) Co-evolution of end-user developers and systems in multi-tiered proxy design problems. In: Dittrich Y, Burnett M, Mørch A, Redmiles D (eds) End-user development. Springer, Berlin Heidelberg, pp 153–168. https://doi.org/10.1007/978-3-642-38706-7_12
56. Freiesleben T, Grote T (2023) Beyond generalization: a theory of robustness in machine learning. Synthese 202:109. https://doi.org/10.1007/s11229-023-04334-9
57. Friedman B, Hendry D (2019) Value sensitive design: shaping technology with moral imagination. The MIT Press, Cambridge, Massachusetts London
58. Garg S, Pundir P, Rathee G, Gupta PK, Garg S, Ahlawat S (2021) On continuous integration / continuous delivery for automated deployment of machine learning models using MLOps. In: 2021 IEEE fourth international conference on artificial intelligence and knowledge engineering (AIKE). IEEE, Laguna Hills, pp 25–28. https://doi.org/10.1109/AIKE52691.2021.00010
59. Gay G, Hembrooke H (2004) Activity-centered design: an ecological approach to designing smart tools and usable systems. The MIT Press
60. Gelardi M (2012) Atlas of nasal cytology, 14th ed. Edi.Ermes s.r.l.
61. Gless S, Silverman E, Weigend T (2016) If robots cause harm, who is to blame? Self-driving cars and criminal liability. New Crim Law Rev 19:412–436. https://doi.org/10.1525/nclr.2016.19.3.412
62. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press
63. Google PAIR (2019) People + AI guidebook. https://pair.withgoogle.com/guidebook. Accessed 13 Jun 2022
64. Grigsby SS (2018) Artificial intelligence for advanced human-machine symbiosis. In: Schmorrow DD, Fidopiastis CM (eds) Augmented cognition: intelligent technologies. Springer International Publishing, Cham, pp 255–266. https://doi.org/10.1007/978-3-319-91470-1_22
65. Guidotti R, Monreale A, Pedreschi D, Giannotti F (2021) Principles of explainable artificial intelligence. In: Sayed-Mouchaweh M (ed) Explainable AI within the digital transformation and cyber physical systems: XAI methods and applications. Springer International Publishing, Cham, pp 9–31. https://doi.org/10.1007/978-3-030-76409-8_2
66. Guidotti R, Monreale A, Ruggieri S, Turini F, Giannotti F, Pedreschi D (2019) A survey of methods for explaining black box models. ACM Comput Surv 51:1–42. https://doi.org/10.1145/3236009
67. Hardt M, Price E, Price E, Srebro N (2016) Equality of opportunity in supervised learning. In: Advances in neural information processing systems 29 (NIPS 2016). Curran Associates, Inc, Barcelona
68. Hartson HR, Smith EC (1991) Rapid prototyping in human-computer interface development. Interact Comput 3:51–91. https://doi.org/10.1016/0953-5438(91)90005-M
69. Hassenzahl M, Tractinsky N (2006) User experience - a research agenda. Behav Inf Technol 25:91–97. https://doi.org/10.1080/01449290500330331
70. Hewett T, Baecker R, Card S, Carey T, Gasen J, Mantei M, Perlman G, Strong G, Verplank W (1992) ACM SIGCHI curricula for human-computer interaction. Association for Computing Machinery
71. Hinman PG (2005) Fundamentals of mathematical logic. A.K. Peters, Wellesley, Mass
72. Holtzblatt K, Beyer H (2017) Contextual design: design for life, 2nd edn. Elsevier, Amsterdam, Cambridge, MA

73. Huang J, Shlobin NA, Lam SK, DeCuypere M (2022) Artificial intelligence applications in pediatric brain tumor imaging: a systematic review. World Neurosurg 157:99–105. https://doi.org/10.1016/j.wneu.2021.10.068
74. ISO (2018) 9241–11:2018 ergonomics of human-system interaction — Part 11: usability: definitions and concepts
75. ISO (2019) 9241–210:2019 Ergonomics of human-system interaction — part 210: human-centred design for interactive systems
76. Iyengar SS, Lepper MR (2000) When choice is demotivating: can one desire too much of a good thing? J Pers Soc Psychol 79:995–1006. https://doi.org/10.1037/0022-3514.79.6.995
77. Jian J-Y, Bisantz AM, Drury CG (2000) Foundations for an empirically determined scale of trust in automated systems. Int J Cogn Ergon 4:53–71. https://doi.org/10.1207/S15327566IJCE0401_04
78. Kaelbling LP, Littman ML, Moore AW (1996) Reinforcement learning: a survey. J Artif Intell Res 4:237–285. https://doi.org/10.1613/jair.301
79. Karna AK, Chen Y, Yu H, Zhong H, Zhao J (2018) The role of model checking in software engineering. Front Comput Sci 12:642–668. https://doi.org/10.1007/s11704-016-6192-0
80. Kaur D, Uslu S, Rittichier KJ, Durresi A (2022) Trustworthy artificial intelligence: a review. ACM Comput Surv 55:39:1–39:38. https://doi.org/10.1145/3491209
81. Kianpour M, Wen S-F (2020) Timing attacks on machine learning: state of the art. In: Bi Y, Bhatia R, Kapoor S (eds) Intelligent systems and applications. Springer International Publishing, Cham, pp 111–125. https://doi.org/10.1007/978-3-030-29516-5_10
82. Kreuzberger D, Kühl N, Hirschl S (2023) Machine learning operations (MLOps): overview, definition, and architecture. IEEE Access 11:31866–31879. https://doi.org/10.1109/ACCESS.2023.3262138
83. Lanzilotti R, Ardito C, Costabile MF, De Angeli A (2011) Do patterns help novice evaluators? A comparative study. Int J Hum-Comput Stud 69:52–69. https://doi.org/10.1016/j.ijhcs.2010.07.005
84. Larusdottir MK, Lanzilotti R, Piccinno A, Visescu I, Costabile MF (2023) UCD sprint: a fast process to involve users in the design practices of software companies. Int J Human–Computer Interact 1–18. https://doi.org/10.1080/10447318.2023.2279816
85. Lazar J, Feng JH, Hochheiser H (2017) Research methods in human-computer interaction, 2nd edn. Elsevier, Morgan Kaufmann Publishers, Cambridge, MA
86. Lee JD, See KA (2004) Trust in automation: designing for appropriate reliance. Hum Factors 46:50–80. https://doi.org/10.1518/hfes.46.1.50_30392
87. Lee MH, Chew CJ (2023) Understanding the effect of counterfactual explanations on trust and reliance on AI for human-AI collaborative clinical decision making. Proc ACM Hum-Comput Interact 7. https://doi.org/10.1145/3610218
88. Li T, Vorvoreanu M, Debellis D, Amershi S (2023) Assessing human-AI interaction early through factorial surveys: a study on the guidelines for human-AI interaction. ACM Trans Comput-Hum Interact 30. https://doi.org/10.1145/3511605
89. Li Y (2018) Deep reinforcement learning. https://doi.org/10.48550/ARXIV.1810.06339
90. Licklider JCR (1960) Man-computer symbiosis. IRE Trans Hum Factors Electron HFE-1:4–11. https://doi.org/10.1109/THFE2.1960.4503259
91. Lieberman H, Paternò F, Klann M, Wulf V (2006) End-user development: an emerging paradigm. In: Lieberman H, Paternò F, Wulf V (eds) End user development. Springer Netherlands, Dordrecht, pp 1–8. https://doi.org/10.1007/1-4020-5386-X_1
92. Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. In: Proceedings of the 31st international conference on neural information processing systems. Curran Associates Inc., Red Hook, NY, pp 4768–4777
93. Mantelero A (2018) AI and big data: a blueprint for a human rights, social and ethical impact assessment. Comput Law Secur Rev 34:754–772. https://doi.org/10.1016/j.clsr.2018.05.017
94. Mayer RC, Davis JH, Schoorman FD (1995) An integrative model of organizational trust. Acad Manag Rev 20:709. https://doi.org/10.2307/258792
95. McMahan B, Moore E, Ramage D, Hampson S, Arcas BA (2017) Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the 20th international conference on artificial intelligence and statistics. PMLR, pp 1273–1282
96. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A (2022) A survey on bias and fairness in machine learning. ACM Comput Surv 54:1–35. https://doi.org/10.1145/3457607
97. Memarian B, Doleck T (2023) Fairness, accountability, transparency, and ethics (FATE) in artificial intelligence (AI) and higher education: a systematic review. Comput Educ Artif Intell 5:100152. https://doi.org/10.1016/j.caeai.2023.100152
98. Miller T (2019) Explanation in artificial intelligence: insights from the social sciences. Artif Intell 267:1–38. https://doi.org/10.1016/j.artint.2018.07.007

99. Miniukovich A, Scaltritti M, Sulpizio S, De Angeli A (2019) Guideline-based evaluation of web readability. In: Proceedings of the 2019 CHI conference on human factors in computing systems. ACM, Glasgow, pp 1–12. https://doi.org/10.1145/3290605.3300738
100. Munoz C, da Costa K, Modenesi B, Koshiyama A (2024) Evaluating explainability for machine learning predictions using model-agnostic metrics. https://doi.org/10.48550/arXiv.2302.12094
101. Nagao K (2019) Symbiosis between humans and artificial intelligence. In: Artificial intelligence accelerates human learning. Springer Singapore, Singapore, pp 135–151. https://doi.org/10.1007/978-981-13-6175-3_6
102. Namoun A, Alrehaili A, Tufail A (2021) A review of automated website usability evaluation tools: research issues and challenges. In: Design, user experience, and usability: UX research and design: 10th international conference, DUXU 2021, held as part of the 23rd HCI international conference, HCII 2021, virtual event, July 24–29, 2021, proceedings, Part I. Springer-Verlag, Berlin, Heidelberg, pp 292–311. https://doi.org/10.1007/978-3-030-78221-4_20
103. Nielsen J (1993) Usability engineering. AP Professional, Cambridge, Mass
104. Pan SJ, Yang Q (2010) A survey on transfer learning. IEEE Trans Knowl Data Eng 22:1345–1359. https://doi.org/10.1109/TKDE.2009.191
105. Parasuraman R, Sheridan TB, Wickens CD (2000) A model for types and levels of human interaction with automation. IEEE Trans Syst Man Cybern - Part Syst Hum 30:286–297. https://doi.org/10.1109/3468.844354
106. Patton MQ (2015) Qualitative research & evaluation methods: integrating theory and practice, Fourth edition. SAGE Publications, Inc, Thousand Oaks, California
107. Peters D, Vold K, Robinson D, Calvo RA (2020) Responsible AI—two frameworks for ethical design practice. IEEE Trans Technol Soc 1:34–47. https://doi.org/10.1109/TTS.2020.2974991
108. Raji ID, Smart A, White RN, Mitchell M, Gebru T, Hutchinson B, Smith-Loud J, Theron D, Barnes P (2020) Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In: Proceedings of the 2020 conference on fairness, accountability, and transparency. ACM, Barcelona, pp 33–44. https://doi.org/10.1145/3351095.3372873
109. Ramesh A, Dhariwal P, Nichol A, Chu C, Chen M (2022) Hierarchical text-conditional image generation with CLIP latents. https://doi.org/10.48550/arXiv.2204.06125
110. Raschka S (2020) Model evaluation, model selection, and algorithm selection in machine learning. https://doi.org/10.48550/arXiv.1811.12808
111. Retzlaff CO, Angerschmid A, Saranti A, Schneeberger D, Röttger R, Müller H, Holzinger A (2024) Post-hoc vs ante-hoc explanations: xAI design guidelines for data scientists. Cogn Syst Res 86:101243. https://doi.org/10.1016/j.cogsys.2024.101243
112. Ribeiro MT, Singh S, Guestrin C (2016) "Why should I trust you?": Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, San Francisco, pp 1135–1144. https://doi.org/10.1145/2939672.2939778
113. Ricci F, Rokach L, Shapira B (2015) Recommender systems handbook. Springer US, Boston, MA
114. Rogers Y, Sharp H, Preece J (2023) Interaction design: beyond human-computer interaction, 6th edn. John Wiley & Sons, Inc, Hoboken
115. Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1:206–215. https://doi.org/10.1038/s42256-019-0048-x
116. Russell S, Norvig P (2016) Artificial intelligence: a modern approach, 3rd edn. Pearson, Boston
117. Schmidt A (2020) Interactive human centered artificial intelligence: a definition and research challenges. In: Proceedings of the international conference on advanced visual interfaces. Association for Computing Machinery, New York, NY, pp 1–4. https://doi.org/10.1145/3399715.3400873
118. Schmidt A, Herrmann T (2017) Intervention user interfaces: a new interaction paradigm for automated systems. Interactions 24:40–45. https://doi.org/10.1145/3121357
119. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2020) Grad-CAM: visual explanations from deep networks via gradient-based localization. Int J Comput Vis 128:336–359. https://doi.org/10.1007/s11263-019-01228-7
120. Sheth D, Giger ML (2020) Artificial intelligence in the interpretation of breast cancer on MRI. J Magn Reson Imaging 51:1310–1324. https://doi.org/10.1002/jmri.26878
121. Shneiderman B (2020) Human-centered artificial intelligence: three fresh ideas. AIS Trans Hum-Comput Interact 12:109–124. https://doi.org/10.17705/1thci.00131
122. Shneiderman B (2020) Human-centered artificial intelligence: reliable, Safe & Trustworthy. Int. J. Human–Computer Interact. 36:495–504. https://doi.org/10.1080/10447318.2020.1741118
123. Shneiderman B (2020) Bridging the gap between ethics and practice: guidelines for reliable, safe, and trustworthy human-centered AI systems. ACM Trans Interact Intell Syst 10:1–31. https://doi.org/10.1145/3419764

124. Shneiderman B (2022) Human-centered AI, 1st edn. Oxford University Press, Oxford
125. Siau K, Wang W (2018) Building Trust in Artificial Intelligence, machine learning, and robotics. Cut Bus Technol J 31:47–53
126. Siau K, Wang W (2020) Artificial intelligence (AI) ethics: ethics of AI and ethical AI. J Database Manag 31:74–87. https://doi.org/10.4018/JDM.2020040105
127. Smits J, Borghuis T (2022) Generative AI and intellectual property rights. In: Custers B, Fosch-Villaronga E (eds) Law and artificial intelligence. T.M.C. Asser Press, The Hague, pp 323–344. https://doi.org/10.1007/978-94-6265-523-2_17
128. Sommerville I (2016) Software engineering, tenth edition, global edn. Pearson, Boston, Columbus, Indianapolis, New York, San Francisco, Hoboken, Amsterdam, Cape Town, Dubai, London, Madrid, Milan, Munich, Paris, Montreal, Toronto, Delhi, Mexico City, São Paulo, Sydney, Hong Kong, Seoul, Singapore, Taipei, Tokyo
129. Sonderegger A, Sauer J (2009) The influence of laboratory set-up in usability tests: effects on user performance, subjective ratings and physiological measures. Ergonomics 52:1350–1361. https://doi.org/10.1080/00140130903067797
130. Sun R (2001) Introduction to computational cognitive modeling. In: Sun R (ed) The Cambridge handbook of computational psychology, 1st edn. Cambridge University Press, pp 3–20. https://doi.org/10.1017/CBO9780511816772.003
131. Tavallaey SS, Ganz C (2019) Automation to autonomy. In: 2019 24th IEEE international conference on emerging technologies and factory automation (ETFA). IEEE, Zaragoza, pp 31–34. https://doi.org/10.1109/ETFA.2019.8869329
132. Torrey L, Shavlik J (2010) Transfer learning. In: Olivas ES, JDM G, Martinez-Sober M, Magdalena-Benedito JR, Serrano López AJ (eds) Handbook of research on machine learning applications and trends. IGI Global, pp 242–264. https://doi.org/10.4018/978-1-60566-766-9.ch011
133. Ullman D, Malle BF (2018) What does it mean to trust a robot?: steps toward a multidimensional measure of trust. In: Companion of the 2018 ACM/IEEE international conference on human-robot interaction. ACM, Chicago IL, pp 263–264. https://doi.org/10.1145/3173386.3176991
134. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R (eds) Advances in neural information processing systems. Curran Associates, Inc
135. Voigt P, Von Dem Bussche A (2017) The EU general data protection regulation (GDPR). Springer International Publishing, Cham
136. Wachter S, Mittelstadt B, Floridi L (2017) Why a right to explanation of automated decision-making does not exist in the general data protection regulation. Int Data Priv Law 7:76–99. https://doi.org/10.1093/idpl/ipx005
137. Wang L, Yoon K-J (2022) Knowledge distillation and student-teacher learning for visual intelligence: a review and new outlooks. IEEE Trans Pattern Anal Mach Intell 44:3048–3068. https://doi.org/10.1109/TPAMI.2021.3055564
138. Watson DP, Scheidt DH (2005) Autonomous systems. Johns Hopkins APL Tech Dig 26:368–376
139. Widmer G, Kubat M (1996) Learning in the presence of concept drift and hidden contexts. Mach Learn 23:69–101. https://doi.org/10.1007/BF00116900
140. Williams S, Layard Horsfall H, Funnell JP, Hanrahan JG, Khan DZ, Muirhead W, Stoyanov D, Marcus HJ (2021) Artificial intelligence in brain tumour surgery—an emerging paradigm. Cancers 13:5010. https://doi.org/10.3390/cancers13195010
141. Wojton HM, Porter D, Lane ST, Bieber C, Madhavan P (2020) Initial validation of the trust of automated systems test (TOAST). J Soc Psychol 160:735–750. https://doi.org/10.1080/00224545.2020.1749020
142. Wu X, Xiao L, Sun Y, Zhang J, Ma T, He L (2022) A survey of human-in-the-loop for machine learning. Futur Gener Comput Syst 135:364–381. https://doi.org/10.1016/j.future.2022.05.014
143. Xu W (2019) Toward human-centered AI: a perspective from human-computer interaction. Interactions 26:42–46. https://doi.org/10.1145/3328485
144. Yan Z, Dong Y, Niemi V, Yu G (2013) Exploring trust of mobile applications based on user behaviors: an empirical study. J Appl Soc Psychol 43:638–659. https://doi.org/10.1111/j.1559-1816.2013.01044.x
145. Yu F, Moehring A, Banerjee O, Salz T, Agarwal N, Rajpurkar P (2024) Heterogeneity and predictors of the effects of AI assistance on radiologists. Nat Med 30:837–849. https://doi.org/10.1038/s41591-024-02850-w
146. Zaina LAM, Sharp H, Barroca L (2021) UX information in the daily work of an agile team: a distributed cognition analysis. Int J Hum-Comput Stud 147:102574. https://doi.org/10.1016/j.ijhcs.2020.102574
147. Zangerle E, Bauer C (2023) Evaluating recommender systems: survey and framework. ACM Comput Surv 55:1–38. https://doi.org/10.1145/3556536

148. Zhang C, Xie Y, Bai H, Yu B, Li W, Gao Y (2021) A survey on federated learning. Knowl-Based Syst 216:106775. https://doi.org/10.1016/j.knosys.2021.106775
149. Zhou N, Zhang Z, Nair VN, Singhal H, Chen J (2022) Bias, fairness and accountability with artificial intelligence and machine learning algorithms. Int Stat Rev 90:468–480. https://doi.org/10.1111/insr.12492

Giuseppe Desolda Giuseppe Desolda is an Assistant Professor at the Department of Computer Science, University of Bari Aldo Moro, Italy. His research interests focus on usable privacy and security, novel interaction techniques, and Internet of Things. He is coordinating projects on Usable Security and on Human-Centered Artificial Intelligence mechanisms for software engineering.

Andrea Esposito Andrea Esposito is a Ph.D. student at the Department of Computer Science, University of Bari Aldo Moro, Italy. He is a member of the Interaction Visualisation Usability (IVU) and UX Laboratory. His interests lie in Human-Computer Interaction, eXplainable Artificial Intelligence, and Human-AI Interaction. He is committed to advancing the field of Human-Centred AI, working to improve the interaction between humans and AI systems.

Rosa Lanzilotti Rosa Lanzilotti is a Professor at the Department of Computer Science of the University of Bari. She promotes usability and UX practices in companies and public institutions. She coordinated projects aimed at developing eGLU-Box PA, a web platform used by Italian institution staffs to perform usability evaluation of their websites.

Antonio Piccinno Antonio Piccinno is an Associate Professor of the Department of Computer Science of the University of Bari. His research interests focus on Human-Centered Design (HCD) and End-User Development (EUD). He promotes the Interplay between Human-Computer Interaction and Software Engineering and, more recently, Secure Software Analysis and Design.

Maria Francesca Costabile Maria Francesca Costabile is a Professor at the Department of Computer Science of the University of Bari, where she is the director of the Interaction Visualisation Usability (IVU) and UX Lab. She was a pioneer of Human-Computer Interaction in Italy and promoted the ACM SIGCHI Italian Chapter in 1995. She received the ACM SIGCHI Lifetime Service Award and the AVI-SIGCHItaly Lifetime Service Award.