

SURVEY

A Multifaceted Vision of the Human-AI Collaboration: A Comprehensive Review

MAITE PUERTA-BELDARRAIN¹, OIHANE GÓMEZ-CARMONA¹,
RUBÉN SÁNCHEZ-CORCUERA², DIEGO CASADO-MANSILLA², DIEGO LÓPEZ-DE-IPÍÑA²,
AND LIMING CHEN³, (Senior Member, IEEE)

¹Deustotech, University of Deusto, 48007 Bilbao, Spain

²Faculty of Engineering, University of Deusto, 48007 Bilbao, Spain

³School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China

Corresponding author: Maite Puerta-Beldarrain (mpuerta004@deusto.es)

This work has been supported by grant IT1582-22 from the Basque Government, which recognizes Human-centric Computing for Smart Sustainable Communities and Environments (DEUSTEK5) as an excellent research group under the Basque university system. Besides, we acknowledge the Ministerio de Ciencia e Innovación for Internet of People (IoP), under Grant No.: PID2020-119682RB-I00. Finally, this work has been partially supported by the European Commission through the Wearables and drones for City Socio-Environmental Observations and Behavioral Change (SOCIO-BEE) project Under Grant No. 101037648.

ABSTRACT Human-AI collaboration has evolved into a complex, multidimensional paradigm shaped by research in various domains. Key areas such as human-in-the-loop systems, Interactive Machine Learning (IML), Hybrid Intelligence, and Human-Agent Interaction have significantly contributed to this development. However, these fields often lack cohesion, underscoring the need for a cohesive perspective to advance. This work addresses this gap by integrating insights from diverse aspects of collaboration to present a holistic approach to fostering effective and adaptive interactions between humans and artificial agents. It emphasizes empowering end-users with greater control and involvement in decision-making processes, thereby enhancing both the levels of interactivity and adaptability within intelligent systems. Moving beyond a focus on AI training techniques, this paper presents a broader perspective on incorporating human input into AI decision-making and learning processes, highlighting the importance of flexibility in systems and user engagement. The manuscript proposes a framework encompassing five levels of human integration and examines their relationship with core collaboration aspects, including the system purpose, participant expertise, and system proactivity. By synthesizing current knowledge on human-AI collaboration and outlining essential design principles, this work aims to advance the field and foster interdisciplinary collaboration among researchers, practitioners, and designers.

INDEX TERMS Hybrid intelligence, human-AI collaboration, human-robot collaboration, human-machine collaboration, human-in-the-loop, interactive machine learning, human-machine symbiosis, human-centred AI.

I. INTRODUCTION

Technology has rapidly evolved to become an integral component of contemporary life, progressively incorporating humans into its decision-making processes and actively integrating them within the technological framework [1], [2], [3]. The rising emphasis on human-centered technology has been a significant driver behind the advancements in AI's learning and reasoning capabilities.

The associate editor coordinating the review of this manuscript and approving it for publication was Zijian Zhang¹.

Understanding this technological evolution requires a thorough examination of how technological systems have adapted to facilitate human involvement in decision-making processes. This transformation likely began with the development of the Ubiquitous Computing paradigm, which ignited the desire for greater technological adaptability and connectivity [4]. As these technologies became more widespread, the concept of Smart Environments emerged, designed to enhance our surroundings without demanding active human engagement [5], [6]. Building on this foundation, Ambient Intelligence (AmI) and Intelligent Environments (IE) further

advanced the principle of user-centric technology, prioritizing interactive and responsive environments [7], [8]. In more recent developments, the concept of the Internet of People (IoP) has shifted its focus towards recognizing users as active contributors within technological ecosystems, rather than passive recipients of data-driven services [2], [9]. These evolving fields collectively illustrate that technology has been continuously advancing to meet users' increasing demand for a more immersive and integrated relationship with technology, positioning them in a central and participatory role.

The progression of technology has significantly influenced the concept of collaboration across multiple disciplines over the years [10], [11]. In this context, collaboration is defined as joint interactions between technology and its users aimed at achieving a specific outcome. This notion of collaboration has evolved, taking on various terminologies and attributes depending on the particular field of study. This variation has led to a fragmentation of the concept across different domains, with each area focusing on unique dimensions of collaborative processes [12], [13], [14], [15]. Such fragmentation has contributed to a lack of coherence in the literature, creating obstacles to developing a unified understanding of the different types of collaboration, their requirements, and the nuances in integrating users into the decision-making process.

This paper aims to bridge these diverse viewpoints by presenting a unified framework that facilitates the development of human-AI collaborative technologies based on a holistic understanding of the types and requirements of collaboration. It strives to synthesize insights from various research areas to establish a balanced, productive, and sustainable model of collaboration between humans and artificial intelligence (AI). The core contributions of this paper include:

- A comprehensive overview of AI techniques and subfields that are instrumental in creating interactive technologies.
- A refined classification system categorizing five groups of human-AI collaboration, providing a structured guide to understanding the different types of collaboration and their distinctive features. This system is organized around key aspects of collaboration, such as the degree of human involvement, level of interaction, and collaboration type.
- Design principles for developing future human-centered collaborative technologies. These principles aim to expand the narrow focus of existing research by fostering a broader approach that enhances the comprehension of human needs within collaborative systems.

The remainder of this paper is organized as follows. Section II highlights several research areas related to human-AI collaboration to illustrate the necessity of developing a shared perspective on this concept. Section III describes the Methodology employed for the literature review. Section IV explains the techniques to develop an interactive approach between users and intelligent systems. Section V presents

the five clusters obtained based on the main characteristics of collaboration. Section VI describes the essential design principles users need for a long-lasting collaboration. Sections VII discuss the necessity to integrate human beings' opinions and knowledge in the design phase of future intelligent systems. Finally, in Section VIII the conclusions of this work are presented.

II. A MULTIFACETED OVERVIEW OF THE COLLABORATION

Generally, collaboration is understood as two or more entities working together, sharing responsibility for solving problems, and actively interacting to achieve a common goal [16]. The primary purpose of collaboration is to leverage the strengths of each entity to compensate for the limitations of others, enabling achievements beyond what each could accomplish independently [12], [14], [17], [18]. However, this definition can vary significantly depending on the specific research domain. This section explores the benefits and challenges of collaboration and examines the different perspectives that this concept encompasses across various fields.

Artificial Intelligence (AI) has demonstrated remarkable computational capabilities, excelling in tasks that involve recognizing patterns, handling unforeseen scenarios, and adapting to diverse conditions using advanced techniques like zero-shot learning and pre-training models [19], [20], [21]. These capabilities enable AI to identify complex patterns in large datasets and process information at speeds that far exceed human abilities, all while remaining unaffected by factors such as stress, motivation, or fatigue [22].

However, despite its strengths, AI's performance is heavily dependent on the quality of its training data. Inaccuracies or biases in this data can lead to flawed outcomes, particularly in sensitive areas such as race, gender, ethics, and politics [20], [23]. Furthermore, AI often lacks a deep understanding of language and context, which can result in misinterpretations that undermine the accuracy of decision-making processes [20]. In contrast, humans inherently possess contextual awareness and ethical reasoning, allowing them to navigate ambiguities and make more nuanced decisions than the machines [24].

While the potential for collaboration between humans and AI is substantial due to their complementary strengths, several significant challenges must be addressed to ensure its success. For effective collaboration, both AI systems and human participants need to understand each other's roles, integrate each other's capabilities into their strategies, and communicate clearly to avoid misunderstandings [25]. Technical barriers also emerge from the fundamentally different ways humans and AI process information, which can create obstacles to seamless communication [25], [26]. Additionally, the complexity and opacity of many AI models—often referred to as black-box systems—pose challenges in terms of transparency, trust, and acceptance within collaborative environments [27], [28]. Nevertheless,

fostering collaboration between humans and AI remains one of the most promising pathways for enhancing the capabilities of both entities.

This approach, despite its inherent challenges, offers considerable advantages by leveraging the unique capabilities of both humans and AI technologies. A major issue, however, lies in achieving a unified understanding of the diverse research fields that apply the concept of collaboration, especially in pinpointing the specific attributes that characterize collaboration within each field:

- *Human-in-the-loop (HiTL)*: Focuses on a human-centered approach to collaboration by directly involving individuals in AI systems' decision-making processes [29], [30], [31].
- *Interactive Machine Learning (IML)*: Highlights the importance of creating interactive communication channels between users and AI models, emphasizing user engagement throughout various stages of AI development [32].
- *Human-Robot Collaboration*: Centers on developing robots designed to work collaboratively with humans in physical spaces to achieve common objectives [33].
- *Collaborative Interactive Learning (CIL)*: Aims to create intelligent systems that seamlessly assist humans in everyday tasks through continuous interaction and learning [13], [34].
- *Human-Agent Symbiosis*: Inspired by biological symbiosis, this concept emphasizes the mutual benefits arising from the dynamic interaction between humans and AI systems [35], [36].
- *Human-Computer Interaction*: Seeks to optimize the design of intuitive and efficient interaction methods between users and autonomous systems [37], [38].
- *Collective Hybrid Intelligence*: Advocates the notion that true intelligence emerges from the sustained collaboration and co-learning between human and machine counterparts [39].
- *Cyber-Physical-Human Systems (CPHSs)*: Focuses on the seamless integration of digital, physical, and human components to facilitate efficient information exchange and task execution [40].
- *Hybrid Intelligence Systems*: Examines socio-technical frameworks that synergize human and AI capabilities to address complex problems through collaborative efforts [39], [41], [42].
- *Artificial Intelligence Generative Content (AIGC)*: Involves AI systems that generate creative outputs like text, images, or music in a process that promotes collaborative learning and user involvement [19], [20], [43], [44].
- *Human-AI team*: Suggests that humans and AI function not as separate entities but as integrated parts of a cohesive team, working symbiotically towards shared goals [45], [46].

Each of these research fields embodies unique characteristics and focuses on specific facets of human-technology

interaction. To develop a comprehensive understanding of the various forms of collaboration, this paper will analyze the degrees of human involvement within these collaborative approaches and identify the key design principles necessary for enabling effective human participation.

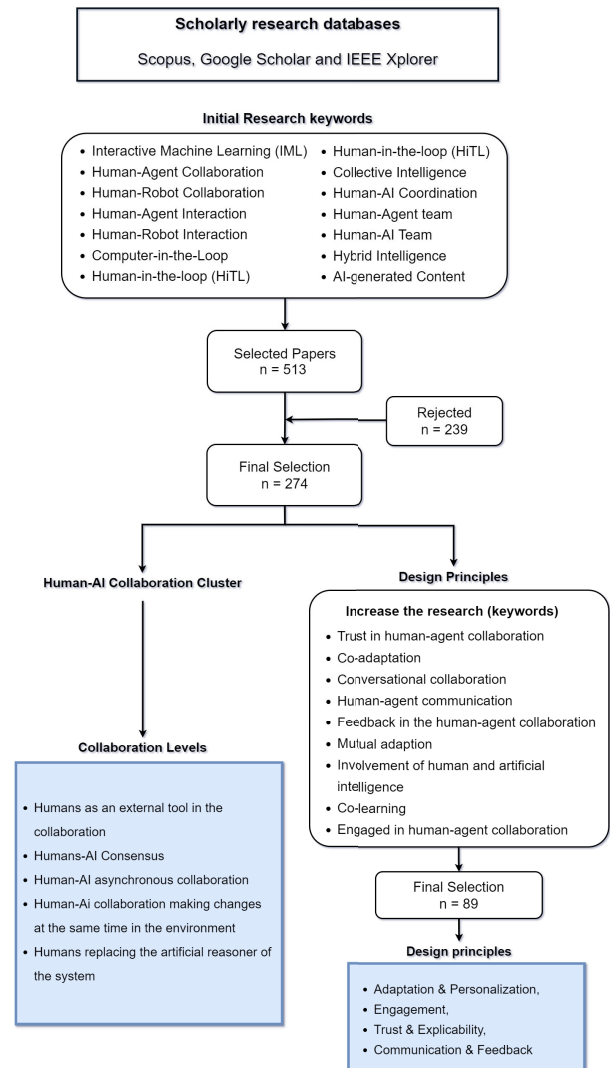


FIGURE 1. The methodology employed for the literature review that resulted in the Sections V and VI.

III. METHODOLOGY

The methodology employed in this academic paper involved a comprehensive exploration of the current landscape of human-AI collaboration and its related fields, as illustrated in Figure 1. This study followed a systematic review approach, drawing information from primary scholarly research databases, including Google Scholar, Scopus, and IEEE Xplore.

To gather relevant literature, the databases were queried using a combination of targeted keywords, such as Interactive Machine Learning (IML), Human-Agent Collaboration, Human-Robot Collaboration, Human-Agent Interaction, Human-Robot Interaction, Computer-in-the-Loop,

Human-in-the-Loop (HiTL), Collective Intelligence, Human-AI Coordination, Human-Agent Team, Human-AI Team, Hybrid Intelligence, and AI-generated Content. Articles containing these keywords in the title, abstract, or keywords were initially selected.

Following the initial selection, a rigorous screening process was conducted. The abstracts of relevant articles were meticulously evaluated to identify works closely aligned with our research objectives. This approach ensured that only studies involving user participation as active component were considered. Additionally, the most important references from key papers were reviewed using the Snowballing method, applying the same exclusion criteria as described.

The selected works provided two key insights: the classification of Human-AI Collaboration and the formulation of Collaborative Design Principles. For the Human-AI Collaboration classification, we specifically focused on studies where collaboration involved humans as active participants, rather than merely passive data collectors. To capture the full range of collaborative possibilities, we selected various types of collaboration that illustrate the diverse nature of this concept and its applicability across different contexts, aiming to create a comprehensive cluster of collaboration models.

For the development of collaborative design principles, an extensive investigation was conducted using additional keywords such as trust, engagement, communication, feedback, and explicability. This phase involved synthesizing insights from various sources to formulate design guidelines that facilitate effective human-AI interaction. By combining these findings with the outcomes of the screening process, we were able to build a robust framework for designing human-AI collaboration systems.

This methodology encompasses a comprehensive classification of collaboration levels informed by the selected works and a set of collaborative design principles. This study aims to provide a thorough understanding of the current state of human-AI collaboration, laying the foundation for future advancements in this dynamic field.

IV. INTERACTIVE TECHNIQUES FOR COLLABORATION

This section provides a concise introduction to collaborative and interactive techniques, acknowledging the current gap in a comprehensive and systematic review of interactive collaboration methods and AI training techniques. To address this gap, the section explores Interactive Machine Learning and its subfields, including Active Learning and Reinforcement Learning. Additionally, the discussion extends to the domain of Artificial Intelligence-Generated Content, reflecting the growing significance and popularity of this emerging approach.

A. INTERACTIVE MACHINE LEARNING

As has been introduced, the evolution of future smart environments and systems is accelerated by technological breakthroughs, resulting in an eventual tighter interaction with users. The next step of this process is to create

interactive solutions that collaborate with people to build dynamic places and solutions in cooperation. To achieve this collaboration, the end-users role must be empowered by giving them an active role and increased control (power decision) over intelligent systems, creating more interactive spaces and services [47]. Nevertheless, traditional systems based on AI and ML techniques (also called automatic ML systems or aML) are not conceived with this vision in mind. They are designed to automatically learn from data and produce outcomes, resolving problems without human intervention. In such scenarios, the initial strategy for building some collaboration was relegated to allowing humans to alter a predefined parameter of an aML algorithm to improve its performance [48]. As aML techniques are still extensively employed, they relegate human interaction to be extremely limited and concrete, preventing the integration of humans into the systems and long-lasting interaction and collaboration. For this reason, the concept of Interactive Machine Learning (IML) emerges as an alternative to this classical approach and, in particular, in evolving scenarios where users can actively collaborate, providing its capacity for flexible and rapid adaptation to new circumstances in new emerging smart solutions. That is to say, IML models are designed to integrate human input into the ML solutions interactively, steering the supervised Learning process of the model in different stages [49]. Thus, through these techniques, the system can incorporate human expertise into account and exploit their strengths, together with ethical, legal, and societal human considerations [14].

To better understand IML's role play, it is important to know its evolution and the aspects that have led it to its current perception. To begin with, the idea of IML was initially approached by Ware et al. [50]. In their work, they presented a graphical concept to ML with explicit representations of data in which experts could participate in defining decision boundaries to improve the construction of ML classifiers by incorporating human input. In the same line, later, Fails and Olsen introduced a more mature representation of the concept of IML and its comparison against classical ML approaches [51]. They explored the convenience of interactive approaches to guide the training of a classifier until the desired results are met, leveraging the capabilities of a domain expert during the learning process. In this line, Smith et al., [52] studied this collaboration from a two-fold perspective.

On the one hand, it addresses how systems can learn interactively from non-ML experts and, on the other hand, how appropriate tools can also enrich non-experts' experience to guide them in the process. This involves those approaches that take advantage of human agents (i.e., domain experts) to create improved models and solutions that can optimize their learning behaviour based on human knowledge and experience [53], [54], [55]. Indeed, taking this ML model-building procedure repeatedly with end-user input, other authors settled on more participatory approaches and started taking advantage of human knowledge and, more specifically,

non-ML experts to build and improve ML solutions [56]. For that, visual exploratory tools [55], [57], [58], [59], data analysis and interactive exploration approaches have been used to improve ML models [60], [61].

Considering the relevant role that humans may take in the ML process and the benefit of this collaboration, Jiang et al. [62] presented a task-oriented taxonomy with nine categories to determine the most prominent application fields of IML. Their work identifies two main phases of typical IML workflow: model preparation and the continuous improvement of deployed models. To begin with, model preparation is related to the training process. It encompasses those actions in which the user can examine the intermediate outcomes of the model and the data graphically. That is, to gain understanding and insights into the data and the model, and then provide feedback to the model, e.g. providing training parameters or contributing to the examination of model outputs [63]. This training process involves users in assessing model results and making adjustments. To continue in the second phase, when continuous model improvement is envisaged, the model is gradually modified by including feedback from humans, e.g., giving continuous feedback to the model about its performance or providing new sources of truth to refine it after training continuously. These ideas and works founded a new and more explicit definition of IML that Dudley et al. [64] present: “*an interaction paradigm in which a user or user group iterative builds and refines a mathematical model to describe a concept through iterative cycles of input and review.*”. Taking as guidelines the previous statements, IML solutions should also be a driver to promote this human-AI collaboration in different phases of the model design, creating long-lasting interactions to make ML algorithms more useful, accessible and usable for every particular individual [65].

Under this definition, applications as varied as predictive maintenance [66], the design of smart spaces [67], the analysis of insurance claims [68], sentient analysis [69] have emerged with this collaboration in mind in domains such as healthcare [53], education [70] or cyber-security [71], among others. All these examples have demonstrated that they can benefit from IML approaches by optimizing their learning behaviour through these interactions. One very illustrative example of this interactive paradigm for incorporating such non-experts’ knowledge in ML models is healthcare [72]. In such domains, some challenges like the uncertainty of the clinical diagnosis or the lack of quality data, the identification of rare events are hard to solve with fully automated approaches without human intervention [73], [74]. For example, human expertise can be utilized to select samples heuristically, reducing the exponential search space and integrating human knowledge into machine intelligence to discover novel, previously unknown insights into data [60], [75]. Given this relevance, Sun et al. studied users’ role in better understanding the available data by a system called label-and-learn, designed to help people without machine

learning expertise understand the classifier’s behaviour in labelling tasks [76]. At the same time, Yang et al. investigated how non-experts could build ML solutions for themselves in real life [56]. Nevertheless, it is recognized that a human-centred design is crucial for creating new technology [77].

Thus, as relevant as it is to understand how to build better models, it is to evaluate the human-centred perspective of ML [78] and how to frame this interaction [79]. For this reason, McCallum et al. evaluated how users could be supported to contribute to the model [80]. In their work, they examined the use case of a system in which a user-driven feature representation for new ML problems needed to be assessed. The conducted evaluation demonstrates that users’ have a low perception of their actions. This idea directly connects to improving the interaction between human agents and learning models to make ML more accessible for participants, enabling humans to efficiently and effectively interact with the system [81]. In parallel, the relevance of integrating explanations into interactive ML models to develop a robust and bi-directional communication channel between human stakeholders and machine learning models has also been a matter of study by the literature [32]. This way, understanding the machine’s behaviour through explanations can provide better-informed feedback that is advantageous to all parties involved. That is, better mechanisms are needed to examine the impact of their actions and outcomes, as understanding ML models and interpreting their behaviours is crucial for user interaction in IML applications.

In light of this, users’ participation has also been addressed from a Human-Computer Interaction (HCI) standpoint that seeks to facilitate the iterative design, improvement, and dissemination process of learning systems [82]. Considering this, Vishwarupe et al. [83] analyzed the challenges of getting a robust, transparent, ethical, intelligent and interactive mechanism for the co-existence of AI and HCI system that may be the basis of this collaboration. In this work, accessibility and user-friendliness are identified among the cornerstone aspects of a user-centred design process that can help develop a new paradigm for advancing technology, including a more seamless collaboration. At the same time, explainability, usability, or transparency are identified areas of improvement for IML systems [84]. Thus, the IML process should be characterized by the user being the primary driver of an interactive bi-directional process to achieve desired system behaviour [85]. For instance, providing the user control over the system’s high-level behaviour aligns with the idea of empowering users in collaboration with intelligent systems [86].

To do so, Inkpen et al. [87] examined how to integrate artificial and human intelligence in decision-making better, addressing the different human roles and relationships in human-AI systems. Similarly, the reliability of the user providing new data to the interactive model was also studied by Tegen et al., [88] and by Amershi et al., [86], which concluded that further evaluation of novel interaction methods is needed

to determine whether the user influence over the model does or not result in better systems. To facilitate those applications, IML can also involve the design of intelligent user interface (IUI) frameworks that drive this interaction, increasing its accessibility [64].

However, we should not limit humans' role to the model's creation and definition. For instance, in classification tasks, IML approaches may propose end-users collaborate with the learning system, adapting and customizing systems by tuning them according to their preferences, validating its outcomes or even feeding it with more data [89]. One case of the last example is customizing Human Activity Recognition (HAR) models with user-dependent data once the model is already in use. This partnership aims to enhance the categorization of human motions and behaviors by combining embodied and non-embodied sensory data [90]. Traditionally, AI and ML systems start with a basic model trained on general data. However, these models may not be enough to generalize when classifying new data acquired in real contexts. Furthermore, it may also not be acceptable for new users whose data has never been viewed before [91]. Consequently, this base model may not perform well considering the particularities of the end-user. For instance, when a left-handed person uses a model trained on data from right-handed people, it may perform poorly. This would result in a bad user experience for the end-user, who sees that the system cannot recognize their movements accurately. When HAR solutions fail or can not classify users' behaviors, users may become frustrated, leading to future disengagement or discomfort with AI models and intelligent solutions [92], [93]. This emphasizes the need for considering each person's unique circumstances and data when creating intelligent systems dedicated to monitoring users' activities and behaviors, as the accuracy of the same intelligent system is perceived very differently by each user depending on their expectations [94]. Thus, their expertise and willingness to interact with the system for model calibration and customization can be useful to improve the system's accuracy, and this collaboration involving the user in assisting the learning system could imply improving its detection skills over time [95].

In essence, the described works demonstrate interaction's relevance in improving ML models, either during their conception or by improving their performance later on. Specifically, this encompasses both the user's role in creating and defining the model and iterative interactions with learning systems over time [96]. These interactive paradigms are revolutionizing collaboration by permitting unrestricted interaction between contributors and a more sustainable approach that generates superior results. When it comes to the application of this interaction, Active learning (AL) [30] and Reinforcement Learning (RL), also named Interactive reinforcement learning [97], are the most popular interactive learning approaches devised to facilitate performance improvement in ML models with the help of human knowledge. Accordingly, in the following, we will describe them to give an overview of how these approaches

might be utilized to promote collaboration. In addition, we must not forget that these techniques can be employed without humans. However, as this work focuses on human-AI collaboration, we exclusively focus on those that involve humans and agents in the system.

1) ACTIVE LEARNING

Active learning (AL) is a sub-field of machine learning in which the model can query a user or other source of information, called an oracle, to label new data points in order to improve model performance, as shown in Figure 2 [55], [98], [99], [100]. The points to label are selected using different strategies that follow some heuristics to improve the model's performance by using the minimum number of queries possible. In this article, we will focus on the human labelling of the data; thus, we have to consider the saturation of the user if the proposed model is constantly asking them to label data.

The strategies used in AL can be divided into three types following the division proposed by [101]: *pool-based*, *stream-based* or *membership query-synthesis based*. *Pool-based* strategies are used when the proposed problem contains a pool of unlabelled data from which points are extracted and labelled if they fall within the requirements of the proposed strategy. Subsequently, the set of labelled instances is used to train the model and check if it has reached a performance threshold set by the user. In *stream-based* tasks, unlike pool ones, the data comes through a stream, and the proposed strategy decides whether to annotate the received point or move on to the next point coming through the stream. Finally, in *membership query-synthesis based* tasks, the model should be able to the whole data distribution. For example, the most important strategy in these tasks is uncertainty sampling, which selects the instance with the highest entropy or the one the model is most uncertain about.

Researchers in the area have conducted a prolific investigation into the strategies used to select the next data point to label depending on the needs of the task proposed. In the following, we will explain common strategies used to decide which data will be labelled from the pool following the classification proposed by [101]. These are the main strategies for classification or regression problems when using AL:

- *Uncertainty Sampling* [102]: This is one of the most common sampling strategies. The data are selected considering which are the least certain when evaluating them with the classification function proposed for the problem. The data with the least certainty are selected because they provide the most information to the model; however, the data with the most certainty are treated as redundant data that do not provide information.
- *Query by committee* [103]: For this strategy, different models are presented that evaluate the data and decide the following data to be labelled by looking at which one creates the highest disagreement between the different models. For this method to work most effectively, each

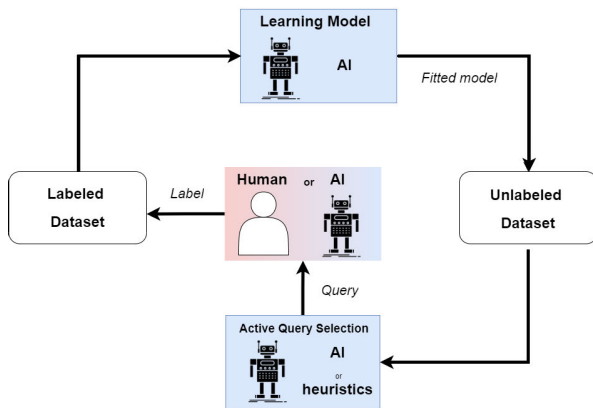


FIGURE 2. A depiction of active learning process from data retrieving, labelling and feeding to the learning model.

of the models in the committee has to represent a part of the space that all the data samples represent.

- **SVM-Based [104]:** The SVM-based strategy uses this algorithm to search for the points closest to the decision boundary. By labelling these points, the classification boundary can be defined more precisely.
- **Density-Based [105]:** This strategy seeks to label the data that is in an area of high data density as it is believed that areas of high data density have high data representation. To do this, the distance of all unlabelled data to each other is calculated and given a distance score. The datum with the lowest distance to each other will be scored next.
- **Variance reduction [106]:** The variance-reduction strategy makes use of statistics to calculate which instances of the data will most reduce the final model error and choose them to be annotated by the oracle. This strategy is used for classification or regression problems with AL.
- **Maximizing Expected Model Change [107]:** This strategy, based on a stochastic gradient, selects the points that it estimates to generate the maximum change in the current model parameters.

Apart from these strategies, which are some of the most widely used, the scientific community has created more to adapt the proposed solutions to the tasks that have arisen [101]. Some of them are:

- **Diversity-Based [108]:** In this strategy, labellers choose multiple unlabeled samples for simultaneous labelling. Labelling multiple instances at once using consistent criteria reduces redundancy within the dataset, enhancing the efficiency of the learning process.
- **Probabilistic Active Learning (PAL) [109]:** This strategy operates on the premise that if two instances share close proximity in the feature space, their labels are likely to be similar. Therefore, the instances considered most interesting are those that have a significant impact on the classification performance. To assess this, a probabilistic estimate is employed to examine

the neighbourhood information of a candidate instance and measure the overall improvement in classification performance.

In addition, in recent years, meta-active learning (meta-AL) strategies have been developed and have gained popularity. As stated by [110], these strategies are intended to replace traditional ones by selecting the optimal set of unlabeled items for labelling [99], [111]. These can involve developing algorithms that dynamically adjust their querying strategies based on the characteristics of the current dataset or creating methods to combine different active learning strategies for optimal performance across diverse scenarios.

Nowadays, a popular meta-active learning strategy is based on formulating the AL problem in the reinforcement learning framework, where the query strategy is expressed as a policy to be learned by reinforcement learning [112]. For example, Fang et al. described the active learning scenario as a Markov decision process by considering a state as an unlabeled sample and the action as whether a label is required [113]. Thus, the optimal policy, or query strategy in the case of AL, was learned by setting the parameters of the prediction model. Woodward et al. used deep reinforcement learning with Long Short-Term Memory (LSTM) to design a function that determines if a data point label needs to be queried for stream-based active learning [114].

Meta-active learning, a subset of the broader meta-learning research field, represents an emerging frontier in machine learning [115], [116]. We briefly explain it here to provide readers with insight into creating AI reasoners in the AL problem. However, our primary focus in this work is on the interaction and integration of users within the AI's learning and decision-making processes. As a result, we will not delve deeply into this area within the scope of this work. If you are interested, further exploration can be found in upcoming surveys [115], [116], [117], [118], [119], [120], [121].

2) REINFORCEMENT LEARNING

Reinforcement learning (RL) is a technique based on learning from experience to achieve a goal through a series of discrete time steps of interaction with its environment [122]. Exploring the current situation or condition in which an agent finds itself, seeing the reaction of the environment to the agent's decision and using that knowledge to make better decisions in the future is the foundation of reinforcement learning. Reinforcement learning is distinguished from other computational approaches by emphasizing an agent's direct interaction with the environment without explicit supervision or exhaustive environmental models to learn the best way to act in each situation [123], [124].

The RL learning process, depicted in Figure 3, operates within a loop denoted as the t -th iteration. In this loop, the agent selects and executes an action (denoted as A_t) based on its understanding of the environment, represented as S_t called the current state. This action is determined by a policy, a set of rules governing the agent's decisions. Following the agent's

action, the environment responds by providing two crucial pieces of information: 1) the reward R_{t+1} , indicating the effectiveness of the chosen action in achieving the ultimate goal, and 2) the new state of the environment after the agent's action, denoted as S_{t+1} .

The agent must continually update its policy based on the obtained rewards to optimise its decision-making process across all possible states [123]. The agent learns the optimal behaviour to achieve the desired goal through this iterative process. Importantly, when humans are involved in the RL problem, the resulting state emerges from the collaborative actions of both the agent and the human within the environment. Consequently, the agent's policy updates and learning are profoundly influenced by the behaviour of the human agent.

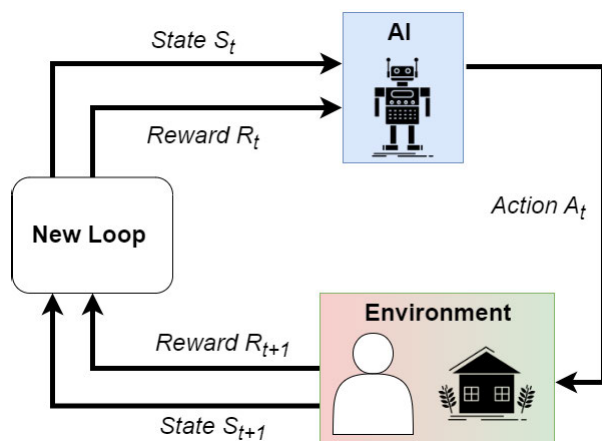


FIGURE 3. A depiction of the reinforcement learning loop.

The set of actions that the agent can perform can be discrete or continuous. For example, if an agent is learning to play the Tic Tac Toe game, their actions determine the places for the agent's symbol, i.e. they are discrete. On the contrary, if the agent's actions are based on knowing how much it has to accelerate to drive, the agent's action is represented by a continuous variable that defines the acceleration. In these cases, the continuous variables are usually segmented into different groups to form a discrete variable. In addition, RL algorithms can be composed of discrete and continuous actions, but many algorithms are limited to one [125].

Another critical characteristic of RL problems is whether they are episodic or continuous. An example of an episodic problem is arcade games, where the agent receives the final reward at the end of the game. The agent needs to play several times to learn how to win. Conversely, the problem can also be continuous, where the task does not have a final step after a predetermined number of iterations, such as when the agent is learning to drive [125].

The reward is the primary criterion for altering the policy; if a low reward follows an action selected, then the policy may be changed to select some other action in the future for this situation. In addition, actions could affect not only the immediate reward but also the subsequent situation and,

by extension, all subsequent rewards. Therefore, the agent has to consider that performing an action that results in a low reward may be necessary at a particular time to obtain a higher reward later. Consequently, the agent's goal is to choose an action that will allow the system to maximise the reward obtained for each action over time [125], [126].

Reinforcement learning problems can be modelled within the mathematical framework of the Markov decision problem (MDP) [127]. It defines how the agent and the environment interact in each time step. However, to use the MDP framework, the current state must be sufficiently representative to not depend on previous states or agent actions [128], [129]. For example, suppose we want to learn to fly a plane. In that case, the state of the plane can carry the trajectory and velocity instead of having the records of the previous historical positions. In this way, historical states are unnecessary in knowing which positions the plane was previously in. That is, the current state comprises a set of variables independent of the previous state and actions.

As mentioned above, the agent has to update its behavioural policy to learn which actions are best for each situation [123]. In general, there are three strategies for updating the agent's knowledge [122]:

- *Policy-based*: This strategy is based on creating a mapping between the state and actions.
- *Value functions*: This strategy calculates a prediction of how good the action or the state is. These evaluation functions will update based on the experience, and then the agent's behaviour improve to the optimal one. Examples: value interaction algorithm, Q-learning [130] and Sarsa [131].
- *Actor-Critic*: Combine the value-based and policy-based methods by using two different RL networks: The Actor uses a policy-based method to propose a set of possible actions given a state, and the Critic estimated value function, which evaluates actions taken by the Actor based on the given policy. The Actor then uses the feedback from the Critic to adjust its policy and make more informed decisions, leading to improved overall performance [132], [133].

A rapidly growing sub-field within reinforcement learning is Deep Reinforcement Learning, which leverages deep neural networks to learn the policy for RL problems [134], [135], [136], [137]. This approach has gained substantial traction in recent years and is now widely applied across various fields, including computational mechanics [138], chemical industry design processes, astronomy [122], urban water systems [139], and autonomous highway vehicle control [140]. Its success is largely attributed to its effectiveness in real-world scenarios that demand learning directly from experience. However, it also faces significant challenges, such as instability during training and difficulties in accurately defining the environment [139].

Reinforcement learning (RL) methods can be broadly categorized into two types: *model-based* and *model-free*

methods. This classification depends on whether the method uses a model of the environment. **Model-based** methods incorporate a model that simulates the environment's behaviour or allows predictions about how the environment will respond [141]. For example, given a state and an action, the model can predict the resulting next state and reward. These models are typically used for planning, where actions are selected by evaluating possible future scenarios before they are encountered. Well-known model-based methods include policy evaluation [142], Pegasus [143], and reinforcement learning for partially observable MDPs [144]. **Model-free methods**, in contrast, do not involve any model of the environment. Instead, they rely purely on trial-and-error learning, often considered the opposite of planning. Model-free systems are unable to predict how their environment will change in response to a single action; they learn directly from experience. Popular model-free methods include Monte Carlo [123], Temporal Difference Learning [145], Q-learning [146], SARSA [131], and Dyna-Q [147]. Both model-based and model-free approaches aim to find the most optimal policy, but they differ significantly in their strategies for learning and decision-making.

B. ARTIFICIAL INTELLIGENCE GENERATED CONTENT (AIGC)

Artificial Intelligence Generated Content (AIGC) creates high-quality and quick content, such as images, text [148], music [149], 3D models, and natural language, utilizing Generated Artificial Intelligence (GAI) algorithms to meet users' requirements. AIGC generates the content according to its knowledge, using the human prompt as a guide to complete the task. The most familiar AIGCs are ChatGPT for text creation [150], [151], [152] and DALL-E for creating unique and high-quality images from textual descriptions [153], [154], [155].

Nowadays, the quality of AIGC content is significantly better than before. A few years ago, the most common content creation strategies were Professional Generated Content (PGC) [156], [157] and User Generated Content (UGC) [158]. However, in recent years, the increasing volume of high-quality data and computational power and the development of large-scale pre-training models and new GAI models have propelled AIGC approaches [20]. That is, GAI is not a novel technique [159], [160]. Previous generative models, such as Restricted Boltzmann Machines [161], Deep Belief Networks [162], and Deep Boltzmann Machines [163], had limitations due to their lack of generalization power [164]. In contrast, since the development of Generative Adversarial Networks (GANs) in 2014 [165], new GAI algorithms have been developed, such as Transformer models [166], Generative Diffusion Models (GDM) [167], Nerf [168], or CLIP [169]. These new methods have gained attention for their ability to leverage the current high data volume and computational power to create more realistic, sophisticated generative models, enabling the creation of higher-quality content [19], [152]. However, new GAI methods are not the

only factors contributing to AIGC's success. Technological developments over the last few decades have increased computational power for training these models while enhancing the capacity to store high-quality data. Both factors are essential for the evolution of GAI methods as they are crucial for training the algorithms [19], [20]. Additionally, large pre-training models [21], such as BERT, GPT, etc., constitute another fundamental pillar of this evolution. A pre-trained model refers to a meticulously crafted and trained model or network developed on an extensive dataset to manage a similar problem. Instead of constructing a model from scratch, AI algorithms can use these pre-trained models as initial templates [170]. Transfer learning has inspired and formalized a two-phase learning framework for these algorithms: an initial phase involves pre-training to accumulate knowledge from one or more source tasks, followed by a fine-tuning stage where this acquired knowledge is transferred to target tasks [21]. Through this process, GAI can enhance its performance and generalization abilities by leveraging the knowledge embedded in pre-trained models. Large-scale pre-trained models offer distinct advantages to AIGC, including improved generalization abilities, reduced training costs, enhanced training efficiency, support for multiple tasks, and continuous optimization [20].

Another fundamental pillar in the evolution of GAI models was the integration of new techniques into them, such as the capability to learn from human feedback [171], [172], [173] to determine the most appropriate response for a given instruction, improving the model's reliability and accuracy over time. This is the case with "Chat Generative Pre-trained Transformer" or ChatGPT [151], [174], a language model developed by OpenAI for building conversational AI systems that can efficiently understand and respond to human language inputs meaningfully. This feedback integration allows ChatGPT to better comprehend human preferences in long dialogues [150]. By combining these advancements, models have made significant progress in AIGC tasks and have been adopted in various industries, including health [175], [176], [177], [178], art [179], and education [180]. In the near future, AIGC will continue to be a significant area of research in machine learning.

The literature exploring Generative Artificial Intelligence (GAI), Artificial Intelligence-Generated Content (AIGC), pre-training models, and Transfer Learning is vast and varied. However, this current study focuses on integrating users into the AI learning loop rather than delving into the intricacies of training and model development for collaboration. Therefore, we do not extensively delve into these technical aspects within the scope of this work, but interested readers can find in-depth explorations in forthcoming surveys [19], [20], [21], [170], [181]. This section exclusively delves into AIGC, their benefits and challenges, with the goal of offering a clear vision of these methods.

AIGC has gained significant popularity owing to its robust capabilities [152]. It is highly efficient and cost-effective and liberates human resources for more strategic tasks. For

instance, AIGC offers numerous advantages over traditional human writing, such as speed and language localisation [182]. This proficiency enables AI tools to generate vast amounts of content quickly. Additionally, AIGC excels at crafting personalized social media posts tailored for diverse platforms. Moreover, it is a valuable resource for writers grappling with creative blocks, offering inspiration, assistance, and refinement [183]. Furthermore, integrating AI-generated content into research endeavours enhances accuracy and efficiency, ultimately saving valuable time and resources.

However, despite this technology's success, this new technology still needs consequences and challenges. AIGC often lacks the emotional depth and authenticity that humans naturally possess, particularly in creative endeavours like music composition and writing. This absence of genuine human touch extends to the intended tone and personality, impacting the trust and mutual understanding between humans and AI [20]. Moreover, AIGC and Generative Artificial Intelligence (GAI) significantly rely on their training data. This dependence can lead to inaccuracies, primarily due to deviations and inaccurate information that hinder the ability of GAI to discern the credibility of sources or assign appropriate weights to different information channels [184]. Sensitivity in topics like race, gender, politics, and crucial decision-making in healthcare exacerbates this issue, making human oversight indispensable. Legal, moral, and ethical concerns also loom large, encompassing areas such as copyright infringement in AI-generated artworks [185], cheating and plagiarism in educational institutions, data privacy, security [186], and the malicious use of deepfakes [187]. To address these concerns, promoting the ethical development of AI and implementing appropriate laws to govern their use are imperative [152], [188]. Additionally, there is the issue of the digital divide, defined as the gap between those with access to computers and the internet and those without [189]. Emerging technologies like generative AI may inadvertently exacerbate this societal divide [190], [191]. To bridge this gap, fostering accessible AI and providing AI literacy training would prove invaluable.

Like any other technology, AIGC presents challenges and benefits to our society. A crucial factor in overcoming these challenges lies in integrating human feedback. As previously discussed, some technologies use interactive machine-learning techniques to incorporate human knowledge within their reasoning loops. Although AIGCs are not interactive, these methods could transform the integration of interactive elements and become the most widely used interactive machine-learning techniques in the coming years.

Besides knowing what techniques exist to interact and learn with the user, it is also necessary to understand how to integrate humans into a collaboration, which will be explained in section V.

V. HUMAN-AI COLLABORATION CLUSTER

This section outlines five clusters of human-AI collaboration based on various collaboration features and the degree of

human integration in AI decision-making. The classification is built upon the following collaboration features:

- *AI Technique (AL, RL, aML, AIGC)*: This feature describes the AI technique used in the collaboration. *AL* refers to Active Learning, *RL* denotes Reinforcement Learning, *aML* stands for automatic Machine Learning algorithms with interactive techniques, and *AIGC* represents Artificial Intelligence Generated Content techniques. We consider automatic machine learning algorithms that have been modified to include human intervention at specific stages of the workflow, treating them as interactive machine learning techniques.
 - *Expert*: This feature indicates whether a human expert is required for effective collaboration. For instance, certain complex tasks may necessitate expert knowledge, while others can be performed with non-expert human.
 - *Interaction (Interface, Physical)*: This feature describes the nature of the collaboration environment, which can take one of two forms: an *interface* interaction (e.g., software applications) or a *physical* interaction (e.g., collaborative robots working alongside humans).
 - *Aim of Collaboration (Cooperation, Improvement, Customization, Consensus, Replace AI)*: This feature categorizes the collaboration's objectives into five options:
 - *Cooperation*: Agents must continuously cooperate by jointly altering the environment to achieve the final goal of the collaboration. The aim of these works is to enhance communication, physical cooperation, or mutual understanding between the agents.
 - *Improvement*: The objective is to achieve a better result than that obtained by both users separately.
 - *Customization*: This collaboration aims to obtain improved outputs tailored to the user's specific needs.
 - *Consensus*: The goal is to reach a unanimous decision or action plan.
 - *Replace AI*: This category applies if the goal is to replace the AI reasoner with human intelligence.
- Multiple aims can be selected for one collaborative system, reflecting the complexity of human-AI interactions.
- *Team Composition (1:1 or 1:n)*: This feature describes the composition of the system's members: *1:1* indicates a system consisting of one AI and one human, while *1:n* denotes a system with one AI and several humans. This distinction is important for understanding dynamics in teamwork and decision-making.
 - *Initiative*: This feature describes whether the human agent can initiate communication with the system when desired. A lack of initiative means that human interaction is confined to specific workflow points of the system, limiting the collaboration's flexibility.

These collaboration features serve as the foundation for the clusters described in the following subsections.

TABLE 1. Some examples of Human-AI collaboration works in which the user is an optional workflow element used to improve or customize the AI output. This table highlighting the main common characteristics of these group. (□: No; ☒: Yes).

Paper - Topic	Techniques					Expert	Interaction		Aim of collaboration					AI-human		Initiative
	AL	RL	aML	AIGC	Interface		Physical	Customization	Improvement	Customization	Consensus	Replace AI	1:1	1:n		
Allen et al. [192] - Decision-making Tree	□	□	☒	□	□	☒	□	□	□	☒	□	□	□	☒	□	□
Cai et al. [193] - video recommendations	□	☒	□	□	□	☒	□	□	□	□	☒	□	□	☒	□	□
Dawar et al. [194] - Music recommendation	☒	□	□	□	□	☒	□	□	□	☒	□	□	□	☒	□	□
Han et al. [195] - Topic detection	□	□	☒	□	□	☒	□	□	□	☒	☒	□	□	☒	□	□
Honeycutt et al. [196] - Object detection	□	□	☒	□	□	☒	□	□	□	☒	□	□	□	☒	□	□
Keikhosrokiani et al.[197]- health product recommendation	☒	□	□	□	□	☒	□	□	□	☒	□	□	□	☒	□	□
Lee et al. [198] - Health	□	□	☒	□	☒	☒	□	□	□	☒	□	□	□	☒	□	□
Liu et al. [199] - Classification task	☒	□	□	□	□	☒	□	□	□	☒	□	□	□	□	☒	□
Murthy et al.[200] - video recommendations	□	☒	□	□	□	☒	□	□	□	☒	□	□	□	☒	□	□
Schroder et al. [201] - Vegetarian recipe	☒	□	□	□	□	☒	□	□	□	☒	□	□	□	☒	□	□
Smith et al. [52] - Topic modeling	□	□	☒	□	□	☒	□	□	□	☒	☒	□	□	☒	□	□
Yang et al.[202] - Text classification	□	□	☒	□	☒	☒	□	□	□	☒	□	□	□	☒	□	□
Yeh et al. [203] - Human-AI Writing	□	□	☒	☒	□	☒	□	□	□	☒	□	□	□	☒	□	□

A. HUMANS AS AN EXTERNAL TOOL IN THE COLLABORATION

Incorporating human input as an optional element to enhance or customize AI outputs is a defining characteristic of the works within this group. The goal of human intervention is to refine future AI outputs or tailor the results, with feedback consisting of a limited set of actions to modify the system’s outcomes. The collaborative workflow for this group is illustrated in Figure 4, where users are positioned downstream from the AI output. Table 1 summarizes several examples of collaborative works in this group, highlighting their shared characteristics.

Customization is the primary objective of these works. Dawar et al. developed a recommender system that considers music habits, preferences, and facial expressions to suggest suitable songs and playlists to users [194]. Users can provide feedback on these recommendations to improve future outputs. Even though face recognition occurs before the AI generates its output, the user’s involvement remains passive, as their cognitive abilities do not actively influence the result. Various recommendation systems are utilized in this group, with AL [197] and RL [204] being the most common techniques. Another example is the system created by Cai et al., which employs reinforcement learning (RL) for video recommendations [193]. In this system, users provide feedback, such as likes and follows, to define their preferences and express agreement with the recommendations. A similar approach is taken in the system developed by Murthy et al. [200]. These systems are part of our daily lives, embedded in platforms like Spotify, YouTube, and others, offering options to customize recommendations based on user preferences.

Han et al. and Smith et al. developed a system that identifies document topics based on word frequency [52], [195]. These topics can be modified and customized through user input to enhance system performance in subsequent iterations. Another illustrative example is provided by Schroder et al., who analyzed the Plant Jammer system [201]. This

system recommends vegetarian recipes based on the user’s selected ingredients and preferences, with the customization process relying on a rating scale from ‘great’ to ‘okay’ for the recipes. A similar approach is utilized by Allen et al., whose system learns users’ qualitative preferences to facilitate car selection [192]. In this case, user preferences are solicited, and the algorithm adapts to their decision-making process. When the model of the user’s preferences is presented, users can provide feedback to refine the recommendations. Another example is the work created by Yeh et al. [203], which developed an AI-enhanced writing tool that allows users to exercise greater agency and personalization in the writing process. The system offers suggestions, which users can accept or reject.

Moreover, system customization can also occur through expert human input. For instance, Lee et al. developed a feature selection system that provides quantitative analysis of patients’ conditions to enhance rehabilitation assessment practices [198]. This analysis is presented on a graphical user interface, allowing experts to give feedback on adjusting features to better fit patients’ needs and personalize rehabilitation.

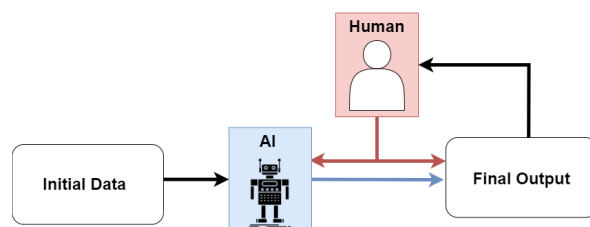


FIGURE 4. Illustration of the workflow in which users serve as an optional element to improve and customize the AI output.

Another goal of this level of integration is to leverage human qualities to improve areas where humans excel compared to AIs. For instance, Liu et al. utilized non-experts to label system classification outputs and identify *unknown-unknowns*, learning from them to enhance initial

classification [199]. In this case, a crowd-sourcing platform gathers non-expert labels for large volumes of data without overburdening individuals.

When the system seeks to employ human feedback to enhance outputs, a learning component is typically incorporated within the AI reasoning framework. However, not all works that integrate humans include this learning element. For example, Yang et al. developed the RulesLearner algorithm, which expresses ML text-analysis rules that experts may modify as needed [202]. The output of this modification becomes the final result. The importance of learning from human feedback was studied by Honeycutt et al., who developed two systems for detecting human faces in images: one with human feedback and one without [196]. The human feedback system utilized an iterative workflow in which a human was instructed to correct each AI output. However, the system did not learn from this feedback, even though users were led to believe it did. They aimed to examine how this lack of learning impacts user trust. This work highlighted the necessity of a learning component in iterative user-integrated loops to learn from feedback and maintain user trust.

Therefore, it is evident from the preceding examples that both human experts and non-experts can enhance system performance or customization. The critical aspect of these works is that a previous output is available, making user integration aimed at improving system performance or personalizing the experience optional. Users have become an increasingly integral component of systems in subsequent levels of integration, with expanded freedom of action, communication, and decision-making.

B. HUMANS-AI CONSENSUS

Works within this integration level aim to establish task allocation agreements and facilitate consensus decision-making among human participants. At this stage, users wield greater decision-making authority compared to previous levels. Typically, the team consists of one artificial intelligence agent and several humans, each of whom may possess unique information regarding the problem at hand. The primary objective is to reach a consensus decision or devise a strategic plan. Table 2 summarizes several examples of collaborative projects within this group, emphasizing their key shared characteristics.

One rationale for employing this workflow is that a human coordinator's decision-making can be swayed by personal motives or biases, whereas an artificial intelligence remains consistently objective. For instance, Pescetelli et al. compared the performance of human and AI coordination in discerning the crowd's intent in an online game [206]. They developed an online platform called BeeMe, where users could vote, socialize, and suggest new actions. The coordinator was responsible for interpreting the participants' intent to determine the next action an actor would take in the game, ensuring that this choice accurately reflected the users' intentions. A collaborative decision-making algorithm

was created that demonstrated a higher level of impartiality compared to a human coordinator.

This workflow is particularly useful in disaster scenarios, where the objectivity of the AI coordinator helps eliminate human bias and its potentially detrimental effects. Additionally, AI excels in efficiently resolving optimization challenges and processing large datasets, which can significantly enhance the effectiveness of interventions in such situations [206]. The initial efforts to incorporate an AI coordinator into human teams were based on shared control systems, which allowed humans to take control when the system encountered failures [208], [209]. However, in disaster contexts, a single lapse in planning can lead to catastrophic outcomes, rendering these approaches less effective.

Consequently, more recent research has focused on developing systems that facilitate communication and promote flexible interactions between humans and AI coordinators, rather than relying solely on shared control techniques. For example, Ramchurn et al. created a flexible human-AI interaction system for generating task allocation plans in emergent scenarios [207], [211]. This system empowers users to reject tasks assigned by the AI coordinator, request additional information, and articulate their preferences. Furthermore, the system can inquire about the status of tasks at any time, considering human preferences in the assignment process. As a result, a more adaptable plan emerges from the enhanced communication among team members.

In parallel, other studies advocate for incorporating a third component responsible for facilitating effective communication between humans and artificial agents [210]. Given the fundamental differences in processing between humans and AI, the primary aim of this component is to translate one model into the other and vice versa [212]. Zakershahra developed a collaboration system comprising one AI agent, four warehouse managers, and a facilitator agent to achieve a consensual plan in a disaster scenario [210]. The facilitator utilized the "Wizard of Oz" (WoZ) technique [213], a common method for comparing perceptions between human-human and human-agent teams. In this scenario, each participant holds different information about the situation, requiring the facilitator to assist team members in adapting their decision-making to forge a consensus, even when complete information is unavailable or direct communication is limited.

A crucial element for effective communication within teams is the ability of members to share their perspectives. The initiative serves as a communication tool that empowers users to express their views, enhancing their trust and enabling system customization. This initiative involves providing users with tools they can use flexibly, such as customizing the system, declining tasks, or proposing new actions without requiring approval. While the initiative can take various forms depending on the system, it consistently grants users autonomy and fosters a stronger relationship between the agent and the AI.

TABLE 2. Collaboration features of human-AI collaboration papers where the goal of the collaboration is to create a consensus decision. This table highlighting the main common characteristics of these group. (□: No; ☒: Yes).

Paper - Topic	Techniques				Expert	Interaction		Aim of collaboration					AI-human		Initiative
	AL	RL	aML	AIGC		Interface	Physical	Cooperation	Improvement	Customization	Consensus	Replace AI	1:1	1:n	
Li et al. [205] - AI assistants	□	□	□	☒	□	☒	□	□	□	☒	□	☒	☒	☒	
Pescetelli et al.[206] - Online game	□	□	☒	□	□	☒	□	□	□	☒	□	□	☒	☒	
Ramchurn et al. [207] -Task allocation	□	□	☒	□	☒	☒	□	□	□	☒	□	□	☒	☒	
Schurr et al. [208, 209] - Planning in disaster scenario	□	□	☒	□	□	☒	□	□	□	☒	□	□	☒	□	
Zakershahrahk at al. [210] - Planning in disaster scenario	□	□	☒	□	□	☒	□	□	□	☒	□	□	☒	☒	

For instance, in the BeeMe project developed by Pescetelli et al., users have the option to propose new actions that the system might not consider [206]. Similarly, in the approach taken by Ramchurn et al., users can request additional information and express their preferences [207]. Lin et al. designed an AI assistant named Decision-Oriented Dialogue (DoD) to aid one or multiple humans in making complex decisions across three scenarios: assigning reviewers to conference papers, planning a multi-step itinerary in a city, and negotiating travel plans for a group of friends [205]. In this case, user interaction with the AI occurs through a dialogue environment.

In essence, the most critical aspect of this integration level is the coordination objective of the system. Humans and AI do not necessarily need to work collaboratively on the same task; rather, they must focus on establishing a consensus, which can be viewed as the initial phase of the collaborative process.

C. HUMAN-AI ASYNCHRONOUS COLLABORATION

The integration of users as a fundamental component of AI systems distinguishes the works that comprise this integration level, aimed at enhancing or personalizing performance. Users are essential to the process—not only to correct the system but also to actively participate in obtaining the desired output. A defining characteristic of this group is that agents must engage without simultaneously altering the environment. This means human participation is often restricted to specific points in the interaction workflow, or the artificial agent operates in parallel with the user. As such, the collaboration is typically asynchronous. The goal of this collaboration is to work together to achieve improved results, rather than for one agent simply to assist another. However, the user’s actions are predetermined and must occur at specific moments within the interaction workflow. Table 3 summarizes examples of collaborative works within this group, highlighting their key common characteristics.

Consequently, human requirements, such as trust, engagement, and usability, are as crucial as artificial requirements in this integration process [234]. Moreover, a human-centered perspective is employed to incorporate these human requirements. In this framework, humans can assume various roles in the AI system’s learning process, as illustrated in Figure 5. For instance, in the study by Hekler et al., both the expert and the AI operate in parallel, independently classifying

images of skin cancer, with their classifications combined to arrive at a final diagnosis based on the accuracy of the AI’s predictions [216]. In this scenario, both agents exert equal effort to achieve the system’s output by performing the same task independently. Another health-related example is provided by Zhang et al., who developed a method for diagnosing sepsis [233]. Similar to the previous study, both experts and AI work in parallel, with the experts making the final decision.

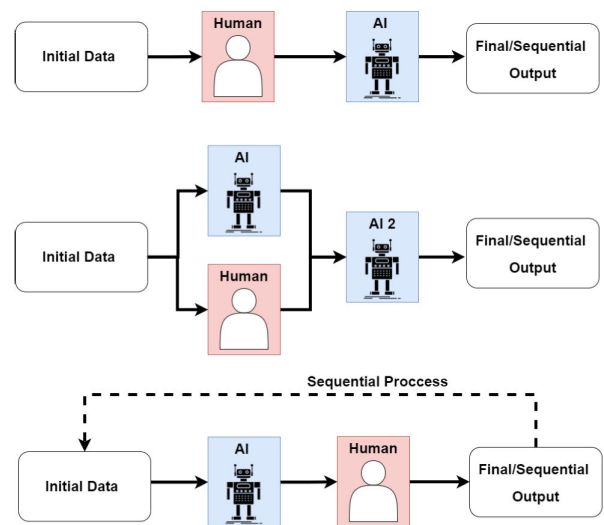


FIGURE 5. Three different human positions for this integration level. In the first workflow, the human position precedes the AI. In the second, humans and AI work in parallel. In the last, the user follows the AI.

A comparable methodology was employed by Reverberi et al., who conducted two distinct experiments to explore this collaboration [235]. In the first experiment, endoscopists diagnosed and classified lesions independently, while in the second, they received AI assistance. The study found that endoscopists were influenced by the AI’s recommendations, regardless of its accuracy. However, in cases of disagreement, the endoscopists maintained their original opinions without being swayed by the AI’s input. This exclusive reliance on AI as the sole decision-maker in the healthcare sector can lead to patient aversion [236]. Conversely, a collaborative approach—where AI assists in the decision-making process—allows us to leverage the

TABLE 3. Some examples of Human-AI collaboration works and their main collaboration characteristic where users are an essential part of the system and where the interaction is thought an interface. The goal of the collaboration is to work together to achieve better results rather than for one agent to assist the other. This table highlighting the main common characteristics of these group. (□: No; ☒: Yes).

Paper - Topic	Techniques				Expert	Interaction		Aim of collaboration					AI-human		Initiative
	AL	RL	aML	AIGC		Interface	Physical	Cooperation	Improvement	Customization	Consensus	Replace AI	1:1	1:n	
Gillotte et al. [185] - Generate artworks	□	□	□	☒	□	☒	□	□	☒	☒	□	□	☒	□	☒
Guimaraes et al. [214] - Similarity detection	□	□	☒	□	□	☒	□	□	☒	□	□	□	□	☒	□
Hasegawa et al. [215] - Audio descriptions of images	□	□	□	☒	□	☒	□	□	☒	☒	□	□	☒	□	☒
Hekler et al.[216] - Medicine, Health	□	□	☒	□	☒	☒	□	□	☒	□	□	□	☒	□	□
Holzinger et al.- [217] - Optimization algorithm	□	□	☒	□	□	☒	□	□	☒	□	□	□	☒	□	□
Huang et al. [218] - Creation of audio content	□	□	□	☒	□	☒	□	□	☒	☒	□	□	☒	□	☒
Kobis et al. [219] - Poems text-generation	□	□	□	☒	□	☒	□	□	☒	☒	□	□	☒	□	☒
Liu et al. [220] - Audio storytelling generated	□	□	□	☒	□	☒	□	□	☒	☒	□	□	☒	□	☒
Liu et al. [221] - Magnetograms of the Sun	□	□	□	☒	□	☒	□	□	☒	☒	□	□	☒	□	☒
Liu et al. [222] - Re-identification tasks	☒	☒	□	□	□	☒	□	□	☒	□	□	□	☒	□	□
Loeschcke et al. [223] - Video generated	□	□	□	☒	□	☒	□	□	☒	☒	□	□	☒	□	☒
Mozafari et al. [224] - Classification tasks	☒	□	□	□	□	☒	□	□	☒	☒	□	□	☒	□	☒
Pollefeys et al. [225] - 2D images into 3D models	□	□	□	☒	□	☒	□	□	☒	☒	□	□	☒	□	☒
Schmitt et al. [226] - Human-ai co-writing	□	□	□	☒	□	☒	□	□	☒	□	□	□	☒	□	☒
Schulman et al. [151] - ChatGPT	□	□	□	☒	□	☒	□	□	☒	☒	□	□	☒	□	☒
Sharma et al. [227] - Text assistant	□	□	□	☒	□	☒	□	□	☒	☒	□	□	☒	□	☒
Tegen et al. [228] - Virtual Sensors	☒	□	□	□	□	☒	□	□	☒	☒	□	□	☒	□	☒
Texler et al. [229] - Video generated	□	□	□	☒	□	☒	□	□	☒	☒	□	□	☒	□	☒
Thoppilan et al. [230, 231] - Question-answering chatbot	□	□	□	☒	□	☒	□	□	☒	☒	□	□	☒	□	☒
Whittaker et al. [187] - Create synthetic faces	□	□	□	☒	□	☒	□	□	☒	☒	□	□	☒	□	☒
Yang et al. [232] - Human-AI Fiction Co-writing	□	□	□	☒	□	☒	□	□	☒	☒	□	□	☒	□	☒
Zhang et al. [233] - Sepsis Diagnosis	□	□	☒	□	☒	☒	□	□	☒	☒	□	□	☒	☒	☒
Zhang et al.[188] - Cartographic design processes	□	□	□	☒	□	☒	□	□	☒	☒	□	□	☒	□	☒

capabilities of AI in healthcare, thereby fostering greater acceptance and integration of AI in this domain [237].

In this system workflow, the human position may precede the AI reasoner, as demonstrated in the study by Guimaraes et al. [214], which focuses on identifying document pair similarity. In this case, users utilize a crowdsourcing platform to group documents by their authors, after which the AI classifies them based on content similarity. However, users are often positioned after the AI to minimize human effort while maximizing the AI’s data processing capabilities. Figure 5 illustrates these different workflows at this integration level.

Common techniques within this integration level include Active Learning and the modification of automatic machine learning algorithms. Another notable technique gaining traction is AI-Generated Content (AIGC), which warrants a separate subsection due to its growing significance and unique method of integrating users.

Crowdsourcing platforms are also widely used since AI algorithms can learn from collective human behavior, enhancing accuracy over time. This approach helps distribute the human labeling effort among multiple users, as a single user typically cannot label an entire dataset. Some researchers investigate ways to integrate machine learning into crowdsourced databases, aiming to combine the accuracy of human labeling with the speed and cost-effectiveness of machine-learning classifiers. In this vein, Mozafari et al. studied the application of active learning strategies in crowdsourced databases to reduce the number of instances needing labeling

by humans, thereby minimizing human effort and enabling the scalability of this approach [224].

This challenge is relevant across various contexts. For example, Liu et al. focused on reducing human annotation efforts while maximizing system performance in person re-identification tasks. They developed a deep reinforcement active learning (DRAL) method to guide an agent in selecting training samples in real-time with human annotators.

Another compelling reason to explore ways to reduce human labeling efforts is the susceptibility of humans to mental state influences, such as stress and boredom. In this context, Tegen et al. demonstrated that designing applications where users provide feedback only when the system’s output is incorrect is more efficient than requiring users to provide correct labels consistently [228]. Their research indicates that human performance declines when users are forced to exert excessive effort over prolonged periods. In some experiments, they allowed users to actively tag without prompts, empowering users with initiative, which increases freedom in collaboration and significantly contributes to meeting human requirements. However, this flexibility is not present in all studies.

For instance, Holzinger et al. addressed the Traveling Salesman Problem using a snake game to facilitate interaction between the user and the AI [217]. They modified an Ant Colony Optimization (ACO) algorithm, specifically the MMAS (MAX-MIN Ant System), allowing users to select the next target for the snake during interactions. This choice affects the AI’s behavior, enabling humans to correct

deficiencies in the algorithm. However, human interactions are time-restricted, which may not align with the optimal timing for adjustments; addressing issues may be more critical in subsequent iterations or earlier phases.

1) ARTIFICIAL INTELLIGENT GENERATED CONTENT

One notable subset of this group is the AIGC (Artificial Intelligent Generated Content) algorithms. These algorithms have gained significant popularity in recent years and have revolutionized various fields and applications, such as medicine [44] and tourism [238]. As mentioned earlier, the performance of these methods is quite straightforward: the user provides prompt input, and the AI generates content that uses this human input as a guide. This interaction, however, is not inherently interactive. Currently, we are at the forefront of developing this technology, with interactive methods increasingly being integrated into generative AI (GAI) models. For example, ChatGPT has integrated human feedback to enhance performance [151], [174]. Moreover, some AIGC systems allow for the repetition of this input-output loop, occasionally incorporating memory as a dialogue feature, which enables improvement and customization based on multiple iterations and/or specifications. These algorithms belong to this integration level because the scope of human input is primarily limited to the type of prompt that the AIGC accepts. However, human input is not restricted in other respects. Additionally, human intervention is crucial for collaboration, as users are the most beneficial contributors.

Although these methods have recently gained traction and transformed the way many tasks are performed, particularly in education [239], [240], [241], they also hold potential influence over the academic peer review process [242].

The diverse forms of AI-generated content can be classified into three categories: text, image, and audio generation:

a: TEXT GENERATION

Creative writing and dialogue writing are the primary subfields of AI-generated text [148], [243].

- *Creative Writing*: This involves generating text with greater openness, creativity, and nuance. For instance, in the work of Yang et al., users can co-write a short sci-fi story with a GPT-2-based text generation model [232]. Another application includes generating poems, as demonstrated by Kobis et al. in their research [219]. Additionally, Sharma et al. developed a tool for recommending co-written text messages for mental health support [227].
- *Dialogue Writing*: This category encompasses chatbots that interact with users through text. These bots are designed to answer questions and provide information. Such AIGC algorithms are considered interactive due to the inherent nature of dialogue [244], [245]. For instance, ChatGPT is designed for conversational usage, producing human-like responses by drawing on its extensive knowledge base [174], [246]. ChatGPT

can remember previous interactions within the same conversation, thereby facilitating a continuous dialogue. Notably, LaMDA is a question-answering chatbot focused on two domains: Mount Everest education and music content recommendations [230], [231]. Another example is the software developed by Schmitt et al., which supports the creation of fictional characters through interactive conversation and gradual development [226]. AI-generated text remains the most expansive category of AIGC, partly due to the influence and success of ChatGPT.

b: IMAGE GENERATION

By leveraging AIGC, users can modify existing images or generate new ones that meet specific requirements based on given prompts [247], [248], all without needing advanced skills or knowledge. AI-generated images have various applications, ranging from artwork [185] to synthetic faces [187] and complex magnetograms of the Sun [221]. DALL-E, a widely used AIGC application, generates images from textual descriptions [150], [154], [155], [249]. Zhang et al. have developed a novel application that integrates cartographic design processes based on DALL-E [188]. Additionally, some applications can translate 2D images into 3D models [225]. Further examples can be found in the survey conducted by Zhang et al. [250]. These AI-generated images can also be used to create videos. AIGCs operate similarly in video generation, processing each frame individually and utilizing AI algorithms to generate video clips, which can be employed to create trailers and promotional videos [223], [229].

c: AUDIO GENERATION

AIGCs can also generate audio, which falls into two primary categories: text-to-speech synthesis and voice cloning [251], [252].

- *Text-to-Speech Synthesis*: This process converts input text into speech that mimics a specific speaker's voice, commonly used in robotics and voice broadcasting applications [253].
- *Voice Cloning*: This technique takes specific target speech as input and transforms it to match the speech patterns of the designated speaker. Some applications automatically generate audio descriptions of images [215]. For instance, WavJourney, developed by Liu et al., creates structured scripts for audio storytelling from text descriptions of auditory scenes [220]. Another innovative application, AudioGPT, created by Huang et al., excels in generating diverse audio content, including speech, music, sound effects, and talking head tasks [218].

Despite having a restricted role within this workflow, users remain essential. As observed, the human position is not static; instead, it depends on the objectives of collaboration and the balance of design effort within the system. Users'

TABLE 4. Some examples of Human-AI collaboration works where user and AI agents interact in a physical environment. This table highlighting the main common characteristics of these group. (□: No; ☒: Yes).

Paper - Topic	Techniques				Expert	Interaction		Aim of collaboration					AI-human		Initiative
	AL	RL	aML	AIGC		Interface	Physical	Cooperation	Improvement	Customization	Consensus	Replace AI	1:1	1:n	
Cao et al. [254] - Robot plan assistant	□	☒	□	□	□	☒	□	☒	□	□	□	□	☒	□	☒
Gómez-Carmona et al. [255] - Smart drink monitoring	☒	□	□	□	□	☒	☒	☒	□	□	□	□	☒	□	☒
Holzinger et al. [256] - Optimization algorithm	□	□	☒	□	□	☒	□	☒	□	□	□	□	☒	□	□
Li et al. [257] - Emergency Indoor Patrolling	□	□	☒	□	□	□	☒	☒	□	□	□	□	☒	□	☒
Lou et al. [258] - Game	□	☒	□	□	□	☒	□	☒	□	□	□	□	☒	□	☒
Mccamish at al. [259, 260] - Mutual understanding	□	☒	□	□	□	☒	□	☒	□	□	□	□	☒	□	☒
Mehak et al. [261] - Human-AI assemble Task	□	□	☒	□	☒	☒	□	☒	□	□	□	□	☒	□	☒
Nikolaidis et al. [262] - Collaboration task	□	☒	□	□	□	□	☒	☒	□	☒	□	□	☒	□	☒
Nikolaidis et al. [263] - Collaboration task	□	☒	□	□	□	□	☒	☒	□	☒	□	□	☒	□	☒
Schelble et al. [264] - Game; RL	□	☒	□	□	□	☒	□	☒	□	□	□	□	☒	□	□
Strouse et al. [265] - Game	□	☒	□	□	□	☒	□	☒	□	□	□	□	☒	□	☒
Tao et al. [266] - Collaboration task	□	☒	□	□	□	□	☒	☒	□	□	□	□	☒	□	☒
Wang et al. [267] - Human-AI coordination	□	□	☒	□	□	☒	□	☒	□	□	□	□	☒	□	☒
Xing et al. [268] - Leadership transition	□	□	☒	□	□	☒	□	☒	□	□	□	□	☒	□	☒
Xu et al. [269] - Game	□	□	☒	□	□	☒	□	☒	□	☒	□	□	☒	□	☒
Yu et al. [270] - Game	□	☒	□	□	□	☒	□	☒	□	□	□	□	☒	□	☒
Zhang et al. [271] - Collaborative tasks	□	□	☒	☒	□	☒	□	☒	□	□	□	□	☒	□	☒
Zhou et al. [272] - Tracking a trajectory	□	□	☒	□	□	□	☒	☒	□	□	□	□	☒	□	☒

flexibility and decision-making power tend to increase in subsequent integration levels.

D. HUMAN-AI COLLABORATION MAKING CHANGES AT THE SAME TIME IN THE ENVIRONMENT

This integration level shares the same objectives as the previous one; however, in this stage, agents have greater freedom to modify the environment through their actions, leading to more equitable collaboration, as illustrated in Figure 6. This strategy affords users enhanced flexibility. Furthermore, humans possess the autonomy to initiate actions without requiring explicit instructions. This results in a more balanced distribution of roles between agents, as the activities performed by humans closely resemble those executable by AI. Consequently, there is an increased freedom for humans and a greater capacity for decision-making, which enhances the level of human integration. To foster collaboration, all participants must be aware of each other’s actions to inform their subsequent responses, thus increasing overall engagement and awareness. Table 4 summarizes examples of collaborative projects in this group, highlighting the key characteristics shared among them.

One notable example of this integration level is the algorithm proposed by Holzinger et al. They introduced a solution for optimizing the Traveling Salesman Problem [256] using the behavior of the Ant Colony Optimization algorithm [273]. In this study, the authors provided users with two innovative tools for modifying the ants’ behavior: a human interaction matrix that enables control over the ants’ movements between nodes and a human impact factor that allows users to define the probability that an ant will follow a human suggestion. This empowers users to act autonomously within the system’s environment, as they can adjust the matrix and factors at their discretion.

Gómez-Carmona et al. explored how to develop a system that empowers users to take initiative [255]. They designed an intelligent drink monitoring system linked to a customizable interface. This system records all user interactions with the smart bottle to analyze hydration patterns. Users are asked to label their actions to ensure the system accurately tracks drinking gestures. To improve accuracy, the system allows users to record examples of their specific drinking movements at any time, augmenting the training dataset. Following an action, users can also verify whether the gesture was correctly classified. This two-way communication enhances system performance, while the customization options improve user comfort during the learning process and regulate their effort.

Cao et al. developed a strategy for robot actions to assist humans using the Theory of Mind (ToM) and reinforcement learning. In this approach, the robot observes user behavior and communicates with them to facilitate task completion more effectively [254].

The primary goal of these works is to achieve cooperation between agents in real-world environments, rather than simply creating a plan or reaching a consensus. Another example of this objective is the work by Zhou et al., who developed a human-robot Cartesian co-manipulation task that navigates random obstacles, promoting mutual communication and accommodation in unstructured environments [272]. Due to the unpredictable barriers, robots face challenges in managing environmental uncertainty, and the task requires tracking specified trajectories while avoiding unexpected obstacles. This system maintains a high level of performance while minimizing cognitive load on users.

A similar study by Xu et al. builds upon previous research on human-human collaboration [269], [274]. They

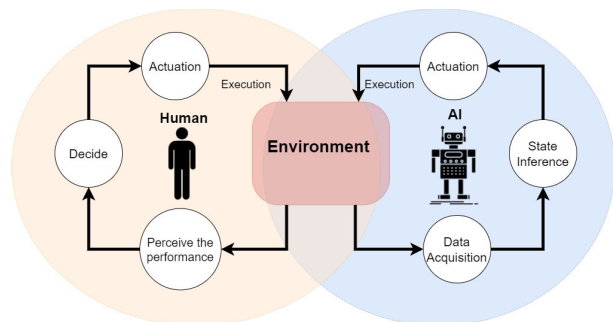


FIGURE 6. A depiction of the integration level of humans in human-AI collaboration. Both humans and AI are free to modify the environment and must take action based on each other's contributions.

developed a Waiter Agent Interactive Training Experimental Restaurant (WAITER) robot that learns to collaborate with human trainers to provide customer service in a virtual restaurant. In this context, both agents (the robot and human) access different data and must communicate effectively to complete the task. Through this dialogue, agents can adapt their behaviors to align with those of their counterparts, employing various machine learning algorithms to facilitate this adaptability, including Bayesian networks and linear prediction.

However, collaboration members sometimes possess disparate information about the environment. For instance, Tao et al. developed a physical collaborative task involving a human and a robot, wherein both participants shared the same information [266]. In their experiment, a human and a robot worked together to manipulate a ball on a platform, aiming to place it in the center hole without letting it fall out. While both agents could observe the environment, they had to learn to anticipate each other's intentions and coordinate their actions without direct communication.

When defining human-robot collaboration, it is crucial for AI to be trained using human data to understand how to respond to various situations. However, this training can impose a significant burden on humans. One method to streamline this process and reduce the human workload involves employing reinforcement learning (RL) strategies. For example, Strouse et al. explored how to train robots for effective collaboration with human partners without relying on human data [265]. They utilized a self-play strategy [275] to replace the human effort during training. Their study employed a cooking simulator game called *Overcooked*, where both agents work together to prepare recipes within a set time. Coordination is critical, as both agents share information and can select their subsequent actions. They utilized a V-MPO [276], an RL algorithm featuring a ResNet plus LSTM architecture, to facilitate this coordination. Another example of human-AI collaboration within the *Overcooked* framework is the work by Zhang et al. [271], which developed a chat system enabling users and AI to discuss strategies for resolving the game.

Furthermore, human corrections to robot actions can demand substantial mental and physical resources [34],

[277]. Consequently, some researchers focus on reducing human involvement while enhancing the robot's capability to adapt its role in collaboration. For instance, Wang et al. developed a dynamic role-shifting system for human-robot collaboration in assembly tasks, aiming to alleviate human fatigue [267]. Similarly, Xing et al. devised a fuzzy logic-based arbitration rule to regulate the transition of the robot's role in collaboration [268]. Another approach by Mehak et al. involves creating a system capable of recognizing human actions and predicting forthcoming actions to foster improved team coordination and communication [261]. In addition to addressing human fatigue, maintaining the trust of human agents is critical in collaboration. For example, Li et al. developed a human-AI collaboration system for indoor patrolling during sudden power outages, focusing on preserving human trust in the agent [257].

Moreover, Yu et al. employed self-play strategies to mitigate human biases in decision-making [270]. They introduced the Hidden-Utility Self-Play (HSP) method, which explicitly models human biases as hidden reward functions. Similarly, Lou et al. utilized self-play strategies in conjunction with a policy ensemble method to enhance partner diversity within the population. Their context-aware approach enables agents to analyze and identify potential policy primitives of their partners, allowing for appropriate action selection [258].

The *Overcooked* game has been used for various research purposes, including studying human collaboration across different task loads in human-AI environments [278], [279]. Despite its popularity, other games have also been employed in similar loops, utilizing RL techniques. The combination of gaming and RL is advantageous due to its inherent qualities [280], [281]. Therefore, it is crucial to examine how the selection of RL and game theory can foster varying levels of cooperation within hybrid teams. Schelble et al. conducted an empirical study to explore how different modern RL algorithms and game theory scenarios could influence cooperation levels in human-AI teams [264]. Their findings highlight the importance of selecting appropriate RL techniques to enhance team effectiveness and composition.

Additionally, Nikolaidis et al. designed a physical collaborative task in which a human and a robot must move a table through a door [262]. Initially, both agents must decide on their actions. This study emphasizes mutual adaptation, where each member must modify their behavior based on the actions of the other [282]. They developed a Bounded-Memory Adaptation Model (BAM), an RL algorithm, alongside a Mixed Observability Markov Decision Process (MOMDP) policy to manage adaptability functions. In subsequent work, they improved mutual adaptability to preserve human trust, although the most efficient task resolution was not achieved [263].

In addition to mutual adaptability, other researchers focus on fostering mutual understanding. For example, Mccamish et al. aim to establish a mutual comprehension of human intent through queries [259], [260]. Using an RL method, they strive to create a shared understanding of human intent [260].

TABLE 5. Some examples of collaborative works where human replacing the AI reasoner component. This table highlighting the main common characteristics of these group. (□: No; ☒: Yes).

Paper - Topic	Techniques				Expert	Interaction		Aim of collaboration					AI-human		Initiative
	AL	RL	aML	AIGC		Interface	Physical	Cooperation	Improvement	Customization	Consensus	Replace AI	1:1	1:n	
Amershi et al.[283] - Build a ML algorithm	□	□	☒	□	□	☒	□	□	□	□	☒	☒	□	□	
Ankerst et al. [284] - Build Decision Tree	□	□	☒	□	□	☒	□	□	□	□	☒	☒	□	□	
Kartoun et al. [285] - Task classification	☒	□	□	□	□	☒	□	□	□	□	☒	☒	□	□	
Talbot et al. [286] - Build ensemble classifier	□	□	☒	□	□	☒	□	□	□	□	☒	☒	□	□	
Wiethof et al. [287] - Customer service	□	□	☒	□	□	☒	□	□	□	□	☒	☒	□	□	

In this context, a query serves as a vital communication element between users and the database system, facilitating mutual understanding. Both humans and AI are rewarded based on their agreement, positioning them as equals with their respective methodologies for understanding and reasoning.

These examples illustrate that, in contrast to earlier groups, the degree of human freedom is significantly greater in this integration level. While all members share the collaboration’s goals, they actively cooperate to achieve them. The next level of integration involves utilizing humans as the reasoning element within the collaboration.

E. HUMANS REPLACING THE ARTIFICIAL REASONER OF THE SYSTEM

In this collaboration group, users act as the primary reasoning element of the system, effectively replacing the AI’s reasoning capabilities through an interface that equips them with the necessary tools to solve problems. The design of this tool varies depending on the specific problem, and the interface is automatically updated when users modify the data. This decision-update process is iterative, enabling users to engage in their own reasoning work. In this group, humans are the sole members with decision-making authority, creating an unbalanced dynamic within the system. Table 5 summarizes several examples of collaborative projects within this group, highlighting their key shared characteristics.

For instance, Kartoun et al. developed a classification system for clinical narrative notes to determine patients’ smoking status [285]. In this project, the AI pre-processed the data by removing all non-smoking narratives and then utilized a random sample to allow humans to evaluate the selected instances.

Another example is provided by Wiethof et al., who created an interface that alerts customer service employees to tasks that may cause delays or be overlooked, thereby improving work efficiency [287]. This agent learns from users’ behaviors to provide timely alerts, but ultimate decision-making authority remains with the human agents. Both expert and non-expert customer service employees evaluate the agent’s recommendations.

In addition, Talbot et al. engaged users in constructing an ensemble classifier, which combines several classifiers to create one that outperforms its individual components [286].

To facilitate this process, they designed a visually interactive interface using the confusion matrix technique, enabling users to explore and combine individual classifiers effectively.

Similarly, Ankerst et al. developed a system that allows non-expert users to create a decision tree [284]. They employed a pixel-oriented technique known as Square Segments [288] for intuitive visualization of high-dimensional data, which enhances users’ understanding of the classifications and the data involved. Additionally, they designed a system where users can build a machine-learning algorithm through a visual interface [283]. In this case, users have the ability to label data, edit and select features, and view various statistical metrics, such as the confusion matrix, to assess model performance. While this approach enhances users’ comprehension of the output and the underlying data, it is important to note that human capacity to detect patterns in large datasets is limited compared to artificial intelligence [38], [236]. This limitation underscores the necessity of achieving the right balance in collaboration, where each participant contributes according to their strengths.

Overall, this integration level is characterized by positioning humans as the reasoning element within the collaboration, granting them the highest level of decision-making authority compared to all other levels discussed in this work. Furthermore, an interface facilitates communication among members at this integration level. This framework emphasizes the importance of human agency in collaborative systems, highlighting that effective collaboration requires not only advanced tools but also a clear understanding of the roles and responsibilities of each participant.

VI. DESIGN PRINCIPLES

This section defines the essential design principles that must be considered during the design phase of a human-AI collaboration system. The design principles that are not directly relevant to promoting such collaborations are excluded. These principles are described based on their impact on collaboration within the system. It is important to note that these principles are part of a broader set of considerations for effective, sustainable, and ethically sound technology design [186], [289]. However, the focus of this work is solely on those principles that are vital for establishing mutually beneficial collaboration where

all participants feel comfortable, integrated, and actively engaged in the process.

Based on a review of the literature, the following design principles have been identified: Adaptation & Customization, Trust & Explainability, Engagement, and Communication & Feedback. A detailed description of each principle is provided, emphasizing their relevance to the integration of human participants. Moreover, practical examples illustrating the application of each principle in real-world scenarios are included to enhance understanding and implementation.

A. ADAPTATION & CUSTOMIZATION

Adaptability is the capacity to alter behavior based on an assessment of the environment. In a collaborative context, this involves modifying the initial approach according to the actions of other agents and the capabilities of each team member. Consequently, adaptability encompasses communication, perception, and an interactive process [290]. According to Li et al., the success of a collaborative system depends on the members' ability to work together and adjust their strategies to achieve mutually beneficial outcomes while completing tasks [291]. This flexibility is essential for members to adapt to one another in a complementary manner. To demonstrate this, Li et al. developed an adaptive agent capable of identifying and selecting the most complementary policy based on the actions observed from the human partner. Their findings indicate that performance improved with the adaptive approach compared to non-adaptive agents. This experiment supports the notion that agents must complement each other by adapting their behaviors to enhance collaboration.

Furthermore, adaptability is essential for collaboration as it enhances human performance, reduces fatigue, and increases the effectiveness of outputs [34], [277]. This assertion was evaluated by Buehler et al., who explored the relationship between human reciprocal cooperation and its correlation with collaborative performance [292]. The results suggest that the cooperative behaviors of an automated agent can significantly influence human cooperativeness. Participants demonstrated greater resource sharing and improved performance when working with a high-cooperation agent compared to a low-cooperation agent. However, it is crucial for the automated agent to calibrate cooperation within the system to avoid inefficient use of resources.

The adaptation of one agent depends on the adaptability of the other. Many authors postulate that adaptation should occur for both human and AI agents [262]. A common approach is the mutual adaptation technique, which asserts that both agents must continuously adjust their behavior based on each other's actions [293]. Nikolaidis et al. focus on achieving mutual adaptation without compromising user trust, aiming to balance task execution with computational accuracy and the maintenance of user trust [262], [263]. Other authors offer different perspectives on this concept. For instance, Xu et al. examine adaptation from the agent's

viewpoint, suggesting that mutual adaptation could enable the agent to manage flexible mappings between human instructions and agent actions [269]. They developed formal conditions and demonstrated that effective collaboration requires enabling the mutual adaptation phenomenon. Key elements include the continuous adaptation of each agent to the actions of the other and the necessity for agents to express their internal states. Both experiments, along with their respective developments, have proven superior to one-way adaptation strategies while maintaining user confidence in the progress of collaboration. This aligns with the work of Zhou et al., which establishes mutual adaptation in unstructured environments that promote mutual communication and adaptation [272].

Despite these considerations, effective collaboration between humans and machines requires not only behavioral changes but also the accuracy of those changes. This knowledge is acquired through the co-learning process inherent in their interactions [294]. Consequently, an approach based on the co-learning process is essential for fostering effective human-AI collaboration. In this context, Van Zoelen et al. examine co-learning patterns in human-robot collaboration [295]. They assert that the initial step of co-learning is co-adaptation, defined as partners adjusting their actions to align with both the task and each other. This research identifies a list of recurring interaction patterns applicable to stable situations as well as those requiring sudden adaptation (changes in behavior occurring over short periods). In stable situations, patterns include actively coordinating actions or working while waiting for a team member. In contrast, sudden adaptation situations may involve confusion regarding the agent's behavior or strategy, attempts to communicate, or feelings of isolation without providing assistance to the agent. Alternatively, Tahboub et al. present a different perspective, advocating that co-adaptation and co-learning occur simultaneously during collaboration [280]. They propose and demonstrate that reinforcement learning (RL) strategies, particularly those utilizing policy gradients, effectively fulfill the requirements of co-adaptation due to their capacity for iterative learning from interactions in dynamic environments.

To facilitate mutual adaptation, it is essential to study humans' ability to adapt rather than solely focusing on the benefits of achieving mutual adaptation. Although several studies have concentrated on how robots should adapt during collaboration [296], a new paradigm is emerging that seeks to understand how human adaptation evolves through interactive collaboration [293]. The primary objective is to learn how this process occurs to incorporate this knowledge into the artificial agent, resulting in more adaptive behavior during collaboration.

For instance, in addition to developing a navigation collaboration system in which each agent possesses partial information, Mohammad et al. investigate the evolution of human adaptation in collaborative settings [297]. Their findings demonstrate that the human adaptation rate is not

fixed; it begins with a non-adaptive phase, during which users attempt to discover the robot's capabilities. Following this initial phase, users enter an adaptation phase, which is influenced by their perception of the robot's ability to complete the task. This second stage is particularly evident among users with high expectations. Consequently, the time required for users to adjust their behavior is contingent upon their perception of the robot's capabilities. Similarly, Nikolaidis et al. develop a robot model that accounts for human perceptions of the robot's capabilities, allowing the robot to choose between actions that reveal its abilities to the user and actions that are optimal based on the available information [282]. This model adapts its actions based on the understanding that user adaptation is influenced by perceptions of the robot's capabilities. Their research indicates that considering the human perspective enhances human-AI collaboration compared to policies that assume the human will fully adapt to the robot. In another collaboration paradigm, Van Zoelen et al. study how humans adapt to continuous leadership shifts in human-robot collaboration and how this influences their trust and understanding [298]. They demonstrate that, over time, humans develop a greater appreciation for collaboration when they assume less of a leadership position. This decision is influenced by the human's perception and confidence in the robot's capacity to complete the assignment.

Customization is a fundamental aspect of human-AI collaboration, enabling the system to adapt to user preferences and requirements. This personalization can be achieved through various techniques, including user feedback, user modeling, and reinforcement learning, among others [299]. For example, ChatGPT utilizes prior user input to generate more customized responses [151], [174], [300], [301]. The level of customization varies according to the context, goals, and learning techniques employed. Personalization has been extensively explored in the development of recommender systems [302], [303], [304]. However, predicting user acceptance of novel recommendations remains a significant challenge. Additionally, customization in recommender systems may raise concerns regarding user privacy [305]. Balancing diversity, user engagement, and privacy constraints is therefore essential. Techniques proposed by Kelly et al. focus on enhancing exploration without deviating from the system's core objectives [306]. In this context, recommender systems can adjust aspects such as suggestion timing, the number of recommendations, and other factors to maintain user engagement while ensuring a balance between exploring new preferences and adhering to the system's goals [307]. This adaptive strategy helps ensure that recommendations align closely with user preferences and system objectives, ultimately enhancing the overall user experience.

B. TRUST & EXPLICABILITY

Understanding the role of trust in engaging humans in collaboration, as well as how trust is influenced during

interactions, is crucial for establishing and promoting effective and productive collaborative relationships between humans and AI [308], [309]. This understanding is vital throughout all phases of collaboration, beginning with the design of the system and continuing through its ongoing interactions with the user. For instance, Yang et al. indicate that the level of human involvement in the machine learning (ML) development process significantly impacts their trust in the models [56]. Additionally, in the interactive collaboration process, initial interactions play a critical role, as demonstrated by Tolmeijer et al. [310]. Their findings suggest that if the initial interactions between humans and these systems fall below expectations, it can negatively affect user adoption. Trust, however, is not a static concept; it evolves over the course of interactions. In this process, it is essential for the system to maintain human trust without diminishing it [263], [282] or exceeding it beyond reasonable limits throughout the collaboration.

For these reasons, human trust is a fundamental factor in effective collaboration, particularly when assessing the appropriate level of trust required by users in the decision-making process [311], [312]. A lack of user trust can hinder their willingness to participate, negatively affecting the adoption of intelligent systems [313]. For instance, Yang et al. analyze how users perceive different tracking devices, noting that perceptions of accuracy can vary depending on individual expectations and personal circumstances [94]. Furthermore, Zerilli et al. highlight that a user's limited understanding of how these devices function, including data collection and measurement calculations, can significantly influence their perception of accuracy [314]. Therefore, human-AI systems should prioritize fostering effective and transparent collaboration and communication.

The concept of explicability and the emphasis on interactive processes stand in contrast to the "black box" nature that AI and machine learning (ML) systems may exhibit when human knowledge is not integrated into their problem-solving processes [212], [315], [316]. This lack of transparency has significantly influenced AI development by increasing human aversion and diminishing user trust [317], [318]. Providing explanations enables users to gain a clearer understanding of the machine's behavior, identify potential limitations and errors in its reasoning patterns, establish a sense of control, and modulate their level of trust in the technology [27], [32]. Empirical findings from Yang et al. suggest that incorporating interactive processes can lead to greater confidence among non-experts in learning outcomes, particularly when explainability is prioritized [56]. This indicates that when users engage with ML systems, it is crucial for them to comprehend how the system reaches its conclusions in order to trust its decision-making processes and provide accurate feedback [32]. Enhancing explicability without compromising computational performance presents a significant challenge for the development of AI-generated content (AIGC) technologies [27].

Achieving an appropriate level of explicability in AI systems is also essential. Overly detailed explanations can distract users and may negatively impact their trust in the system. Conversely, the inability to understand a given explanation can lead users to lose trust in the system prematurely and unjustifiably [196]. Therefore, finding the right balance between user engagement, the amount of explanation provided, and the necessary human effort is crucial to maintaining user trust within appropriate boundaries [319].

Additionally, it is essential for communication tools within interactive systems to establish a bi-directional communication channel and to consider how these tools are presented to users [32]. In this context, McCallum et al. investigated the role of user-driven feature representation in systems, revealing that users often have a limited understanding of the impact of their contributions to the model [80]. Similarly, Amershi et al. examined the potential advantages of collaborative approaches that allow users to assess the quality of the model and decide on further inputs, enhancing both user experience and the learning system itself [86]. They concluded that further research into new interaction methods is necessary to determine whether user influence on the model can indeed lead to more effective systems. Meza et al. also emphasized the importance of evaluating different forms of interaction and their impact on user trust, highlighting the need for a deeper understanding of these dynamics at this stage [320].

In summary, the interactive nature of collaborative solutions contributes to building trust in AI systems by increasing user involvement in the decision-making process and enhancing their understanding of how AI functions [321]. Overall, factors such as user perception of accuracy, system learning progress, explainability, the level of user integration, and the interactive nature of collaboration significantly influence user trust in these systems. These aspects should be carefully considered in the development of future collaborative technologies.

C. ENGAGEMENT

Lowering the barriers that hinder effective interaction between humans and technology is a primary challenge in promoting collaboration between them [322]. Indeed, the collaboration between humans and AI involves active participation to create a joint environment where they work together to achieve a common goal. However, motivating users to collaborate over an extended period can be challenging and requires considerable involvement. In addition**,** there may not always be a perceived value produced. For instance, in an experimental evaluation by Masson et al. on user involvement with AI systems, participants initially took an active role in data capture [323]. Still, eventually, most of them quit due to the significant time required, poor accuracy, and perceived lack of reward. In this line, Ramos et al. delved into user willingness to interact with a deployed system and determined that it may vary with context, task, and individual characteristics [324].

These concepts of involvement, interest, and attention relate to the idea of engagement, which is crucial in human-agent interaction as it is usually a prerequisite for the objective of the interaction [325]. Hence, engagement with AI systems is vital for comprehending the user's role and, consequently, enriching this relationship and enhancing user experience [86]. In this context, user engagement refers to the degree of involvement, interaction, and emotional connection a user has with an AI system or technology [326]. From the AI agent's perspective, engagement can be viewed as the goal of maintaining user interactions. From the human perspective, engagement goes beyond just the functional use of the system; it refers to the user's active participation and willingness to continue using the system [327]. Oertel et al. reviewed previous efforts addressing the relevance of engagement in human-AI interaction from both perspectives [328]. According to them, knowing the level of user engagement can be advantageous for customizing agent behavior and gauging the quality of interaction and user satisfaction with the system.

At the same time, the perception of relevance is also pivotal for promoting user engagement, as users need to feel that their inputs contribute to the system's value [64]. In this regard, reducing the effort to interpret outputs and express feedback about the potential outcomes of the provided inputs may enhance user engagement. For this reason, it is essential to maintain a good trade-off between the efforts made and the perceived value of such efforts. Therefore, in the context of human-AI collaboration, user engagement also encompasses the level of communication and transparency perceived from the AI system [329]. These conclusions highlight the need for increased collaboration between AI/ML and HCI design research fields. In this line, other scholars have evaluated such engagement practices to reduce the effort needed to cooperate with the system. Those efforts aim to support the iterative design, development, and dissemination processes of learning systems oriented toward conversational AI agents [330] or to define a human-centered thinking approach to applying IML methods [82].

In essence, user engagement is critical in ensuring that the collaboration between humans and AI is effective, efficient, and beneficial for both parties. An effective system needs to actively engage users in the task being performed to motivate them to achieve the desired outcomes of the human-AI collaboration.

D. COMMUNICATION & FEEDBACK

The more effectively agents comprehend each other's capabilities, limitations, inputs, outputs, and context, the more efficiently they can collaborate to achieve their shared objectives [331], [332]. As Marathe et al. concluded, AI systems must be capable of understanding the state and intent of their human co-workers; simultaneously, humans must clearly articulate their intentions [333]. Furthermore, it is crucial for humans to recognize the limitations of AI; otherwise,

they may develop unrealistic expectations regarding AI's capabilities, which can negatively impact their engagement, adaptation, and trust [334], [335].

Effective collaboration necessitates bi-directional communication to establish a mutual understanding that enables the participants to perceive each other through natural, fluid interactions [336], [337]. In a formal sense, communication refers to a reciprocal exchange in which team members share messages to coordinate ideas, norms, and strategies to achieve their collective goals [338], [339].

Several scholars have explored strategies to foster this bi-directional communication in human-AI systems. For example, Holder et al. provide an overview of best practices and challenges in establishing bi-directional transparency at various stages of the human-AI-robot systems (HARTs) lifecycle [331], [340]. Similarly, Koop et al. apply principles of human-to-human communication as the foundation for developing effective human-AI interactions [341]. They argue that agents should engage in cooperative and incremental dialogue construction with human users, supported by AI's coordination mechanisms. They advocate for the next phase in human-AI communication research to focus on developing agents with the capacity to intertwine incremental metacognitive processes with socio-communicative behaviors, enabling them to co-create seamless interactions and rectify misunderstandings, misinterpretations, and dialogue interruptions.

Building on this concept, McCamish et al. have introduced a collaborative communication approach that implements a formal framework for representing and understanding information needs in database querying, aiming to develop a mutual language for representing these needs [259], [260]. This mutual language allows users to articulate their intentions precisely when querying databases and simultaneously provides feedback mechanisms that help users understand whether their intentions align with the AI's responses. Their findings indicate that users tend not to explore alternative methods once they discover a reliable way to express their intentions. Moreover, Hanna et al. developed a communication model called Speech Act Theory (SAT), which explores how utterances can facilitate actions in collaborative dialogues [342]. This model also examines the influence of human perceptions of AI speech acts on collaborative performance, showing that a positive human perspective is associated with the AI's ability to articulate its intentions in a comprehensible manner, thereby promoting mutual communication.

When addressing communication scenarios, the agent's initiative, or proactive communication capabilities, is also a critical factor. Zoelen et al. developed proactive communication techniques to enable more fluid human-AI collaboration by using context and adaptability to guide communication strategies [343]. They demonstrated that technology trained to communicate proactively outperforms those without such capabilities, especially in dynamic environments like disaster scenarios or military decision-making, where adaptability

and re-planning are essential [344]. Ramchurn et al. further exemplified this concept by showing that proactive agents empower users to decline tasks, request information, and communicate preferences, while also allowing the AI to inquire about task statuses as needed [207]. This dynamic enables both parties to initiate communication when appropriate, fostering a deeper understanding of each other's perspectives.

Chatbots, as conversational agents, have evolved significantly in recent years, enhancing their ability to simulate human interactions through text, voice, and visual communication [27], [345]. These advancements have led to personalized and engaging conversational experiences that can provide on-demand health interventions, among other services [346]. As AI technologies continue to progress, we can anticipate even greater improvements in chatbot capabilities, enabling them to handle increasingly complex tasks and offer more valuable user assistance [227].

Feedback is a crucial element in human-AI collaboration, serving as the primary mechanism through which each agent communicates its internal state in response to its partner's actions. Selecting the appropriate feedback type for the human collaborator is essential for fostering co-adaptation and co-learning within the partnership [347]. Knox et al. illustrated the significance of robot feedback through their work on TAMER, a robot that learns behaviors based on human numerical feedback in navigation tasks [348]. Other studies have examined the role of feedback in dialogue, emphasizing the importance of providing users with clear indications of progress, which helps to meet their expectations during collaborative interactions [349]. These studies identify several feedback states—ready, processing, reporting, busy-no-response, and busy-delayed-response—that AI systems can simulate to manage communication effectively.

Finally, it is important to consider how feedback influences an AI's behavior relative to its current contextual knowledge versus the information provided by the human collaborator. Misinterpretations in feedback can arise due to the user's personal circumstances rather than the AI's actual performance. Therefore, effective communication strategies are vital for resolving these misunderstandings and correctly interpreting the feedback to enhance the collaborative process.

VII. DISCUSSION

The rapid advancement of technology has propelled AI progress far beyond our wildest imaginations. We now possess the necessary components to shape a future where human-AI collaboration becomes remarkably feasible. However, this evolution demands thoughtful consideration. Humans must be integral, active participants with decision-making authority to ensure responsible integration. Consequently, this collaboration cannot be approached haphazardly. Careful consideration of essential human factors is vital to our collaborative approach. These factors underpin the

design, emphasizing the need for meticulous and mindful implementation of this alliance.

Traditional research in various facets of collaboration has often occurred in isolated silos, leading to inherent biases in adaptive system design and user evaluation. These biases emerge from overlooking the impact of the system's communication style and interaction methods on user comfort. A narrow focus on adaptation alone creates an incomplete picture; users assess not only the system's adaptability but also its communication style, proactive engagement, and integration within the broader system context. Adopting a global perspective is paramount for a comprehensive understanding of human-AI systems.

This review paper aspires to foster collaborative development across distinct research domains, exploring their interconnections and influence on human users. New inquiries will emerge, delving into the selection of interaction types and AI system adaptability based on user integration levels. A joint view will generate fresh research queries, shedding light on how each collaboration facet impacts user integration and engagement. This approach allows us to probe how the interplay among diverse collaboration components shapes the user experience. Additionally, this understanding will guide future advancements, informing decisions regarding user integration and engagement strategies while maintaining awareness of their interrelated nature.

It is crucial to recognize that the influence of these collaborative advancements extends beyond the immediate characteristics of collaboration. Enhanced user comfort with these systems will foster increased crowd-sourcing and participation in computer science research, empowering users to engage more actively.

This unified vision can serve as a guiding principle across diverse disciplines, encouraging active user involvement in collaborations. Such an approach facilitates the harmonious merging of AI and human knowledge, fostering a comfortable, trustworthy, and enduring partnership. By embracing this perspective, we pave the way for a future where human-AI collaboration transcends boundaries, creating seamless synergy that empowers both parties and fuels innovation across various domains.

VIII. CONCLUSION

In this work, we have explored the different roles and levels that lead to the integration of humans in the learning and decision loop of intelligent systems. In addition, we presented several design principles that must be considered in the design phase to create a balanced collaboration. This work focuses exclusively on these principles, which are vital to establishing a mutually advantageous collaboration where all participants experience a sense of comfort, integration, and active engagement in the collaborative process: trust, engagement, communication, feedback, penalization, and adaptation. All these factors significantly impact human perception and willingness to participate in collaborative efforts.

Furthermore, we have generated our insights based on a comprehensive review of research fields related to human-AI collaboration, aiming to create a holistic view of this concept and the essential (human and system) requirements. Based on the reviewed works, future implementations of intelligent systems must integrate interactive, inclusive, engaging, adaptive, bidirectional communication, and trust mechanisms to promote equal and seamless collaboration. To achieve this, future interactive intelligent systems should be designed in accordance with a global vision of all aspects of human-AI collaboration. With this work, we aim to contribute to establishing the foundation for a more human-centered vision of collaboration between humans and AI, making it more effective, efficient, and beneficial for both parties.

ACKNOWLEDGMENT

During the preparation of this work the authors used ChatGPT in order to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and took full responsibility for the publication's content.

REFERENCES

- [1] E. L. Zheng, W. Jin, G. Hamarneh, and S. S.-J. Lee, "From human-in-the-loop to human-in-power," *Amer. J. Bioethics*, vol. 24, no. 9, pp. 84–86, Sep. 2024.
- [2] F. Shi, W. Wang, H. Wang, and H. Ning, "The internet of people: A survey and tutorial," 2021, *arXiv:2104.04079*.
- [3] K. Stephens, A. Harris, A. Hughes, C. Montagnolo, K. Nader, S. A. Stevens, T. Tasuji, Y. Xu, H. Purohit, and C. Zobel, "Human-AI teaming during an ongoing disaster: How scripts around training and feedback reveal this is a form of human-machine communication," *Hum.-Mach. Commun.*, vol. 6, pp. 65–85, Jul. 2023.
- [4] M. E. Frisse, "Ubiquitous computing," *Academic Med.*, vol. 67, no. 10, pp. 642–644, 1992.
- [5] D. Cook and S. K. Das, *Smart Environments: Technology, Protocols, and Applications*, vol. 43. Hoboken, NJ, USA: Wiley, 2004.
- [6] A. Khang, S. K. Gupta, S. Rani, and D. A. Karras, *Smart Cities: IoT Technologies, Big Data Solutions, Cloud Platforms, and Cybersecurity Techniques*. Boca Raton, FL, USA: CRC Press, 2023.
- [7] J. C. Augusto, H. Nakashima, and H. Aghajan, "Ambient intelligence and smart environments: A state of the art," in *Handbook of Ambient Intelligence and Smart Environments*, H. Nakashima, H. Aghajan, and J. C. Augusto, Eds. Boston, MA, USA: Springer, 2010, doi: [10.1007/978-0-387-93808-0](https://doi.org/10.1007/978-0-387-93808-0).
- [8] J. C. Augusto, V. Callaghan, D. Cook, A. Kameas, and I. Satoh, "Intelligent environments: A manifesto," *Hum.-Centric Comput. Inf. Sci.*, vol. 3, no. 1, pp. 1–18, Dec. 2013.
- [9] R. Salama, F. Al-Turjman, M. Aeri, and S. P. Yadav, "Internet of Intelligent Things (IoT)—An overview," in *Proc. Int. Conf. Comput. Intell., Commun. Technol. Netw. (CICITN)*, Apr. 2023, pp. 801–805.
- [10] H. Li, Y. Wang, and H. Qu, "Where are we so far? Understanding data storytelling tools from the perspective of human-AI collaboration," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2024, pp. 1–19.
- [11] Z. Li, "A design trajectory map of human-AI collaborative reinforcement learning systems: Survey and taxonomy," 2024, *arXiv:2405.10214*.
- [12] J. Sherson, J. Rafner, and S. Büyükgüzel, "Operational criteria of hybrid intelligence for generative AI virtual assistants," in *HAAI 2024: Hybrid Human AI Systems for the Social Good*. Amsterdam, The Netherlands: IOS Press, 2024, pp. 475–477.
- [13] T. Hanika, M. Herde, J. Kuhn, J. Marco Leimeister, P. Lukowicz, S. Oeste-Reiß, A. Schmidt, B. Sick, G. Stumme, S. Tomforde, and K. Anna Zweig, "Collaborative interactive learning—A clarification of terms and a differentiation from other research fields," 2019, *arXiv:1905.07264*.

- [14] Z. Akata, D. Balliet, M. De Rijke, F. Dignum, V. Dignum, G. Eiben, A. Fokkens, D. Grossi, K. Hindriks, and H. Hoos, "A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence," *Computer*, vol. 53, no. 8, pp. 18–28, Aug. 2020.
- [15] F. Shi, F. Zhou, H. Liu, L. Chen, and H. Ning, "Survey and tutorial on hybrid human-artificial intelligence," *Tsinghua Sci. Technol.*, vol. 28, no. 3, pp. 486–499, Jun. 2023.
- [16] L. Petri, "Concept analysis of interdisciplinary collaboration," *Nursing Forum*, vol. 45, no. 2, pp. 73–82, Apr. 2010.
- [17] A. Ajoudani, A. M. Zanchettin, S. Ivaldi, A. Albu-Schäffer, K. Kosuge, and O. Khatib, "Progress and prospects of the human-robot collaboration," *Auto. Robots*, vol. 42, no. 5, pp. 957–975, Jun. 2018.
- [18] D. Dellermann, P. Ebel, M. Söllner, and J. M. Leimeister, "Hybrid intelligence," *Bus. Inf. Syst. Eng.*, vol. 61, no. 5, pp. 637–643, 2019.
- [19] Y. Cao, S. Li, Y. Liu, Z. Yan, Y. Dai, P. S. Yu, and L. Sun, "A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT," 2023, *arXiv:2303.04226*.
- [20] J. Wu, W. Gan, Z. Chen, S. Wan, and H. Lin, "AI-generated content (AIGC): A survey," 2023, *arXiv:2304.06632*.
- [21] X. Han, Z. Zhang, N. Ding, Y. Gu, X. Liu, Y. Huo, J. Qiu, Y. Yao, A. Zhang, and L. Zhang, "Pre-trained models: Past, present and future," *AI Open*, vol. 2, pp. 225–250, Jan. 2021.
- [22] K. van den Bosch, T. Schoonderwoerd, R. Blankendaal, and M. A. Neerincx, "Six challenges for human-AI co-learning," in *Proc. Int. Conf. Hum.-Comput. Interact.* Cham, Switzerland: Springer, Jan. 2019, pp. 572–589.
- [23] L. Hofeditz, M. Mirbabaie, and M. Ortmann, "Ethical challenges for human-agent interaction in virtual collaboration at work," *Int. J. Hum.-Comput. Interact.*, vol. 40, no. 23, pp. 8229–8245, Dec. 2023.
- [24] L. Vicente and H. Matute, "The inherited bias effect: The propagation of artificial intelligence biases to human decisions," *Tech. Rep.*, 2023.
- [25] R. W. Andrews, J. M. Lilly, D. Srivastava, and K. M. Feigh, "The role of shared mental models in human-AI teams: A theoretical review," *Theor. Issues Ergonom. Sci.*, vol. 24, no. 2, pp. 129–175, Mar. 2023.
- [26] N. C. Krämer, A. von der Pütten, and S. C. Eimler, "Human-agent and human-robot interaction theory: Similarities to and differences from human-human interaction," in *Human-Computer Interaction: Agency Perspective* (Studies in Computational Intelligence), vol. 396, M. Zaccarias and J. V. de Oliveira, Eds. Berlin, Germany: Springer, 2012, doi: 10.1007/978-3-642-25691-2_9.
- [27] W. Saeed and C. Omlin, "Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities," *Knowl.-Based Syst.*, vol. 263, Mar. 2023, Art. no. 110273.
- [28] E. Jussupow, I. Benbasat, and A. Heinzl, "Why are we averse towards algorithms? a comprehensive literature review on algorithm aversion," in *Proc. ECIS*, Marrakech, Morocco, 2020.
- [29] J. S. Andersen and W. Maalej, "Design patterns for machine learning-based systems with humans in the loop," *IEEE Softw.*, vol. 41, no. 4, pp. 151–159, Jul. 2024.
- [30] E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán, and Á. Fernández-Leal, "Human-in-the-loop machine learning: A state of the art," *Artif. Intell. Rev.*, vol. 56, no. 4, pp. 3005–3054, Apr. 2023.
- [31] B. Shneiderman, *Human-Centered AI*. London, U.K.: Oxford Univ. Press, 2022.
- [32] S. Teso, Ö. Alkan, W. Stammer, and E. Daly, "Leveraging explanations in interactive machine learning: An overview," *Frontiers Artif. Intell.*, vol. 6, Feb. 2023, Art. no. 1066049.
- [33] A. Baratta, A. Cimino, M. G. Gnani, and F. Longo, "Human robot collaboration in Industry 4.0: A literature review," *Proc. Comput. Sci.*, vol. 217, pp. 1887–1895, Jan. 2023.
- [34] Y. Liu, G. Caldwell, M. Rittenbruch, M. Belek Fialho Teixeira, A. Burden, and M. Guertler, "What affects human decision making in human-robot collaboration: A scoping review," *Robotics*, vol. 13, no. 2, p. 30, Feb. 2024.
- [35] T. Miletic, "Human-artificial intelligence symbiosis: The possibility of moral augmentation," Ph.D. dissertation, Dept. Philosophy, Univ. Rijeka. Fac. Humanities Social Sci., Rijeka, Croatia, 2021.
- [36] J. Inga, M. Ruess, J. H. Robens, T. Nelius, S. Rothfuß, S. Kille, P. Dahlinger, A. Lindenmann, R. Thomaschke, G. Neumann, S. Matthiesen, S. Hohmann, and A. Kiesel, "Human-machine symbiosis: A multivariate perspective for physically coupled human-machine systems," *Int. J. Hum.-Comput. Stud.*, vol. 170, Feb. 2023, Art. no. 102926.
- [37] R. Zhen, W. Song, Q. He, J. Cao, L. Shi, and J. Luo, "Human-computer interaction system: A survey of talking-head generation," *Electronics*, vol. 12, no. 1, p. 218, Jan. 2023.
- [38] M. Jirgl, Z. Bradac, and P. Fiedler, "Human-in-the-loop issue in context of the cyber-physical systems," *IFAC-PapersOnLine*, vol. 51, no. 6, pp. 225–230, 2018.
- [39] M. M. M. Peeters, J. van Diggelen, K. van den Bosch, A. Bronkhorst, M. A. Neerincx, J. M. Schraagen, and S. Raaijmakers, "Hybrid collective intelligence in a human-AI society," *AI Soc.*, vol. 36, no. 1, pp. 217–238, Mar. 2021.
- [40] A. M. Annaswamy, P. P. Khargonekar, and S. K. Spurgeon, *Cyber-Physical-Human Systems: Fundamentals and Applications*. Hoboken, NJ, USA: Wiley, 2023.
- [41] S. F. Pileggi, "Ontology in hybrid intelligence: A concise literature review," *Future Internet*, vol. 16, no. 8, p. 268, Jul. 2024.
- [42] D. Dellermann, A. Calma, N. Lipusch, T. Weber, S. Weigel, and P. Ebel, "The future of human-AI collaboration: A taxonomy of design knowledge for hybrid intelligence systems," 2021, *arXiv:2105.03354*.
- [43] S. Wang, "Generative AI: An in-depth exploration of methods, uses, and challenges," *Highlights Sci., Eng. Technol.*, vol. 85, pp. 196–202, Mar. 2024.
- [44] L. Shao, B. Chen, Z. Zhang, Z. Zhang, and X. Chen, "Artificial intelligence generated content (AIGC) in medicine: A narrative review," *Math. Biosci. Eng.*, vol. 21, no. 1, pp. 1672–1711, 2024.
- [45] W. Xu and Z. Gao, "Applying HCAI in developing effective human-AI teaming: A perspective from human-AI joint cognitive systems," *Interactions*, vol. 31, no. 1, pp. 32–37, Jan. 2024.
- [46] J. B. Schmutz, N. Outland, S. Kerstan, E. Georganta, and A.-S. Ulfert, "AI-teaming: Redefining collaboration in the digital era," *Current Opinion Psychol.*, vol. 58, Aug. 2024, Art. no. 101837.
- [47] S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, "Security, privacy and trust in Internet of Things: The road ahead," *Comput. Netw.*, vol. 76, pp. 146–164, Jan. 2015.
- [48] S. Robert, S. Büttner, C. Röcker, and A. Holzinger, "Reasoning under uncertainty: Towards collaborative interactive machine learning," in *Machine Learning for Health Informatics*. Cham, Switzerland: Springer, 2016, pp. 357–376.
- [49] X. Wu, L. Xiao, Y. Sun, J. Zhang, T. Ma, and L. He, "A survey of human-in-the-loop for machine learning," *Future Gener. Comput. Syst.*, vol. 135, pp. 364–381, Oct. 2022.
- [50] M. Ware, E. Frank, G. Holmes, M. Hall, and I. H. Witten, "Interactive machine learning: Letting users build classifiers," *Int. J. Hum.-Comput. Stud.*, vol. 55, no. 3, pp. 281–292, Sep. 2001.
- [51] J. A. Fails and D. R. Olsen, "Interactive machine learning," in *Proc. 8th Int. Conf. Intell. User Interface*, Jan. 2003, pp. 39–45.
- [52] A. Smith, V. Kumar, J. Boyd-Graber, K. Seppi, and L. Findlater, "Closing the loop: User-centered design and evaluation of a human-in-the-loop topic modeling system," in *Proc. 23rd Int. Conf. Intell. User Interface*, Mar. 2018, pp. 293–304.
- [53] A. Holzinger, "Interactive machine learning for health informatics: When do we need the human-in-the-loop?" *Brain Informat.*, vol. 3, no. 2, pp. 119–131, Jun. 2016.
- [54] M. Maadi, H. A. Khorshidi, and U. Aickelin, "A review on human-AI interaction in machine learning and insights for medical applications," *Int. J. Environ. Res. Public Health*, vol. 18, no. 4, p. 2121, Feb. 2021.
- [55] S. Budd, E. C. Robinson, and B. Kainz, "A survey on active learning and human-in-the-loop deep learning for medical image analysis," *Med. Image Anal.*, vol. 71, Jul. 2021, Art. no. 102062.
- [56] Q. Yang, J. Suh, N.-C. Chen, and G. Ramos, "Grounding interactive machine learning tool design in how non-experts actually build models," in *Proc. Designing Interact. Syst. Conf.*, Jun. 2018, pp. 573–584.
- [57] S. Berg, D. Kutra, T. Kroeger, C. N. Straehle, B. X. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, and M. Rudy, "Ilastik: Interactive machine learning for (bio)image analysis," *Nature Methods*, vol. 16, no. 12, pp. 1226–1232, 2019.
- [58] N. Andrienko, G. Andrienko, L. Adilova, and S. Wrobel, "Visual analytics for human-centered machine learning," *IEEE Comput. Graph. Appl.*, vol. 42, no. 1, pp. 123–133, Jan. 2022.
- [59] I. Krak, O. Barmak, and E. Manziuk, "Using visual analytics to develop human and machine-centric models: A review of approaches and proposed information technology," *Comput. Intell.*, vol. 38, no. 3, pp. 921–946, Jun. 2022.
- [60] A. Holzinger and I. Jurisica, "Knowledge discovery and data mining in biomedical informatics: The future is in integrative, interactive machine learning solutions," in *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*. Cham, Switzerland: Springer, 2014, pp. 1–18.

- [61] N.-C. Chen, J. Suh, J. Verwey, G. Ramos, S. Drucker, and P. Simard, "AnchorViz: Facilitating classifier error discovery through interactive semantic data exploration," in *Proc. 23rd Int. Conf. Intell. User Interface*, Mar. 2018, pp. 269–280.
- [62] L. Jiang, S. Liu, and C. Chen, "Recent research advances on interactive machine learning," *J. Visualizat.*, vol. 22, no. 2, pp. 401–417, Apr. 2019.
- [63] R. Porter, J. Theiler, and D. Hush, "Interactive machine learning in data exploitation," *Comput. Sci. Eng.*, vol. 15, no. 5, pp. 12–20, Sep. 2013.
- [64] J. J. Dudley and P. O. Kristensson, "A review of user interface design for interactive machine learning," *ACM Trans. Interact. Intell. Syst.*, vol. 8, no. 2, pp. 1–37, Jun. 2018.
- [65] M. Gillies, R. Fiebrink, A. Tanaka, J. Garcia, F. Bevilacqua, A. Héloir, F. Nunnari, W. E. Mackay, S. Amershi, B. Lee, N. d'Alessandro, J. Tilmanne, T. Kulesza, and B. Caramiaux, "Human-centred machine learning," in *Proc. CHI Conf. Extended Abstr. Hum. Factors Comput. Syst.*, May 2016, pp. 3558–3565.
- [66] A. Nikitin and S. Kaski, "Human-in-the-loop large-scale predictive maintenance of workstations," in *Proc. 28th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2022, pp. 3682–3690.
- [67] A. M. Chirkin and R. König, "Concept of interactive machine learning in urban design problems," in *Proc. SEACHI Smart Cities Better Living With HCI UX*, May 2016, pp. 10–13.
- [68] R. Ghani and M. Kumar, "Interactive learning for efficiently detecting errors in insurance claims," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2011, pp. 325–333.
- [69] S.-W. Huang, P.-F. Tu, W.-T. Fu, and M. Amanzadeh, "Leveraging the crowd to improve feature-sentiment analysis of user reviews," in *Proc. Int. Conf. Intell. User Interface*, Mar. 2013, pp. 3–14.
- [70] H. Shang, C. B. Sivaparthipan, and ThanjaiVadivel, "Interactive teaching using human-machine interaction for higher education systems," *Comput. Electr. Eng.*, vol. 100, May 2022, Art. no. 107811.
- [71] M. Chignell, M.-H. Chung, Y. Yang, G. Cento, and A. Raman, "Human factors in interactive machine learning: A cybersecurity case study," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, Sep. 2021, vol. 65, no. 1, pp. 1495–1499.
- [72] A. Holzinger, *Machine Learning for Health Informatics*. Cham, Switzerland: Springer, 2016.
- [73] J. Waring, C. Lindvall, and R. Umeton, "Automated machine learning: Review of the state-of-the-art and opportunities for healthcare," *Artif. Intell. Med.*, vol. 104, Apr. 2020, Art. no. 101822.
- [74] Y. Y. M. Aung, D. C. S. Wong, and D. S. W. Ting, "The promise of artificial intelligence: A review of the opportunities and challenges of artificial intelligence in healthcare," *Brit. Med. Bull.*, vol. 139, no. 1, pp. 4–15, Sep. 2021.
- [75] A. Holzinger, "On knowledge discovery and interactive intelligent visualization of biomedical data," in *Proc. Int. Conf. Data Technol. Appl. DATA*, 2012, pp. 5–16.
- [76] Y. Sun, E. Lank, and M. Terry, "Label-and-learn: Visualizing the likelihood of machine learning classifier's success during data labeling," in *Proc. 22nd Int. Conf. Intell. User Interface*, Mar. 2017, pp. 523–534.
- [77] N. Schwalbe and B. Wahl, "Artificial intelligence and the future of global health," *Lancet*, vol. 395, no. 10236, pp. 1579–1586, May 2020.
- [78] F. Sperrle, M. El-Assady, G. Guo, R. Borgo, D. H. Chau, A. Endert, and D. Keim, "A survey of human-centered evaluations in human-centered machine learning," *Comput. Graph. Forum*, vol. 40, no. 3, pp. 543–568, Jun. 2021.
- [79] S. Stumpf, V. Rajaram, L. Li, M. Burnett, T. Dietterich, E. Sullivan, R. Drummond, and J. Herlocker, "Toward harnessing user feedback for machine learning," in *Proc. 12th Int. Conf. Intell. User Interface*, Jan. 2007, pp. 82–91.
- [80] L. McCallum and R. Fiebrink, "Supporting feature engineering in end-user machine learning," in *Proc. CHI Workshop Emerg. Perspect. Hum.-Centered Mach. Learn.* Glasgow, U.K., May 2019.
- [81] K. W. Mathewson and P. M. Pilarski, "A brief guide to designing and evaluating human-centered interactive machine learning," 2022, [arXiv:2204.09622](https://arxiv.org/abs/2204.09622).
- [82] K. W. Mathewson, "A human-centered approach to interactive machine learning," 2019, [arXiv:1905.06289](https://arxiv.org/abs/1905.06289).
- [83] V. Vishwarupe, S. Maheshwari, A. Deshmukh, S. Mhaisalkar, P. M. Joshi, and N. Mathias, "Bringing humans at the epicenter of artificial intelligence: A confluence of AI, HCI and human centered computing," *Proc. Comput. Sci.*, vol. 204, pp. 914–921, Jan. 2022.
- [84] C. M. Cutillo, K. R. Sharma, L. Foschini, S. Kundu, M. Mackintosh, K. D. Mandl, T. Beck, E. Collier, C. Colvis, K. Gersing, V. Gordon, R. Jensen, B. Shabestari, and N. Southall, "Machine intelligence in healthcare—Perspectives on trustworthiness, explainability, usability, and transparency," *npj Digit. Med.*, vol. 3, no. 1, pp. 1–5, Mar. 2020.
- [85] A. van der Stappen and M. Funk, "Towards guidelines for designing human-in-the-loop machine training interfaces," in *Proc. 26th Int. Conf. Intell. User Interface*, Apr. 2021, pp. 514–519.
- [86] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza, "Power to the people: The role of humans in interactive machine learning," *AI Mag.*, vol. 35, no. 4, pp. 105–120, Dec. 2014.
- [87] K. Inkpen, S. Chancellor, M. De Choudhury, M. Veale, and E. P. S. Baumer, "Where is the human: Bridging the gap between AI and HCI," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, May 2019, pp. 1–9.
- [88] A. Tegen, P. Davidsson, and J. A. Persson, *The Effects of Reluctant and Fallible Users in Interactive Online Machine Learning*. Malmö, Sweden: Malmö Universitet, 2022, p. 209.
- [89] F. Bernardo, M. Zbyszynski, R. Fiebrink, and M. Grierson, "Interactive machine learning for end-user innovation," in *Proc. AAAI Spring Symp. Ser.*, Dec. 2016, pp. 1–7.
- [90] A. Tegen, P. Davidsson, and J. A. Persson, "Activity recognition through interactive machine learning in a dynamic sensor setting," *Pers. Ubiquitous Comput.*, vol. 28, no. 1, pp. 273–286, Feb. 2024.
- [91] C.-Y. Lin and R. Marculescu, "Model personalization for human activity recognition," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops (PerCom Workshops)*, Mar. 2020, pp. 1–7.
- [92] C. B. Fausset, T. L. Mitzner, C. Price, B. Jones, B. Fain, and W. A. Rogers, "Older adults' use of and attitudes toward activity monitoring technologies," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, Sep. 2013, vol. 57, no. 1, pp. 1683–1687.
- [93] P. C. Shih, K. Han, E. S. Poole, M. B. Rosson, and J. M. Carroll, "Use and adoption challenges of wearable activity trackers," in *Proc. IConf.*, Mar. 2015, pp. 1–12.
- [94] R. Yang, E. Shin, M. W. Newman, and M. S. Ackerman, "When fitness trackers don't 'fit': End-user difficulties in the assessment of personal tracking device accuracy," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, Sep. 2015, pp. 623–634.
- [95] A. Wijekoon, N. Wiratunga, S. Sani, and K. Cooper, "A knowledge-light approach to personalised and open-ended human activity recognition," *Knowl.-Based Syst.*, vol. 192, Mar. 2020, Art. no. 105651.
- [96] G. Schirner, D. Erdogmus, K. Chowdhury, and T. Padir, "The future of human-in-the-loop cyber-physical systems," *Computer*, vol. 46, no. 1, pp. 36–45, Jan. 2013.
- [97] C. Arzate Cruz and T. Igarashi, "A survey on interactive reinforcement learning: Design principles and open challenges," in *Proc. ACM Designing Interact. Syst. Conf.*, Jul. 2020, pp. 1195–1209.
- [98] A. Marchand Martella and D. Schneider, "A reflection on the current state of active learning research," *J. Scholarship Teaching Learn.*, vol. 24, no. 3, pp. 119–136, Sep. 2024.
- [99] A. Tharwat and W. Schenck, "A survey on active learning: State-of-the-art, practical challenges and research directions," *Mathematics*, vol. 11, no. 4, p. 820, Feb. 2023.
- [100] P. Doolittle, K. Wojdak, and A. Walters, "Defining active learning: A restricted systemic review," *Teaching Learn. Inquiry*, vol. 11, 2023.
- [101] P. Kumar and A. Gupta, "Active learning query strategies for classification, regression, and clustering: A survey," *J. Comput. Sci. Technol.*, vol. 35, no. 4, pp. 913–945, Jul. 2020.
- [102] D. Lewis and W. A. Gale, "A sequential algorithm for training text classifiers," in *Proc. SIGIR*. Cham, Switzerland: Springer, Aug. 1994, pp. 3–12.
- [103] H. S. Seung, M. Opper, and H. Sompolinsky, "Query by committee," in *Proc. 5th Annual Workshop Comput. Learn. Theory*, 1992, pp. 287–294.
- [104] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
- [105] Z. Xu, R. Akella, and Y. Zhang, "Incorporating diversity and density in active learning for relevance feedback," in *Proc. Eur. Conf. Inf. Retr.* Cham, Switzerland: Springer, Jun. 2007, pp. 246–257.
- [106] S. Rouhani, "Variance reduction analysis," *Water Resour. Res.*, vol. 21, no. 6, pp. 837–846, Jun. 1985.
- [107] W. Cai, Y. Zhang, and J. Zhou, "Maximizing expected model change for active learning in regression," in *Proc. IEEE 13th Int. Conf. Data Mining*, Dec. 2013, pp. 51–60.

- [108] K. Brinker, "Incorporating diversity in active learning with support vector machines," in *Proc. 20th Int. Conf. Mach. Learn. (ICML)*, Aug. 2003, pp. 59–66.
- [109] G. Kreml, D. Kottke, and M. Spiliopoulou, "Probabilistic active learning: Towards combining versatility, optimality and efficiency," in *Proc. 17th Int. Conf. Discovery Sci. (DS)*, Bled, Slovenia, Jan. 2014, pp. 168–179.
- [110] G. Contardo, L. Denoyer, and T. Artières, "A meta-learning approach to one-step active-learning," in *Proc. Int. Workshop Autom. Selection, Configuration Composition Mach. Learn. Algorithms*, 1998, pp. 28–40.
- [111] K. Konyushkova, R. Sznitman, and P. Fua, "Learning active learning from data," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, Jan. 2017, pp. 4225–4235.
- [112] O. Saadallah and Z. Rouissi, "Interpretable meta-active learning for regression ensemble learning," in *Proc. IAL@ PKDD/ECML*, 2023, pp. 1–15.
- [113] M. Fang, Y. Li, and T. Cohn, "Learning how to active learn: A deep reinforcement learning approach," 2017, *arXiv:1708.02383*.
- [114] M. Woodward and C. Finn, "Active one-shot learning," 2017, *arXiv:1702.06559*.
- [115] S. Flesca, D. Mandaglio, F. Scala, and A. Tagarelli, "A meta-active learning approach exploiting instance importance," *Expert Syst. Appl.*, vol. 247, Aug. 2024, Art. no. 123320.
- [116] H. Dong, A. S. Barnard, and A. J. Parker, "Online meta-learned gradient norms for active learning in science and technology," *Mach. Learn., Sci. Technol.*, vol. 5, no. 1, Mar. 2024, Art. no. 015041.
- [117] A. Rafiei, R. Moore, S. Jahromi, F. Hajati, and R. Kamaleswaran, "Meta-learning in healthcare: A survey," *Social Netw. Comput. Sci.*, vol. 5, no. 6, p. 791, Aug. 2024.
- [118] J. X. Wang, "Meta-learning in natural and artificial intelligence," *Current Opinion Behav. Sci.*, vol. 38, pp. 90–95, Apr. 2021.
- [119] R. Vilalta and Y. Drissi, "A perspective view and survey of meta-learning," *Artif. Intell. Rev.*, vol. 18, no. 2, pp. 77–95, 2002.
- [120] J. Vanschoren, "Meta-learning," in *Automated Machine Learning: Methods, Systems, Challenges* (The Springer Series on Challenges in Machine Learning). Cham, Switzerland: Springer, 2019, pp. 35–61.
- [121] N. Schweighofer and K. Doya, "Meta-learning in reinforcement learning," *Neural Netw.*, vol. 16, no. 1, pp. 5–9, Jan. 2003.
- [122] S. Yatawatta, "Reinforcement learning," *Astron. Comput.*, vol. 48, Jul. 2024, Art. no. 100833.
- [123] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
- [124] D. Ernst and A. Louette, *Introduction to Reinforcement Learning*, S. Feuerriegel, J. Hartmann, C. Janiesch, and P. Zschech, Eds., 2024, pp. 111–126.
- [125] P. Winder, *Reinforcement Learning*. Sebastopol, CA, USA: O'Reilly Media, 2020.
- [126] L. P. Kaelbling, M. L. Littman, and A. Moore, "Reinforcement learning: A survey," *J. Artif. Intell. Res.*, vol. 4, pp. 237–285, May 1996.
- [127] E. V. Denardo, "A Markov decision problem," in *Mathematical Programming*. Amsterdam, The Netherlands: Elsevier, 1973, pp. 33–68.
- [128] R. Bellman, "A Markovian decision process," *Indiana Univ. Math. J.*, vol. 6, no. 4, pp. 679–684, 1957.
- [129] R. A. Howard, "Dynamic Programming and Markov Processes," *Annu. Rev. Statist. Appl.*, vol. 7, 2020.
- [130] J. Clifton and E. B. Laber, "Q-learning: Theory and applications," *Annu. Rev. Statist. Appl.*, vol. 7, no. 1, pp. 279–301, Mar. 2020.
- [131] D. Zhao, H. Wang, K. Shao, and Y. Zhu, "Deep reinforcement learning with experience replay based on SARSA," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Dec. 2016, pp. 1–6.
- [132] M. M. Afsar, T. Crump, and B. Far, "Reinforcement learning based recommender systems: A survey," *ACM Comput. Surv.*, vol. 55, no. 7, pp. 1–38, Jul. 2023.
- [133] X. Chen, L. Yao, J. McAuley, G. Zhou, and X. Wang, "Deep reinforcement learning in recommender systems: A survey and new perspectives," *Knowl.-Based Syst.*, vol. 264, Mar. 2023, Art. no. 110335.
- [134] Q. Gao and A. M. Schweidtmann, "Deep reinforcement learning for process design: Review and perspective," *Current Opinion Chem. Eng.*, vol. 44, Jun. 2024, Art. no. 101012.
- [135] J. Shuford, "Deep reinforcement learning unleashing the power of AI in decision-making," *J. Artif. Intell. Gen. Sci. (JAIGS)*, vol. 1, no. 1, 2024.
- [136] S. S. Mousavi, M. Schukat, and E. Howley, "Deep reinforcement learning: An overview," in *Proc. SAI Intell. Syst. Conf. (IntelliSys)*, vol. 2, Cham, Switzerland: Springer, Aug. 2017, pp. 426–440.
- [137] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017.
- [138] L. Herrmann and S. Kollmannsberger, "Deep learning in computational mechanics: A review," *Comput. Mech.*, vol. 74, no. 2, pp. 281–331, Aug. 2024.
- [139] A. Negm, X. Ma, and G. Aggidis, "Deep reinforcement learning challenges and opportunities for urban water systems," *Water Res.*, vol. 253, Apr. 2024, Art. no. 121145.
- [140] A. Irshayid, J. Chen, and G. Xiong, "A review on reinforcement learning-based highway autonomous vehicle control," *Green Energy Intell. Transp.*, vol. 3, no. 4, Aug. 2024, Art. no. 100156.
- [141] F.-M. Luo, T. Xu, H. Lai, X. Chen, W. Zhang, and Y. Yu, "A survey on model-based reinforcement learning," *Sci. China Inf. Sci.*, vol. 67, no. 2, Jan. 2024, Art. no. 121101.
- [142] S. J. Russell, *Artificial Intelligence a Modern Approach*. London, U.K.: Pearson, 2010.
- [143] A. Y. Ng and M. I. Jordan, "PEGASUS: A policy search method for large MDPs and POMDPs," in *Proc. 16th Conf. Uncertainty Artif. Intell.*, Jun. 2000, pp. 406–415.
- [144] T. Jaakkola, S. Singh, and M. I. Jordan, "Reinforcement learning algorithm for partially observable Markov decision problems," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 7, Jan. 1994, pp. 345–352.
- [145] G. Tesauro, "Temporal difference learning and td-gammon," *Commun. ACM*, vol. 38, no. 3, pp. 58–68, 1995.
- [146] C. J. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, May 1992.
- [147] R. S. Sutton, "Dyna, an integrated architecture for learning, planning, and reacting," *ACM SIGART Bull.*, vol. 2, no. 4, pp. 160–163, Jul. 1991.
- [148] B. Li, P. Yang, Y. Sun, Z. Hu, and M. Yi, "Advances and challenges in artificial intelligence text generation," *Frontiers Inf. Technol. Electron. Eng.*, vol. 25, no. 1, pp. 64–83, 2024.
- [149] L. Wang, Z. Zhao, H. Liu, J. Pang, Y. Qin, and Q. Wu, "A review of intelligent music generation systems," *Neural Comput. Appl.*, vol. 36, no. 12, pp. 6381–6401, Apr. 2024.
- [150] K. Vayadande, C. B. Pednekar, P. A. Khune, V. S. Prabhavalkar, and V. R. Dange, "GPT-3- and DALL-E-powered applications: A complete survey," in *How Machine Learning is Innovating Today's World: A Concise Technical Guide*. Wiley, 2024, pp. 329–341.
- [151] J. Schulman, B. Zoph, C. Kim, J. Hilton, J. Menick, J. Weng, J. F. C. Uribe, L. Fedus, L. Metz, and M. Pokorný, "ChatGPT: Optimizing language models for dialogue," *OpenAI blog*, vol. 2, no. 4, 2022.
- [152] F. Fui-Hoon Nah, R. Zheng, J. Cai, K. Siau, and L. Chen, "Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration," *J. Inf. Technol. Case Appl. Res.*, vol. 25, no. 3, pp. 277–304, Jul. 2023.
- [153] K.-Q. Zhou and H. Nabus, "The ethical implications of DALL-E: Opportunities and challenges," *Mesopotamian J. Comput. Sci.*, vol. 2023, pp. 17–23, Jan. 2023.
- [154] A. Ramesh, M. Pavlov, G. Goh, S. Gray, C. Voss, A. Radford, M. Chen, and I. Sutskever, "Zero-shot text-to-image generation," in *Proc. Int. Conf. Mach. Learn.*, Jan. 2021, pp. 8821–8831.
- [155] A. Ramesh, P. Dhariwal, A. Nichol, C. Chu, and M. Chen, "Hierarchical text-conditional image generation with CLIP latents," 2022, *arXiv:2204.06125*.
- [156] G. Li, B. Chen, L. Zhu, Q. He, H. Fan, and S. Wang, "PUGCQ: A large scale dataset for quality assessment of professional user-generated content," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 3728–3736.
- [157] J. Kim, "The institutionalization of YouTube: From user-generated content to professionally generated content," *Media, Culture Soc.*, vol. 34, no. 1, pp. 53–67, Jan. 2012.
- [158] C. Wyrwoll and C. Wyrwoll, *User-Generated Content*. Cham, Switzerland: Springer, 2014.
- [159] M. Stefanini, M. Cornia, L. Baraldi, S. Cascianelli, G. Fiameni, and R. Cucchiara, "From show to tell: A survey on deep learning-based image captioning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 539–559, Jan. 2023.
- [160] P. Pu Liang, A. Zadeh, and L.-P. Morency, "Foundations and trends in multimodal machine learning: Principles, challenges, and open questions," 2022, *arXiv:2209.03430*.
- [161] P. Smolensky, "Information processing in dynamical systems: Foundations of harmony theory," *Parallel Distrib. Process.*, vol. 1, pp. 194–281, Jan. 1986.

- [162] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006.
- [163] R. Salakhutdinov and H. Larochelle, "Efficient learning of deep Boltzmann machines," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, Mar. 2010, pp. 693–700.
- [164] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng, "Recent progress on generative adversarial networks (GANs): A survey," *IEEE Access*, vol. 7, pp. 36322–36333, 2019.
- [165] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020.
- [166] N. Kitaev, L. Kaiser, and A. Levskaya, "Reformer: The efficient transformer," *CoRR*, vol. abs/2001.04451, 2020. [Online]. Available: <https://arxiv.org/abs/2001.04451>
- [167] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning image restoration without clean data," in *Proc. Int. Conf. Mach. Learn.*, Jan. 2018, pp. 4620–4631.
- [168] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: Representing scenes as neural radiance fields for view synthesis," *Commun. ACM*, vol. 65, no. 1, pp. 99–106, Jan. 2022.
- [169] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever, "Learning transferable visual models from natural language supervision," in *Proc. Int. Conf. Mach. Learn.*, Jan. 2021, pp. 8748–8763.
- [170] X. Wang, G. Chen, G. Qian, P. Gao, X.-Y. Wei, Y. Wang, Y. Tian, and W. Gao, "Large-scale multi-modal pre-trained models: A comprehensive survey," *Mach. Intell. Res.*, vol. 20, no. 4, pp. 447–482, Aug. 2023.
- [171] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, P. Gao, J. Schulman, J. Hilton, F. Kelton, L. E. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, and R. Lowe, "Training language models to follow instructions with human feedback," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2022, pp. 27730–27744.
- [172] P. Christiano et al., "Deep reinforcement learning from human preferences," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
- [173] N. Stiennon, L. Ouyang, J. Wu, D. Ziegler, R. Lowe, C. Voss, A. Radford, D. Amodei, and P. F. Christiano, "Learning to summarize with human feedback," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 3008–3021.
- [174] T. Wu, S. He, J. Liu, S. Sun, K. Liu, Q.-L. Han, and Y. Tang, "A brief overview of ChatGPT: The history, status quo and potential future development," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 5, pp. 1122–1136, May 2023.
- [175] J. Gunawan, "Exploring the future of nursing: Insights from the ChatGPT model," *Beltung Nursing J.*, vol. 9, no. 1, pp. 1–5, Feb. 2023.
- [176] M. Cascella, J. Montomoli, V. Bellini, and E. Bignami, "Evaluating the feasibility of ChatGPT in healthcare: An analysis of multiple clinical and research scenarios," *J. Med. Syst.*, vol. 47, no. 1, p. 33, Mar. 2023.
- [177] S. S. Biswas, "Role of chat GPT in public health," *Ann. Biomed. Eng.*, vol. 51, no. 5, pp. 868–869, May 2023.
- [178] C. M. Boßelmann, C. Leu, and D. Lal, "Are AI language models such as ChatGPT ready to improve the care of individuals with epilepsy?" *Epilepsia*, vol. 64, no. 5, pp. 1195–1199, May 2023.
- [179] N. Anantrasirichai and D. Bull, "Artificial intelligence in the creative industries: A review," *Artif. Intell. Rev.*, vol. 55, no. 1, pp. 589–656, Jan. 2022.
- [180] G. Cooper, "Examining science education in ChatGPT: An exploratory study of generative artificial intelligence," *J. Sci. Educ. Technol.*, vol. 32, no. 3, pp. 444–452, Jun. 2023.
- [181] M. Xu, H. Du, D. Niyato, J. Kang, Z. Xiong, S. Mao, Z. Han, A. Jamalipour, D. I. Kim, X. Shen, V. C. M. Leung, and H. V. Poor, "Unleashing the power of edge-cloud generative AI in mobile networks: A survey of AIGC services," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 2, pp. 1127–1170, 2nd Quart., 2024.
- [182] D. Adams and K.-M. Chuah, "Artificial intelligence-based tools in research writing: Current trends and future potentials," in *Artificial Intelligence in Higher Education*. Wiley, 2022, pp. 169–184.
- [183] D. Ippolito, A. Yuan, A. Coenen, and S. Burnam, "Creative writing with an AI-powered writing assistant: Perspectives from professional writers," 2022, [arXiv:2211.05030](https://arxiv.org/abs/2211.05030).
- [184] Y. Shen, L. Heacock, J. Elias, K. D. Hentel, B. Reig, G. Shih, and L. Moy, "ChatGPT and other large language models are double-edged swords," *Radiology*, vol. 307, no. 2, Apr. 2023, Art. no. e230163.
- [185] J. L. Gillotte, "Copyright infringement in AI-generated artworks," *UC Davis L. Rev.*, vol. 53, p. 2655, Jun. 2019.
- [186] K. Siau and W. Wang, "Artificial intelligence (AI) ethics: Ethics of AI and ethical AI," *J. Database Manage.*, vol. 31, no. 2, pp. 74–87, Apr. 2020.
- [187] L. Whittaker, T. C. Kietzmann, J. Kietzmann, and A. Dabirian, "All around me are synthetic faces': The mad world of AI-generated media," *IT Prof.*, vol. 22, no. 5, pp. 90–99, 2020.
- [188] Q. Zhang, Y. Kang, and R. Roth, "The ethics of AI-generated maps: DALL-E 2 and AI's implications for cartography (short paper)," in *Proc. 12th Int. Conf. Geographic Inf. Sci. (GISci.)*, 2023, pp. 93:1–93:6.
- [189] J. A. Van Dijk, "Digital divide research, achievements and shortcomings," *Poetics*, vol. 34, nos. 4–5, pp. 221–235, Aug. 2006.
- [190] L. Carter, D. Liu, and C. Cantrell, "Exploring the intersection of the digital divide and artificial intelligence: A hermeneutic literature review," *AIS Trans. Hum.-Comput. Interact.*, vol. 12, no. 4, pp. 253–275, Dec. 2020.
- [191] A. Bozkurt and R. C. Sharma, "Challenging the status quo and exploring the new boundaries in the age of algorithms: Reimagining the role of generative AI in distance education and online learning," *Asian J. Distance Educ.*, vol. 18, no. 1, Mar. 2023.
- [192] J. H. Allen, A. Moussa, and X. Liu, "Human-in-the-loop learning of qualitative preference models," in *Proc. 32nd Int. Flairs Conf.*, Sep. 2019, pp. 108–111.
- [193] Q. Cai, S. Liu, X. Wang, T. Zuo, W. Xie, B. Yang, D. Zheng, P. Jiang, and K. Gai, "Reinforcing user retention in a billion scale short video recommender system," in *Proc. Companion ACM Web Conf.*, Apr. 2023, pp. 421–426.
- [194] S. Dawar, S. Chatterjee, M. F. Hossain, and S. Malarvizhi, "Music recommendation system using real time parameters," in *Proc. Int. Conf. Recent Adv. Electr., Electron., Ubiquitous Commun., Comput. Intell. (RAEEUCCI)*, Apr. 2023, pp. 1–6.
- [195] W. Han, Z. Tian, C. Zhu, Z. Huang, Y. Jia, and M. Guizani, "A topic representation model for online social networks based on hybrid human-artificial intelligence," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 1, pp. 191–200, Feb. 2021.
- [196] D. R. Honeycutt, M. Nourani, and E. D. Ragan, "Soliciting human-in-the-loop user feedback for interactive machine learning reduces user trust and impressions of model accuracy," in *Proc. AAAI Conf. Hum. Comput. Crowdsourcing*, vol. 8, Oct. 2020, pp. 63–72.
- [197] P. Keikhosrokiani and G. M. Fye, "A hybrid recommender system for health supplement e-commerce based on customer data implicit ratings," *Multimedia Tools Appl.*, vol. 83, no. 15, pp. 45315–45344, Oct. 2023.
- [198] M. H. Lee, D. P. Siewiorek, A. Smailagic, A. Bernardino, and S. B. B. I. Badia, "A human-AI collaborative approach for clinical decision making on rehabilitation assessment," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2021, pp. 1–14.
- [199] A. Liu, S. Guerra, I. Fung, G. Matute, E. Kamar, and W. Lasecki, "Towards hybrid human-AI workflows for unknown unknown detection," in *Proc. Web Conf.*, Apr. 2020, pp. 2432–2442.
- [200] S. Murthy and S. S. Murthy, "Sapphire vs conventional YouTube video ranking," *Authorea*, Jun. 2024, doi: [10.22541/au.171894017.74660532/v1](https://doi.org/10.22541/au.171894017.74660532/v1).
- [201] A. Schroder, I. Constantiou, V. Tuunainen, and R. D. Austin, "Human-AI collaboration—Coordinating automation and augmentation tasks in a digital service company," in *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, 2022, pp. 1–10.
- [202] Y. Yang, E. Kandogan, Y. Li, P. Sen, and W. S. Lasecki, "A study on interaction in human-in-the-loop machine learning for text analytics," in *Proc. IUI Workshops*, Jan. 2019.
- [203] C. Yeh, G. Ramos, R. Ng, A. Huntington, and R. Banks, "GhostWriter: Augmenting collaborative human-AI writing experiences through personalization and agency," 2024, [arXiv:2402.08855](https://arxiv.org/abs/2402.08855).
- [204] X. Chen, S. Wang, J. McAuley, D. Jannach, and L. Yao, "On the opportunities and challenges of offline reinforcement learning for recommender systems," *ACM Trans. Inf. Syst.*, vol. 42, no. 6, pp. 1–26, Nov. 2024.
- [205] J. Lin, N. Tomlin, J. Andreas, and J. Eisner, "Decision-oriented dialogue for human-AI collaboration," *Trans. Assoc. Comput. Linguistics*, vol. 12, pp. 892–911, Aug. 2024.
- [206] N. Pescetelli, M. Cebrian, and I. Rahwan, "BeeMe: Real-time internet control of situated human agents," *Computer*, vol. 53, no. 8, pp. 49–58, Aug. 2020.

- [207] S. D. Ramchurn, F. Wu, W. Jiang, J. E. Fischer, S. Reece, S. Roberts, T. Rodden, C. Greenhalgh, and N. R. Jennings, "Human-agent collaboration for disaster response," *Auto. Agents Multi-Agent Syst.*, vol. 30, no. 1, pp. 82–111, Jan. 2016.
- [208] N. Schurr, J. Marecki, N. Kasinadhuni, M. Tambe, J. P. Lewis, and P. Scerri, "The DEFACTO system for human omnipresence to coordinate agent teams: The future of disaster response," in *Proc. 4th Int. Joint Conf. Auto. Agents Multiagent Syst.*, Jul. 2005, pp. 1229–1230.
- [209] N. Schurr, J. Marecki, M. Tambe, P. Scerri, N. Kasinadhuni, and J. Lewis, "The future of disaster response: Humans working with multiagent teams using DEFACTO," in *Proc. AAAI Spring Symp., AI Technol. Homeland Secur.*, Jan. 2005, pp. 9–16.
- [210] M. Zakershahrahk, "Towards human-machine symbiosis: Design for effective AI facilitation," Ph.D. dissertation, Arizona State Univ., Tempe, AZ, USA, 2021.
- [211] C. Maathuis, "Human-centered AI in military cyber operations," in *Proc. Int. Conf. Cyber Warfare Secur.*, Mar. 2024, vol. 19, no. 1, pp. 121–128.
- [212] M. R. Endsley, "Supporting human-AI teams: Transparency, explainability, and situation awareness," *Comput. Hum. Behav.*, vol. 140, Mar. 2023, Art. no. 107574.
- [213] G. Musick, T. A. O'Neill, B. G. Schelble, N. J. McNeese, and J. B. Henke, "What happens when humans believe their teammate is an AI? An investigation into humans teaming with autonomy," *Comput. Hum. Behav.*, vol. 122, Sep. 2021, Art. no. 106852.
- [214] D. Guimariães, D. Paulino, A. Correia, L. Trigo, P. Brazdil, and H. Paredes, "Towards a human-AI hybrid framework for inter-researcher similarity detection," in *Proc. IEEE 2nd Int. Conf. Hum.-Mach. Syst. (ICHMS)*, Sep. 2021, pp. 1–4.
- [215] M. Hasegawa-Johnson, A. W. Black, L. Ondel, O. Scharenborg, and F. Ciannella, "Image2Speech: Automatically generating audio descriptions of images," in *Proc. Casablanca*, Jan. 2017, p. 65.
- [216] A. Hekler, J. S. Utikal, A. H. Enk, A. Hauschild, M. Weichenthal, R. C. Maron, C. Berking, S. Haferkamp, J. Klode, and D. Schadendorf, "Superior skin cancer classification by the combination of human and artificial intelligence," *Eur. J. Cancer*, vol. 120, pp. 114–121, Sep. 2019.
- [217] A. Holzinger, M. Plass, M. Kickmeier-Rust, K. Holzinger, G. C. Crişan, C.-M. Pintea, and V. Palade, "Interactive machine learning: Experimental evidence for the human in the algorithmic loop," *Appl. Intell.*, vol. 49, no. 7, pp. 2401–2414, Jul. 2019.
- [218] R. Huang, M. Li, D. Yang, J. Shi, X. Chang, Z. Ye, Y. Wu, Z. Hong, J. Huang, J. Liu, Y. Ren, Y. Zou, Z. Zhao, and S. Watanabe, "AudioGPT: Understanding and generating speech, music, sound, and talking head," in *Proc. AAAI Conf. Artif. Intell.*, Mar. 2024, vol. 38, no. 21, pp. 23802–23804.
- [219] N. Köbis and L. D. Mossink, "Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry," *Comput. Hum. Behav.*, vol. 114, Jan. 2021, Art. no. 106553.
- [220] X. Liu, Z. Zhu, H. Liu, Y. Yuan, M. Cui, Q. Huang, J. Liang, Y. Cao, Q. Kong, M. D. Plumbley, and W. Wang, "WavJourney: Compositional audio creation with large language models," 2023, *arXiv:2307.14335*.
- [221] J. Liu, Y. Wang, X. Huang, M. B. Korsós, Y. Jiang, Y. Wang, and R. Erdélyi, "Reliability of AI-generated magnetograms from only EUV images," *Nature Astron.*, vol. 5, no. 2, pp. 108–110, Feb. 2021.
- [222] Z. Liu, J. Wang, S. Gong, D. Tao, and H. Lu, "Deep reinforcement active learning for human-in-the-loop person re-identification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6121–6130.
- [223] S. Loeschke, S. Belongie, and S. Benaim, "Text-driven stylization of video objects," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, Jan. 2023, pp. 594–609.
- [224] B. Mozafari, P. Sarkar, M. Franklin, M. Jordan, and S. Madden, "Scaling up crowd-sourcing to very large datasets: A case for active learning," *Proc. VLDB Endowment*, vol. 8, no. 2, pp. 125–136, Oct. 2014.
- [225] M. Pollefeys and L. V. Gool, "From images to 3D models," *Commun. ACM*, vol. 45, no. 7, pp. 50–55, Jul. 2002.
- [226] O. Schmitt and D. Buschek, "CharacterChat: Supporting the creation of fictional characters through conversation and progressive manifestation with a chatbot," in *Proc. Creativity Cognition*, Jun. 2021, pp. 1–10.
- [227] A. Sharma, I. W. Lin, A. S. Miner, D. C. Atkins, and T. Althoff, "Human-AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support," *Nature Mach. Intell.*, vol. 5, no. 1, pp. 46–57, Jan. 2023.
- [228] A. Tegen, P. Davidsson, R.-C. Mihailescu, and J. A. Persson, "Collaborative sensing with interactive learning using dynamic intelligent virtual sensors," *Sensors*, vol. 19, no. 3, p. 477, Jan. 2019.
- [229] O. Textler, D. Futschik, M. Kučera, O. Jamriška, Š. Sochorová, M. Chai, S. Tulyakov, and D. Šýkora, "Interactive video stylization using few-shot patch-based training," *ACM Trans. Graph.*, vol. 39, no. 4, pp. 1–73, Aug. 2020.
- [230] R. Thoppilan, D. De Freitas, J. Hall, N. Shazeer, A. Kulshreshtha, H.-T. Cheng, A. Jin, T. Bos, L. Baker, and Y. Du, "LaMDA: Language models for dialog applications," 2022, *arXiv:2201.08239*.
- [231] T. Le, T. Nguyen, N. Ho, H. Bui, and D. Phung, "LaMDA: Label matching deep domain adaptation," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2021, pp. 6043–6054.
- [232] D. Yang, Y. Zhou, Z. Zhang, T. J.-J. Li, and R. LC, "AI as an active writer: Interaction strategies with generated text in human-AI collaborative fiction writing," in *Proc. Joint ACM IUI Workshops*, vol. 10, 2022, pp. 1–11.
- [233] S. Zhang, J. Yu, X. Xu, C. Yin, Y. Lu, B. Yao, M. Tory, L. M. Padilla, J. Caterino, P. Zhang, and D. Wang, "Rethinking human-AI collaboration in complex medical decision making: A case study in sepsis diagnosis," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2024, pp. 1–18.
- [234] M. Puerta-Beldarrain, O. Gómez-Carmona, D. Casado-Mansilla, and D. López-de-Ipiña, "Human-AI collaboration to promote trust, engagement and adaptation in the process of pro-environmental and health behaviour change," in *Proc. Int. Conf. Ubiquitous Comput. Ambient Intell.*, Nov. 2022, pp. 381–392.
- [235] C. Reverberi, T. Rigon, A. Solari, C. Hassan, P. Cherubini, and A. Cherubini, "Experimental evidence of effective human-AI collaboration in medical decision-making," *Sci. Rep.*, vol. 12, no. 1, p. 14952, Sep. 2022.
- [236] B. Berger, M. Adam, A. Rühr, and A. Benlian, "Watch me improve—Algorithm aversion and demonstrating the ability to learn," *Bus. Inf. Syst. Eng.*, vol. 63, no. 1, pp. 55–68, Feb. 2021.
- [237] P. Hemmer, M. Schemmer, L. Riefler, N. Rosellen, M. Vössing, and N. Kühl, "Factors that influence the adoption of human-AI collaboration in clinical decision-making," 2022, *arXiv:2204.09082*.
- [238] I. Carvalho and S. Ivanov, "ChatGPT for tourism: Applications, benefits and risks," *Tourism Rev.*, vol. 79, no. 2, pp. 290–303, Feb. 2024.
- [239] T. Luther, J. Kimmerle, and U. Creß, "Teaming up with an AI: Exploring human-AI collaboration in a writing scenario with ChatGPT," *AI*, vol. 5, pp. 1357–1376, Feb. 2024.
- [240] A. Nguyen, F. Ilesanmi, B. Dang, E. Vuorenmaa, and S. Järvelä, "Hybrid intelligence in academic writing: Examining self-regulated learning patterns in an AI-assisted writing task," in *Frontiers in Artificial Intelligence and Applications*. Amsterdam, The Netherlands: IOS Press, 2024.
- [241] P. S. Dhillon, S. Molaei, J. Li, M. Golub, S. Zheng, and L. P. Robert, "Shaping human-AI collaboration: Varied scaffolding levels in co-writing with language models," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2024, pp. 1–18.
- [242] S. Sarker, A. Susarla, R. Gopal, and J. B. Thatcher, "Democratizing knowledge creation through human-AI collaboration in academic peer review," *J. Assoc. Inf. Syst.*, vol. 25, no. 1, pp. 158–171, 2024.
- [243] D. Sefeni, M. Johnson, and J. Lee, "Game-theoretic approaches for step-wise controllable text generation in large language models," *Authoria*, Sep. 2024, doi: [10.22541/au.172538388.88528596/v1](https://doi.org/10.22541/au.172538388.88528596/v1).
- [244] S. Marri, "12 conversational archetypes for human-AI interaction," *Int. J. Multidisciplinary Res.*, vol. 6, no. 3, May/Jun. 2024.
- [245] T. Ait Baha, M. El Hajji, Y. Es-Saady, and H. Fadili, "The impact of educational chatbot on Student learning experience," *Educ. Inf. Technol.*, vol. 29, no. 8, pp. 10153–10176, Jun. 2024.
- [246] J. Li, A. Dada, B. Puladi, J. Kleesiek, and J. Egger, "ChatGPT in healthcare: A taxonomy and systematic review," *Comput. Methods Programs Biomed.*, vol. 245, Mar. 2024, Art. no. 108013.
- [247] P. Stock, A. Fan, B. Graham, E. Grave, R. Gribonval, H. Jegou, and A. Joulin, "Training with quantization noise for extreme model compression," in *Proc. Int. Conf. Learn. Represent.*, 2021.
- [248] H. Ren, H. Dai, Z. Dai, M. Yang, J. Leskovec, D. Schuurmans, and B. Dai, "Combiner: Full attention transformer with sparse computation cost," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2021, pp. 22470–22482.
- [249] A. J. Adetayo, "Reimagining learning through AI art: The promise of DALL-E and MidJourney for education and libraries," *Library Hi Tech News*, 2024, doi: [10.1108/LHTN-01-2024-0005](https://doi.org/10.1108/LHTN-01-2024-0005).

- [250] C. Zhang, C. Zhang, M. Zhang, I. So Kweon, and J. Kim, "Text-to-image diffusion models in generative AI: A survey," 2023, *arXiv:2303.07909*.
- [251] C. Zhang, C. Zhang, S. Zheng, M. Zhang, M. Qamar, S.-H. Bae, and I. S. Kweon, "A survey on audio diffusion models: Text to speech synthesis and enhancement in generative AI," 2023, *arXiv:2303.13336*.
- [252] H. Liu, Y. Yuan, X. Liu, X. Mei, Q. Kong, Q. Tian, Y. Wang, W. Wang, Y. Wang, and M. D. Plumbley, "AudioLDM 2: Learning holistic audio generation with self-supervised pretraining," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 32, pp. 2871–2883, 2024.
- [253] Y. Yuan, H. Liu, X. Liu, Q. Huang, M. D. Plumbley, and W. Wang, "Retrieval-augmented text-to-audio generation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2024, pp. 581–585.
- [254] C. Cao, Y. Fu, S. Xu, R. Zhang, and S. Li, "Enhancing human-AI collaboration through logic-guided reasoning," in *Proc. 12th Int. Conf. Learn. Represent.*, 2024.
- [255] O. Gómez-Carmona, J. García-Zubia, and D. Diego Casado-Mansilla, "Promoting the perception of emerging technologies in work environments through edge computing and hybrid intelligence," Ph.D. dissertation, Univ. Deusto, Bilbao, Spain, 2021.
- [256] A. Holzinger, M. Plass, K. Holzinger, G. C. Crisan, C.-M. Pintea, and V. Palade, "A glass-box interactive machine learning approach for solving NP-hard problems with the human-in-the-loop," 2017, *arXiv:1708.01104*.
- [257] Y. Li, J. Xu, D. Guo, and H. Liu, "Trust-aware human-robot fusion decision-making for emergency indoor patrolling," *IEEE Trans. Autom. Sci. Eng.*, early access, Jan. 11, 2024, doi: [10.1109/TASE.2024.3350639](https://doi.org/10.1109/TASE.2024.3350639).
- [258] X. Lou, J. Guo, J. Zhang, J. Wang, K. Huang, and Y. Du, "PECAN: Leveraging policy ensemble for context-aware zero-shot human-AI coordination," 2023, *arXiv:2301.06387*.
- [259] B. McCamish, A. Termehchy, and B. Touri, "A game-theoretic approach to data interaction: A progress report," in *Proc. 2nd Workshop Human Loop Data Anal.*, May 2017, pp. 1–4.
- [260] A. Termehchy and B. Touri, "A signaling game approach to databases querying and interaction," in *Proc. Int. Conf. Theory Inf. Retr.*, Sep. 2015, pp. 361–364.
- [261] S. Mehak, J. D. Kelleher, M. Guilfoyle, and M. C. Leva, "Action recognition for human-robot teaming: Exploring mutual performance monitoring possibilities," *Machines*, vol. 12, no. 1, p. 45, Jan. 2024.
- [262] S. Nikolaidis, D. Hsu, and S. Srinivasa, "Human-robot mutual adaptation in collaborative tasks: Models and experiments," *Int. J. Robot. Res.*, vol. 36, nos. 5–7, pp. 618–634, Jun. 2017.
- [263] S. Nikolaidis, Y. X. Zhu, D. Hsu, and S. Srinivasa, "Human-robot mutual adaptation in shared autonomy," in *Proc. 12th ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI)*, Mar. 2017, pp. 294–302.
- [264] B. Schelble, C. Flathmann, L.-B. Canonico, and N. Mcneese, "Understanding human-AI cooperation through game-theory and reinforcement learning models," in *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, 2021, pp. 348–357.
- [265] D. Strouse, K. R. McKee, M. Botvinick, E. Hughes, and R. Everett, "Collaborating with humans without human data," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2021, pp. 14502–14515.
- [266] L. Tao, M. Bowman, J. Zhang, and X. Zhang, "Forming real-world human-robot cooperation for tasks with general goal," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 762–769, Apr. 2022.
- [267] C. Wang and J. Zhao, "Role dynamic assignment of human-robot collaboration based on target prediction and fuzzy inference," *IEEE Trans. Ind. Informat.*, vol. 20, no. 1, pp. 471–481, Jan. 2024.
- [268] X. Xing, W. Li, S. Yuan, and Y. Li, "Fuzzy logic-based arbitration for shared control in continuous human-robot collaboration," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 7, pp. 3979–3991, Jul. 2024.
- [269] Y. Xu, Y. Ohmoto, S. Okada, K. Ueda, T. Komatsu, T. Okadome, K. Kamei, Y. Sumi, and T. Nishida, "Formation conditions of mutual adaptation in human-agent collaborative interaction," *Appl. Intell.*, vol. 36, no. 1, pp. 208–228, Jan. 2012.
- [270] C. Yu, J. Gao, W. Liu, B. Xu, H. Tang, J. Yang, Y. Wang, and Y. Wu, "Learning zero-shot cooperation with humans, assuming humans are biased," in *Proc. 11th Int. Conf. Learn. RRepresent.*, 2023, pp. 1–11.
- [271] S. Zhang, X. Wang, W. Zhang, Y. Chen, L. Gao, D. Wang, W. Zhang, X. Wang, and Y. Wen, "Mutual theory of mind in human-AI collaboration: An empirical study with LLM-driven AI agents in a real-time shared workspace task," 2024, *arXiv:2409.08811*.
- [272] H. Zhou, D. Wei, Y. Chen, and F. Wu, "Promoting mutual adaptation in haptic negotiation using adaptive virtual fixture," *Ind. Robot. Int. J. Robot. Res. Appl.*, vol. 48, no. 2, pp. 313–326, Jul. 2021.
- [273] W. Deng, J. Xu, Y. Song, and H. Zhao, "An effective improved co-evolution ant colony optimisation algorithm with multi-strategies and its application," *Int. J. Bio-Inspired Comput.*, vol. 16, no. 3, p. 158, 2020.
- [274] Y. Xu, K. Ueda, T. Komatsu, T. Okadome, T. Hattori, Y. Sumi, and T. Nishida, "WOZ experiments for understanding mutual adaptation," *AI Soc.*, vol. 23, no. 2, pp. 201–212, Mar. 2009.
- [275] J. Heinrich, M. Lanctot, and D. Silver, "Fictitious self-play in extensive-form games," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2015, pp. 805–813.
- [276] H. Francis Song, A. Abdolmaleki, J. T. Springenberg, A. Clark, H. Soyer, J. W. Rae, S. Noury, A. Ahuja, S. Liu, D. Tirumala, N. Heess, D. Belov, M. Riedmiller, and M. M. Botvinick, "V-MPO: On-policy maximum a posteriori policy optimization for discrete and continuous control," 2019, *arXiv:1909.12238*.
- [277] Y. Nassar, G. Albeaino, I. Jeelani, M. Gheisari, and R. R. A. Issa, "Human-robot collaboration levels in construction: Focusing on Individuals' cognitive workload," in *Proc. Construct. Res. Congr.*, Mar. 2024, pp. 639–648.
- [278] A. Rosero, F. Dinh, E. J. de Visser, T. Shaw, and E. Phillips, "Two many cooks: Understanding dynamic human-agent team communication and perception using overcooked 2," 2021, *arXiv:2110.03071*.
- [279] S. A. Wu, R. E. Wang, J. A. Evans, J. B. Tenenbaum, D. C. Parkes, and M. Kleiman-Weiner, "Too many cooks: Bayesian inference for coordinating multi-agent collaboration," *Topics Cognit. Sci.*, vol. 13, no. 2, pp. 414–432, 2021.
- [280] K. A. Tahboub, "Human-machine coadaptation based on reinforcement learning with policy gradients," in *Proc. 8th Int. Conf. Syst. Control (ICSC)*, Oct. 2019, pp. 247–251.
- [281] R.-J. Qin and Y. Yu, "Learning in games: A systematic review," *Sci. China Inf. Sci.*, vol. 67, no. 7, Jul. 2024, Art. no. 171101.
- [282] S. Nikolaidis, S. Nath, A. D. Procaccia, and S. Srinivasa, "Game-theoretic modeling of human adaptation in human-robot collaboration," in *Proc. 12th ACM/IEEE Int. Conf. Hum.-Robot Interact. (HRI)*, Mar. 2017, pp. 323–331.
- [283] S. Amershi, M. Chickering, S. M. Drucker, B. Lee, P. Simard, and J. Suh, "ModelTracker: Redesigning performance analysis tools for machine learning," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, Apr. 2015, pp. 337–346.
- [284] H. Song, A. Abdolmaleki, J. T. Springenberg, A. Clark, H. Soyer, J. W. Rae, S. Noury, A. Ahuja, S. Liu, D. Tirumala, N. Heess, D. Belov, M. Riedmiller, and M. Botvinick, "V-MPO: On-policy maximum a posteriori policy optimization for discrete and continuous control," in *Proc. Int. Conf. Learn. Represent.*, Jan. 2019.
- [285] U. Kartoun, "Text nailing: An efficient human-in-the-loop text-processing method," *Interactions*, vol. 24, no. 6, pp. 44–49, Oct. 2017.
- [286] J. Talbot, B. Lee, A. Kapoor, and D. S. Tan, "EnsembleMatrix: Interactive visualization to support machine learning with multiple classifiers," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Apr. 2009, pp. 1283–1292.
- [287] C. Wiethof and E. A. Bittner, "Toward a hybrid intelligence system in customer service: Collaborative learning of human and AI," ECIS, Research Papers 66, 2022. [Online]. Available: https://aisel.aisnet.org/ecis2022_rp66
- [288] D. A. Keim, "Designing pixel-oriented visualization techniques: Theory and applications," *IEEE Trans. Vis. Comput. Graphics*, vol. 6, no. 1, pp. 59–78, Jan. 2000.
- [289] M. Manion, "Ethics, engineering, and sustainable development," *IEEE Technol. Soc. Mag.*, vol. 21, no. 3, pp. 39–48, Nov. 2002.
- [290] M. Zhao, R. Simmons, and H. Admoni, "The role of adaptation in collective human-AI teaming," *Topics Cognit. Sci.*, Nov. 2022.
- [291] H. Li, T. Ni, S. Agrawal, F. Jia, S. Raja, Y. Gui, D. Hughes, M. Lewis, and K. Sycara, "Individualized mutual adaptation in human-agent teams," *IEEE Trans. Human-Mach. Syst.*, vol. 51, no. 6, pp. 706–714, Dec. 2021.
- [292] M. C. Buehler and T. H. Weisswange, "Theory of mind based communication for human agent cooperation," in *Proc. IEEE Int. Conf. Human-Machine Syst. (ICHMS)*, Sep. 2020, pp. 1–6.
- [293] S. Nikolaidis, "Mathematical models of adaptation in human-robot collaboration," Ph.D. dissertation, Carnegie Mellon Univ., Pittsburgh, PA, USA, 2017.
- [294] T. Sawaragi, "Dynamical and complex behaviors in human-machine co-adaptive systems," *IFAC Proc. Volumes*, vol. 38, no. 1, pp. 94–99, 2005.
- [295] E. M. van Zoelen, K. van den Bosch, and M. Neerinx, "Becoming team members: Identifying interaction patterns of mutual adaptation for human-robot co-learning," *Frontiers Robot. AI*, vol. 8, Jul. 2021, Art. no. 692811.

- [296] C. Zeng, C. Yang, and Z. Chen, "Bio-inspired robotic impedance adaptation for human-robot collaborative tasks," *Sci. China Inf. Sci.*, vol. 63, no. 7, pp. 1–10, Jul. 2020.
- [297] Y. Mohammad and T. Nishida, "Human adaptation to a miniature robot: Precursors of mutual adaptation," in *Proc. RO-MAN 17th IEEE Int. Symp. Robot Human Interact. Commun.*, Aug. 2008, pp. 124–129.
- [298] E. M. van Zoelen, E. I. Barakova, and M. Rauterberg, "Adaptive leader-follower behavior in human-robot collaboration," in *Proc. 29th IEEE Int. Conf. Robot Hum. Interact. Commun. (RO-MAN)*, Aug. 2020, pp. 1259–1265.
- [299] L. Paripati, V. R. Hajari, N. Narukulla, N. Prasad, J. Shah, and A. Agarwal, "AI algorithms for personalization: Recommender systems, predictive analytics, and beyond," *Darpan Int. Res. Anal.*, vol. 12, no. 2, pp. 51–63, 2024.
- [300] S. Hua, S. Jin, and S. Jiang, "The limitations and ethical considerations of ChatGPT," *Data Intell.*, vol. 6, no. 1, pp. 201–239, Feb. 2024.
- [301] N. Tabari, A. A. Deshmukh, W.-C. Kang, H. Zamani, R. Gangadharaiah, J. McAuley, and G. Karypis, "First workshop on generative AI for recommender systems and personalization," in *Proc. 30th ACM SIGKDD Conf. Knowl. Discovery Data Mining*, Aug. 2024, pp. 6737–6738.
- [302] J. Li, W. Zhang, T. Wang, G. Xiong, A. Lu, and G. Medioni, "GPT4Rec: A generative framework for personalized recommendation and user interests interpretation," 2023, *arXiv:2304.03879*.
- [303] S. Mysore, M. Jasim, A. McCallum, and H. Zamani, "Editable user profiles for controllable text recommendations," in *Proc. 46th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2023, pp. 993–1003.
- [304] F. Radlinski, K. Balog, F. Diaz, L. Dixon, and B. Wedin, "On natural language user profiles for transparent and scrutable recommendation," in *Proc. 45th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2022, pp. 2863–2874.
- [305] Y. Wang, J. Zhu, R. Liu, and Y. Jiang, "Enhancing recommendation acceptance: Resolving the personalization–privacy paradox in recommender systems: A privacy calculus perspective," *Int. J. Inf. Manage.*, vol. 76, Jun. 2024, Art. no. 102755.
- [306] J. P. Kelly and D. Bridge, "Enhancing the diversity of conversational collaborative recommendations: A comparison," *Artif. Intell. Rev.*, vol. 25, nos. 1–2, pp. 79–95, Nov. 2007.
- [307] R. Rafter and B. Smyth, "Conversational collaborative recommendation—An experimental analysis," *Artif. Intell. Rev.*, vol. 24, nos. 3–4, pp. 301–318, Nov. 2005.
- [308] J. Ostheimer, S. Chowdhury, and S. Iqbal, "An alliance of humans and machines for machine learning: Hybrid intelligent systems and their design principles," *Technol. Soc.*, vol. 66, Aug. 2021, Art. no. 101647.
- [309] S. Mehrotra, C. Degachi, O. Vereschak, C. M. Jonker, and M. L. Tielman, "A systematic review on fostering appropriate trust in human-AI interaction: Trends, opportunities and challenges," *ACM J. Responsible Comput.*, vol. 1, no. 4, pp. 1–45, Dec. 2024.
- [310] S. Tolmeijer, U. Gadiraju, R. Ghantasala, A. Gupta, and A. Bernstein, "Second chance for a first impression? Trust development in intelligent system interaction," in *Proc. 29th ACM Conf. User Modeling, Adaptation Personalization*, Jun. 2021, pp. 77–87.
- [311] E. Janhunen, T. Toivikko, K. Blomqvist, and D. Siemon, "Trust in digital human-AI team collaboration: A systematic review," in *Proc. AMCIS*, 2024, pp. 287–294. [Online]. Available: <https://aisel.aisnet.org/amcis2024/cnow/cnow/3>
- [312] O. Lane and K. Klave, "Quantifying ethics and trust in human-AI collaboration," Tech. Rep., 2024.
- [313] M. Eiband, D. Buschek, and H. Hussmann, "How to support users in understanding intelligent systems? Structuring the discussion," in *Proc. 26th Int. Conf. Intell. User Interface*, Apr. 2021, pp. 120–132.
- [314] J. Zerilli, U. Bhatt, and A. Weller, "How transparency modulates trust in artificial intelligence," *Patterns*, vol. 3, no. 4, Apr. 2022, Art. no. 100455.
- [315] V. Schmitt, L.-F. Villa-Arenas, N. Feldhus, J. Meyer, R. P. Spang, and S. Möller, "The role of explainability in collaborative human-AI disinformation detection," in *Proc. ACM Conf. Fairness, Accountability, Transparency*, Jun. 2024, pp. 2157–2174.
- [316] B. Chandler, C. John, L. Warriar, and K. Gopalakrishnan, "Toward trustworthy artificial intelligence (TAI) in the context of explainability and robustness," *ACM Comput. Surv.*, Jun. 2024, doi: [10.1145/3675392](https://doi.org/10.1145/3675392).
- [317] N. D. Scharowski, "Trust in artificial intelligence: Understanding and calibrating trust and distrust in the human-AI interaction," Ph.D. dissertation, Univ. Basel, Basel, Switzerland, 2024.
- [318] P. Wang and H. Ding, "The rationality of explanation or human capacity? Understanding the impact of explainable artificial intelligence on human-AI trust and decision performance," *Inf. Process. Manage.*, vol. 61, no. 4, Jul. 2024, Art. no. 103732.
- [319] B. Ling, B. Dong, and F. Cai, "Applicants' fairness perception of human and AI collaboration in resume screening," *Int. J. Hum.-Comput. Interact.*, pp. 1–12, 2024.
- [320] M. A. Meza Martínez, M. Nadj, and A. Maedche, "Towards an integrative theoretical framework of interactive machine learning systems," in *Proc. 27th Eur. Conf. Inf. Syst. (ECIS)*, Stockholm, Sweden, Jun. 2019. [Online]. Available: https://aisel.aisnet.org/ecis2019_rp/172
- [321] M. Vössing, N. Kühn, M. Lind, and G. Satzger, "Designing transparency for effective human-AI collaboration," *Inf. Syst. Frontiers*, vol. 24, no. 3, pp. 877–895, Jun. 2022.
- [322] W. Abdelghani, C. A. Zayani, I. Amous, and F. Sèdes, "User-centric IoT: Challenges and perspectives," in *Proc. UBICOMM 12th Int. Conf. Mobile Ubiquitous Comput., Syst., Services Technol.*, Jun. 2019, pp. 27–34.
- [323] C. B. Masson, D. Martin, T. Colombino, and A. Grasso, "'The device is not well designed for me' on the use of activity trackers in the workplace?" in *Proc. COOP 12th Int. Conf. Design Cooperat. Syst.*, Trento, Italy. Cham, Switzerland: Springer, May 2016, pp. 39–55.
- [324] G. Ramos, J. Suh, S. Ghorashi, C. Meek, R. Banks, S. Amershi, R. Fiebrink, A. Smith-Renner, and G. Bansal, "Emerging perspectives in human-centered machine learning," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, May 2019, pp. 1–8.
- [325] N. Glas and C. Pelachaud, "Definitions of engagement in human-agent interaction," in *Proc. Int. Conf. Affect. Comput. Intell. Interact. (ACII)*, Sep. 2015, pp. 944–949.
- [326] H. L. O'Brien and E. G. Toms, "What is user engagement? A conceptual framework for defining user engagement with technology," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 59, no. 6, pp. 938–955, Apr. 2008.
- [327] D. Wang, E. Churchill, P. Maes, X. Fan, B. Shneiderman, Y. Shi, and Q. Wang, "From human-human collaboration to human-AI collaboration: Designing AI systems that can work together with people," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2020, pp. 1–6.
- [328] C. Oertel, G. Castellano, M. Chetouani, J. Nasir, M. Obaid, C. Pelachaud, and C. Peters, "Engagement in human-agent interaction: An overview," *Frontiers Robot. AI*, vol. 7, p. 92, Aug. 2020.
- [329] J. Rezwana and M. L. Maher, "Understanding user perceptions, collaborative experience and user engagement in different human-AI interaction designs for co-creative systems," in *Proc. Creativity Cognition*, Jun. 2022, pp. 38–48.
- [330] Q. Zheng, Y. Tang, Y. Liu, W. Liu, and Y. Huang, "UX research on conversational human-AI interaction: A literature review of the ACM digital library," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2022, pp. 1–24.
- [331] E. Holder, L. Huang, E. K. Chiou, M. Jeon, and J. B. Lyons, "Designing for bi-directional transparency in human-AI-robot-teaming," in *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, Sep. 2021, vol. 65, no. 1, pp. 57–61.
- [332] R. Zhang, W. Duan, C. Flathmann, N. McNeese, G. Freeman, and A. Williams, "Investigating AI teammate communication strategies and their impact in human-AI teams for effective teamwork," *Proc. ACM Hum.-Comput. Interact.*, vol. 7, pp. 1–31, Sep. 2023.
- [333] A. R. Marathe, K. E. Schaefer, A. W. Evans, and J. S. Metcalfe, "Bidirectional communication for effective human-agent teaming," in *Proc. Int. Conf. Virtual, Augmented Mixed Reality*. Cham, Switzerland: Springer, Jan. 2018, pp. 338–350.
- [334] M. Riveiro and S. Thill, "'That's (not) the output I expected!' on the role of end user expectations in creating explanations of AI systems," *Artif. Intell.*, vol. 298, Sep. 2021, Art. no. 103507.
- [335] S. Milani, N. Topin, M. Veloso, and F. Fang, "Explainable reinforcement learning: A survey and comparative review," *ACM Comput. Surveys*, vol. 56, no. 7, pp. 1–36, Jul. 2024.
- [336] K. E. Schaefer, B. S. Perelman, R. W. Brewer, J. L. Wright, N. Roy, and D. Aksaray, "Quantifying human decision-making: Implications for bidirectional communication in human-robot teams," in *Proc. Int. Conf. Virtual, Augmented Mixed Reality*, Jan. 2018, pp. 361–379.
- [337] K. Breckner, T. Neumayr, M. Streit, and M. Augstein, "Personalized complementarity in human-ai collaboration," in *Proc. Mensch und Comput. Workshopband*, 2024, p. 18420.

[338] E. Salas, M. L. Shuffler, A. L. Thayer, W. L. Bedwell, and E. H. Lazzara, "Understanding and improving teamwork in organizations: A scientifically based practical guide," *Hum. Resource Manage.*, vol. 54, no. 4, pp. 599–622, Jul. 2015.

[339] N. J. Cooke, J. C. Gorman, C. W. Myers, and J. L. Duran, "Interactive team cognition," *Cognit. Sci.*, vol. 37, no. 2, pp. 255–285, Mar. 2013.

[340] W. Duan, N. McNeese, and R. Zhang, "Communication in human-AI teaming," in *Group Communication*. Evanston, IL, USA: Routledge, 2023, pp. 340–352.

[341] S. Kopp and N. Krämer, "Revisiting human-agent communication: The importance of joint co-construction and understanding mental states," *Frontiers Psychol.*, vol. 12, Mar. 2021, Art. no. 580955.

[342] N. Hanna and D. Richards, "Speech act theory as an evaluation tool for human-agent communication," *Algorithms*, vol. 12, no. 4, p. 79, Apr. 2019.

[343] E. M. Zoelen, A. Cremers, F. Dignum, J. Diggelen, and M. Peeters, "Learning to communicate proactively in human-agent teaming," in *Proc. Int. Conf. Practical Appl. Agents Multi-Agent Syst.*, Jan. 2020, pp. 238–249.

[344] K. Van Den Bosch and A. Bronkhorst, *Human-AI Cooperation to Benefit Military Decision Making*. Brussels, Belgium: NATO, 2018.

[345] A. Aggarwal, C. C. Tam, D. Wu, X. Li, and S. Qiao, "Artificial intelligence-based chatbots for promoting health behavioral changes: Systematic review," *J. Med. Internet Res.*, vol. 25, Feb. 2023, Art. no. e40789.

[346] M. A. Kuhail, N. Alturki, S. Alramlawi, and K. Alhejori, "Interacting with educational chatbots: A systematic review," *Educ. Inf. Technol.*, vol. 28, no. 1, pp. 973–1018, Jan. 2023.

[347] M. Zbyszynski, B. Di Donato, and A. Tanaka, "The effect of co-adaptive learning & Feedback in interactive machine learning," in *Proc. ACM CHI, Hum.-Centered Mach. Learn. Perspect. Workshop*, Glasgow, U.K., May 2019. Tech. Rep., 2019.

[348] W. B. Knox, P. Stone, and C. Breazeal, "Training a robot via human feedback: A case study," in *Proc. Int. Conf. Social Robot.* Cham, Switzerland: Springer, Jan. 2013, pp. 460–470.

[349] M. A. Pérez-Quiñones and J. L. Sibert, "A collaborative model of feedback in human-computer interaction," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. Common Ground (CHI)*, 1996, pp. 316–323.



RUBÉN SÁNCHEZ-CORCUERA received the Ph.D. degree from the University of Deusto, in 2023. He is currently an Associate Researcher with the University of Deusto. His main research interests include analysis of online social networks using natural language processing and graph-based methods.



DIEGO CASADO-MANSILLA received the Ph.D. degree from the University of Deusto, in 2016, on smart object-human interaction for pro-environmental behavior change. He is currently a Lecturer with the University of Deusto and a Postdoctoral Researcher with DeustoTech. He is also a member of MORElab Research Group, Faculty of Engineering, University of Deusto. His research interests include hybrid intelligence, where the IoT devices and humans collaborate, physical interaction with everyday objects, sustainable human-computer interaction (SHCI), and persuasive and behavioral change technologies. He has more than 100 scientific publications on HCI, behavior change, and the Internet of Things (IoT). He has participated in or led H2020 projects, basic-research national projects, and various regional projects in collaboration with companies. He is a member of several IEEE and ACM scientific committees and has been part of the organizing committee of several international workshops and conferences. He is a reviewer of journals on the top-tier JCRs. He has carried out research stays in the U.K. and France.



MAITE PUERTA-BELDARRAIN received the bachelor's degree in mathematics from the Universidad del País Vasco and the master's degree in artificial intelligence from the Universidad Politécnica de Madrid. She is currently a Research Assistant, specializing in artificial intelligence and possessing a solid foundation in citizen science. She also holds a position with the University of Deusto, actively contributing to cutting-edge research projects.



OIHANE GÓMEZ-CARMONA received the Ph.D. degree from the University of Deusto, in 2021, on the topics of the Internet of Things (IoT) and artificial intelligence. She is currently an Associate Researcher with the University of Deusto. Grounded in the general principles of IoT, her research delves into the intersection between IoT and machine learning, specifically oriented toward resource-constrained devices and the interplay between humans and the broader spectrum of IoT technologies. She has participated in various projects, both regional, national, and European, related to the creation of intelligent environments, boasting several publications associated with her line of research.



DIEGO LÓPEZ-DE-IPÍÑA received the Ph.D. degree from the University of Cambridge, U.K., in 2002, with a dissertation titled "Visual Sensing and Middleware Support for Sentient Computing." He is currently a Full Professor and a Principal Researcher of the "DEUSTEK/MORElab—ICT for Good" Group (<http://morelab.deusto.es/>), associated to the Faculty of Engineering, University of Deusto, Bilbao, Spain. His main research interests include pervasive computing, the Internet of Things, semantic service middleware, open linked data, big data management, and mobile-mediated and tangible human-environment interaction. He is currently focusing his work on the role of citizens as active data contributors to the knowledge of smart communities modeled as knowledge graphs.



LIMING CHEN (Senior Member, IEEE) received the B.Eng. and M.Eng. degrees from Beijing Institute of Technology, China, in 1985 and 1988, respectively, and the Ph.D. degree from De Montfort University, U.K., in 2003. He is currently a Chair Professor with the School of Computer Science and Technology, Dalian University of Technology, China. His research interests include pervasive computing, data analytics, artificial intelligence, user-centred intelligent cyber-physical systems and their applications in smart healthcare, and cyber security. He has over 300 publications in the aforementioned areas. He is an IET Fellow.

...