IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 28 April 2024, accepted 12 May 2024, date of publication 15 May 2024, date of current version 4 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3401547

SURVEY

Applications, Challenges, and Future Directions of Human-in-the-Loop Learning

SUSHANT KUMAR[1], SUMIT DATTA[2], (Senior Member, IEEE), VISHAKHA SINGH[1], DEEPANWITA DATTA[3], SANJAY KUMAR SINGH[1], (Senior Member, IEEE), AND RITESH SHARMA[4]

[1]Department of Computer Science and Engineering, Indian Institute of Technology (BHU) Varanasi, Varanasi 221005, India
[2]School of Electronic Systems and Automation, Digital University Kerala (Formerly IIITM Kerala), Thiruvananthapuram 695317, India
[3]Department of Information System Management, Indian Institute of Management (IIM) Sambalpur, Sambalpur 768025, India
[4]Department of Information and Communication Technology, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

Corresponding authors: Ritesh Sharma (ritesh.sharma@manipal.edu) and Sumit Datta (sumit.datta@iiitmk.ac.in)

ABSTRACT Machine learning (ML) has become a popular technique for various automation tasks in the era of Industry 4.0, such as the analysis and synthesis of visual data such as images and videos, natural language and speech, financial data, and biomedical applications. However, ML-based automation techniques are facing difficulties like decision-making, thus incorporating user expertise into the system might be advantageous. The goal of adding human domain expertise with ML-based automation is to provide more accurate prediction models. Human-in-the-loop (HITL) systems that integrate human expertise with ML algorithms are becoming more and more common in various industries. However, there are a number of methodological, technical, and ethical difficulties with the development and application of HITL systems. This paper aims to explore the methodologies, challenges, and opportunities associated with HITL systems implementations. We also discuss a number of issues that must be resolved for HITL systems to be effective, including data quality, bias, and user engagement. Besides, we also explored several approaches that can be utilized to enhance the performance of HITL systems, such as active learning (AL), iterative ML, and reinforcement learning, as well as the current state of the art in HITL systems. We also selectively highlighted the advantages of HITL systems, such as their potential to increase decision-making process accountability and transparency by utilizing human experience to improve ML decision-making capability. The paper will be very useful for researchers, practitioners, and policymakers.

INDEX TERMS Human-in-the-loop (HITL), machine learning algorithms, accountability, transparency.

I. INTRODUCTION

Deep learning (DL) has garnered impressive achievements across a range of domains, encompassing tasks such as the interpretation and generation of visual data like images and videos, understanding and processing natural language and speech, applications in the medical field, as well as enhancing the capabilities of intelligent transportation systems [1]. This success can be attributed to the use of larger models with numerous parameters, providing greater flexibility and descriptive power [2]. However, the effectiveness of DL relies on a substantial amount of labeled training data [3]. Acquiring and annotating such data is a challenging and time-consuming task, as the growth rate of data is much slower compared to the rate of model parameter expansion. To overcome this challenge, researchers are employing techniques such as generating new datasets, speeding up model iteration, and decreasing expenses related to data annotation [4], [5], [6], [7]. Additionally, pre-trained models and transfer learning methods like Transformers [8], BERT [9], and GPT [10] have demonstrated impressive outcomes. Although the generated data initializes the model, specific data labeling and updates are often necessary to achieve a high-precision usable model. Weak supervision techniques and few-shot

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

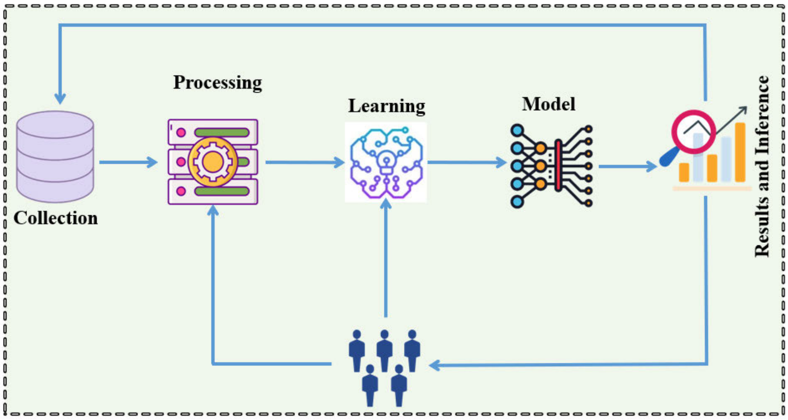


FIGURE 1. Human-in-the-loop learning framework.

learning have been proposed to address data scarcity issues, with researchers actively exploring these approaches [11], [12], [13].

Integrating prior knowledge into the learning framework has emerged as a valuable approach to tackling sparse data problems. By incorporating human wisdom and knowledge [14] the machine can learn from existing knowledge sources. Notably, researchers have been increasingly incorporating pre-existing knowledge into their learning frameworks, particularly in medical fields such as clinical diagnosis, where data availability may be limited [15], [16], [17]. Utilizing pre-training knowledge can enhance performance and address data limitations effectively [18], [19], [20].

Recent studies have highlighted the importance of human-related aspects, including emotional state and practical capability, in influencing the results of teaching and learning. To address these challenges, the concept of "human-in-the-loop" has been introduced, involving the incorporating human knowledge into the process of modeling [21]. This multidisciplinary approach intersects computer science, cognitive science, and psychology.

A typical ML framework with Human-in-the-loop (HITL) learning is shown in Fig. 1 which consists of three components: data pre-processing, data modeling, and modifying the process to improve performance [22]. However, the results and performance of ML models can be unpredictable, making it uncertain which aspect of human-machine interaction yields the best learning outcomes. Researchers concentrate on introducing manual intervention at various stages to address this issue. This article explores prevailing research on HITL technology and explores diverse implementations from a practical standpoint. These implementations encompass data processing, model training and inference, as well as system construction and application. The authors aim to explore the impact of different types of human interaction on learning outcomes in intelligent systems and its interaction with other components of the HITL pipeline. Additionally, researchers are developing independent systems to enhance the model improvement process. The paper discusses methods to improve model performance through data processing, intervention-based model training, and the configuration of system-independent "human-in-the-loop" setups [23], [24], [25], [26], [27].

HITL systems play a crucial role in the implementation of Industry 5.0, which focuses on enhancing collaboration between humans and machines to improve productivity, efficiency, and safety in various industries, particularly manufacturing. Industry 5.0 emphasizes the utilization of "weak AI" that is understandable and manageable by humans, highlighting the importance of the HITL concept for transparent man-machine cooperation, ethical decision-making, and resilience. In contrast, Industry 4.0 relies on "Black Box AI" that offers limited human control [28]. The concepts of Operator 4.0 and Operator 5.0 introduce a human-centric perspective and a symbiotic relationship between humans and automation in the fourth industrial revolution. The vision of a resilient Operator 5.0 aims to create self-resilience for the workforce and system resilience to ensure optimal operation through smarter collaboration between operators and machines. Training such professionals necessitates the use of mixed reality frameworks and platforms [29]. The HITL offers several advantages, such as:

Improving precision: As humans continue to refine the model's responses to different scenarios, the algorithm becomes more accurate and consistent. In fields like content moderation, there are limits to how much analysis can be automated, and humans are crucial in interpreting context, multilingual text, and cultural nuances.

Enhancing data acquisition: ML models require large amounts of data to be effective, and a HITL can generate data and ensure its accuracy in situations where there is a lack of data. P *Mitigating bias:* AI programs designed by humans based on historical data can perpetuate inequalities, and a HITL can identify and correct bias early on.

Increasing efficiency: Machine intelligence can save significant time and costs by processing and filtering large amounts of data, and the remaining tasks can be performed by humans. While not all aspects of the process can be automated, a substantial portion of it can be, resulting in time savings.

Our investigation focuses on the following inquiries concerning HITL for ML:

1) What is the role of human expertise in improving ML models?
2) How can HITL approaches be integrated into the machine-learning pipeline?
3) What are the most effective methods for incorporating human feedback into the learning process?
4) How do different types of human feedback affect the learning outcomes of the model?
5) What are the ethical implications of using HITL approaches in ML?
6) How can we ensure that human feedback is unbiased and does not reinforce existing biases in the data?
7) What are the trade-offs between using HITL approaches and fully automated ML?
8) How can we design HITL systems that are user-friendly and accessible to non-experts?
9) What are the implications of HITL approaches for data privacy and security?
10) What are the implications of HITL approaches for the scalability and efficiency of the ML process?

A. CONTRIBUTION

This paper provides a comprehensive review and analysis of the research on HITL, with a particular focus on the following key aspects:

- Present different methodologies for incorporating human feedback into ML, including Active learning (AL), interactive ML, and crowd-sourcing.
- Classify and compare different methods of HITL, highlighting the challenges faced by this approach and proposing potential solutions. Also, provide qualitative evaluations and comparisons of these methods to help readers choose the appropriate one for their specific problem.
- Identify a series of important milestones achieved by various methods in this area.
- Discusses the challenges associated with HITL systems, such as the need for unbiased human feedback, ensuring privacy and security, and designing systems that are accessible to non-experts.
- Highlight the opportunities that HITL systems provide, such as improving model accuracy and generalization, reducing bias, and enabling domain experts to provide their knowledge of the model.
- Included several case studies that demonstrate the effectiveness of HITL systems in different domains, such as healthcare, education, and natural language processing.

II. METHODOLOGIES

HITL methodologies are collaborative approaches that involve both humans and machines working together to enhance the performance of ML algorithms. Popular HITL methodologies include AL, Reinforcement learning, and Explainable AI. These methodologies have been shown to improve the accuracy and efficiency of ML algorithms. However, HITL approaches require a significant investment of time and resources to implement and maintain, making them more suitable for real-world applications.

A. ACTIVE LEARNING

Active learning is an ML technique that involves selecting the most informative or representative data points to label and adding them to the training set [2]. This approach aims to reduce the cost of data annotation, as it requires fewer labeled data points than traditional supervised learning methods. Cloud computing services have played a significant role in facilitating AL by providing access to scalable and cost-effective computational resources. In addition to computational resources, cloud computing services also provide access to crowdsourcing frameworks, which can be used to further reduce the cost of data annotation. Crowdsourcing involves outsourcing small tasks to a large group of people, typically through an online platform. This allows researchers to leverage the collective intelligence of a diverse group of individuals to perform tasks such as data annotation. By combining AL with crowdsourcing, researchers can create a powerful and cost-effective framework for building high-quality ML models. This approach has been used in a wide range of applications, including natural language processing, computer vision, and healthcare, among others.

AL for a general classification task can be defined based on Mitchell's [30], [31] concept of a well-formed ML problem:

Task function: $f : \upsilon \rightarrow \{1, 2, \ldots .p\}$ where υ represents the set of data items with their corresponding true labels, and $\{1, 2, \ldots .p\}$ represents the available classes.

Performance Measures: Precision, recall, F1-score, etc., serve as performance metrics.

Active learning Process: In the active learning (AL) process, unlabeled samples $X_t \subseteq D_t^U$ are chosen at each time step t, using a query strategy S_q. Subsequently, labels are requested for each selected sample from an oracle. Following the query and labeling at time step t, the labeled dataset becomes $D_{t+1}^L = D_t^L \cup X_t$, and the unlabeled dataset becomes $D_{t+1}^U = D_t^U - X_t$.

AL loops often involve incorporating the input of human annotators or non-expert contributors, particularly in subjective domains. Over the years, extensive research has been conducted on integrating human input into AL loops to explore a range of classical problems. Fig. 2 depicts a typical AL framework, which involves determining the informativeness of each unannotated data point based on the query type chosen. The chosen data points are then annotated, and the AL framework utilizes the newly acquired

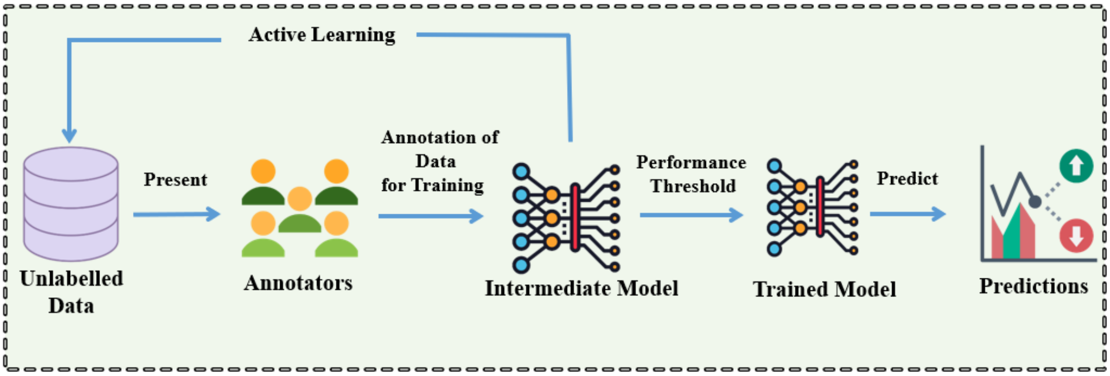


FIGURE 2. Human-in-the-loop active learning frameworks.

annotations to improve the model. This enhancement can be attained by either completely retraining the model using all existing annotated data or by fine-tuning the network with the most recently labeled data points. By adopting this approach, it is possible to attain state-of-the-art performance using fewer annotations for various biomedical image analysis tasks. Consequently, this widens the annotation bottleneck and reduces the costs associated with developing DL-enabled systems from unannotated data. AL has been successfully applied in many fields, including Natural Language Processing (NLP), data annotation, and image classification tasks [32], [33], [34]. In NLP, for example, AL can be used to improve the accuracy of natural language understanding systems by selecting the most informative examples for annotation, for example, the samples that are difficult to classify or the representative samples of a particular domain or topic. Similarly, in image classification, AL can be used to select the most representative images for annotation, which can improve the accuracy of the model and reduce the amount of labeled data required.

AL involves eliciting ground truth labels for uncertain data instances to enhance the model's performance [35], [36]. Different approaches have been proposed, including prioritizing the inspection of uncertain samples [37], [38], enhancing diversity involves a strategic collection of samples to accurately depict the entire data distribution, all while considering both uncertainty and diversity aspects simultaneously [39]. However, the existing AL models have some fundamental limitations for practical use. Initially, a significant portion of evaluations relies on static test data, which does not account for concept drifts occurring in real active learning scenarios [40], [41]. Second, manual inspection or annotation is costly, and the budget is often limited, making it difficult for exploration-oriented AL algorithms to succeed.

B. REINFORCEMENT LEARNING

Reinforcement learning (RL) is a type of ML that enables an agent to learn from its own interactions with an environment by receiving feedback in the form of rewards or penalties [42], [43], [44], [45]. RL has been proposed as a potential use case of AL in medical image analysis, where an agent can learn to decide which examples are worth labeling. The combination of RL and AL has been explored in several works, with promising results in improving prediction accuracy or data selection policy. An overview of the HITL RL framework [46], in which a human provides new actions in response to state queries can be seen in Figure 3. Recently, effective generalization in reinforcement learning over large, high-dimensional state spaces, extending the set of actions makes this task even more challenging. Therefore, the concentrates on the more traditional reinforcement learning setting with a discrete, relatively small state space [47].

In [48], authors used RL to reframe the data selection process as an RL problem and learned a data selection policy that is independent of the heuristics commonly used in AL frameworks. They showed improvements in entity recognition. Although RL methods offer a different approach to AL and HITL problems, can facilitate real-time feedback between a DL-enabled application and its end-users. It should be noted that RL requires task-specific goals that may not be generalizable across various medical image analysis tasks. Therefore, it is essential to clearly define the goals of the RL agent to ensure they align with the specific task at hand. The combination of RL and AL has the potential to improve the efficiency and effectiveness of medical image analysis tasks. However, further research is necessary to investigate the generalizability and scalability of this approach to different tasks and domains. The human-in-the-loop low-shot (HILL) [49] framework utilizes uncertainty assessment to identify challenging samples, human intervention to label them, and reinforcement learning to adapt the model based on this feedback. Knox et al. [50] introduce a new "regret preference modeling" approach for learning reward functions in reinforcement learning from human feedback.

C. EXPLAINABLE AI

Explainable AI (XAI) is a rapidly growing field that aims to create ML systems that are more transparent and interpretable. XAI techniques can help users to understand

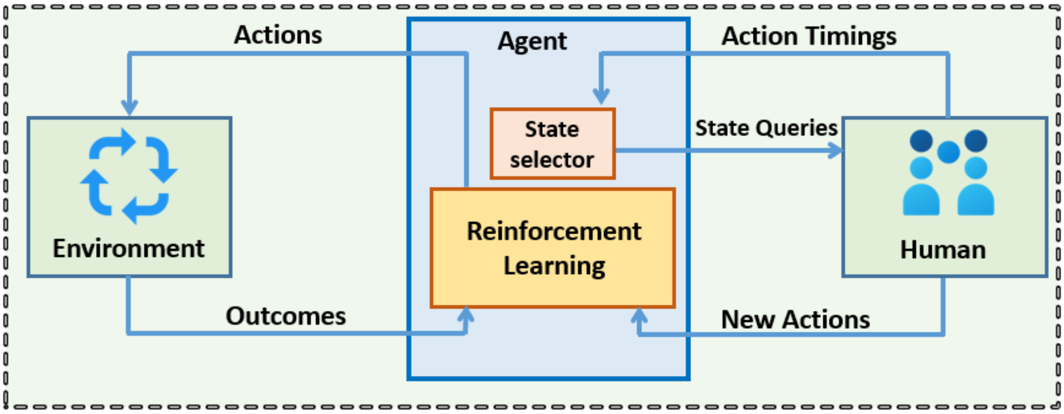


FIGURE 3. Human-in-the-loop reinforcement learning framework.

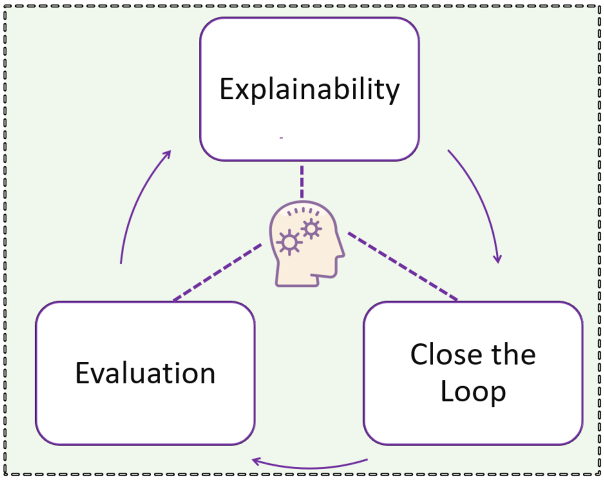


FIGURE 4. Human-in-the-loop in XAI framework.

how decisions are made by AI systems as shown in Fig. 4, which can be critical in domains where trust, accountability, and human oversight are important, such as medicine, law, and defense [51].

XAI techniques include a range of methods such as model visualization, feature attribution, and natural language explanations, which can help users to understand the behavior of complex AI models. By providing more transparent and interpretable models, XAI can also help to improve the trustworthiness and reliability of AI systems and can enable more effective collaboration between humans and machines [52]. However, it is important to note that there are trade-offs between explainability and performance in AI systems, and achieving both can be challenging. XAI research is therefore an ongoing effort to find the right balance between performance and interpretability in ML systems.

The effectiveness of explanations in XAI heavily relies on the ability of humans to understand and interpret them accurately. Poor explanations can mislead users and generate undesired bias. Therefore, the XAI research field expands from IT-related fields to human-centered disciplines, such as psychology and decision-making. Current research directs its attention toward assessing human behaviors during the exploration, interpretation, and utilization of explanations. Adequately framing and evaluating the interpretability and efficacy of these explanations necessitates a profound comprehension of how humans perceive and comprehend them. Achieving equilibrium between broad and detailed explanations aids users in grasping and forecasting the model's operations. Designing adaptable explanation strategies and explainability techniques that effectively convey model behavior, contingent on the specific user group, is crucial [53]. Engaging humans in XAI fundamentally enhances the formulation and advancement of explanations.

ML models become more complex and capable, it becomes increasingly challenging for humans to interpret their behavior and make decisions based on their output. This is where XAI comes in, which aims to make ML models more transparent and understandable to humans. XAI methods involve developing algorithms and models that not only provide accurate predictions but also offer clear explanations for their decisions. This can help users, including both experts and laypeople, to better understand how the model works and why it is making certain decisions. XAI techniques can also help to identify potential biases and errors in the model and allow for more effective and ethical use of AI in various domains, such as healthcare, finance, and security. By incorporating XAI methods into the development and deployment of ML models, we can ensure that humans remain in control of the decision-making process, while also taking advantage of the incredible power and efficiency of AI systems [54].

III. APPLICATIONS

Various applications of HITL learning are summarized in Fig. 5 and explained in detail in the following subsections.

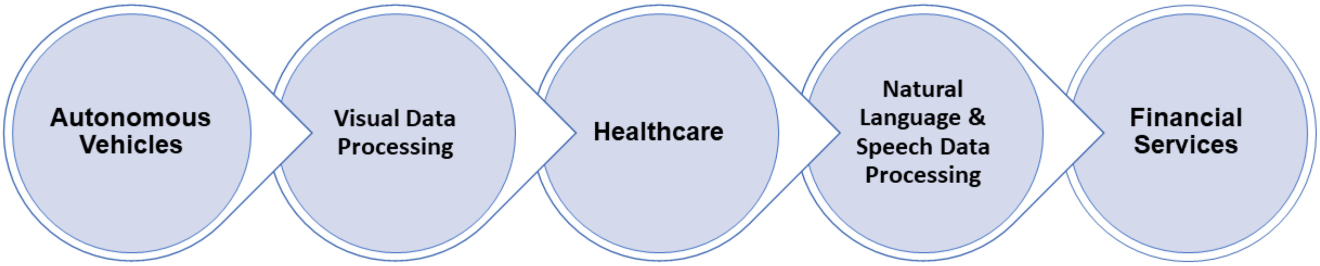


FIGURE 5. Applications of Human-in-the-loop learning.

A. AUTONOMOUS VEHICLES

The potential benefits of autonomous vehicles are road safety and efficient transportation. However, the need for human factors in autonomous vehicles is vital to gain public trust and ensure operational safety from any type of malfunction. HITL-based autonomous driving could help to address these concerns and pave the way for the widespread adoption of autonomous vehicles. Wang et al. [55] present a human-centered feed-forward control (HCFC) system designed for enhancing vehicle steering performance and minimizing driver workload. The proposed system employs a dynamic control strategy that considers factors like vehicle speed, lateral deviation, yaw error, and steering angle as inputs, generating the anticipated steering ratio for the driver. For determining the control strategy's parameters, drivers are categorized into three groups based on their sensitivity to errors. The HCFC system introduces a human-centered steering system (HCSS) that features an adjustable steering gain, which adapts to the driver's path-following tendencies. This adaptation diminishes driver workload by assisting in maintaining a prescribed path with reduced steering wheel angles and rate of angle change. A series of experiments are executed to assess the system's performance, with a focus on tracking the centerline during a double lane change (DLC). These experiments utilize CarSim in conjunction with a portable driving simulator, all within a fixed-speed environment. The selected drivers encompass three distinct profiles for testing purposes. The outcomes of both simulations and experiments highlight the superiority of the proposed HCSS with the dynamic control strategy over the conventional steering ratio-based control strategy. Task performance is shown to improve by approximately 7%. This improvement could be due to various factors, such as, clearer guidance, reduced distractions, improved decision-making. Furthermore, the new system results in a reduction of approximately 35% in physical workload and about 50% in mental workload for drivers following the prescribed path. This reduction could be due to automation, improved ergonomics, reduced decision-making burden, reduced information overload, improved situational awareness and reduced monitoring. In [56], the authors develop a 3D virtual simulation platform that incorporates human input into the evaluation of autonomous vehicle performance. The system is designed to provide end-to-end support for the development of autonomous vehicle controllers, including the generation and collection of human driving data, as well as the testing of controllers in critical scenarios in a virtual environment. The study focuses on the configuration of the system, which is intended to provide a comprehensive platform for the development and testing of autonomous vehicle technology. This includes the use of human driving data to improve the performance of autonomous vehicles and the testing of autonomous vehicle controllers in a simulated environment to ensure their safety in real-world scenarios. The system is designed to be flexible and adaptable, allowing for the integration of new technologies and approaches as they emerge in the field of autonomous vehicle development. Ultimately, the goal of the study is to provide a powerful and effective tool for the continued advancement of autonomous vehicle technology, with a focus on safety and performance.

In [57], the authors introduce an intelligent haptic interface aimed at improving the driving state recovery and control performance of human drivers during takeover situations in automated vehicles. The foundation of this interface rests on a novel two-phase human-machine interaction model. This model applies haptic torque to the steering wheel and dynamically switches its function between predictive guidance and haptic assistance, contingent upon the human driver's state and control capabilities. To ascertain the effectiveness of their approach, the authors conducted vehicle experiments involving 26 participants. The results demonstrate that the proposed method substantially enhances the driving state recovery and control performance of human drivers during takeover scenarios. Importantly, this approach surpasses an existing method in terms of both the safety and smoothness of human-machine interaction. Hang et al. [58] presents a collaborative control framework where the human driver assumes the role of the primary controller. An active rear steering (ARS) system is integrated to counterbalance any control deviations introduced by the human driver, using real-time front steering angle information. The primary objective of this approach is to enhance the vehicle's handling performance, lateral stability, rollover prevention, and path-tracking capabilities. The framework establishes two distinct driving profiles: experienced and inexperienced, in order to validate the efficacy and adaptability of the proposed methodology. Additionally, the approach is subjected to testing under challenging conditions, such as scenarios involving low-adhesion roads and highly curved paths. This evaluation process serves to assess the viability and resilience of the linear time-varying model predictive control (LTV-MPC) algorithm, taking into account inherent parametric uncertainties.

Liu et al. [59] propose a strategy that aims to optimize the driving authority allocation between the driver and the automated driving system (ADS) in real time, based on the driver's behavior and intention. By taking into account the driver's preferences and behavior, the ADS can adjust the driving authority assigned to the driver to improve the overall driving performance and safety. The proposed method uses an online optimization algorithm that considers the driver's intention and behavior prediction and dynamically adjusts the driving authority allocation accordingly.

Wang et al. [60] consider the dynamics of both connected autonomous vehicles (CAVs) and the following human-driven vehicles (HDVs) in an optimization framework to improve holistic energy efficiency, along with car-following safety and traffic efficiency. The HDVs are modeled using the intelligent driver model (IDM) based on the Next Generation Simulation (NGSIM) dataset, which covers a broad range of stochastic and realistic driving behaviors. The study analyzes the quantitative impacts of various driving behaviors on energy efficiency improvement using the proposed control algorithm.

Wu et al. [61] present an innovative approach involving a HITL deep reinforcement learning (DRL) framework. This framework integrates human intelligence in real time during the process of model training. They introduce the Real-Time Hug-DRL method, which they apply to train agents in autonomous driving scenarios, with the primary aim of enhancing both learning efficiency and the overall performance of an off-policy DRL agent. The proposed architecture introduces a dynamic learning process that incorporates human expertise. At each step of the learning process, an evaluation module assesses the relative utilities of the actions guided by humans and those taken by the DRL agent. This dynamic assessment ensures a balanced integration of human insights and the DRL agent's learning trajectory.

Lee and Park [62] presents a new method for autonomous drone control using deep reinforcement learning (DRL) to improve navigation in difficult environments. It focuses on real-time obstacle avoidance through advanced trajectory optimization and sensor-aware control. A unique feature is incorporating human feedback (HCI) to adapt to unforeseen situations. Human as AI mentor-based deep reinforcement learning (HAIM-DRL) [63] framework for mixed traffic platoons shows promise for advancing safe and efficient autonomous driving by incorporating human expertise into the learning process.

B. VISUAL DATA PROCESSING

HITL systems are widely used in visual data processing applications, which involve the processing and analysis of image or video data. Training models with humans in the loop has a long history within the field of computer vision [64], [65], [66], [67]. The most recent approach for handling visual data is the different types of DL-based methods [68], [69]. In order to further improve this, researchers have been looking into how to incorporate human feedback into DL architecture. Interactive labeling, AL, or HITL strategies can be used to provide this input. The system as a whole can become more intelligent and capable of handling complex cases by incorporating human input. This incorporation of human input could help DL techniques perform better on visual data processing tasks.

Object detection is a foundational and complex challenge within computer vision [70]. In recent times, it has garnered substantial attention [71]. The objective is to identify instances of specific object classes within digital images. Despite notable advancements in object detection techniques, the detection of occluded, diminutive, or unclear objects remains a formidable task. To surmount these obstacles, researchers have ventured into incorporating human feedback into the object detection process. In their work [72], authors introduce an interactive object detection framework that enlists human involvement to rectify annotations proposed by a detection system. However, this approach can be both time-intensive and costly. Alternatively, in [73], authors propose a HITL object detection strategy that combines bi-directional deep SORT [74] with annotation-free segment identification. Here, human validation is sought for object candidates that the model cannot autonomously detect. Subsequently, the model is retrained using the additional objects annotated by humans. These methodologies hold promise in enhancing the overall performance of object detection.

Image restoration entails the endeavor to recuperate damaged images [75] to their original condition. There exist two principal avenues in image restoration: exemplar-based approaches [76] and methods rooted in DL [77]. While DL-based techniques hold current prominence and efficacy, they may grapple with overfitting in scenarios with limited training data. Additionally, restored images can contain unfamiliar artifacts due to the absence of semantic information in severely corrupted regions. To confront these challenges, Weber et al. [78] present an interactive ML system for image restoration, built upon the foundation of deep image prior (DIP) [79]. This approach involves the infusion of human insights into the training process. In the domain of Electron Microscopy, Roels et al. [80] propose a hybrid HITL framework that synergizes expert microscopy knowledge with image restoration algorithms, culminating in an enhancement of image quality. These HITL paradigms offer the potential to heighten the precision and efficacy of image restoration by harnessing human operators' expertise alongside automated algorithms.

Image segmentation constitutes a pivotal stage in numerous computer vision applications [81], involving the assignment of class labels to individual pixels within an image [82]. While this field has witnessed remarkable growth, scant attention has been directed towards effectively identifying and rectifying shortcomings in high-performing semantic segmentation models. To tackle this concern, Wang et al. [83] propose a dual-stage hybrid system that leverages human involvement to troubleshoot pixel-level image labeling models. This approach commences by autonomously selecting informative unlabeled images, thereby revealing vulnerabilities in the target model's performance. Subsequently, human experts filter this set to create a refined subset, which is then employed for fine-tuning and retraining the target model. Within the realm of medical image processing, data annotation can be intricate and resource-intensive [84]. Ravanbakhsh et al. [85] introduce a training protocol that synergistically combines conditional generative adversarial networks (cGAN) with human annotators. Human experts annotate complex cases, and the newly annotated images perpetuate the training and inference process. These strategies exhibit the potential of human-machine synergy in elevating the accuracy and efficacy of image segmentation models.

Image enhancement (IE) is a significant task in computer vision, aimed at improving image quality for specific applications [86]. However, most existing IE methods ignore user preferences, resulting in sub-optimal results that may not meet the user's needs. To overcome this issue, several researchers have proposed user-guided image enhancement methods. For instance, Murata and Dobashi [87] introduced a framework that learns the user's preference via user-provided examples and score-based feedback. Similarly, Fischer et al. [88] developed a system called neural image correction and enhancement routine (NICER), which allows users to modify image manipulation parameters and guide the optimization process towards satisfying local optima. These user-guided IE methods can provide end-users with more personalized and satisfactory image enhancement results.

Video object segmentation (VOS) stands as a pivotal computer vision task, focused on segmenting an object instance throughout an entire video sequence based on either a manually or automatically selected first frame [89]. Despite the growing prominence of the VOS research field, its intricacy persists due to the inherent characteristics of videos, such as motion blur and occlusion. To grapple with these challenges, HITL frameworks have been embraced to secure precise outcomes. Benard and Gygli [90] introduce an inventive interactive VOS approach that treats the prevailing segmentation mask as an additional input. Similarly, Oh et al. [91] bring forth Interaction-and-Propagation Networks (IPN), permitting iterative interactions between individuals and the proposed model. In this process, feedback is provided through scribbles across multiple frames during the interactive phase. These frameworks hold promise in delivering accurate VOS outcomes, aided by human input, thereby mitigating the necessity for manual annotation and curtailing time consumption.

Some researchers [92], [93] have been directed toward diminishing the burden of annotation costs in the context of HITL model learning. In the study by Abad et al. [92], an effort was made to train crowdsource workers through an iterative human-machine collaboration mechanism. This process involves a classifier—essentially, a machine learning model—that selects the most superior examples to train the crowdsource workers (i.e., humans). Subsequently, these workers annotate the lower-quality examples, and the resulting annotated data is employed to retrain the classifier with more precise examples. This iterative approach consistently enhances the quality of the training data. The effectiveness of this method was demonstrated through its application to two distinct tasks: Relation Extraction and Community Question Answering, conducted in English and Arabic languages, respectively. The experimental outcomes exhibited a substantial improvement in creating Gold Standard data in comparison to utilizing distant supervision or resorting to crowdsourcing devoid of worker training. Notably, the method approached the performance levels of state-of-the-art techniques employing costly Gold Standard for worker training. In another work by Ravanbakhsh et al. [93], a novel strategy is presented for executing image segmentation within a semi-supervised setup, leveraging a HITL framework alongside a conditional Generative Adversarial Network (cGAN). This method harnesses the discriminator in the cGAN to identify slices with questionable reliability, warranting expert annotation. Meanwhile, the generator synthesizes segmentations for unlabeled data that the model deems confident about. Evaluation on a widely recognized benchmark demonstrates the approach's comparable performance to state-of-the-art fully supervised methods in slice-level evaluation, demanding notably less annotated data. These findings indicate that the proposed approach has the potential to considerably curtail the volume of expert annotation required, all while upholding elevated levels of segmentation precision.

Wang et al. [94] introduces the human-in-the-loop based deep neural networks (H-DNNs), where human input helps achieve better performance with optimized resource usage. Enhancing various mobile applications powered by DNNs, such as voice assistants, augmented reality, and image recognition. In [95], the authors present a new framework called HITL Video Semantic segmentation Auto-annotation (HVSA) that utilizes a HITL approach to generate semantic segmentation annotations for videos while only requiring a small annotation budget. The framework incorporates an active sample selection algorithm to choose the most important samples for manual annotations, and a test-time fine-tuning algorithm to propagate the annotations to the entire video. The results of real-world experiments demonstrate that the proposed approach generates highly accurate and consistent annotations while keeping annotation costs low. Controllable fine-grained text2face (CFTF) [96] offers a promising approach for generating faces with a high degree of user control, opening up possibilities in various fields beyond the initial suspect portrait application. Deng et al. [97] developed the potential of human-in-the-loop learning approaches for developing intelligent robot vision systems that can effectively handle complex tasks, paving the way for advancements in various applications requiring robot-human collaboration.

C. HEALTHCARE

Fully automatic DL-based approaches become the leading technique in various healthcare applications, including computer-aided detection, diagnosis, treatment planning, interventions, and therapy. Despite its advanced capabilities, medical image/signal analysis poses unique challenges, making the involvement of a human end-user advantageous in any DL-enabled system. Kim et al. [98] present concept activation vectors (CAVs) that enable neural networks' internal states to be interpreted in a way that is easy for humans to understand. Instead of being a hindrance, the high-dimensional internal state of a neural network can be leveraged to gain insights. Testing with CAVs (TCAV) is a technique that utilizes directional derivatives to measure how crucial a user-defined concept is to a classification outcome. For instance, it quantifies the extent to which the presence of stripes influences the prediction of a zebra. By applying CAVs in the realm of image classification and demonstrating how they can be used to investigate theories and produce insights for a standard image classification network and a medical application. Cai et al. [99] discuss the development and evaluation of tools designed to assist pathologists in searching for similar images retrieved using a DL algorithm. The authors identified the needs of pathologists in this context and developed tools that allow users to refine their search on the fly, communicating what types of similarity are most important at different moments in time. The assessment of these tools encompassed two studies involving pathologists, revealing that the refinement tools amplified the diagnostic effectiveness of identified images and augmented user confidence in the algorithm's outcomes. Moreover, these tools were favored over a conventional interface, without any discernible compromise in diagnostic precision. The authors additionally noted a shift in user strategies when engaging with refinement tools, employing them to explore and grasp the underlying algorithm's functioning, and to distinguish between errors stemming from machine learning and those originating from their own evaluations.

In [100], the authors delve into the enhancement of human performance in deception detection through machine-learning models while preserving human agency. They introduce a spectrum encompassing complete human agency and full automation, then proceed to devise various levels of machine assistance that gradually augment the influence of machine predictions. The study unveils that presenting explanations alone slightly enhances human performance, while displaying predicted labels yields a significant improvement (>20% relative improvement). Additionally, explicitly suggesting strong machine performance further elevates human effectiveness. Interestingly, when predicted labels are showcased, explanations of machine predictions yield accuracy levels akin to an explicit declaration of robust machine performance. These findings underscore the equilibrium between human performance and agency, emphasizing that explanations of machine predictions can mediate this balance. Bansal et al. [101] focalize on two pivotal attributes of AI error boundaries—parsimony and stochasticity—and their influence on human comprehension of AI capabilities and collaborative team performance. Their investigation delves into these attributes within the context of task dimensionality, scrutinizing their interplay with existing research. The authors advocate considering objectives beyond accuracy during model selection and optimization to optimize human-AI team performance. Their experimental results demonstrate the impacts of these attributes on team performance and the shaping of mental models regarding AI capabilities. Overall, these findings emphasize the necessity of comprehending the attributes of an AI's error boundary to effectively enhance human-AI team performance. In [102], the authors explore the informational requirements of medical experts when initially introduced to a diagnostic AI assistant. They conducted interviews with 21 pathologists before, during, and after exposing them to deep neural network predictions for prostate cancer diagnosis. The results underscore that clinicians seek upfront information regarding fundamental, overarching aspects of the model, including its known strengths, limitations, subjective viewpoint, and overall design intent. Participants likened these informational needs to the collaborative mental models they developed when seeking second opinions from medical colleagues. This study underscores the importance of furnishing medical experts with information that surpasses the localized, case-specific rationale behind individual model decisions. These findings contribute to discussions surrounding AI transparency for collaborative decision-making, providing insights into what experts deem crucial when acquainting themselves with AI assistants prior to their integration into routine practice. In [103], the authors new approach to chest radiograph diagnosis that combines swarm-based technology to amplify the diagnostic accuracy of networked human groups with AI capabilities.

Beede et al. [104], conducted a human-centered study of a DL system used in clinics for the detection of diabetic eye disease. The study involved interviews and observations across eleven clinics in Thailand to understand the current eye-screening workflows, user expectations for an AI-assisted screening process, and post-deployment experiences. The authors found that several socio-environmental factors impact model performance, nursing workflows, and patient experience. For instance, the availability of a quiet and well-lit screening area significantly impacted the accuracy of the model. Additionally, nursing workflows were found to be affected by the integration of the DL system, leading to changes in how nurses conducted eye screenings. The study highlights the importance of conducting human-centered evaluative research alongside prospective evaluations of model accuracy. By doing so, researchers can better understand how the system is being used in real-world contexts and identify areas for improvement beyond technical aspects such as model accuracy. Tschand et al. [105] reported the effects of varied representations of AI-based support on diagnostic accuracy across different levels of clinical expertise and multiple clinical workflows, using skin cancer diagnosis as a case study. The study found that good quality AI-based support of clinical decision-making improved diagnostic accuracy over that of either AI or physicians alone and that the least experienced clinicians gained the most from AI-based support. The study also compared AI-based multiclass probabilities to content-based image retrieval (CBIR) representations of AI in the mobile technology environment and found that AI-based multiclass probabilities outperformed CBIR representations. Additionally, the study demonstrated the utility of AI-based support in simulations of second opinions and telemedicine triage. The study highlights the potential benefits associated with good quality AI in the hands of non-expert clinicians and the potential for faulty AI to mislead clinicians across the spectrum of expertise. Furthermore, the study shows how insights derived from AI class-activation maps can inform improvements in human diagnosis.

Steyvers et al. [106] devised a Bayesian framework that harmonizes the forecasts and confidence scores generated by both humans and machines, enhancing the performance of image classification tasks. The study showcases that a hybrid amalgamation of human and machine predictions can yield superior outcomes compared to relying solely on either human or machine predictions. This holds true as long as the accuracy disparities between the two realms fall within a specific range dictated by the underlying correlation between their confidence scores. Furthermore, the authors unveil that refining hybrid human-machine performance can be achieved by distinguishing between the errors made by human and machine classifiers across distinct class labels. Notably, the study emphasizes the benefits of incorporating human confidence ratings into the Bayesian fusion model, leading to heightened performance for the hybrid approach. Gu et al. [107] introduced NaviPath, a collaborative human-AI navigation system tailored for pathologists to enhance their navigation processes. This system seamlessly integrates domain knowledge and workflow considerations into practice. NaviPath alleviates pathologists' burdens by employing an AI-assisted algorithm to automate navigation, while simultaneously enhancing their work through a collaborative workflow. By intertwining domain expertise and practical workflow integration, NaviPath exhibits the potential to both automate navigation and enhance pathologists' tasks. The collaborative workflow within NaviPath is poised to reduce pathologists' workloads and elevate their navigation efficiency, consequently bolstering the quality of examinations. The commendable aspect is how NaviPath bridges the divide between medical professionals and AI, embedding doctors' domain knowledge and affording them the ability to delegate tasks to AI based on their preferences. Equally impressive is the outcome of the user evaluation study, where medical professionals affirmed that the human-AI system enhanced navigation efficiency and bolstered examination quality. The findings of this study hold the promise of substantial implications for medical decision-making and provide valuable insights for HCI researchers in crafting collaborative AI systems for medical professionals.

Sharma et al. [108] engineered an AI-in-the-loop agent named HAILEY, designed to furnish timely feedback to peer supporters on the TalkLife online peer-to-peer support platform. A randomized controlled trial involving 300 participants was conducted, revealing a noteworthy 19.6% elevation in conversational empathy between peers overall through their approach. Notably, within the subgroup of peer supporters facing challenges in providing support, a more substantial increase of 38.9% was observed. An analysis of human-AI collaboration patterns indicated that peer supporters adeptly utilized AI feedback, both directly and indirectly, without undue dependence on AI. Participants also reported heightened self-efficacy post-feedback. This study effectively underscores the potential of feedback-driven, AI-in-the-loop writing systems to empower individuals in socially significant tasks such as empathetic conversations. Cabitza et al. [109] scrutinize human-AI collaboration protocols, a design construct focused on evaluating how humans and AI harmoniously collaborate in cognitive tasks. This study encompassed two user studies involving 12 radiologists and 44 ECG readers who evaluated cases across diverse collaboration configurations. The study underscores that AI support offers assistance, yet the deployment of eXplainable AI (XAI) might trigger a "white-box paradox," leading to neutral or counterproductive effects. The sequence of presentation also emerges as a critical factor, with AI-first protocols correlating with higher diagnostic accuracy in comparison to human-first protocols, as well as standalone humans and AI. The study unveils the optimal conditions for AI to enhance human diagnostic capabilities while circumventing cognitive biases that could impair decision-making effectiveness. Zhou et al. [110] advocate a video-based augmented reality system for HITL assessment of muscle strength in children with Juvenile Dermatomyositis (JDM). This system incorporates an Automatic Quantitative Assessment (AQA) algorithm for JDM muscle strength assessment, relying on contrastive regression and trained on a dedicated JDM dataset. The AQA results manifest as a virtual character via a 3D animation dataset, allowing users to juxtapose real-world patients with the virtual character to comprehend and verify AQA outcomes. In [111], the authors presented a proof-of-concept system to demonstrate its feasibility for building on user needs and their willingness to participate in Interactive Machine Learning (IML) solutions. This system was evaluated through a prototype Internet of Things (IoT) application called "The Smart Drink Monitoring System".

Human Correction of AI-Generated Labels (H-COAL) [112] framework allows selective human correction of AI-generated labels based on their confidence scores. Significantly closes the gap between AI and human-labeled model performance up to 86% improvement with 20% correction. Liang et al. [113] present by combining transfer learning, active learning, and human-in-the-loop interaction. Extracts crucial biomedical information from under-resourced languages. Reduces dependence on large amounts of manually annotated data, saving time and resources. Agile Modeling [114] groundbreaking method empowers everyday users, not just experts, to create vision models for subjective concepts. Forget labeling countless images or grappling with complex algorithms. Imagine shaping a model's understanding through intuitive interaction, providing feedback and examples in real-time.

D. NATURAL LANGUAGE AND SPEECH DATA PROCESSING

AI has achieved remarkable strides in the realm of natural language and speech data processing. Researchers have harnessed diverse techniques to train and deduce experimental outcomes. Notably, recognizing the varying levels of human creativity has emerged as a key factor in enhancing the accuracy of these methodologies. The evolving landscape of data availability and computational prowess suggests that AI's momentum in these domains will persist, driving significant advancements and breakthroughs in the times ahead. Within the domain of Opinion Mining (OM), Sentiment Analysis (SA) holds a pivotal position, focusing on computationally dissecting individuals' attitudes and viewpoints towards entities mentioned within text. In recent times, a profusion of neural network-based strategies has gained prominence, showcasing their efficacy in addressing sentiment analysis tasks [115], [116], [117]. DL based techniques have notably dominated the SA landscape, showcasing robust accuracy and F1 scores in sentiment prediction. However, these metrics offer limited insight into the rationale behind erroneous predictions [118]. To bridge this gap, Liu et al. [119] introduced an explainable HITL SA framework. This framework orchestrates a data perturbation process to dissect local feature contributions, aggregates these local features to derive comprehensible global-level attributes, and involves human assessment of top-ranked global features to gauge their relevance to the ground truth and identify errors. Subsequently, the system computes an error score based on both global and local sentimental features. Predictions bearing scores surpassing a predefined threshold are classified as inaccurate predictions. This proposed framework offers a holistic assessment of sentiment analysis models, while its inherent explainability holds the potential to foster trust and widespread adoption of these models in practical applications.

Text Classification (TC) is another important NLP task that involves categorizing text into specific categories. Researchers have orchestrated diverse methodologies to elevate the precision of TC systems. For instance, Karmakharm et al. [120] have put forth a rumor classification system that taps into feedback from journalists. This feedback is employed to retrain a machine-learning model, enhancing its accuracy. Given that many contemporary TC approaches pivot on deep neural networks [115], [121], which are often perceived as "black boxes," Arous et al. [122] have ventured into developing an augmented human-AI framework named MARTA. The core objective of MARTA is to render these models more interpretable. This novel framework embraces a Bayesian paradigm that iteratively refines model parameters and human reliability, fostering mutual enrichment until alignment is achieved between labels and rationales.

In addition to text classification, there is a growing interest in HITL approaches for syntactic and semantic parsing. Syntactic parsing entails extracting the valid syntactic structure from input sentences, whereas semantic parsing involves mapping natural language to formal, domain-specific semantic representations. A noteworthy instance in this realm is the HITL parsing method [123] designed to elevate the precision of Combinatory Categorial Grammar (CCG) parsing. This ingenious approach leverages non-experts to respond to straightforward questions generated from the parser's output. These responses are then assimilated as soft constraints during the model's retraining phase. Yet, parsing technologies continue to grapple with assorted challenges, encompassing user input ambiguity, suboptimal performance of contemporary parsers, and the dearth of elucidation within neural network-based models. Addressing these quandaries, researchers have introduced the concept of interactive semantic parsing [124]. Furthermore, a model-based interactive semantic parsing framework [125] has been formulated as a universal principle for interactive semantic parsing. These innovative strategies present encouraging potential in enhancing the accuracy and explicability of syntactic and semantic parsing systems.

HITL paradigms are making inroads into the realm of text summarization, a process centered on crafting concise versions of texts while upholding their underlying significance [126]. Notably, in a study by Ziegler et al. [127], endeavors were made to refine pre-trained language models through reinforcement learning, employing a reward model derived from human preferences. This methodology has been wielded to engender summaries across diverse datasets, encompassing Reddit TL, DR, and CNN/DM. Nonetheless, disparities in perspectives between labelers and researchers can yield low concordance rates, curtailing the efficacy of this approach. To surmount this challenge, another strategy proposed in Stiennon et al. [128] entails constructing a dataset of human preferences by juxtaposing pairs of summaries. This dataset serves as fodder to train a reward model via supervised learning. Subsequently, a policy is honed through reinforcement learning, aiming to maximize the reward model's generated score. This innovative methodology displays promise in heightening the accord between labelers and researchers, whilst also segregating policy and value networks to enhance the efficiency of text summarization processes.

Recently, there has been a growing interest in designing frameworks that incorporate human feedback into dialogue and Question Answering (QA) systems. These frameworks, referred to as HITL intelligent systems, can be broadly categorized into two types: online and offline feedback loops [129]. In the context of the online feedback loop, as exemplified by Hancock et al. [130], human feedback is persistently harnessed to iteratively refine the model through reinforcement learning. This dynamic approach has evidenced its potency in enhancing the efficacy of chatbots and agents. By perpetually incorporating human input, these systems continually generate novel instances and undergo self-retraining, leading to performance improvements. Conversely, the offline feedback loop entails amassing a substantial corpus of human feedback for model enhancement. Wallace et al. [131] have pursued this path to fortify system robustness, recognizing that some end-user feedback may be misleading. Through judiciously integrating such feedback, they endeavored to bolster system resilience. Across the spectrum of these frameworks, researchers have substantiated their efficacy, poised to markedly contribute to the advancement of more natural and adept conversational agents.

HITL techniques have been widely applied to various NLP tasks for better performance, interpretability, and usability. Empirical outcomes from these endeavors underscore that a relatively modest corpus of human feedback can substantially augment model efficacy in text classification [120], dialogue interactions, and question-answering domains [130]. Moreover, these techniques have demonstrated prowess in enhancing model robustness and generalization [128]. Beyond their performance-enhancing attributes, HITL approaches have also exhibited the capacity to render models more interpretable and comprehensible to humans. To exemplify, Arous et al. [122] ingeniously wove human rationales into an attention-based Bayesian framework, thus yielding a classification interpretation that is more human-intelligible. Similarly, Liu et al. [119] leveraged uni-grams as explainable features for LIME [132], aiding end-users in comprehending the impact of each word on the final sentiment classification rendered by the model. Further showcasing the breadth of application, Wallace et al. [131] engaged "trivia enthusiasts" to imaginatively devise specific adversarial questions, offering insights into the inherent attributes of intelligent question-answering systems.

Stiennon et al. [128] presented a quartet of significant contributions in their study: (1) the efficacy of training summarization models using human feedback, (2) the robust generalizability of human feedback models across diverse domains, (3) an extensive empirical exploration of their policy and reward model, and (4) the public release of their human feedback dataset for further exploration and research. The authors posit that their methods address growing concerns regarding the alignment of AI systems with human values and preferences, particularly as AI systems wield increasing power and undertake more critical tasks. By amassing an extensive dataset of human comparisons between summaries, they trained a model to predict human-preferred summaries, employing this as a reward mechanism to refine summarization policies through reinforcement learning. Their approach was applied to the TL; DR dataset from Reddit posts, where their models notably outperformed both human reference summaries and larger models that underwent supervised learning-based fine-tuning exclusively. Moreover, the authors demonstrated the transference of their models to CNN/DM news articles, yielding summaries almost as proficient as human reference summaries without necessitating news-specific fine-tuning.

Fan et al. [133] introduced the Nested HITL Reward Learning algorithm NANO, designed to utilize human feedback for generating text adhering to diverse distributions. The NANO algorithm features an outer loop that comprises three distinct phases: generation, human feedback collection, and model training. Within this framework, an inner loop conducts a tree search, with nodes sampled from a language model. The study demonstrates that integrating human feedback into the training process enhances performance across quantified and unquantified distributions, and attains personalization using only a limited 64 labels per individual. NANO also attains state-of-the-art outcomes in governing quantified distributions and generating content with specific topics/attributes even with few-shot data. The significance of multi-iteration human feedback is underscored through ablation studies. This approach finds applications in domains such as chatbots, automated writing, and tailored recommendation systems. Dong et al. [134] proposed a novel technique termed HITL-based swarm learning (HBSL) to amplify the effectiveness of swarm learning in counterfeit news detection. The HBSL method incorporates user feedback into the swarm learning process, forming a loop consisting of three stages: local learning, model updating, and human feedback. The loop continues until predefined stopping criteria are met. During local learning, nodes independently learn detection models on their respective local data. In the model updating phase, the primary node updates models by averaging model weights. In the human feedback stage, user-provided feedback on test data predictions augments training data to enhance detection accuracy. Experimental results on the LIAR dataset reveal that the proposed HBSL method outperforms traditional swarm learning for decentralized counterfeit news detection. The study's contributions encompass the introduction of HBSL and its successful validation on the LIAR dataset, showcasing notable enhancement in counterfeit news detection performance for each node through local training coupled with user feedback.

Bonet-Jover et al. [135] proposed a semi-automatic annotation methodology that combines summarization and human-in-the-loop techniques to create resources for disinformation detection.Achieves high accuracy in both reliability detection (0.95) and veracity detection (0.78). Introduce MArBLE, a hierarchical multi-armed bandit method for human-in-the-Loop set expansion [136]. With MArBLE, an expert sees suggestions from different models one by one and decides whether to accept or reject them. MArBLE learns which models are best for this specific task and suggests accordingly. The Human-in-the-Loop methodology for semi-automatic annotation combines human expertise with machine assistance to accelerate the process [137]. It is applied to construct a Spanish news dataset for assessing content reliability, known as the RUN dataset. Reduced annotation time up to 64% faster compared to fully manual methods.

E. FINANCIAL SERVICES

The significance of human involvement and supervision is to ensure that artificial intelligence (AI) operates responsibly and in line with desired objectives in the financial domain. Asthana et al. [138] proposed a novel methodology called DaME (Data Mapping Engine) that utilizes HITL techniques and trains a data mapping engine to perform data mapping. The results from propped method evaluation on a financial services dataset in an industrial application have been promising, as it has significantly reduced the manual effort required for data mapping and enabled learning reuse. In comparison to the existing state-of-the-art, our dataset achieved a much higher accuracy rate of 69%, compared to the previous accuracy rate of 34%. DaME has helped improve the productivity of industry practitioners by saving them 14,000 hours of manual mapping work over a period of ten months.

Yasir et al. [139] discover the obstacles that prevent people from participating in the financial sector and suggest AI as a solution to these problems. To achieve this, the literature on AI and financial inclusion was examined, and examples of AI implementation in finance were provided to support the argument. The study concentrates on the barriers that have a potential solution in AI. Additionally, the authors discuss the challenges of adopting AI to enhance financial inclusion. Ultimately, this research serves as a guide for economies to recognize the importance of AI in achieving financial inclusion. Ding et al. [140], introduce ALARM1, short for Analyst-in-the-Loop Anomaly Reasoning and Management, as an end-to-end framework. ALARM1 facilitates the entire anomaly mining cycle, spanning from anomaly detection to subsequent actions. Beyond its capacity for autonomously detecting emerging anomalies, the framework provides explanations for anomalies and incorporates an interactive graphical user interface (GUI) to engage HITL processes. These processes encompass visual exploration, comprehension enhancement, and the formulation of new detection rules. These new rules supplement the rule-based supervised detection methods commonly utilized in various operational systems, thus closing the loop in the anomaly management process. The study effectively showcases the capabilities of ALARM1 through diverse case studies involving fraud analysts from the financial sector.

Buckley et al. [141] addressed the growing significance of AI in finance by emphasizing human accountability. The primary concern is the AI ''black box'' problem, where the AI may generate unexpected or unwanted outcomes that are not recognized or predicted due to the complexity of its internal workings or its independent operation without human supervision. The article explores the various applications of AI in finance, its rapid progress, and the potential issues and regulatory challenges it poses. It argues that effective regulatory measures involve personal responsibility regimes that incorporate human oversight, eliminating the black box defense for AI decision-making and operations. Zetzsche et al. [142] proposed a regulatory roadmap for the implementation of AI in finance that prioritizes human responsibility and involvement. They provide examples of AI usage in the finance industry and outline potential challenges that may arise. The article discusses the regulatory challenges involved and the tools that may be utilized. Key concerns include information asymmetries, data dependencies, and system interdependencies, which can result in unexpected outcomes.

Truby et al. [143] suggested that to promote a sustainable future in AI innovation within the financial sector, it's crucial to adopt a proactive regulatory approach that emphasizes preventive measures rather than reactive ones. This approach should involve creating rational regulations that align with jurisdiction-specific guidelines and carefully crafted international principles. The objective is to prevent any financial harm before it happens.

IV. CHALLENGES

HITL systems are used in a variety of domains and they offer several benefits over traditional ML systems, such as increased accuracy, transparency, and the ability to handle complex tasks. However, HITL systems also pose several challenges, ranging from human factors to technical challenges, as shown in Fig. 6. In this section, we will discuss the challenges of HITL and potential solutions to address them.

A. HUMAN FACTORS

One of the most significant challenges of HITL is the involvement of humans in the decision-making process [27], [144], [145], [146], [147]. Humans can introduce bias, subjectivity, and inconsistency into the system, which can negatively

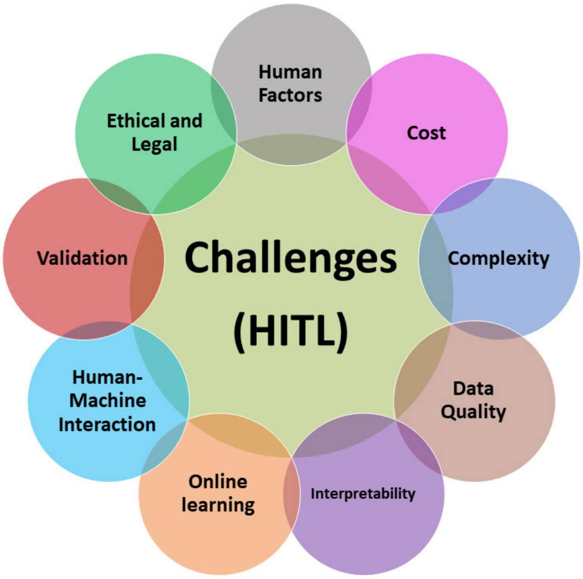


FIGURE 6. Challenges of Human-in-the-loop learning.

impact the performance of the ML system. For instance, if a human annotator labels a dataset incorrectly, the model's performance will be affected, resulting in poor predictions. To address this challenge, organizations can adopt several strategies. First, they can use multiple human annotators to label data and aggregate the results to improve the quality of annotations. Second, they can use quality control measures, such as inter-annotator agreement, to ensure that the annotations are consistent and accurate. Third, they can use ML algorithms to detect and correct errors in the labeled data. Another challenge related to human factors is the scalability of human input. As the amount of data increases, it becomes increasingly difficult for humans to provide input at the same rate. This can lead to bottlenecks in the learning process, slowing down the system's performance. To address this challenge, organizations can use specialized tasks, where humans only provide input on specific parts of the data, allowing them to focus on their areas of expertise. They can also automate some tasks, reducing the burden on human annotators. Crowdsourcing is another option, where organizations can tap into a global pool of workers to provide input.

B. COST

The cost of HITL can be high due to several factors, including the need for human experts, training costs, and the time required for human input [27], [148], [149]. One of the primary costs associated with HITL is the need for human experts. In some cases, HITL may require the involvement of highly skilled and experienced experts to provide input and ensure that the learning system produces accurate and relevant results. These experts may have high salaries or require specialized training, which can significantly increase the overall cost of the HITL system.

Another factor that can impact the cost of HITL is the training required for human input. In some cases, HITL systems may require extensive training for humans to provide input that is accurate and consistent. This can be especially true for complex tasks or processes that require specialized knowledge or skills. Training costs can include not only the cost of training materials but also the cost of the trainer's time and expertise.

The time required for human input is another significant cost associated with HITL. As data volumes increase, the time required for humans to provide input can become a bottleneck in the learning process, delaying the delivery of results and increasing the cost of the system. Additionally, the need for human input can make the learning process more time-consuming and slower than fully automated ML systems.

Fortunately, there are several strategies that organizations can use to mitigate the cost of HITL. One approach is to use automation wherever possible to reduce the amount of human input required. For example, automated pre-processing and feature extraction can reduce the amount of manual data preparation required, while automated quality control and feedback mechanisms can ensure that human input is accurate and consistent. Another strategy is to specialize tasks between humans and machines. This approach involves identifying tasks that are best performed by humans and tasks that are best performed by machines, and then assigning those tasks accordingly. For example, humans may be better suited to tasks that require judgment or reasoning, while machines may be better suited to tasks that require processing large volumes of data quickly and accurately.

Crowdsourcing is another strategy that can help reduce the cost of HITL [150], [151]. Crowdsourcing involves soliciting input from a large number of individuals, often via online platforms, to complete small, discrete tasks. This approach can be especially useful for tasks that require human input but do not require specialized knowledge or expertise. Crowdsourcing can be an effective way to tap into a large pool of talent, reduce costs, and speed up the learning process. Continuous training is another strategy that can help reduce the cost of HITL. By continuously updating the learning system with new data and feedback, organizations can improve the accuracy and relevance of the system over time. Continuous training can also help reduce the need for extensive initial training, as the system can learn and adapt over time.

C. COMPLEXITY

HITL learning involves the integration of human expertise and judgment into the ML process. However, this integration also adds complexity to the learning process. In this section, we will discuss the challenges of complexity for HITL learning.

One of the primary challenges of complexity in HITL learning is the need for specialized skills and knowledge [148], [152]. Unlike traditional ML, HITL learning requires the involvement of experts in various fields to provide domain-specific knowledge and feedback. For example, in the healthcare industry, HITL learning may require the input of medical professionals to provide insights into the effectiveness of treatments or the diagnosis of diseases. Similarly, in the financial industry, experts in risk management may be required to provide input on the identification and mitigation of potential risks.

This need for specialized skills and knowledge can create challenges in the recruitment and retention of experts for HITL learning. It may be difficult to find and attract experts who are willing to devote their time and effort to the learning process. Additionally, experts may require compensation for their participation, adding to the cost of HITL learning. Another challenge of complexity in HITL learning is the need to manage the feedback provided by experts. Unlike traditional ML, HITL learning involves the integration of feedback from multiple experts, which can be conflicting or inconsistent. Managing and reconciling this feedback can be a complex and time-consuming process. Furthermore, the feedback provided by experts may not always be accurate or reliable. Experts may have biases or make errors, which can negatively impact the learning process. For example, in the legal industry, HITL learning may require input from lawyers on the interpretation of legal statutes. However, lawyers may have differing opinions on the interpretation of the law, which can make it difficult to determine the correct answer.

In addition to the challenges of managing feedback from experts, HITL learning also requires the integration of feedback from end-users. End-users provide feedback on the usability and effectiveness of the ML system. However, this feedback can be subjective and inconsistent, making it challenging to incorporate into the learning process. Another challenge of complexity in HITL learning is the need for interpretability and transparency. As discussed earlier, HITL learning requires the use of interpretable ML models that can be easily understood and interpreted by experts and end-users. This interpretability is necessary to ensure that the ML system is making decisions that align with the expectations and needs of experts and end-users.

However, achieving interpretability and transparency in HITL learning can be challenging. ML models can be complex and difficult to interpret, especially when they involve the integration of feedback from multiple experts and end-users. Additionally, the integration of feedback from experts and end-users can make it difficult to explain how the machine-learning system arrived at a particular decision.

D. QUALITY OF THE DATA

One of the primary challenges of HITL learning is the quality of the data [27], [151], [153]. HITL systems rely on input from humans to improve the performance of the ML models. However, the quality of human input can be subjective, biased, or inconsistent, which can negatively impact the performance of the ML system. One major issue is that humans may have different interpretations of the same data, leading to inconsistencies in labeling or categorization. For example, in a medical diagnosis system, different doctors may interpret the same symptom differently, leading to inconsistent labels for the training data. This can lead to inaccurate models that do not generalize well. Another issue is that humans may introduce biases into the training data, whether intentionally or unintentionally. This can occur when the human labelers have prior beliefs or stereotypes that influence their labeling decisions. For example, in a hiring algorithm, if the human labelers have a bias towards certain demographics, the resulting model will perpetuate that bias.

Additionally, humans may not be able to provide high-quality data consistently as the amount of data increases. This can lead to a bottleneck in the learning process [154], as the system cannot learn as quickly as new data is generated. To address these challenges, several strategies can be employed. One approach is to use multiple human labelers and aggregate their responses to reduce the impact of individual biases or inconsistencies. This approach can also help identify problematic labelers and remove their responses from the dataset. Another approach is to use domain experts as labelers [155], as they have more knowledge and experience in the specific domain and can provide more accurate labels. In the medical diagnosis example, using doctors as labelers can improve the quality of the training data.

Automated quality control mechanisms can also be implemented to identify and correct inconsistencies or errors in the training data. These mechanisms can flag questionable labels for human review or automatically correct obvious errors. Finally, continuous training can help address the issue of scalability by enabling the system to learn and improve over time as more data becomes available. In this approach, humans provide initial labels, but the system can retrain and adjust its model as it receives new data.

E. INTERPRETABILITY

Model interpretability refers to the ability to understand how a machine-learning model arrives at a decision [156], [157]. This is essential for domains such as healthcare and finance, where the decisions made by the system have significant consequences. To address this challenge, organizations can use interpretable ML models, such as decision trees, linear models, and rule-based models. These models are easier to interpret than black-box models such as deep neural networks. Another approach is to use post-hoc methods [158], such as feature importance and local explanations, to explain the model's decisions.

F. ONLINE LEARNING

Online learning refers to the ability of the system to learn continuously from new data, updating the model's parameters as new data arrives. This is critical for domains such as fraud detection and cybersecurity, where the system needs to adapt to new threats. To address this challenge, organizations can use online learning algorithms such as stochastic gradient descent and online linear regression [148], [159], [160]. These algorithms update the model's parameters as new data arrives, allowing the system to adapt to new patterns.

G. HUMAN-MACHINE INTERACTION

Within the HITL framework, the challenge of human-machine interaction is rooted in the intricate nature of collaboration. Despite an expanding body of technical research on human-machine synergy in ML, the fundamental collaboration model and mechanisms remain enigmatic. An array of uncertainties prevails, encompassing the delineation of the collaboration mode, the amalgamation of machine and human outputs (e.g., machine-extracted and human-nominated features) across parallel and sequential contexts, and the orchestration of transitions between both sides for diverse learning tasks. Furthermore, the assignment of roles to human collaborators, whether lay individuals or domain experts, along with optimizing the requisite number of human contributors, introduces heightened complexity. Successfully addressing these pivotal queries is imperative for unveiling the ML collaboration model. Although human intelligence can enrich interpretable features, merging them with machine-generated attributes holds potential for a more potent learning paradigm. Yet, unresolved hurdles persist, chiefly in assigning values to human-nominated features. This task's resource-intensive nature, attributed to the nuanced qualities of many features, necessitates human engagement in labeling. In contrast to the single-label-per-entry scenario in classification tasks, value acquisition involves multiple human interventions per entry, prompting the need for strategies that balance cost against model efficacy. A promising avenue involves curating information-rich features, striving for equilibrium between feature labeling costs and model precision. Furthermore, the integration of human and machine attributes presents a conundrum. While a straightforward approach involves concatenating features into an elongated vector, an alternate strategy entails training distinct machine and human classifiers and subsequently amalgamating them through ensemble learning techniques to bolster predictive accuracy [161], [162].

H. VALIDATION

In the context of HITL learning, validation becomes more challenging due to the involvement of humans in the learning process [152], [163], [164]. This section will discuss some of the challenges associated with validating HITL systems. One of the primary challenges of HITL validation is the subjectivity of human input. Humans can have different opinions and interpretations of the same data, which can lead to inconsistencies in the validation process. For example, in a medical diagnosis system, different doctors may have different opinions on a patient's condition, leading to inconsistencies in the system's performance evaluation. To overcome this challenge, it is essential to have clear and objective criteria for evaluating the performance of HITL systems. These criteria should be defined in collaboration with domain experts and should be based on measurable outcomes [165].

Another challenge of HITL validation is the lack of a ground truth for human input. In traditional ML systems, a ground truth is used to evaluate the system's performance. The ground truth is a set of data that is known to be correct, and the system's output is compared against this ground truth to assess its performance. However, in HITL systems, the ground truth is often missing, as the input provided by humans may not be entirely accurate. For example, in a language translation system, the translation provided by a human may not be the only correct translation. To overcome this challenge, it is essential to have multiple human annotators to provide input and to use statistical methods to evaluate the system's performance. Another challenge of HITL validation is related to the scalability of human input. As the amount of data increases, it becomes increasingly challenging for humans to provide input at the same rate. This can lead to bottlenecks in the learning process and can affect the system's performance. To overcome this challenge, it is essential to use efficient data sampling techniques and to specialize tasks between humans and machines [54].

Furthermore, HITL systems may require ongoing training and adaptation, making validation an ongoing process. As the system continues to learn and adapt, the criteria for validation may also change. To overcome this challenge, it is essential to have a feedback loop in place that allows for continuous evaluation and improvement of the system. Another challenge of HITL validation is related to the ethical implications of human input. In some cases, the input provided by humans may contain biases that can lead to discrimination and unfairness in the system's output. For example, in a hiring system that uses HITL, human input may contain biases based on gender, ethnicity, or age. To overcome this challenge, it is essential to have a diverse group of human annotators and to use statistical methods to identify and correct biases in the system's output.

I. ETHICAL AND LEGAL ISSUES

HITL systems also pose several ethical and legal challenges, such as data privacy and bias [141]. Data privacy refers to the protection of personal information, such as medical records and financial information. HITL systems can pose a risk to data privacy, especially if human annotators have access to sensitive data. To address this challenge, organizations can use techniques such as differential privacy to protect sensitive data. Differential privacy adds noise to the data to ensure that the data cannot be traced back to an individual. Bias is another ethical challenge in HITL systems. Bias can occur when the dataset used to train the model is biased, leading to biased predictions. Leverage the HITL approach, combining human and machine intelligence throughout the entire ethical AI design process [166]. For instance, if a facial recognition system is trained on a dataset.

V. FUTURE DIRECTIONS

HITL systems refer to a type of ML system that involves human input and oversight. These systems are becoming increasingly popular, and as a result, there are many research opportunities in this area. Here are some of them:

- The feedback individuals can provide for systems like chatbots or automatic summarization tools is sparse due to the size of the output space [167]. It's important to explore intelligent questions for syntactic parsing tasks [123] and to consider trust and confidence in user-centered design and evaluation of topic modeling tasks [168]. However, some HITL techniques may allow malicious individuals to train models for their purposes, potentially causing harm to society by exploiting human feedback. This could result in language models being used to manipulate human beliefs, instill radical ideas, commit fraud, and more [128].
- In order to improve image restoration, it is important to optimize predictive parameters using supervised regression models and scientifically analyze correlations between different algorithms using HITL methods [80]. Additionally, in image enhancement tasks, AL can be used to help users estimate cluster membership with the fewest image enhancements. These approaches are outlined in studies on HITL computer vision systems [169].
- Human supervision is preferred due to varying levels of expertise and potential for erroneous decisions as workload increases [170]. To improve NLP and CV, more human feedback datasets should be collected and shared [171]. User credibility should also be considered to evaluate the quality of feedback provided [120]. In addition to model performance, more rigorous user studies should be designed and conducted to assess the effectiveness and robustness of HITL frameworks [168]. For generative tasks, an explicit function can be defined with user feedback to collect and evaluate generated signals [167]. It is important to find an efficient method to dynamically select the most representative and valuable feedback [172]. Finally, a more user-friendly way of displaying the model's learning and the feedback process should be explored through visualization [173].
- To ensure reliability and safety, it is crucial to choose the right time for artificial intervention, particularly for tasks that require high levels of safety and security [174]. When it comes to human-computer interaction systems, users' experience expectations are usually given more importance than performance. Therefore, it is crucial to model sensor signals and create a unified code for both abstract and concrete information to make this process smoother [175]. While human intervention usually revolves around shallow judgments such as acceptance, rejection or direction, it is important to explore more complex feedback for HITL applications [176].
- To improve the efficiency and performance of robotic systems, the research aims to gain a comprehensive understanding of evaluative feedback and preference learning in PlacingBall-Simulation, Reaching, and two real robot tasks. The objective is to develop robotic systems that can effectively assist humans in various tasks and operate in a human-like manner. To achieve this, the findings will be validated by running all experiments with actual human trainers who can provide feedback and help improve the systems' performance [177].
- Developing methods to improve the generation ability of large language models is necessary because the present model is limited by the generation ability of GPT-3 [178]. Future research could focus on developing methods to improve the generation ability of large language models, such as training models on more diverse and representative datasets, developing more effective fine-tuning methods, or designing new architectures that can better capture long-term dependencies.
- Investigating finer-grained human feedback for text generation because the current approach relies on sentence-level feedback, which may not be sufficient for generating high-quality sentences that fully meet the desired attribute/distribution. Exploring finer-grained human feedback, such as rating or rewriting part of a sentence. Future research could investigate the effectiveness of such feedback mechanisms, as well as develop new methods for incorporating them into text generation models. This could involve designing new interfaces for collecting feedback, developing algorithms for processing and incorporating feedback into models, or exploring the use of reinforcement learning to optimize generation based on feedback.
- Developing tutorials for ML models and their explanations could help relieve some of the cognitive burdens from humans. Such tutorials could summarize the model as a list of rules, add heatmaps in examples, or provide a sequence of training examples with explanations and sufficient coverage. Future research could focus on developing effective tutorials for ML models and their explanations, and investigating how they can best support human decision-making [100].
- Providing narratives to improve the effectiveness of explanations because feature-based and example-based explanations could further improve the trust of humans in machine predictions. Future research could investigate the effectiveness of different forms of narratives in enhancing the interpretability and trustworthiness of ML models. This could involve designing new approaches for generating narratives that are tailored to the needs of different user groups or exploring the use of storytelling techniques to communicate complex ML concepts [100].

- Studying the ethical concerns of providing assistance from ML models in human decision-making because of raises important ethical concerns, such as the risk of removing human agency and the potential for unfairness in algorithmic decision-making. Future research could investigate these concerns in greater detail, and explore ways to mitigate potential harms. This could involve developing new approaches for designing fair and transparent ML models or investigating the legal and social implications of using ML in decision-making contexts [100].
- Investigating the impact of the system on the patient experience: The authors note that further research is needed to understand how the system affects the patient experience, particularly in terms of patient trust and likelihood to act on the system's predictions. Future research could involve conducting user studies with patients to better understand their perceptions of the system and investigating ways to enhance patient trust and engagement.
- Understanding the impact of the system on ophthalmologists' practices: The authors suggest that additional research is needed to understand how the system may alter the practices of ophthalmologists who evaluate patients who have received a prediction from the DL system. Future research could involve conducting surveys or interviews with ophthalmologists to understand their perceptions of the system and exploring ways to integrate the system into their clinical workflows.
- Designing study protocols for human-centered prospective studies: The authors suggest that an important area of future work is the design of study protocols for conducting human-centered prospective studies and studies on end-to-end service design of AI-based clinical products. Future research could focus on developing new methods and best practices for conducting user studies in clinical settings and investigating the design of AI-based clinical products from a human-centered perspective. This could involve exploring ways to involve patients and clinicians in the design process and developing methods for evaluating the impact of AI-based clinical products on patient outcomes and clinical workflows.
- Designing and testing onboarding materials for AI Assistants based on the findings of this research. This includes exploring how onboarding materials can shape work practices, instill accurate and actionable mental models, and impact assessments and attitudes toward the AI Assistant, such as user trust [102].
- Investigating how the AI Assistant can be used in different collaborative decision-making scenarios, including those that involve complex and uncertain information. This research could explore how the AI Assistant can be used to help people develop more effective strategies faster, and how it can support decision-making processes that involve multiple stakeholders.
- Exploring how AI Assistants can be customized for different users and contexts. This could include designing personalized onboarding materials, tailoring the AI Assistant's capabilities to different decision-making contexts, and considering how user preferences and needs can be incorporated into the design of the AI Assistant.
- Examining the long-term effects of using an AI Assistant for collaborative decision-making. This research could explore how the use of an AI Assistant impacts team dynamics and collaboration over time, and how it influences the development of decision-making skills and expertise among team members.
- Investigating ethical considerations related to the use of AI Assistants for collaborative decision-making. This could include exploring issues related to bias, fairness, transparency, and accountability, as well as considering how to ensure that the AI Assistant is aligned with the values and goals of the organization and its stakeholders.
- Utilizing machine learning algorithms for automated mapping enhances model accuracy and mitigates challenges associated with Data Outdated and Ambiguous Mapping. Broadening the dataset's coverage across different geographies and business categories serves to validate the applicability of the proposed approach in diverse environments [138].
- Designing a steering assistance system when the human driver model is inaccurate and the curvature is time-varying [179].
- Future research work could include investigating the relationship between human driver attention, scenario contextual information, and other surrounding agents to develop an intelligent and adaptive driver attention estimation system. Additionally, the semantic segmentation technology could be further improved, and it could be adopted to obtain semantic information [180].
- Future research work mentioned in the given paragraph is "fixed-time formation control for HiTL UAV systems with prescribed performance [181], [182], [183], [184], [185], [186].
- Improving the performance of the brain-control behavior to address the challenge of inaccurate or untimely output of desired commands by the drivers. This can be achieved by redesigning the stimulus presentation paradigm to help users to produce the discriminative EEG signals or developing some hybrid Brain-computer interfaces (BCIs) to improve the decoding performance of BCIs. Enhancing the driving ability of the brain-control drivers by practicing and developing a more user-friendly interface. Increasing the number of acceleration commands and driving speed to enhance the brain-controlled vehicles (BCVs) [187].
- To improve the computation efficiency and the adaptivity of the algorithm to various complex traffic scenarios and validate the performance of the proposed

algorithms in real-time using hardware-in-the-loop simulations [188].
- Development of the decision-making strategy based on the human-demonstration-aided reinforcement learning (RL) method. This could involve exploring different types of safe demonstrations or modifying the Double Dueling Deep Q Network (D3QN)-RL algorithm to further enhance its performance. Collaborating with different research domains to advance decision-making and autonomous driving technologies. This could involve working with experts in computer vision, robotics, and human factors to develop more comprehensive and integrated approaches to autonomous vehicles [57], [189], [190], [191], [192], [193], [194].
- Collecting data from diverse age groups and increasing the number of participants to improve the reliability and generalizability of the study's findings. Studying a more in-depth human driving model to identify potential areas of integration into the performance-based approach. Conducting experiments in real traffic environments as autonomous functions become more accessible and legal to use when driving on the road [195].
- Conducting annotations on larger scales to evaluate the presented error detection framework from more angles other than precision. Investigating how the identified erroneous instances or features could be used for further fixing or debugging the pre-trained models [119].
- Conducting further research on developing rational and intuitive interfaces for human-machine (brain) interfaces to improve explainable AI (XAI). Developing an explainable twin AI system to work in parallel to the DL systems that are designed for optimization performance [196]. Developing defense mechanisms that can recognize targeted attacks against DL and XAI engines. Conducting further research on understanding DL modules that contribute to physical (PHY) and medium access control (MAC) layer roles. Embarking on research to incorporate XAI into future wireless systems [197].
- Establishing automatic assessment algorithms for each type of action based on different childhood myositis assessment scale (CMAS) rules. Generating more realistic motions by combining the motion-style transfer technique for the corresponding character for a given action. Conducting further user studies to assess the repeatability and usability of the system. Further optimization and validation of the system will be necessary before clinical use [110].
- Improving the accuracy of the causal model by addressing limitations like sampling biases and missing proxy variables. This can be done through the use of non-linear structural equation models (SEMs) and causal discovery algorithms like Fast Causal Inference (FCI). Enhancing the scalability of the tool to handle larger datasets by exploring GPU-based parallel implementation of PC algorithms like cuPC or using inherently faster causal discovery algorithms like F-GES. Optimizing graph layout algorithms and exploring other visual analytics techniques like node aggregation to help navigate larger graphs better. Extending the HITL methodology to tackle biases in other domains such as word embeddings. Addressing human factors involved in the tool's operation, such as user bias and misuse, by choosing responsible users, checking system logs, and holding users accountable for their actions.
- Exploring the utility of higher-order information such as curvature, which could facilitate the search process by allowing exploration in a curved subregion. Studying utility for choosing other ($l \neq 1$) dimensional subspaces, as well as its combination with conditional generative models or domain-specific approaches. Testing on different types of generative models such as body shapes and hairstyles to see how well it works [198].
- Conducting multi-disciplinary studies that contain human factors and ergonomics (HF/E) as an essential discipline to support the shift in focus from a technology-centric view to a systems perspective when designing and developing AI applications for healthcare. Developing reporting guidelines for AI studies that include rigorous HF/E practices and evidence to ensure that AI-enabled healthcare is trustworthy and sustainable. Exploring how HF/E considerations such as situation awareness, workload, automation bias, explanation and trust, human-AI teaming, training, the relationships between staff and patient, and attention to ethical issues can be integrated into the design and use of AI applications in healthcare to improve patient safety, patient experience, staff well-being, and the efficiency of health systems. Reinforcing regulatory expectations that HF/E best practices have been followed when designing and developing AI applications for healthcare. Providing education and support in HF/E for healthcare professionals and organizations to effectively apply HF/E theories and approaches in practice and embed AI successfully in health systems. Extending the scope of research beyond the limited evidence base of AI in real-world use and exploring how well AI-enabled healthcare works for other medical applications and settings. Investigating how higher-order information, such as curvature, can facilitate the search process in exploring high-dimensional latent spaces for deep generative models [199].
- Investigating the effectiveness of democratic AI as a method for value alignment. While the current study demonstrates that an AI system can be trained to satisfy a democratic objective, it is important to further explore the strengths and limitations of this approach. For example, future research could investigate the potential for the "tyranny of the majority" to arise with democratic AI, and explore ways to protect the

interests of minority groups. Examining the potential for AI systems to replace human decision-makers in various contexts. The study raises questions about whether people would trust AI systems to design mechanisms in place of humans, and whether such systems could be used in the public sphere without human intervention. Future research could explore these issues in greater depth, as well as investigate the potential benefits and risks of using AI systems in place of human decision-makers [200].

- Exploring the long-term effects of metacognitive monitoring feedback and the self-regulated fuzzy index (SRF) index in a HITL simulation, to better understand the relationship between metacognitive judgments and human performance in different environments and age groups. Additionally, the study could investigate further the SRF index to find various relationships between the SR learning process and human performance. This research will be valuable in developing more advanced feedback learning algorithms to improve an operator's situation awareness (SA) [201].
- Evaluating the effectiveness of the HITL control strategy for a larger group of patients with various lower limb impairments. Additionally, the use of other imaging techniques, such as electromyography (EMG), could be explored to further understand the muscle activity and coordination involved in the rehabilitation training. Furthermore, the adaptive controller could be improved to better account for varying levels of impairment and to provide personalized assistance for each patient [202].
- Introducing functional magnetic resonance imaging (fMRI) to evaluate the performance of lower extremity motor function in patients in the context of a unilateral exoskeleton system for rehabilitation training. Exploring the potential of using an ML estimation technique called approximate expectation-maximization (EM) and a multiple-model estimator that switches between multiple nonlinear human motion models to infer human motion trajectories and estimate reaching goal intentions for assistive robots in joint tasks with humans. Investigating how to ensure the safety of humans in the presence of actuator and sensor failures and exploring the use of sensors such as ultrasound imaging to estimate human intention in human-robot collaboration (HRC) tasks involving assistive robots [203], [204], [205].
- Conducting a larger user study to involve other categories of computer users and explore their distraction patterns, as well as considering computing professionals in the Wall Street business enterprises to capture their distraction behavior, are future research directions. While using a camera for collecting ground truth is more reliable, it may be privacy-invasive in a computing environment [206].
- Validation of the HITL weight compensation method on a real upper limb exoskeleton device with myoelectric sensors. Minimization of whole muscles' effort using partial observation of four mono-articular muscles. Investigation of the adaptation rule's robustness to sensory noise in EMG sensors. Consideration of the adaptation rule as a cognitive model for mass estimation in static conditions. Exploration of the potential of the method as a deep sensing method for individuals with hand amputations who still have sensory neurons [207].
- HITL learning is an emerging paradigm that holds great promise for solving complex problems in various domains. Developing formal computational frameworks that incorporate humans in the loop is critical for the success of HITL systems. These frameworks should address the challenges of HITL learning, such as the quality of human input, scalability, and ethical issues, while also predicting the future of HITL platforms. Continued research and development in HITL learning will help to maximize the benefits of human-machine collaboration and create more intelligent, adaptive, and efficient systems.

VI. CONCLUSION

HITL systems are a promising area of research that seeks to leverage the strengths of both humans and machines to accomplish complex tasks. Despite the possible advantages, creating HITL systems has its own set of difficulties, including creating efficient human-machine interfaces, dealing with ambiguity and uncertainty, controlling biases, and addressing ethical issues. This paper has reviewed the current state-of-the-art methodologies for developing HITL systems and identified the major challenges that need to be addressed. We have also talked about the possible uses of HITL systems in a number of industries, including healthcare, finance, and education. Moving forward, the development of HITL systems requires a collaborative effort between experts in ML, human-computer interaction, and domain-specific knowledge. By working together, we can advance the development of HITL systems that can effectively leverage the strengths of both humans and machines to improve decision-making, automate tedious tasks, and enhance overall performance.

ACKNOWLEDGMENT

The authors express their gratitude to Manipal Academy of Higher Education, Manipal, for their essential role in facilitating the publication of this article.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

[1] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," *Comput. Sci. Rev.*, vol. 40, May 2021, Art. no. 100379.

[2] A. Brutzkus and A. Globerson, "Why do larger models generalize better? A theoretical perspective via the XOR problem," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 822–830.

[3] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[4] J. Li, J. Yang, A. Hertzmann, J. Zhang, and T. Xu, "LayoutGAN: Synthesizing graphic layouts with vector-wireframe adversarial networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2388–2399, Jul. 2021.

[5] S. Zhao, Z. Liu, J. Lin, J.-Y. Zhu, and S. Han, "Differentiable augmentation for data-efficient GAN training," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 7559–7570.

[6] H. T. Shen, X. Zhu, Z. Zhang, S.-H. Wang, Y. Chen, X. Xu, and J. Shao, "Heterogeneous data fusion for predicting mild cognitive impairment conversion," *Inf. Fusion*, vol. 66, pp. 54–63, Feb. 2021.

[7] S. Li, C. H. Liu, Q. Lin, Q. Wen, L. Su, G. Huang, and Z. Ding, "Deep residual correction network for partial domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2329–2344, Jul. 2021.

[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–15.

[9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.

[10] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," Tech. Rep., 2018.

[11] M. Habermann, W. Xu, M. Zollhöfer, G. Pons-Moll, and C. Theobalt, "DeepCap: Monocular human performance capture using weak supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5051–5062.

[12] Y. Wang, W. Yang, F. Ma, J. Xu, B. Zhong, Q. Deng, and J. Gao, "Weak supervision for fake news detection via reinforcement learning," in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 1, pp. 516–523.

[13] S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, and S. Yu, "A survey: Deep learning for hyperspectral image classification with few labeled samples," *Neurocomputing*, vol. 448, pp. 179–204, Aug. 2021.

[14] M. Diligenti, S. Roychowdhury, and M. Gori, "Integrating prior knowledge into deep learning," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2017, pp. 920–923.

[15] S. Chen, Y. Leng, and S. Labi, "A deep learning algorithm for simulating autonomous driving considering prior knowledge and temporal information," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 35, no. 4, pp. 305–321, Apr. 2020.

[16] Y. Lin, S. L. Pintea, and J. C. van Gemert, "Deep Hough-transform line priors," in *Proc. Eur. Conf. Comput. Vis.*, Glasgow, U.K., Aug. 2020, pp. 323–340.

[17] G. Hartmann, Z. Shiller, and A. Azaria, "Deep reinforcement learning for time optimal velocity control using prior knowledge," in *Proc. IEEE 31st Int. Conf. Tools Artif. Intell. (ICTAI)*, Nov. 2019, pp. 186–193.

[18] X. Zhang, S. Wang, J. Liu, and C. Tao, "Towards improving diagnosis of skin diseases by combining deep neural network and human knowledge," *BMC Med. Informat. Decis. Making*, vol. 18, no. S2, pp. 69–76, Jul. 2018.

[19] 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: A case study on Ant colony optimization," *Appl. Intell.*, vol. 49, no. 7, pp. 2401–2414, Jul. 2019.

[20] Y.-T. Zhuang, F. Wu, C. Chen, and Y.-H. Pan, "Challenges and opportunities: From big data to knowledge in AI 2.0," *Frontiers Inf. Technol. Electron. Eng.*, vol. 18, no. 1, pp. 3–14, Jan. 2017.

[21] V. Kumar, A. Smith-Renner, L. Findlater, K. Seppi, and J. Boyd-Graber, "Why didn't you listen to me? Comparing user control of human-in-the-loop topic models," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, Florence, Italy, A. Korhonen, D. Traum, and L. Màrquez, Eds., 2019, pp. 6323–6330.

[22] D. Xin, L. Ma, J. Liu, S. Macke, S. Song, and A. Parameswaran, "Accelerating human-in-the-loop machine learning: Challenges and opportunities," in *Proc. 2nd Workshop Data Manage. End-End Mach. Learn.*, Jun. 2018, pp. 1–4.

[23] M. Golovianko, V. Terziyan, V. Branytskyi, and D. Malyk, "Industry 4.0 vs. industry 5.0: Co-existence, transition, or a hybrid," *Proc. Comput. Sci.*, vol. 217, pp. 102–113, Jan. 2023.

[24] S. Aheleroff, H. Huang, X. Xu, and R. Y. Zhong, "Toward sustainability and resilience with industry 4.0 and industry 5.0," *Frontiers Manuf. Technol.*, vol. 2, Oct. 2022, pp. 1–20.

[25] 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.

[26] A. Endert, M. S. Hossain, N. Ramakrishnan, C. North, P. Fiaux, and C. Andrews, "The human is the loop: New directions for visual analytics," *J. Intell. Inf. Syst.*, vol. 43, no. 3, pp. 411–435, Dec. 2014.

[27] 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.

[28] J. Vogt, "Where is the human got to go? Artificial intelligence, machine learning, big data, digitalisation, and human–robot interaction in industry 4.0 and 5.0: Review comment on: Bauer, M. (2020). preise Kalkulieren mit KI-Gestützter onlineplattform BAM GmbH, Weiden, Bavaria, Germany," *AI Soc.*, vol. 36, no. 3, pp. 1083–1087, Sep. 2021.

[29] D. Mourtzis, J. Angelopoulos, and N. Panopoulos, "Operator 5.0: A survey on enabling technologies and a framework for digital manufacturing based on extended reality," *J. Mach. Eng.*, vol. 22, no. 1, pp. 43–69, 2022.

[30] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 1997.

[31] T. Liu, "Human-in-the-loop learning from crowdsourcing and social media," Thesis, Rochester Inst. Technol., 2020. [Online]. Available: https://repository.rit.edu/theses/10462

[32] H. Su, Z. Yin, T. Kanade, and S. Huh, "Active sample selection and correction propagation on a gradually-augmented graph," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1975–1983.

[33] S. Paul, J. H. Bappy, and A. K. Roy-Chowdhury, "Non-uniform subset selection for active learning in structured data," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 830–839.

[34] W. H. Beluch, T. Genewein, A. Nurnberger, and J. M. Kohler, "The power of ensembles for active learning in image classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9368–9377.

[35] B. Settles, "Active learning literature survey," Dept. Comput. Sci., Univ. Wisconsin-Madison, Madison, WI, USA, 2009.

[36] P. Ren, Y. Xiao, X. Chang, P.-Y. Huang, Z. Li, B. B. Gupta, X. Chen, and X. Wang, "A survey of deep active learning," *ACM Comput. Surv.*, vol. 54, no. 9, pp. 1–40, Oct. 2021.

[37] Y. Gal, "Uncertainty in deep learning. University of Cambridge," Ph.D. thesis, Dept. Eng., Univ. Cambridge, Cambridge, U.K., 2016.

[38] D. Yoo and I. S. Kweon, "Learning loss for active learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 93–102.

[39] J. T. Ash, C. Zhang, A. Krishnamurthy, J. Langford, and A. Agarwal, "Deep batch active learning by diverse, uncertain gradient lower bounds," in *Proc. Int. Conf. Learn. Represent.*, 2020, pp. 1–26.

[40] A. Kirsch, J. Van Amersfoort, and Y. Gal, "BatchBALD: Efficient and diverse batch acquisition for deep Bayesian active learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–16.

[41] H. Song, S. Kim, M. Kim, and J.-G. Lee, "Ada-boundary: Accelerating DNN training via adaptive boundary batch selection," *Mach. Learn.*, vol. 109, nos. 9–10, pp. 1837–1853, Sep. 2020.

[42] P. F. Christiano, J. Leike, T. Brown, M. Martic, S. Legg, and D. Amodei, "Deep reinforcement learning from human preferences," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–9.

[43] J. Li, A. H. Miller, S. Chopra, M. Ranzato, and J. Weston, "Dialogue learning with human-in-the-loop," in *Proc. 5th Int. Conf. Learn. Represent.*, 2017, pp. 1–23.

[44] C. O. Retzlaff, S. Das, C. Wayllace, P. Mousavi, M. Afshari, T. Yang, A. Saranti, A. Angerschmid, M. E. Taylor, and A. Holzinger, "Human-in-the-loop reinforcement learning: A survey and position on requirements, challenges, and opportunities," *J. Artif. Intell. Res.*, vol. 79, pp. 359–415, Jan. 2024.

[45] S. Casper et al., "Open problems and fundamental limitations of reinforcement learning from human feedback," 2023, *arXiv:2307.15217*.

[46] T. Mandel, Y.-E. Liu, E. Brunskill, and Z. Popović, "Where to add actions in human-in-the-loop reinforcement learning," in *Proc. AAAI Conf. Artif. Intell.*, 2017, vol. 31, no. 1, pp. 2322–2328.

[47] T. Kaufmann, P. Weng, V. Bengs, and E. Hüllermeier, "A survey of reinforcement learning from human feedback," 2023, *arXiv:2312.14925*.

[48] M. Fang, Y. Li, and T. Cohn, "Learning how to active learn: A deep reinforcement learning approach," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 595–605.

[49] S. Wan, Y. Hou, F. Bao, Z. Ren, Y. Dong, Q. Dai, and Y. Deng, "Human-in-the-loop low-shot learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 3287–3292, Jul. 2021.

[50] W. Bradley Knox, S. Hatgis-Kessell, S. Booth, S. Niekum, P. Stone, and A. Allievi, "Models of human preference for learning reward functions," 2022, *arXiv:2206.02231*.

[51] B. Goodman and S. Flaxman, "European union regulations on algorithmic decision making and a 'right to explanation,'" *AI Mag.*, vol. 38, no. 3, pp. 50–57, Sep. 2017.

[52] D. Gunning, "Explainable artificial intelligence (XAI)," *Defense Adv. Res. Projects Agency*, vol. 2, no. 2, p. 1, 2017.

[53] A. Tocchetti and M. Brambilla, "The role of human knowledge in explainable AI," *Data*, vol. 7, no. 7, p. 93, Jul. 2022.

[54] 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.

[55] W. Wang, J. Xi, C. Liu, and X. Li, "Human-centered feed-forward control of a vehicle steering system based on a driver's path-following characteristics," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 6, pp. 1440–1453, Jun. 2017.

[56] C. J. Hong and V. R. Aparow, "System configuration of human-in-the-loop simulation for level 3 autonomous vehicle using IPG CarMaker," in *Proc. IEEE Int. Conf. Internet Things Intell. Syst. (IoTaIS)*, Nov. 2021, pp. 215–221.

[57] C. Lv, Y. Li, Y. Xing, C. Huang, D. Cao, Y. Zhao, and Y. Liu, "Human–machine collaboration for automated driving using an intelligent two-phase haptic interface," *Adv. Intell. Syst.*, vol. 3, no. 4, Apr. 2021, Art. no. 2000229.

[58] P. Hang, X. Chen, and W. Wang, "Cooperative control framework for human driver and active rear steering system to advance active safety," *IEEE Trans. Intell. Vehicles*, vol. 6, no. 3, pp. 460–469, Sep. 2021.

[59] J. Liu, Q. Dai, H. Guo, J. Guo, and H. Chen, "Human-oriented online driving authority optimization for driver-automation shared steering control," *IEEE Trans. Intell. Vehicles*, vol. 7, no. 4, pp. 863–872, Dec. 2022.

[60] Q. Wang, H. Dong, F. Ju, W. Zhuang, C. Lv, L. Wang, and Z. Song, "Adaptive leading cruise control in mixed traffic considering human behavioral diversity," 2022, *arXiv:2210.02147*.

[61] J. Wu, Z. Huang, Z. Hu, and C. Lv, "Toward human-in-the-loop AI: Enhancing deep reinforcement learning via real-time human guidance for autonomous driving," *Engineering*, vol. 21, pp. 75–91, Feb. 2023.

[62] H. Lee and S. Park, "Sensing-aware deep reinforcement learning with HCI-based human-in-the-loop feedback for autonomous nonlinear drone mobility control," *IEEE Access*, vol. 12, pp. 1727–1736, 2024.

[63] Z. Huang, Z. Sheng, C. Ma, and S. Chen, "Human as AI mentor: Enhanced human-in-the-loop reinforcement learning for safe and efficient autonomous driving," 2024, *arXiv:2401.03160*.

[64] M. Daum, E. Zhang, D. He, M. Balazinska, B. Haynes, R. Krishna, A. Craig, and A. Wirsing, "VOCAL: Video organization and interactive compositional analytics," in *Proc. 12th Annu. Conf. Innov. Data Syst. Res.*, 2022, pp. 1–11.

[65] R. Krishna, D. Lee, L. Fei-Fei, and M. S. Bernstein, "Socially situated artificial intelligence enables learning from human interaction," *Proc. Nat. Acad. Sci. USA*, vol. 119, no. 39, Sep. 2022, Art. no. e2115730119.

[66] R. Krishna, M. Gordon, L. Fei-Fei, and M. Bernstein, "Visual intelligence through human interaction," in *Artificial Intelligence for Human Computer Interaction: A Modern Approach*. Springer, 2021, pp. 257–314.

[67] J. Park, R. Krishna, P. Khadpe, L. Fei-Fei, and M. Bernstein, "AI-based request augmentation to increase crowdsourcing participation," in *Proc. AAAI Conf. Hum. Comput. Crowdsourcing*, 2019, vol. 7, no. 1, pp. 115–124.

[68] X. Wu, Y. Zheng, T. Ma, H. Ye, and L. He, "Document image layout analysis via explicit edge embedding network," *Inf. Sci.*, vol. 577, pp. 436–448, Oct. 2021.

[69] X. Wu, B. Xu, Y. Zheng, H. Ye, J. Yang, and L. He, "Fast video crowd counting with a temporal aware network," *Neurocomputing*, vol. 403, pp. 13–20, Aug. 2020.

[70] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[71] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proc. IEEE*, vol. 111, no. 3, pp. 257–276, Mar. 2023.

[72] A. Yao, J. Gall, C. Leistner, and L. Van Gool, "Interactive object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3242–3249.

[73] K. Madono, T. Nakano, T. Kobayashi, and T. Ogawa, "Efficient human-in-the-loop object detection using bi-directional deep SORT and annotation-free segment identification," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Dec. 2020, pp. 1226–1233.

[74] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3645–3649.

[75] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Process. Mag.*, vol. 14, no. 2, pp. 24–41, Mar. 1997.

[76] A. Criminisi, P. Perez, and K. Toyama, "Object removal by exemplar-based inpainting," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2003, pp. 1–8.

[77] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 85–100.

[78] T. Weber, H. Hußmann, Z. Han, S. Matthes, and Y. Liu, "Draw with me: Human-in-the-loop for image restoration," in *Proc. 25th Int. Conf. Intell. User Interface*, Mar. 2020, pp. 243–253.

[79] D. Ulyanov, A. Vedaldi, and S. Victor, "Deep image prior," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, 2018, pp. 1–23.

[80] J. Roels, F. Vernaillen, A. Kremer, A. Gonçalves, J. Aelterman, H. Q. Luong, B. Goossens, W. Philips, S. Lippens, and Y. Saeys, "A 'human-in-the-loop' approach for semi-automated image restoration in electron microscopy," *BioRxiv*, 2019, Art. no. 644146.

[81] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[82] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, Jul. 2022.

[83] H. Wang, T. Chen, Z. Wang, and K. Ma, "Troubleshooting image segmentation models with human-in-the-loop," *Mach. Learn.*, vol. 112, no. 3, pp. 1033–1051, Mar. 2023.

[84] A. Taleb, C. Lipperit, T. Klein, and M. Nabi, "Multimodal self-supervised learning for medical image analysis," in *Proc. Int. Conf. Inf. Process. Med. Imag.*, 2021, pp. 661–673.

[85] M. Ravanbakhsh, V. Tschernezki, F. Last, T. Klein, K. Batmanghelich, V. Tresp, and M. Nabi, "Human–machine collaboration for medical image segmentation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 1040–1044.

[86] K. N. Shukla, A. Potnis, and P. Dwivedy, "A review on image enhancement techniques," *Int. J. Eng. Appl. Comput. Sci*, vol. 2, no. 7, pp. 232–235, 2017.

[87] Y. Murata and Y. Dobashi, "Automatic image enhancement taking into account user preference," in *Proc. Int. Conf. Cyberworlds (CW)*, Oct. 2019, pp. 374–377.

[88] M. Fischer, K. Kobs, and A. Hotho, "NICER: Aesthetic image enhancement with humans in the loop," in *Proc. 13th Int. Conf. Adv. Comput.-Hum. Interact.*, 2020, pp. 357–362.

[89] R. Yao, G. Lin, S. Xia, J. Zhao, and Y. Zhou, "Video object segmentation and tracking: A survey," *ACM Trans. Intell. Syst. Technol.*, vol. 11, no. 4, pp. 1–47, 2020.

[90] A. Benard and M. Gygli, "Interactive video object segmentation in the wild," 2017, *arXiv:1801.00269*.

[91] S. W. Oh, J.-Y. Lee, N. Xu, and S. J. Kim, "Fast user-guided video object segmentation by interaction-and-propagation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5242–5251.

[92] A. Abad, M. Nabi, and A. Moschitti, "Autonomous crowdsourcing through human-machine collaborative learning," in *Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 2017, pp. 873–876.

[93] M. Ravanbakhsh, T. Klein, K. Batmanghelich, and M. Nabi, "Uncertainty-driven semantic segmentation through human-machine collaborative learning," 2019, *arXiv:1909.00626*.

[94] Y. Wang, Z. Yu, S. Liu, Z. Zhou, and B. Guo, "Genie in the model: Automatic generation of human-in-the-loop deep neural networks for mobile applications," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 7, no. 1, pp. 1–29, Mar. 2023.

[95] N. Qiao, Y. Sun, C. Liu, L. Xia, J. Luo, K. Zhang, and C.-H. Kuo, "Human-in-the-loop video semantic segmentation auto-annotation," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2023, pp. 5870–5880.

[96] Z. Hu, J. Wu, D. Gao, Y. Zhou, and Q. Zhu, "CFTF: Controllable fine-grained Text2Face and its human-in-the-loop suspect portraits application," in *Proc. 31st ACM Int. Conf. Multimedia*, Oct. 2023, pp. 9390–9392.

[97] W. Deng, Q. Liu, F. Zhao, D. T. Pham, J. Hu, Y. Wang, and Z. Zhou, "Learning by doing: A dual-loop implementation architecture of deep active learning and human-machine collaboration for smart robot vision," *Robot. Comput.-Integr. Manuf.*, vol. 86, Apr. 2024, Art. no. 102673.

[98] B. Kim, M. Wattenberg, J. Gilmer, C. Cai, J. Wexler, F. Viegas, and R. sayres, "Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (TCAV)," in *Proc. Int. Conf. Mach. Learn. (ICML)*, in Proceedings of Machine Learning Research, vol. 80, J. Dy and A. Krause, Eds., 2018, pp. 2668–2677.

[99] C. J. Cai, E. Reif, N. Hegde, J. Hipp, B. Kim, D. Smilkov, M. Wattenberg, F. Viegas, G. S. Corrado, M. C. Stumpe, and M. Terry, "Human-centered tools for coping with imperfect algorithms during medical decision-making," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2019, pp. 1–14.

[100] V. Lai and C. Tan, "On human predictions with explanations and predictions of machine learning models: A case study on deception detection," in *Proc. Conf. Fairness, Accountability, Transparency*, Jan. 2019, pp. 29–38.

[101] G. Bansal, B. Nushi, E. Kamar, W. S. Lasecki, D. S. Weld, and E. Horvitz, "Beyond accuracy: The role of mental models in human-AI team performance," in *Proc. AAAI Conf. Hum. Comput. Crowdsourcing*, vol. 7, 2019, pp. 2–11.

[102] C. J. Cai, S. Winter, D. Steiner, L. Wilcox, and M. Terry, "'Hello AI': Uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making," *Proc. ACM Hum.-Comput. Interact.*, vol. 3, no. 2, pp. 1–24, Nov. 2019.

[103] B. N. Patel, L. Rosenberg, G. Willcox, D. Baltaxe, M. Lyons, J. Irvin, P. Rajpurkar, T. Amrhein, R. Gupta, S. Halabi, C. Langlotz, E. Lo, J. Mammarappallil, A. J. Mariano, G. Riley, J. Seekins, L. Shen, E. Zucker, and M. P. Lungren, "Human–machine partnership with artificial intelligence for chest radiograph diagnosis," *NPJ Digit. Med.*, vol. 2, no. 1, p. 111, Nov. 2019.

[104] E. Beede, E. Baylor, F. Hersch, A. Iurchenko, L. Wilcox, P. Ruamviboonsuk, and L. M. Vardoulakis, "A human-centered evaluation of a deep learning system deployed in clinics for the detection of diabetic retinopathy," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2020, pp. 1–12.

[105] P. Tschandl, C. Rinner, Z. Apalla, G. Argenziano, N. Codella, A. Halpern, M. Janda, A. Lallas, C. Longo, J. Malvehy, and J. Paoli, "Human–computer collaboration for skin cancer recognition," *Nature Med.*, vol. 26, no. 8, pp. 1229–1234, 2020.

[106] M. Steyvers, H. Tejeda, G. Kerrigan, and P. Smyth, "Bayesian modeling of human–AI complementarity," *Proc. Nat. Acad. Sci. USA*, vol. 119, no. 11, Mar. 2022, Art. no. e2111547119.

[107] H. Gu, C. Yang, M. Haeri, J. Wang, S. Tang, W. Yan, S. He, C. K. Williams, S. Magaki, and X. Chen, "Augmenting pathologists with NaviPath: Design and evaluation of a human-AI collaborative navigation system," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2023, pp. 1–19.

[108] 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, 2023.

[109] F. Cabitza, A. Campagner, L. Ronzio, M. Cameli, G. E. Mandoli, M. C. Pastore, L. M. Sconfienza, D. Folgado, M. Barandas, and H. Gamboa, "Rams, hounds and white boxes: Investigating human–AI collaboration protocols in medical diagnosis," *Artif. Intell. Med.*, vol. 138, Jan. 2023, Art. no. 102506.

[110] K. Zhou, R. Cai, Y. Ma, Q. Tan, X. Wang, J. Li, H. P. H. Shum, F. W. B. Li, S. Jin, and X. Liang, "A video-based augmented reality system for human-in-the-loop muscle strength assessment of juvenile dermatomyositis," *IEEE Trans. Vis. Comput. Graphics*, vol. 29, no. 5, pp. 2456–2466, May 2023.

[111] O. Gómez-Carmona, D. Casado-Mansilla, D. López-de-Ipiña, and J. García-Zubia, "Human-in-the-loop machine learning: Reconceptualizing the role of the user in interactive approaches," *Internet Things*, vol. 25, Apr. 2024, Art. no. 101048.

[112] X. Duan and J. P. Lalor, "H-COAL: Human correction of AI-generated labels for biomedical named entity recognition," 2023, *arXiv:2311.11981*.

[113] S. Liang, M. Hartmann, and D. Sonntag, "Cross-lingual German biomedical information extraction: From zero-shot to human-in-the-loop," 2023, *arXiv:2301.09908*.

[114] O. Stretcu, E. Vendrow, K. Hata, K. Viswanathan, V. Ferrari, S. Tavakkol, W. Zhou, A. Avinash, E. Luo, N. G. Alldrin, M. Bateni, G. Berger, A. Bunner, C.-T. Lu, J. Rey, G. DeSalvo, R. Krishna, and A. Fuxman, "Agile modeling: From concept to classifier in minutes," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2023, pp. 22323–22334.

[115] Y. Song, J. Wang, T. Jiang, Z. Liu, and Y. Rao, "Targeted sentiment classification with attentional encoder network," in *Proc. 28th Int. Conf. Artif. Neural Netw.*, vol. 28, Munich, Germany, 2019, pp. 93–103.

[116] L. Xiao, X. Hu, Y. Chen, Y. Xue, D. Gu, B. Chen, and T. Zhang, "Targeted sentiment classification based on attentional encoding and graph convolutional networks," *Appl. Sci.*, vol. 10, no. 3, p. 957, Feb. 2020.

[117] L. Xiao, X. Hu, Y. Chen, Y. Xue, B. Chen, D. Gu, and B. Tang, "Multi-head self-attention based gated graph convolutional networks for aspect-based sentiment classification," *Multimedia Tools Appl.*, vol. 81, no. 14, pp. 19051–19070, Jun. 2022.

[118] B. Nushi, E. Kamar, and E. Horvitz, "Towards accountable AI: Hybrid human-machine analyses for characterizing system failure," in *Proc. AAAI Conf. Hum. Comput. Crowdsourcing*, 2018, pp. 126–135.

[119] Z. Liu, Y. Guo, and J. Mahmud, "When and why does a model fail? A human-in-the-loop error detection framework for sentiment analysis," in *Proc. Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics*, 2021, pp. 1–8.

[120] T. Karmakharm, N. Aletras, and K. Bontcheva, "Journalist-in-the-loop: Continuous learning as a service for rumour analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP), Syst. Demonstration*, 2019, pp. 115–120.

[121] X. Bai, P. Liu, and Y. Zhang, "Investigating typed syntactic dependencies for targeted sentiment classification using graph attention neural network," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 503–514, 2021.

[122] I. Arous, L. Dolamic, J. Yang, A. Bhardwaj, G. Cuccu, and P. Cudré-Mauroux, "MARTA: Leveraging human rationales for explainable text classification," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, no. 7, pp. 5868–5876.

[123] L. He, J. Michael, M. Lewis, and L. Zettlemoyer, "Human-in-the-loop parsing," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 2337–2342.

[124] Z. Yao, X. Li, J. Gao, B. Sadler, and H. Sun, "Interactive semantic parsing for if-then recipes via hierarchical reinforcement learning," in *Proc. AAAI Conf. Artif. Intell.*, 2019, vol. 33, no. 1, pp. 2547–2554.

[125] Z. Yao, Y. Su, H. Sun, and W.-T. Yih, "Model-based interactive semantic parsing: A unified framework and a text-to-SQL case study," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019, pp. 5447–5458.

[126] S. Chopra, M. Auli, and A. M. Rush, "Abstractive sentence summarization with attentive recurrent neural networks," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Human Lang. Technol.*, 2016, pp. 93–98.

[127] D. M. Ziegler, N. Stiennon, J. Wu, T. B. Brown, A. Radford, D. Amodei, P. Christiano, and G. Irving, "Fine-tuning language models from human preferences," 2019, *arXiv:1909.08593*.

[128] 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.

[129] Z. J. Wang, D. Choi, S. Xu, and D. Yang, "Putting humans in the natural language processing loop: A survey," in *Proc. 1st Workshop Bridging Hum.-Comput. Interact. Natural Lang. Process.*, 2021, pp. 47–52.

[130] B. Hancock, A. Bordes, P.-E. Mazare, and J. Weston, "Learning from dialogue after deployment: Feed yourself, chatbot!" in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 3667–3684.

[131] E. Wallace, P. Rodriguez, S. Feng, I. Yamada, and J. Boyd-Graber, "Trick me if you can: Human-in-the-loop generation of adversarial examples for question answering," *Trans. Assoc. Comput. Linguistics*, vol. 7, pp. 387–401, Nov. 2019.

[132] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 1135–1144.

[133] X. Fan, Y. Lyu, P. Pu Liang, R. Salakhutdinov, and L.-P. Morency, "Nano: Nested human-in-the-loop reward learning for few-shot language model control," 2022, *arXiv:2211.05750*.

[134] X. Dong, S. Sarker, and L. Qian, "Integrating human-in-the-loop into swarm learning for decentralized fake news detection," in *Proc. Int. Conf. Intell. Data Sci. Technol. Appl. (IDSTA)*, Sep. 2022, pp. 46–53.

[135] A. Bonet-Jover, R. Sepúlveda-Torres, E. Saquete, and P. Martínez-Barco, "A semi-automatic annotation methodology that combines summarization and human-in-the-loop to create disinformation detection resources," *Knowl.-Based Syst.*, vol. 275, Sep. 2023, Art. no. 110723.

[136] M. Wahed, D. Gruhl, and I. Lourentzou, "MArBLE: Hierarchical multi-armed bandits for human-in-the-loop set expansion," in *Proc. 32nd ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2023, pp. 4857–4863.

[137] A. Bonet-Jover, R. Sepúlveda-Torres, E. Saquete, P. Martínez-Barco, A. Piad-Morffis, and S. Estevez-Velarde, "Applying human-in-the-loop to construct a dataset for determining content reliability to combat fake news," *Eng. Appl. Artif. Intell.*, vol. 126, Nov. 2023, Art. no. 107152.

[138] S. Asthana and R. Mahindru, "Mapping of financial services datasets using human-in-the-loop," in *Proc. 3rd ACM Int. Conf. AI Finance*, Nov. 2022, pp. 183–191.

[139] A. Yasir, A. Ahmad, S. Abbas, M. Inairat, A. H. Al-Kassem, and A. Rasool, "How artificial intelligence is promoting financial inclusion? A study on barriers of financial inclusion," in *Proc. Int. Conf. Bus. Anal. Technol. Secur. (ICBATS)*, Feb. 2022, pp. 1–6.

[140] X. Ding, N. Seleznev, S. Kumar, C. B. Bruss, and L. Akoglu, "From explanation to action: An end-to-end human-in-the-loop framework for anomaly reasoning and management," 2023, *arXiv:2304.03368*.

[141] R. P. Buckley, D. A. Zetzsche, D. W. Arner, and B. W. Tang, "Regulating artificial intelligence in finance: Putting the human in the loop," *Sydney Law Rev.*, vol. 43, no. 1, pp. 43–81, 2021.

[142] D. A. Zetzsche, D. W. Arner, R. P. Buckley, and B. Tang, "Artificial intelligence in finance: Putting the human in the loop," CFTE Acad. Paper Ser., Centre Finance, Technol. Entrepreneurship, Univ. Hong Kong Fac. Law, Res. Paper 2020/006, Feb. 2020, no. 1. [Online]. Available: https://ssrn.com/abstract=3531711

[143] J. Truby, R. Brown, and A. Dahdal, "Banking on AI: Mandating a proactive approach to AI regulation in the financial sector," *Law Financial Markets Rev.*, vol. 14, no. 2, pp. 110–120, Apr. 2020.

[144] S. Nahavandi, "Trusted autonomy between humans and robots: Toward human-on-the-loop in robotics and autonomous systems," *IEEE Syst., Man, Cybern. Mag.*, vol. 3, no. 1, pp. 10–17, Jan. 2017.

[145] C. Cimini, F. Pirola, R. Pinto, and S. Cavalieri, "A human-in-the-loop manufacturing control architecture for the next generation of production systems," *J. Manuf. Syst.*, vol. 54, pp. 258–271, Jan. 2020.

[146] M. De-Arteaga, R. Fogliato, and A. Chouldechova, "A case for humans-in-the-loop: Decisions in the presence of erroneous algorithmic scores," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2020, pp. 1–12.

[147] J. Y. C. Chen and M. J. Barnes, "Human–agent teaming for multirobot control: A review of human factors issues," *IEEE Trans. Hum.-Mach. Syst.*, vol. 44, no. 1, pp. 13–29, Feb. 2014.

[148] 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.

[149] Y. Yang, N. D. Truong, C. Maher, A. Nikpour, and O. Kavehei, "Continental generalization of a human-in-the-loop AI system for clinical seizure recognition," *Expert Syst. Appl.*, vol. 207, Nov. 2022, Art. no. 118083.

[150] L. Han, "When the human is in the loop: Cost, effort and behavior," in *Proc. 43rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2020, p. 2480.

[151] G. Li, "Human-in-the-loop data integration," *Proc. VLDB Endowment*, vol. 10, no. 12, pp. 2006–2017, Aug. 2017.

[152] R. M. Monarch, *Human-in-the-Loop Machine Learning: Active Learning and Annotation for Human-centered AI*. New York, NY, USA: Simon and Schuster, 2021.

[153] D. Ustalov, "Challenges in data production for AI with human-in-the-loop," in *Proc. 15th ACM Int. Conf. Web Search Data Mining*, Feb. 2022, pp. 1651–1652.

[154] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Apr. 2004, pp. 985–990.

[155] R. W. White, S. T. Dumais, and J. Teevan, "Characterizing the influence of domain expertise on web search behavior," in *Proc. 2nd ACM Int. Conf. Web Search Data Mining*, Feb. 2009, pp. 132–141.

[156] S. R. Hong, J. Hullman, and E. Bertini, "Human factors in model interpretability: Industry practices, challenges, and needs," *Proc. ACM Hum.-Comput. Interact.*, vol. 4, no. 1, pp. 1–26, May 2020.

[157] I. Lage and F. Doshi-Velez, "Human-in-the-loop learning of interpretable and intuitive representations," in *Proc. ICML Workshop Hum. Interpretability Mach. Learn.*, vol. 17, Vienna, Austria, 2020, pp. 1–10.

[158] Z. Zhao, P. Xu, C. Scheidegger, and L. Ren, "Human-in-the-loop extraction of interpretable concepts in deep learning models," *IEEE Trans. Vis. Comput. Graph.*, vol. 28, no. 1, pp. 780–790, Jan. 2022.

[159] X. Gao, J. Si, Y. Wen, M. Li, and H. Huang, "Reinforcement learning control of robotic knee with human-in-the-loop by flexible policy iteration," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 10, pp. 5873–5887, Oct. 2022.

[160] B. Perozzi, R. Al-Rfou, and S. Skiena, "DeepWalk: Online learning of social representations," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2014, pp. 701–710.

[161] J. Cheng and M. S. Bernstein, "Flock: Hybrid crowd-machine learning classifiers," in *Proc. 18th ACM Conf. Comput. Supported Cooperat. Work Social Comput.*, Feb. 2015, pp. 600–611.

[162] J. Wang, B. Guo, and L. Chen, "Human-in-the-loop machine learning: A macro-micro perspective," 2022, *arXiv:2202.10564*.

[163] D. S. Nunes, P. Zhang, and J. Sá Silva, "A survey on human-in-the-loop applications towards an internet of all," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 944–965, 2nd Quart., 2015.

[164] J. Zhang, P. Fiers, K. A. Witte, R. W. Jackson, K. L. Poggensee, C. G. Atkeson, and S. H. Collins, "Human-in-the-loop optimization of exoskeleton assistance during walking," *Science*, vol. 356, no. 6344, pp. 1280–1284, Jun. 2017.

[165] A. Donabedian, "Evaluating the quality of medical care," *Milbank Quart.*, vol. 83, no. 4, pp. 691–729, Dec. 2005.

[166] X. Chen, X. Wang, and Y. Qu, "Constructing ethical AI based on the 'Human-in-the-Loop' system," *Systems*, vol. 11, no. 11, p. 548, Nov. 2023.

[167] J. Kreutzer, S. Riezler, and C. Lawrence, "Offline reinforcement learning from human feedback in real-world sequence-to-sequence tasks," in *Proc. 5th Workshop Structured Predict. NLP*, 2021, pp. 37–43.

[168] 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.

[169] A. Kapoor, J. C. Caicedo, D. Lischinski, and S. B. Kang, "Collaborative personalization of image enhancement," *Int. J. Comput. Vis.*, vol. 108, nos. 1–2, pp. 148–164, May 2014.

[170] J.-S. Jwo, C.-S. Lin, and C.-H. Lee, "Smart technology–driven aspects for human-in-the-loop smart manufacturing," *Int. J. Adv. Manuf. Technol.*, vol. 114, nos. 5–6, pp. 1741–1752, May 2021.

[171] C. Chai and G. Li, "Human-in-the-loop techniques in machine learning," *IEEE Data Eng. Bull.*, vol. 43, no. 3, pp. 37–52, Jun. 2020.

[172] B. Settles, "Closing the loop: Fast, interactive semi-supervised annotation with queries on features and instances," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2011, pp. 1467–1478.

[173] T. Y. Lee, A. Smith, K. Seppi, N. Elmqvist, J. Boyd-Graber, and L. Findlater, "The human touch: How non-expert users perceive, interpret, and fix topic models," *Int. J. Hum.-Comput. Stud.*, vol. 105, pp. 28–42, Sep. 2017.

[174] N. M. Marquand, "Automated modeling of human-in-the-loop systems," Thesis, Purdue Univ. Graduate School, 2021, doi: 10.25394/PGS.16840279.v1.

[175] 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.

[176] W. Xu, M. J. Dainoff, L. Ge, and Z. Gao, "Transitioning to human interaction with AI systems: New challenges and opportunities for HCI professionals to enable human-centered AI," *Int. J. Hum.-Comput. Interact.*, vol. 39, no. 3, pp. 494–518, Feb. 2023.

[177] D. Bansal, Y. Hao, A. Hiranaka, R. Martin-Martin, C. Wang, and R. Zhang, "Dual representation for human-in-the-loop robot learning," Stanford Univ., Stanford, CA, USA, Tech. Rep., 2022.

[178] T. Brown et al., "Language models are few-shot learners," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds. Curran Associates, 2020, pp. 1877–1901. [Online]. Available: https:// proceedings. neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

[179] M. Huang, W. Gao, and Z.-P. Jiang, "A data-based lane-keeping steering control for autonomous vehicles: A human-in-the-loop approach," in *Proc. 35th Chin. Control Conf. (CCC)*, Jul. 2016, pp. 8974–8979.

[180] Z. Hu, Y. Zhang, Q. Li, and C. Lv, "A novel heterogeneous network for modeling driver attention with multi-level visual content," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 24343–24354, Dec. 2022.

[181] G. Lin, H. Li, C. K. Ahn, and D. Yao, "Event-Based finite-time neural control for human-in-the-loop UAV attitude systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 12, pp. 10387–10397, 2023, doi: 10.1109/TNNLS.2022.3166531.

[182] Y. Cao and Y. Song, "Performance guaranteed consensus tracking control of nonlinear multiagent systems: A finite-time function-based approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 4, pp. 1536–1546, Apr. 2021.

[183] S.-L. Dai, S. He, X. Chen, and X. Jin, "Adaptive leader–follower formation control of nonholonomic mobile robots with prescribed transient and steady-state performance," *IEEE Trans. Ind. Informat.*, vol. 16, no. 6, pp. 3662–3671, Jun. 2020.

[184] Y. Liu, X. Liu, Y. Jing, H. Wang, and X. Li, "Annular domain finite-time connective control for large-scale systems with expanding construction," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 10, pp. 6159–6169, Oct. 2021.

[185] W. Zhou, Y. Wang, C. K. Ahn, J. Cheng, and C. Chen, "Adaptive fuzzy backstepping-based formation control of unmanned surface vehicles with unknown model nonlinearity and actuator saturation," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14749–14764, Dec. 2020.

[186] L. Liu, D. Wang, Z. Peng, and Q.-L. Han, "Distributed path following of multiple under-actuated autonomous surface vehicles based on data-driven neural predictors via integral concurrent learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 12, pp. 5334–5344, Dec. 2021.

[187] Y. Lu and L. Bi, "Human behavior model-based predictive control of longitudinal brain-controlled driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1361–1374, Mar. 2021.

[188] P. Hang, C. Lv, Y. Xing, C. Huang, and Z. Hu, "Human-like decision making for autonomous driving: A noncooperative game theoretic approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 4, pp. 2076–2087, Apr. 2021.

[189] J. Wu, W. Huang, N. de Boer, Y. Mo, X. He, and C. Lv, "Safe decision-making for lane-change of autonomous vehicles via human demonstration-aided reinforcement learning," in *Proc. IEEE 25th Int. Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2022, pp. 1228–1233.

[190] X. He, H. Yang, Z. Hu, and C. Lv, "Robust lane change decision making for autonomous vehicles: An observation adversarial reinforcement learning approach," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 1, pp. 184–193, Jan. 2023.

[191] Z. Hu, Y. Xing, W. Gu, D. Cao, and C. Lv, "Driver anomaly quantification for intelligent vehicles: A contrastive learning approach with representation clustering," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 1, pp. 37–47, Jan. 2023.

[192] Z. Hu, Y. Zhang, Y. Xing, Y. Zhao, D. Cao, and C. Lv, "Toward human-centered automated driving: A novel spatiotemporal vision transformer-enabled head tracker," *IEEE Veh. Technol. Mag.*, vol. 17, no. 4, pp. 57–64, Dec. 2022.

[193] H. Chen, J. Zhang, and C. Lv, "RHONN modelling-enabled nonlinear predictive control for lateral dynamics stabilization of an in-wheel motor driven vehicle," *IEEE Trans. Veh. Technol.*, vol. 71, no. 8, pp. 8296–8308, Aug. 2022.

[194] Y. Zhang, P. Hang, C. Huang, and C. Lv, "Human-like interactive behavior generation for autonomous vehicles: A Bayesian game-theoretic approach with Turing test," *Adv. Intell. Syst.*, vol. 4, no. 5, 2022, Art. no. 2100211.

[195] H. J. Kim and J. H. Yang, "Takeover requests in simulated partially autonomous vehicles considering human factors," *IEEE Trans. Hum.-Mach. Syst.*, vol. 47, no. 5, pp. 735–740, Oct. 2017.

[196] J. Mao, C. Gan, P. Kohli, J. B. Tenenbaum, and J. Wu, "The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision," in *Proc. Int. Conf. Learn. Represent.*, 2019, pp. 1–28.

[197] W. Guo, "Explainable artificial intelligence for 6G: Improving trust between human and machine," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 39–45, Jun. 2020.

[198] C.-H. Chiu, Y. Koyama, Y.-C. Lai, T. Igarashi, and Y. Yue, "Human-in-the-loop differential subspace search in high-dimensional latent space," *ACM Trans. Graph.*, vol. 39, no. 4, pp. 1–85, Aug. 2020.

[199] M. Sujan et al., "Human factors and ergonomics in healthcare AI," Sep. 2021, doi: 10.13140/RG.2.2.22455.85924.

[200] R. Koster, J. Balaguer, A. Tacchetti, A. Weinstein, T. Zhu, O. Hauser, D. Williams, L. Campbell-Gillingham, P. Thacker, M. Botvinick, and C. Summerfield, "Human-centred mechanism design with democratic AI," *Nature Hum. Behav.*, vol. 6, no. 10, pp. 1398–1407, Jul. 2022.

[201] J. H. Kim, L. Rothrock, and A. Tharanathan, "Applying fuzzy linear regression to understand metacognitive judgments in a human-in-the-loop simulation environment," *IEEE Trans. Hum.-Mach. Syst.*, vol. 46, no. 3, pp. 360–369, Jun. 2016.

[202] D. Wei, Z. Li, Q. Wei, H. Su, B. Song, W. He, and J. Li, "Human-in-the-Loop control strategy of unilateral exoskeleton robots for gait rehabilitation," *IEEE Trans. Cognit. Develop. Syst.*, vol. 13, no. 1, pp. 57–66, Mar. 2021.

[203] A. P. Dani, I. Salehi, G. Rotithor, D. Trombetta, and H. Ravichandar, "Human-in-the-loop robot control for human–robot collaboration: Human intention estimation and safe trajectory tracking control for collaborative tasks," *IEEE Control Syst. Mag.*, vol. 40, no. 6, pp. 29–56, Dec. 2020.

[204] A. B. Farjadian, B. Thomsen, A. M. Annaswamy, and D. D. Woods, "Resilient flight control: An architecture for human supervision of automation," *IEEE Trans. Control Syst. Technol.*, vol. 29, no. 1, pp. 29–42, Jan. 2021.

[205] Q. Zhang, K. Kim, and N. Sharma, "Prediction of ankle dorsiflexion moment by combined ultrasound sonography and electromyography," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 1, pp. 318–327, Jan. 2020.

[206] S. Munir, J. A. Stankovic, C. M. Liang, and S. Lin, "Reducing energy waste for computers by human-in-the-loop control," *IEEE Trans. Emerg. Topics Comput.*, vol. 2, no. 4, pp. 448–460, Dec. 2014.

[207] R. Nasiri, H. Aftabi, and M. N. Ahmadabadi, "Human-in-the-loop weight compensation in upper limb wearable robots towards total muscles' effort minimization," *IEEE Robot. Autom. Lett.*, vol. 7, no. 2, pp. 3273–3278, Apr. 2022.

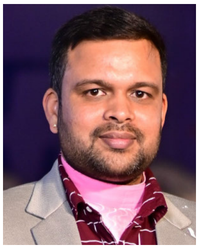
SUSHANT KUMAR was a Visvesvaraya Ph.D. Fellow with the Ministry of Electronics and Information Technology, GoI. He is currently a Research Associate III with the Department of Computer Science and Engineering, Indian Institute of Technology (IIT BHU) Varanasi, Varanasi, India. His research interests include joint compressed sensing, dictionary learning, and machine learning for biomedical signal analysis.

SUMIT DATTA (Senior Member, IEEE) is currently an Assistant Professor with the School of Electronic Systems and Automation, Digital University Kerala (Formerly IIITM Kerala), India. Prior to joining DUK, he was with the Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, India, as a Postdoctoral Fellow. His research interests include biomedical signal/image processing, compressed sensing MRI, super-resolution, and medical image analysis using deep learning.

VISHAKHA SINGH is currently pursuing the Ph.D. degree with the Department of Computer Science and Engineering, IIT (BHU) Varanasi, Varanasi, India. She is a Prime Minister Research Fellow. She has authored and coauthored several articles in peer-reviewed high-impact factor journals. Her research interests include artificial intelligence, deep learning, machine learning, and multi-objective optimization.

DEEPANWITA DATTA received the Ph.D. degree in computer science and engineering from Indian Institute of Technology (BHU) Varanasi, India. She is currently an Assistant Professor of information system management with Indian Institute of Management (IIM) Sambalpur. Before that, she was a full-time Computer Science and Engineering Faculty Member with Wrexham Glyndwr University, U.K. Previously, she was a Research Fellow (data mining & AI) with the University of the West of England, Bristol, U.K. There, she was involved in the Innovate U.K. Project, where her goal was to develop machine learning models for cost and budget prediction for partners, such as Network Rail, Highways England, and Transport for London (TfL). She has been conferred the prestigious ERCIM-Alain Bensoussan Fellowship for postdoctoral research with NTNU, Norway. She has also been a Visiting Scientist with VTT, Finland. Her areas of expertise include data analytics, data science, machine learning, and image retrieval, with a special focus on data analysis and text analysis. With over nine years of academic experience, her skill set includes professional-level data engineering and data management.

SANJAY KUMAR SINGH (Senior Member, IEEE) is currently a Professor with the Department of Computer Science and Engineering, Indian Institute of Technology (BHU) Varanasi, Varanasi, India. He has authored or coauthored more than 150 national and international journal publications, book chapters, and conference papers. His current research interests include machine learning, deep learning, computer vision, medical image analysis, pattern recognition, and biometrics. He is a Senior Member of ACM and the Computer Society of India. He is also a guest editorial board member and a reviewer for many international journals of repute.



RITESH SHARMA is currently an Assistant Professor with the Department of Information and Communication Technology, Manipal Institute of Technology (MIT), Manipal. Prior to joining MIT, he was with the Department of Computer Science and Engineering, Institute of Technology (BHU) Varanasi, Varanasi, India, as a Senior Research Fellow. He has published several peer-reviewed journal articles. His research interests include artificial intelligence, deep learning, and machine learning.

Open Access funding provided by ‘Manipal Academy of Higher Education’ within the CRUI CARE Agreement