Human-Computer Intelligent Interaction: Foundations, Technologies, and Future Perspectives

SurveyForge

Abstract— Human-computer intelligent interaction (HCII) is a transformative paradigm that integrates artificial intelligence into human-computer interaction to create systems capable of adaptive, intuitive engagements. This comprehensive survey examines the landscape of HCII, detailing its evolution, foundational principles, and technological enablers. Key research dimensions include multimodal interaction systems that synthesize inputs like speech, gestures, and eye movements to enhance user experiences, along with computational architectures such as multimodal transformers and cross-modal learning methods which address challenges of scalability and resource intensity. The paper identifies significant challenges in ethical considerations, transparency, and inclusivity, advocating for user-centered approaches through explainable AI and socio-technical designs. Technological advancements such as brain-computer interfaces and extended reality are highlighted as promising pathways for enhancing collaboration and accessibility. Future directions emphasize reducing cognitive load through proactive systems that anticipate user needs while maintaining ethical standards. The survey underscores HCII's potential to redefine human-computer interactions across diverse applications by focusing on adaptive technology integrated seamlessly with human-centric values.

Index Terms—Multimodal Interaction Systems, Adaptive Intelligence Integration, Explainable Artificial Intelligence

1 Introduction

Human-Computer Intelligent Interaction (HCII) represents a transformative paradigm in the evolving landscape of human-computer interaction (HCI), emphasizing systems enriched with artificial intelligence (AI) that facilitate intuitive, adaptive, and effective engagement between humans and machines. Unlike traditional HCI paradigms, HCII aims to dissolve rigid interaction boundaries, allowing systems not merely to respond to user actions but to anticipate, learn, and adapt dynamically. The 21st century has seen the migration from static interfaces to systems capable of real-time personalization and contextual comprehension, redefining usability and utility for computational technologies [1], [2].

Historically, HCI has bridged cognitive science, software engineering, and design to craft usable and purposeful systems. However, HCII amplifies this synergy, integrating advances across areas such as machine learning, natural language processing (NLP), and multimodal interaction. For instance, gaze-tracking systems outlined in [3] show how modern technologies dynamically capture user attention for seamless interactions, while gesture-based HCI explored in [4] demonstrates non-invasive, device-free mechanisms for human-computer engagement. Together, these advancements coalesce into an HCII paradigm that prioritizes seamless, naturalistic communication across modalities.

A critical facet of HCII lies in its interdisciplinary nature. Cognitive science contributes insights into human perception and decision-making, enabling designs that reduce cognitive load and foster intuitiveness [5]. Meanwhile, AI advancements like large language models (LLMs) bolster conversational systems, allowing personalized interactions tailored to user preferences and behaviors [6], [7]. These interdisciplinary integrations reveal trade-offs between system transparency, computational optimization, and user autonomy, highlighting active research challenges such as achieving explainability without compromising effectiveness [8], [9].

Advancing HCII systems brings emerging challenges, which demand nuanced approaches. The ethical balance between pervasive data collection and user agency requires frameworks that integrate fairness, privacy stewardship, and trust [2], [10]. Moreover, shifting the interaction paradigm from user compliance to active partnership necessitates a redefinition of "intelligent systems," as outlined by frameworks like hybrid intelligence systems that harmonize human-AI collaboration [11].

Looking forward, the HCII field stands poised to transform various domains through further breakthroughs in multimodal comprehension, adaptive interaction, and co-creative systems. The rise of more autonomous technologies, as explored in human-AI teaming dynamics [12], signals a future where systems will not only focus on understanding human intentions but meaningfully augment cognitive and creative capabilities. Ultimately, HCII will define the next standard in human-centric computing, shaping intelligent systems that operate not merely as tools but as empathetic collaborators in diverse societal and industrial applications.

2 Foundations and Theoretical Frameworks of Human-Computer Intelligent Interaction

2.1 Core Principles of Human-Centered Design

Human-centered design (HCD) lies at the heart of human-computer intelligent interaction (HCII), serving as a blueprint for developing systems that prioritize user needs, accessibility, and empowerment. Fundamentally, HCD is not merely about optimizing functional performance; it seeks to create intuitive, inclusive, and adaptive interactions that cater to diverse human behaviors and contexts. Grounded in interdisciplinary research spanning human-computer interaction (HCI), artificial intelligence (AI), and cognitive science, the principles of HCD enable intelligent systems to align with both technical goals and societal values.

A core tenet of HCD is ensuring usability and intuitiveness in system design. Intuitive interaction minimizes cognitive overhead, enabling users to perform tasks seamlessly without detailed prior knowledge or instructions [13]. Methodologies such as iterative usability testing and participatory design frameworks ensure systems remain easy to learn and operate, even for non-technical users. However, achieving this presents trade-offs, particularly in balancing system complexity and clarity. Advanced AI models like large language models, which decode naturalistic user input, can enable enhanced intuitiveness but at the cost of requiring robust training datasets and fine-tuned human-system alignment [14].

Accessibility and universal design principles are indispensable in HCD, ensuring equitable interaction experiences for users across different abilities and cultural contexts. Technologies such as gesture recognition and eye-tracking exemplify efforts toward inclusive design [3], [15]. However, implementation challenges persist, especially in adapting systems for situational impairments, as highlighted in approaches like Human I/O's multimodal detection framework [16].

Adaptivity and personalization form another pillar of HCD, enabling intelligent systems to dynamically tailor their behavior based on user context and preferences. Techniques such as reinforcement learning (RL) offer substantial promise in this regard, providing systems with the ability to learn from user feedback and adapt interaction flows [17]. Yet, adaptivity introduces ethical challenges, particularly concerning data privacy and systemic bias [2]. Effective solutions demand federated learning mechanisms to ensure individualization while safeguarding user autonomy [2].

Embedding empathy in system design amplifies emotional engagement and user trust. For instance, conversational AI systems increasingly leverage emotion recognition to foster empathetic interactions [6]. Nonetheless, operationalizing empathy raises questions about over-simulating humanness, as excessive anthropomorphism risks misleading users about systems' actual capabilities, especially in critical domains like healthcare or public safety [18].

In conclusion, achieving the core principles of human-centered design necessitates navigating intricate trade-offs between functional optimization, inclusivity, and user autonomy. While emerging trends in contextual adaptivity and affective computing provide promising opportunities, they require careful integration of ethical considerations and interdisciplinary insights. Future research should explore multimodal interaction systems that holistically address cognitive, physical, and emotional user dimensions, ensuring that intelligent systems remain not only effective but also equitable and meaningful in their interactions.

2.2 Theoretical Models and Cognitive Frameworks

Theoretical models and cognitive frameworks form the backbone of human-computer intelligent interaction (HCII), offering foundational insights into designing systems that enhance usability, trust, and user alignment. These frameworks emphasize the critical importance of cognitive load management, user-centered thinking, and shared autonomy—pillars that directly influence interaction paradigms, system design strategies, and evaluation methodologies.

Cognitive load theories provide a lens for understanding how humans allocate mental resources during interactions, significantly impacting task efficiency and error rates. Interfaces that overload users, whether through excessive informational density or misaligned prioritization, can degrade performance and satisfaction. Research highlights various strategies to mitigate mental strain, including the use of optimized geometrical patterns and icon structures to improve task acquisition times [19]. Furthermore, neuroimaging techniques such as EEG offer real-time insights into cognitive states like workload, attention, and error recognition, enabling dynamic system adaptations [20]. However, implementing such real-time models across diverse user populations presents challenges, particularly under dynamic and noisy environmental conditions.

User-centered and participatory design models complement these cognitive approaches by systematically incorporating user preferences, feedback, and behaviors throughout system development. Participatory frameworks empower users, particularly non-technical collaborators, to shape system functionality, providing practical applications in areas like distributed human computation [21]. Iterative implementations of these models have led to higher system relevance and adoption. Nevertheless, the integration of stakeholder diversity into these frameworks faces hurdles, including navigating cultural, linguistic, and cognitive differences amidst persistent biases in datasets and design practices [22], [23].

Shared autonomy frameworks further advance HCII by balancing human and machine control, ensuring that systems amplify rather than diminish human agency. Research into collaborative ideation and hybrid systems highlights how flexible autonomy can alleviate cognitive overload and improve situational awareness during prolonged interactions [21], [24]. Dynamic autonomy models, which adjust control levels based on real-time contextual parameters and trust dynamics, exemplify how shared autonomy can adapt to diverse needs. However, such systems also introduce ethical questions regarding moral responsibility and transparency—key areas requiring further exploration [25].

The growing emphasis on multi-dimensional user modeling reinforces these frameworks, enabling deeper adaptivity and personalization. By integrating behavioral, intentional, and environmental data, systems can acquire a nuanced understanding of users, which improves predictive performance in applications ranging from information retrieval to mental-health coaching [26], [27]. Yet, cross-cultural benchmarks for model interpretability and scalability across heterogeneous populations remain areas that require significant methodological advancement.

Looking ahead, HCII research must confront these foundational challenges through the synthesis of advanced cognitive models, participatory design principles, and inclusive user modeling practices. Refining cognitive load metrics, addressing cultural sensitivities, and further developing shared autonomy frameworks stand out as pivotal directions. Simultaneously, the integration of neuroimaging and multimodal data analytics offers transformative potential for improving real-time adaptivity and context-awareness, propelling HCII toward more intuitive, inclusive, and ethically grounded interactions.

2.3 Role of Artificial Intelligence in Interactive Systems

Artificial Intelligence (AI) has fundamentally reshaped the landscape of human-computer interaction (HCI) by enabling systems to exhibit intelligence that adapts, learns, and predicts, thereby fostering more intuitive, context-aware, and effective interfaces. This modernization underpins a paradigm shift from static interfaces to dynamic systems capable of addressing individual user needs and environmental complexities in real time. Central to this transformation are AI-driven enhancements in three key domains: context-awareness, decision support, and adaptivity.

Context-awareness is a cornerstone of interactive systems, where AI facilitates the interpretation of user behavior and surrounding environments to deliver tailored experiences. AI systems leverage multimodal data—such as sensor inputs, user preferences, and historical behavior—to infer context dynamically. For instance, shared autonomy mechanisms in human-robot collaboration utilize probabilistic planning and game-theoretic algorithms to model user intention and adapt those models based on situational context [28], [29]. High-dimensional data fusion architectures and energy-function-based safe exploration further enhance these systems' responsiveness, ensuring safety and efficacy in environments with uncertain human behaviors [30]. Despite these advancements, achieving scalable and accurate context interpretation remains a technical challenge, especially under resource constraints in real-time applications.

AI also amplifies decision support within interactive systems by embedding systems with predictive and prescriptive capabilities. Decision aids utilizing machine learning (ML) optimize human-centric outcomes by recommending or refining tasks based on observed user behavior. For example, algorithms incorporating trust-based feedback loops dynamically adjust automation transparency to calibrate user trust and improve decision efficiency [31]. In creative settings, co-creative AI systems not only refine user inputs but also augment ideation processes by iteratively generating variations, fostering a nonlinear synergy between humans and machines [32], [33]. Model interpretability and explainability, however, persist as critical limitations in decision-support systems, potentially undermining user trust without sufficient transparency [2].

Adaptivity in AI-fueled systems is driven by machine learning mechanisms that evolve through user interactions and feedback loops. Techniques such as reinforcement learning (RL) enable continuous system optimization, even in dynamic or ambiguous contexts. Interactive systems employing active preference learning, for instance, iteratively solicit user inputs to fine-tune robotic behaviors, effectively minimizing performance variances and individualizing user experiences [34]. Additionally, adaptive strategies are critical for scaling interaction infrastructures—such as hierarchical cooperation architectures—harmonizing the variability in user actions with system capabilities [35]. Challenges lie in balancing adaptation with stability, avoiding over-adjustment to transient behaviors without compromising long-term system usability.

Emerging trends in AI-driven systems signify significant opportunities and unresolved challenges. Methods such as leveraging generative AI for anticipatory content creation and advanced modeling of human cognitive states offer promising pathways to deepen system-human synergy [36], [37]. Future research must prioritize ethical transparency, scalability in complex environments, and inclusivity to ensure intelligent systems serve diverse populations effectively. The ongoing co-evolution of human and AI roles within interaction paradigms underscores the necessity for systems that are not just reactive tools but active thought partners, enhancing collaboration across contexts and domains [38].

2.4 Trust and Ethical Foundations

Trust and ethical considerations are pivotal in designing, implementing, and adopting human-computer intelligent interaction (HCII) systems. Trust forms the foundation of user acceptance, while ethical design ensures alignment with societal values, fairness, and accountability. This subsection delves into strategies for fostering trust through transparency, fairness, explainability, and privacy, while also addressing key ethical challenges intrinsic to HCII systems.

Establishing trust in HCII systems involves promoting predictability, reliability, and operational transparency. Explainability is fundamental in this context, as it enables users to comprehend system decision-making processes and outputs [2], [39]. Yet, achieving effective explainability requires navigating inherent trade-offs: overly simplified explanations may compromise technical fidelity, whereas highly detailed descriptions can alienate non-expert users. Recent advancements propose user-centric explanation frameworks tailored to varying levels of expertise, striking a balance between accessibility and technical rigor [10], [40].

Fairness in AI-augmented HCII systems poses another pressing challenge, as algorithmic decisions can inadvertently perpetuate biases embedded within training data, resulting in inequitable treatment of different demographic groups. Bias mitigation strategies include pre-processing techniques to cleanse training datasets, designing bias-aware algorithms, and employing post-hoc evaluation frameworks [41], [42]. However, defining consistent fairness metrics and managing trade-offs between individual and collective fairness remain complex issues. For example, efforts to reduce bias for one subgroup might unintentionally exacerbate disparities for another [10].

Privacy and data stewardship are integral to ethical HCII design, particularly as these systems rely on large-scale data collection to enable personalization. Tensions between delivering personalized experiences and safeguarding user privacy necessitate innovative approaches, such as federated learning and differential privacy, which allow insights to be drawn from data while preserving confidentiality [10], [34]. Nonetheless, ensuring clear user consent and accountability in real-time data processing remains an ongoing challenge in managing data responsibly and ethically.

Accountability mechanisms play a critical role in ensuring fairness, assigning responsibility, and fostering user trust. Transparent systems, often characterized as employing "glass-box" methods, seek to provide interpretability at both the algorithmic and outcome levels, which allows stakeholders to identify potential sources of systemic errors or biases [43], [44]. Human-in-the-loop frameworks, particularly in high-stakes domains like healthcare or criminal justice, further underscore the importance of robust accountability measures to ensure both user safety and equitable outcomes [42], [45].

Emerging paradigms emphasize the necessity of translating ethical principles into practical system designs. Participatory design methodologies, which integrate diverse stakeholder viewpoints throughout the system lifecycle—from conception to deployment—have shown promise in embedding ethical considerations while enhancing user engagement and trust. These approaches mitigate biases and align systems with societal and cultural values [46], [47]. Advancing such strategies requires dynamic, context-sensitive evaluation frameworks and adaptive explainability mechanisms to address evolving ethical and practical needs comprehensively.

To conclude, realizing trustworthy and ethical HCII systems mandates interdisciplinary collaboration and iterative improvement across fairness, explainability, privacy, and accountability dimensions. By actively involving users and aligning systems with ethical priorities, intelligent interactions can balance technological advancements with societal responsibility, fostering the development of human-centered AI ecosystems.

3 Multimodal Interaction and Adaptive Interfaces

3.1 Foundations of Multimodal Interaction Systems

Multimodal interaction systems represent a cornerstone of advanced human-computer interaction, leveraging the complementary strengths of diverse sensory modalities—such as vision, speech, gestures, and touch—to enable seamless and natural communication between humans and intelligent systems. This subsection lays the theoretical and technical foundations underlying such systems, addressing their structure, capabilities, integration mechanisms, and their implications for adaptive, user-centric design.

Central to multimodal interaction systems is the concept of leveraging core modalities to compensate for limitations in any singular sensory channel. For example, vision-based systems are highly effective for spatial understanding and visual recognition but may struggle in environments with low visibility or occlusions. Speech, while intuitive for real-time input, suffers from performance drops in noisy environments. By combining redundant and complementary cues from multiple modalities, systems gain robustness, accuracy, and flexibility, enabling novel interaction paradigms [3], [4].

A foundational technical challenge in these systems lies in achieving effective synchronization and fusion of multimodal data streams. Temporal alignment strategies, such as dynamic time warping or cross-attention models, are key to creating coherent interactions. Additionally, machine learning approaches like multimodal transformers and shared latent representations play an increasing role in enabling systems to model synergistic relationships across modalities [17], [48]. For instance, large language models combined with gesture recognition frameworks allow systems to interpret user intent beyond the constraints of traditional natural language processing [49].

The advantages of multimodal integration extend beyond enhanced input-output fidelity—they also foster accessibility by tailoring interactions to diverse user needs. Systems that combine tactile and auditory feedback, for example, facilitate interaction for visually impaired individuals, while brain-computer interfaces (BCIs) open new possibilities for hands-free operation, extending usability for individuals with limited motor functions [4]. However, achieving such personalized adaptability demands addressing computational challenges in managing high-dimensional input streams in real-time while maintaining resource efficiency.

Emergent trends in multimodal systems include the incorporation of embodied AI and human-object interaction frameworks, supporting contexts where physical and virtual interactions coexist, such as augmented reality (AR) or robotics. These applications benefit from defining granular interaction units, such as chains-of-contact models, where tasks are decomposed by interaction regions and goals [50], [51].

Despite these advances, challenges remain. Cross-modal ambiguity, where signals conflict or provide redundant information, complicates meaningful fusion. Error correction and resilience strategies, such as redundancy suppression and trust calibration mechanisms, are vital to managing this complexity [8]. The ethical implications of sensor-based multimodal learning—especially regarding privacy and surveillance concerns—must also inform future developments [46].

In conclusion, the foundational principles and methodologies of multimodal systems underscore their transformative potential in HCI. Ensuring scalability in diverse contexts, explaining system decisions, and integrating creative interaction channels such as emotion sensing present promising areas for exploration. Advances in probabilistic data fusion and human-centered evaluation metrics will further define the state-of-the-art, enabling multimodal systems to close the gap between humans and intelligent agents in increasingly natural, adaptive, and inclusive ways [9], [52].

3.2 Cross-Modal Fusion and Learning

"'markdown Cross-modal fusion and learning form the backbone of robust multimodal interaction systems, enabling seamless integration of data from heterogeneous modalities such as speech, vision, and tactile feedback. Serving as the bridge between the theoretical underpinnings of multimodal systems and their practical, adaptive applications, cross-modal fusion ensures cohesion across diverse sensory inputs and enriches user interaction experiences. This subsection delves into the methodologies, architectures, and key technological trends that characterize this critical domain.

At its core, cross-modal fusion involves aligning disparate modalities into a shared representational space, ensuring coherence across temporal, spatial, and semantic dimensions. Techniques such as canonical correlation analysis (CCA), which identifies shared subspaces by maximizing correlation across modalities, provide a foundational understanding of alignment methodologies. However, the limitations of static approaches in capturing temporal dependencies and contextual dynamics underscore the need for advanced techniques. Deep learning innovations, such as multimodal transformers and cross-attention mechanisms, address this gap by learning shared latent spaces that integrate embeddings while preserving modality-specific nuances. These architectures have achieved significant breakthroughs in tasks like audiovisual speech recognition and gesture-based command prediction, underscoring their relevance in real-world human-computer interaction scenarios [53], [54].

While robust fusion methods enhance system performance, the inherent variability and noise levels across modalities can introduce challenges related to redundancy and ambiguity. Addressing these requires dynamic and selective strategies, such as gating mechanisms, which filter inputs to prioritize informative signals. For instance, in assistive technologies, multimodal systems combining visual and auditory data have leveraged temporal gating to de-emphasize compromised visual inputs during occlusions, while amplifying auditory cues to maintain contextual understanding. Such dynamic interplay of modalities has driven substantial improvements in adaptability and precision for real-time applications [55].

Beyond fusion, cross-modal learning enables one modality to enrich or compensate for another, fostering complementary interactions across modalities. Visual recognition systems augmented by natural language annotations, for example, benefit from shared latent embeddings that enhance robustness when facing noisy datasets [56]. Similarly, reinforcement learning architectures tailored for interactive systems have demonstrated their efficacy in iteratively optimizing cross-modal interactions. By strategically determining when and how to prioritize modalities, these architectures enhance performance in dynamic tasks, reflecting the intricate interplay between human intent and system adaptability [17].

Nevertheless, cross-modal integration is not without significant challenges. Computational inefficiencies and susceptibility to data imbalances remain critical barriers, particularly in resource-constrained environments. Advances in efficient architectures, such as sparse multimodal transformers, offer promising solutions for scalability, while diagnostic tools like cross-modal alignment scores and mutual information measures provide valuable insights into improving fusion coherence [57], [58].

Emerging trends emphasize the role of generalist models, exemplified by large multimodal models capable of capturing inter-modal relationships at scale through expansive pretraining. However, the increasing power of these models underscores the critical need for transparency in fused representations, especially in domains where trust and interpretability are paramount, such as healthcare and education [59], [60].

As cross-modal fusion continues to advance, it provides a vital link between the foundational principles of multimodal interaction explored earlier and the adaptable, user-centered interfaces discussed next. Addressing outstanding challenges, such as ethical data usage, real-time latency optimization, and interpretability of fused representations, will cement cross-modal learning as a cornerstone of sophisticated multimodal systems, ensuring they dynamically adapt to the complex and evolving needs of their users. "'

3.3 Adaptive and Personalized Multimodal Interfaces

Adaptive and personalized multimodal interfaces represent a critical evolution in human-computer interaction, enabling systems to dynamically respond to individual user preferences, behaviors, and contextual variations across environments. These interfaces unify multiple interaction modalities such as speech, gesture, visual inputs, and contextual data to deliver a cohesively tailored user experience. Central to their effectiveness is the integration of adaptive algorithms that leverage user-specific models, combining machine learning, cognition-based frameworks, and real-time analytics.

Key to personalization is the ability to dynamically model user preferences and set behaviors through contextual data analysis. Machine learning approaches, including reinforcement learning and neural collaborative filtering, have played a pivotal role in achieving personalization at scale while considering multimodal input streams. For instance, reinforcement learning frameworks in shared control setups have shown efficacy in enabling human-machine systems to co-adapt to user preferences during collaboration [29]. Such methods facilitate data-driven personalization by allowing systems to modify behavior based on user responses and performance over time.

Context-aware personalization remains one of the most impactful dimensions of adaptive interfaces. These systems tap into spatial, temporal, or activity-driven context to optimize user-system engagement dynamically. For example, context-awareness in autonomous vehicle interactions has demonstrated the ability to balance cognitive load and user confidence by altering explanatory modalities in response to driver expertise and environmental variables [61]. Additionally, incorporating probabilistic modeling into adaptive systems allows for refining predictions about user states, enabling systems to recalibrate their interactions to suit nuanced and evolving preferences [30].

An essential design feature of adaptive multimodal systems is their ability to address accessibility through personalization. By adapting outputs to meet the needs of diverse user bases, including those with varying cognitive or physical abilities, these interfaces extend inclusivity. For example, conversational AI systems tailored for elderly users can detect cognitive impairments while providing meaningful, engaging interactions [62]. Such inclusivity is enabled by explicit user modeling that incorporates variables such as stress levels, environmental noise, and physical disabilities into adaptation pipelines.

However, the design and implementation of adaptive multimodal interfaces present challenges. Scalability is a significant concern, as real-time personalization across diverse modalities introduces computational complexity, particularly when fusing asynchronous data streams [29]. Moreover, issues related to data privacy and ethical considerations, such as ensuring bias-free personalization and enabling users to control how personalization occurs, remain open research areas [46].

Future directions for adaptive multimodal interfaces include enhancing cross-modal learning techniques to improve synergy between interaction channels, such as combining speech with non-verbal cues like gaze and gestures [6]. Furthermore, embedding hybrid cognitive architectures capable of generalized reasoning (e.g., combining symbolic AI with neural methods) can significantly enhance adaptation fidelity and robustness across novel contexts [37]. Researchers must also prioritize transparency, e.g., explainable adaptations tailored to user preferences, to ensure trust and sustained engagement [31].

Overall, adaptive and personalized multimodal interfaces represent both immense potential and substantial challenges. As systems become increasingly embedded in dynamic environments and diverse user landscapes, further research must continue to explore algorithmic scalability, ethical responsibility, and the interplay between personalization, inclusivity, and multimodal consistency.

3.4 Emerging Technologies in Multimodal Interfaces

Emerging technologies in multimodal interfaces are revolutionizing human-computer interaction, enhancing the integration and interpretation of diverse modalities with remarkable precision and depth. These advancements, encompassing cutting-edge sensing mechanisms, brain-computer interfaces (BCIs), and haptic feedback systems, collectively push the boundaries of interaction fidelity and immersive experiences, seamlessly expanding the capabilities of adaptive and personalized multimodal interfaces [63].

Advanced sensing mechanisms play a crucial role in capturing multimodal inputs with higher precision and reliability. By synthesizing various environmental and user-specific data streams—such as gestures, gaze, speech, and tactile inputs—these systems create cohesive, context-aware representations. Leveraging neural-symbolic fusion models and temporal attention mechanisms, they facilitate the real-time alignment of asynchronous signals, thereby overcoming challenges in dynamic, complex environments. For instance, multi-sensor systems have enabled interactive platforms to situate user expressions in their environments, fostering robust situational awareness. However, significant challenges remain, especially in synchronizing heterogeneous sensors and processing data in low-latency, resource-constrained settings [64]. Future advancements must prioritize energy-efficient algorithms and architectures capable of handling noisy or incomplete sensory streams, ensuring scalability and adaptability in multimodal ecosystems.

Brain-computer interfaces (BCIs) herald a paradigm shift by enabling direct neural interaction, bypassing conventional input modalities. By translating neural activity into actionable commands or expressive outputs, BCIs offer transformative solutions for accessibility and naturalistic interaction. Recent strides include machine learning models that integrate neurofeedback loops to refine the accuracy and responsiveness of neural signal interpretation, demonstrating potential in augmentative communication and control applications [24]. Nevertheless, BCIs face critical challenges in the real-time decoding of overlapping neural signals, requiring advancements in computational modeling and robust system adaptability. Furthermore, ethical and privacy concerns surrounding the use of neural data highlight the urgency of developing comprehensive data stewardship frameworks to safeguard against unauthorized use [10].

Haptic feedback systems further enrich multimodal interfaces by simulating tactile sensations, immersing users in both real and virtual environments. These systems extend beyond basic vibration feedback to include nuanced force, texture, and thermal simulations, enhancing interactivity in applications such as surgical training and remote operation. For example, high-fidelity haptic systems have demonstrated significant value in replicating detailed physical interactions, offering precise situational control in complex environments [65]. However, designing scalable and affordable haptic systems without compromising on feedback fidelity presents ongoing challenges that must be addressed to enable broader adoption.

As multimodal systems continue to evolve, the interplay between emerging technologies and adaptive interfaces will play a defining role in shaping next-generation human-computer interactions. Innovations in AR/VR platforms and wearables are already beginning to integrate these modalities, supporting more seamless, embodied, and personalized interactions [63]. To further advance this frontier, research must prioritize generalizable multimodal architectures that facilitate integration across diverse technologies while ensuring ethical implementation and equitable access. This trajectory not only highlights the transformative potential of emerging multimodal interfaces but positions them as key enablers of inclusive, adaptive, and human-centered intelligent systems, preparing the groundwork for their convergence with advances in scalability and context-aware personalization.

3.5 Challenges and Future Directions in Multimodal Interaction

The field of multimodal interaction holds immense promise for enhancing human-computer interfaces by leveraging the complementary strengths of diverse modalities such as vision, speech, gesture, and tactile feedback. However, significant challenges remain in designing, implementing, and deploying these systems. One fundamental barrier is the computational complexity of multimodal fusion, where diverse modalities must be aligned and integrated seamlessly. While advanced architectures such as multimodal transformers have demonstrated success in modeling cross-modal relationships [56], they often encounter scalability issues when processing heterogeneous and high-dimensional data streams. This restricts real-time applications in resource-constrained environments, especially in mobile or embedded devices [66].

Another pressing challenge lies in managing ambiguity and redundancy across modalities. Multimodal systems frequently receive conflicting or redundant data, such as overlapping speech and gesture commands. While probabilistic methods such as Bayesian fusion have shown promise in handling uncertainty [67], these approaches often struggle with dynamic environments where contextual relevance shifts rapidly. This limitation highlights the need for adaptive, context-aware alignment mechanisms that can dynamically prioritize certain modalities based on the interaction context [68].

Ethical and privacy concerns further complicate the deployment of multimodal systems. As these systems increasingly leverage sensitive data from multiple sources, they amplify risks of misuse and bias. Studies [56] emphasize the importance of designing systems that adhere to principles of fairness and transparency. However, achieving bias mitigation remains nontrivial, especially when biased training data from one modality influences overall fusion results. Privacy-preserving techniques, such as federated learning, offer some solutions but demand more research to accommodate the cross-modal nature of interaction systems [10].

Moreover, interoperability across modalities and devices presents notable technical hurdles. Multimodal systems must operate in diverse ecosystems while ensuring consistency and reliability. The lack of standardized frameworks for cross-device and cross-modality interactions hampers seamless integration [21]. Furthermore, empirical studies indicate that the user experience can degrade when multimodal interfaces are not appropriately aligned with cultural and individual differences [22].

Future research opportunities exist in several key areas. First, the development of neural-symbolic multimodal systems could combine the interpretability of symbolic reasoning with the adaptive power of neural models [58]. Second, advancing human-AI collaboration through proactive multimodal interfaces that adapt to user behavior without overstepping user agency holds significant potential [11]. Additionally, techniques such as interactive machine learning for real-time user feedback can improve system robustness and personalization [69]. Finally, addressing scalability through lightweight, energy-efficient multimodal architectures remains critical for their deployment in ubiquitous computing environments [66].

In summation, while advancements in multimodal interaction have achieved substantial milestones, unresolved challenges in scalability, ambiguity management, ethical considerations, and interoperability underscore the necessity for interdisciplinary efforts. By tackling these issues through innovative methods and human-centered approaches, multimodal systems can advance toward delivering seamless, robust, and inclusive interaction paradigms across diverse domains.

4 Computational Techniques for Intelligent Interaction

4.1 Machine Learning for Personalized Interaction

The incorporation of machine learning (ML) into personalized interaction systems has revolutionized human-computer interaction, enabling systems to dynamically adapt to user-specific preferences, behaviors, and contexts. Leveraging supervised, unsupervised, and reinforcement learning, these systems learn to anticipate and respond to individual user needs, thereby enhancing user satisfaction and engagement while fostering intuitive and seamless interactions.

Supervised learning, one of the most widely used paradigms, employs labeled datasets to predict user preferences and behavior patterns. For example, in intelligent tutoring systems, supervised models analyze historical student interaction data to tailor educational content, adapting to a learner's unique pace and comprehension level [70]. Advances in user behavior modeling—through algorithms such as gradient boosting and neural networks—have allowed for increasingly precise predictions of user intent, though these approaches often depend heavily on the availability and quality of annotated training data. This limitation underscores the need for scalable, high-quality data collection strategies.

Unsupervised learning, by contrast, excels in discovering latent user patterns without reliance on labeled data. Techniques such as clustering and dimensionality reduction uncover meaningful user archetypes and preferences within sparse or noisy datasets. For instance, unsupervised methods have been utilized in conversational recommender systems to segment users based on linguistic styles and contextual variables [71]. While inherently flexible, challenges persist in balancing interpretability with generalization, as users may exhibit behaviors that do not conform neatly to predefined clusters.

Reinforcement learning (RL) introduces an adaptive decision-making framework by enabling systems to optimize interaction strategies through trial-and-error processes. Interactive systems leveraging RL are particularly effective in complex, dynamic environments, where user preferences evolve over time. For example, RL-based dialog systems iteratively improve their conversational flow by rewarding behaviors that enhance user engagement and task success [17], [72]. The primary advantage of RL lies in its model-free architecture, which minimizes reliance on extensive domain knowledge. However, challenges such as sample inefficiency and delayed reward structures require ongoing research to refine algorithms and ensure scalability for real-world applications.

Emerging paradigms like federated learning further bolster personalization while addressing critical issues of privacy and computational efficiency. Federated learning enables decentralized training, wherein user-specific models are collaboratively updated without exposing private data, a feature increasingly crucial in healthcare and behavioral analysis domains [73]. Similarly, transfer learning facilitates model adaptation across domains, reducing the computational cost of personalization when vast amounts of user interactions are unavailable [74].

Despite these advancements, developing robust, real-time adaptive systems remains challenging. Scalability issues in large-scale deployment, biases in data and models, and ensuring fairness in diverse user populations are significant hurdles [22]. Future research must prioritize these aspects, along with exploring hybrid models integrating supervised, unsupervised, and reinforcement learning to maximize system adaptability. Generative AI models, particularly those leveraging extensive pre-trained architectures, offer promising avenues for constructing proactive agents that anticipate user needs rather than merely responding to them [53].

In conclusion, machine learning techniques furnish sophisticated computational foundations for personalized interaction systems, facilitating responsiveness, adaptability, and user-aligned decision-making. By addressing technical and ethical challenges, these models hold the potential to advance next-generation human-computer intelligent interaction, ensuring inclusivity, transparency, and enhanced user trust.

4.2 Natural Language Understanding and Conversational Systems

Natural Language Understanding (NLU) and conversational systems stand as cornerstone technologies for facilitating seamless and intuitive human-computer communication. Recent advancements in natural language processing (NLP), driven by large-scale language models, have significantly enriched these systems, enabling context-aware, dynamic, and human-like interactions. This subsection examines the computational approaches underpinning NLU and conversational agents, along with their major breakthroughs, ongoing challenges, and future trajectories.

Central to modern NLU systems are pre-trained language models such as GPT, BERT, and T5, which utilize transformer architectures and attention mechanisms to extract nuanced semantic representations from language data. These models excel in tasks like question answering, information retrieval, and machine translation by fine-tuning on domain-specific datasets, allowing them to cater to specialized applications. For example, dialogue systems demonstrate enhanced contextual understanding through mechanisms that encode sequential dependencies and user intent, empowering more coherent and relevant conversational exchanges [7], [75]. Additionally, advancements in context-aware language modeling, such as the integration of external knowledge graphs or memory modules, further enable these systems to sustain meaningful and contextually accurate dialogue over longer interactions [76], [77].

Another pivotal innovation is the incorporation of emotion recognition into dialogue management. By interpreting sentiment and affective states from user inputs, conversational systems are better positioned to deliver empathetic responses, thereby closing communication gaps between humans and machines. Research [26], [77] underscores the transformative impact of empathetic conversational interfaces in applications such as mental health support and well-being coaching, where emotionally attuned systems show greater efficacy compared to their neutral counterparts.

Despite these remarkable advancements, conversational AI faces notable challenges in ambiguity handling, multilingual processing, and adaptation to diverse linguistic structures. Multilingual conversational systems often confront difficulties in achieving cross-lingual transferability due to inconsistencies in how syntactically and semantically diverse languages are represented [23]. To address this, approaches like zero-shot and few-shot learning, powered by large-scale language models, have emerged to enable generalization across unseen languages without extensive retraining [60]. Furthermore, the development of dialogue systems tailored for multiturn and information-seeking scenarios—where information retrieval is seamlessly integrated into conversational flows—marks a significant step forward in personalizing and deepening user engagement [7], [78].

However, progress in conversational AI raises ethical concerns and challenges regarding fairness, interpretability, and alignment with user expectations. This includes mitigating biases in training data, ensuring inclusivity for underrepresented linguistic and demographic groups, and fostering trust by aligning interactions with societal norms [59], [79]. Future systems must navigate these challenges by balancing proactive conversational capabilities with respectful, unobtrusive interaction designs [78], thereby prioritizing user-centricity alongside technical performance.

In summary, while NLU and conversational systems have achieved significant progress, their future lies in addressing scalability, robustness in real-world applications, and the incorporation of ethical considerations. These efforts will be essential for creating adaptive, inclusive, and context-sensitive AI communication platforms that can meet the needs of a globally diverse user base.

4.3 Multimodal Data Fusion and Hybrid Interaction

Multimodal data fusion and hybrid interaction occupy a crucial domain in enabling human-computer intelligent interaction systems to interpret and act on information from diverse sources like vision, speech, tactile input, or gesture-based modalities. Integrating these disparate data streams into a coherent framework is necessary for achieving context awareness, adaptive responsiveness, and interaction fidelity. The fusion of multimodal inputs must address challenges such as temporal alignment, modality-specific uncertainties, and system scalability.

The cornerstone of multimodal data fusion lies in its ability to combine complementary modalities effectively, leveraging their individual strengths to overcome weaknesses. Methods such as shared latent space learning and cross-modal attention networks have emerged as prominent approaches. For example, techniques that map data into a unified representation improve processing coherence, enabling tasks such as human intention recognition during collaboration [80]. Notably, cross-modal transformers excel at capturing interdependencies between modalities by learning aligned representations that bridge semantic gaps [81].

Sensor fusion techniques play a critical role in human intention recognition, wherein multimodal data streams—such as visual inputs combined with tactile feedback—enable systems to accurately interpret user behaviors. These techniques often incorporate probabilistic models, notably Bayesian inference frameworks, to handle uncertainties and misaligned signals, ensuring robust performance even in dynamic environments [28], [82]. Importantly, emerging energy-based frameworks enable safe exploration of multimodal input spaces, dynamically adapting models based on user interactions and thereby enhancing real-time collaboration [30].

Real-time multimodal data processing necessitates low-latency architectures. The advent of parallelized deep learning pipelines combined with lightweight recurrent modules ensures computational efficiency, particularly in latency-sensitive applications such as autonomous systems or assistive robotics. However, the trade-off lies in balancing computational resource needs with the fidelity of interaction, as highlighted by the complexity of shared control systems [83].

Innovations in cross-modal learning further enable knowledge transfer across modalities, where information deduced from one modality informs predictions in another. The integration of reinforcement learning into hybrid systems exemplifies this synergy, facilitating dynamic system adaptivity [29]. Furthermore, emerging generalist models aim to unify modality processing under common architectures that synthesize diverse forms of interaction data into actionable insights [35].

Despite these advancements, significant challenges persist. Addressing redundancy and ambiguity in multimodal inputs requires sophisticated disambiguation strategies, including employing shared autonomy paradigms to mediate between conflicting data streams [84]. Scalability remains a concern given the high-dimensional nature of multimodal data, while ethical considerations emerge from privacy concerns as systems collect and analyze user data across diverse modalities.

Future directions in this field include the integration of symbolic reasoning with multimodal fusion methodologies, enabling systems to incorporate contextual rules into data-driven models for higher interpretability. The refinement of multimodal generalist models and the inclusion of user-feedback loops in data fusion pipelines are likely to redefine the boundaries of hybrid interaction, tailoring intelligent systems to diverse user needs while ensuring trust and reliability [31]. These advancements will catalyze the design of systems capable of offering seamless, predictive, and personalized experiences in both collaborative and independent interaction scenarios.

4.4 Robustness and Explainability in Real-Time Systems

Robustness and explainability are central to realizing effective and trustworthy real-time systems for intelligent human-computer interaction. These systems must deliver accurate and transparent decisions while maintaining responsiveness in dynamic environments. They confront challenges such as uncertainties in input data, unexpected environmental changes, and computational resource constraints. Simultaneously, ensuring interpretability of outcomes is essential, particularly in mission-critical domains such as healthcare, autonomous driving, and surveillance, where the consequences of errors are profound.

Achieving robustness in real-time intelligent systems necessitates strategies that anticipate and address errors, performance degradation, and operational variances. Central to these efforts are error detection and recovery mechanisms, which identify anomalies in system behavior and apply corrective actions, such as fault-tolerant design and dynamic adaptation techniques [17], [34]. Methods like dynamic model recalibration, often leveraging reinforcement learning and hybrid multimodal architectures, offer solutions to handle environmental shifts and reduce model brittleness when encountering novel conditions [17], [63]. However, maintaining real-time performance while achieving higher levels of robustness poses trade-offs, especially in high-stakes contexts. Predictive resource allocation strategies can mitigate this tension by focusing on critical interactions, thereby sustaining operational continuity even under constrained computational resources [64].

Explainability, a key enabler of trusted human-AI collaboration, adds another layer of complexity to real-time system design. For users to rely on these systems, the decisions made by intelligent models must be interpretable and intuitively presented. Explainable reinforcement learning frameworks integrate model outputs with human-intelligible justifications, narrowing the cognitive gap between system rationale and user understanding [17], [85]. However, balancing explainability with system latency presents challenges, as the reasoning processes behind explanations may slow response times in real-world settings [2]. Furthermore, inaccurate or inconsistent explanations—even paired with correct decisions—can erode user trust and degrade overall system performance, emphasizing the need for precise and well-calibrated explanatory mechanisms [86].

Comprehensive evaluation frameworks tailored to robustness and explainability remain an area of opportunity, necessary for navigating the trade-offs between these dimensions. Promising approaches include combining algorithm-centered stress testing with human-centered usability studies to assess system adaptability, user reliance, and explanation effectiveness under varied conditions [44], [87]. These metrics would offer a holistic means of evaluating systems and guiding iterative improvements.

Emerging research directions are beginning to fuse robustness and explainability with anticipatory system design. Context-sensitive adaptive modeling, for instance, enables systems to preemptively align behaviors with user expectations and environmental dynamics, fostering trust and resilience [17], [88]. Additionally, advances in human-in-the-loop learning—where users directly shape model behavior and explanations—provide groundbreaking opportunities to reduce brittleness and enhance interpretability [43]. Moving forward, striking a balance between computationally efficient robustness and user-aligned interpretability will remain an enduring and essential focus in the development of intelligent human-computer interaction systems.

4.5 Real-Time Interaction and Context-Sensitive Adaptation

Real-time interaction and context-sensitive adaptation have emerged as central capabilities in intelligent human-computer interaction systems, necessitating sophisticated computational frameworks that balance responsiveness, scalability, and adaptivity to dynamic user and environmental contexts. This subsection examines key methodologies and challenges in realizing these capabilities, focusing on temporal modeling, predictive context awareness, resource allocation, and the architecture of scalable real-time decision-making systems.

A critical facet of real-time interaction lies in dynamic resource allocation strategies that optimize computational resources to maintain low-latency responses under real-time constraints. These strategies often adopt reinforcement learning or optimization-based approaches to prioritize tasks dynamically based on contextual relevance [89]. While heuristic methods offer simplicity and efficiency, adaptive algorithms that model system resource consumption using Markov Decision Processes (MDPs) extend robustness by forecasting resource demands with high fidelity in multi-task environments. However, such frameworks face trade-offs between computational expense and real-time guarantees, especially in systems with constrained resources such as mobile or embedded devices.

Temporal interaction modeling is essential for understanding and responding to the evolving nature of user behavior and contextual changes. Recurrent neural networks (RNNs) and temporal convolution networks (TCNs) have been widely adopted to model temporal dependencies in interaction data. However, recent advancements in transformer architectures have further enabled dynamic adaptation by processing global temporal dependencies with attention mechanisms, proving effective in interactive tasks with irregular or long-term temporal patterns [40]. Despite their scalability, transformers necessitate strategies to manage computational complexity, such as sparse attention mechanisms, to maintain real-time performance.

Context-sensitive adaptivity leverages predictive modeling to anticipate user needs or environmental shifts proactively. Predictive context modeling combines multimodal data fusion with probabilistic inference techniques such as Bayesian Networks or Conditional Random Fields to derive actionable insights from incomplete or noisy data streams [56]. These predictions are used to trigger anticipatory responses, enabling systems to move from reactive to proactive interaction modes. While predictive frameworks increase accuracy and user satisfaction, challenges remain in achieving robustness to contextual ambiguities and unexpected user inputs.

Real-time decision engines represent the backbone of scalable interaction systems, requiring architectures tailored for low-latency data processing and inference. Distributed computational paradigms, particularly edge-cloud hybrid systems, have gained traction for reducing interaction latencies by localizing computation nearer to the user. These architectures often incorporate adaptive task offloading frameworks that optimize between high-frequency low-power edge computation and computationally intensive cloud operations. However, maintaining consistency in decision-making across distributed components presents a significant challenge, underscoring the need for consensus protocols and fault-tolerant designs [85].

Emerging trends point toward integrating machine learning paradigms that not only process multimodal data streams in real time but also align interaction dynamics with user mental models to foster trust and usability. For instance, approaches that incorporate user feedback loops, such as Interactive Attention Alignment (IAA), have shown promise in creating human-steerable systems that dynamically adjust their behavior based on user-guided refinements [90]. However, these methods can introduce challenges related to scalability and optimization convergence in dynamically changing environments, necessitating further research into hybrid reinforcement-based and evolutionary adaptation techniques.

Future directions in real-time, context-sensitive systems should emphasize robust evaluation frameworks that couple technical performance metrics—such as latency, throughput, and accuracy—with human-centered metrics, including trust calibration, user satisfaction, and accessibility. Critical advancements are also expected in integrating better predictive modeling with energy-efficient architectures, enabling systems to scale adaptively across diverse demographic and geographic conditions. By bridging computational efficiency with nuanced human-AI alignment, the next generation of these systems holds the potential to redefine intelligent interaction paradigms, enabling seamless, contextually relevant, and user-centric technologies.

4.6 Ethical and User-Centric Considerations in Computational Interaction

Ethical and user-centric considerations constitute essential pillars in the design and deployment of human-computer intelligent interaction (HCI2) systems. These dimensions are critical in ensuring that such systems are not only technologically robust but also socially acceptable and aligned with user needs. Core principles such as fairness, inclusivity, accountability, and trustworthiness govern the societal impact and usability of interaction-centric intelligent systems. This subsection synthesizes the challenges, existing solutions, and emerging trends in addressing these concerns while emphasizing the importance of user-centric design principles to bridge the gap between computational innovation and societal expectations.

Fairness in human-computer interaction aims to eliminate biases in both the data and algorithms that drive intelligent systems. Biased training datasets widely encountered in real-world applications often result in disproportionate performance, unfairly disadvantaging certain demographic groups and perpetuating discrimination [91]. Techniques such as data augmentation, adversarial debiasing, and bias-aware loss functions have been proposed to address these issues. While these methods improve algorithmic equitability, they introduce computational trade-offs, such as decreased interpretability and potential underperformance on underrepresented groups [92]. This underscores the critical need for innovative solutions that ensure fairness without compromising system scalability, accuracy, or inclusivity.

Inclusivity extends the goal of fairness by striving to ensure accessibility and personalization for diverse user populations, including those with linguistic, physical, or situational barriers. Intelligent systems are increasingly adopting multimodal interfaces that integrate voice, gesture, and tactile inputs to enable seamless interaction for users with disabilities or situational impairments [16], [93]. However, designing systems that balance personalization with computational efficiency remains a persistent challenge. Approaches such as clustering user preferences through transfer learning or employing adaptive algorithms that continuously evolve based on user interaction patterns show significant promise in addressing these challenges while promoting inclusivity [94].

The ethical handling of user data is paramount in fostering trust in intelligent interaction systems. The reliance on sensitive user data to enable real-time adaptation and personalization raises significant concerns about privacy and data stewardship. Federated learning has emerged as a key solution to preserve privacy by enabling distributed model training directly on user devices, reducing dependence on centralized data collection [95]. However, federated learning comes with its own limitations, such as communication overhead, device heterogeneity, and susceptibility to adversarial attacks [92]. Ensuring transparent communication of how data is used and obtaining informed user consent are critical design imperatives that must guide future system development.

Accountability is another cornerstone in cultivating user trust and confidence in HCI2 systems. Explainable AI (XAI) techniques offer a tangible means of ensuring accountability by allowing users to understand and evaluate system behavior. Employing multimodal approaches, such as combining textual explanations with visual summaries, has proven effective in bridging the gap between technical complexity and user comprehension [92]. Yet, achieving this balance between explainability and system performance remains a key challenge, as models often trade off accuracy for transparency, leading to suboptimal outcomes in specialized tasks.

Future advances in ethical and user-centric design should align closely with technological developments in real-time, context-sensitive interaction systems. For example, integrating multimodal capabilities with predictive and adaptive interaction models could ensure fairness and inclusivity in dynamic environments [96]. Standardizing ethical evaluation frameworks and methodologies through interdisciplinary collaborations across AI, the social sciences, and user advocacy domains could further enhance accountability and consistency in system assessment [92]. By integrating computational advancements with ethical scrutiny and prioritizing user needs, HCI2 systems stand to redefine intelligent interaction paradigms, fostering not only technological innovation but also societal trust, equity, and inclusion.

5 Applications and Domains of Human-Computer Intelligent Interaction

5.1 Assistive Technologies and Accessibility Solutions

Assistive technologies represent a pivotal application domain in human-computer intelligent interaction, aiming to empower individuals with disabilities by leveraging advanced interaction paradigms to bridge gaps in communication, mobility, and daily life activities. These technologies are increasingly enabled through intelligent systems that incorporate artificial intelligence (AI), natural language processing (NLP), and multimodal interaction frameworks to deliver tailored accessibility solutions.

One significant development is the use of speech-to-text and sign language recognition systems, which allow users with auditory or speech impairments to engage with interactive devices in real-time environmental contexts. Contemporary advancements in NLP and deep learning models have improved the robustness of these systems in diverse usage scenarios [6]. Moreover, sign language recognition now benefits from multimodal sensing techniques, which integrate visual cues and motion tracking to ensure accuracy and inclusivity in dynamic environments [4].

Gesture-controlled devices provide another critical area of innovation, particularly for users with limited mobility. These systems enable seamless control of virtual and physical devices by detecting hand or body movements using computer vision and sensor technologies [15]. Importantly, advancements in this domain emphasize functional flexibility, allowing gestures to be recognized across varied behavioral contexts with minimal calibration requirements. While effective for many users, issues such as motion ambiguity and environmental noise remain technical challenges requiring enhanced noise-robust algorithms for scalable deployment.

The integration of AI in context-aware, sensor-based accessibility tools has revolutionized adaptive assistive solutions, particularly for visually impaired individuals. For instance, sensor-embedded wearables leverage context interpretation—processing environmental data in real time to provide navigation assistance using haptic feedback or auditory signals. The adaptation capabilities in these systems stem from advances in multimodal data fusion and hybrid architectures [4].

AI-driven multimodal systems are also enhancing the functionality of assistive robots and intelligent home systems. Robots equipped with reinforcement learning frameworks can adapt their assistance based on user preferences and feedback over time, ensuring personalized support [17]. This is augmented by AI-enhanced robotic prosthetics, where models integrate biosignals with dexterous actuation to restore fine motor capabilities and facilitate natural interaction.

Despite this progress, challenges remain. Notable limitations include the high computational cost of multimodal recognition systems, scalability issues in diverse real-world environments, and the need for datasets that reflect the full diversity of user conditions and impairments [22]. Furthermore, ethical concerns related to data privacy in sensor-based tools and biases within training datasets underscore the importance of inclusive, participatory design methodologies [46]. Future research should prioritize developing scalable, low-power algorithms and fostering interdisciplinary collaboration to address these challenges.

As the field advances, the emphasis on user-centric design, context-awareness, and multimodal integration is expected to redefine the role of assistive technologies, ensuring greater autonomy and inclusion for individuals with disabilities while addressing critical technical and ethical concerns.

5.2 Intelligent Educational Systems and Personalized Learning Environments

The integration of intelligent technologies into educational systems has profoundly reshaped traditional teaching and learning paradigms, transitioning from one-size-fits-all, teacher-centered approaches to adaptable, learner-driven models. AI-powered tools in education are designed to personalize learning experiences, address diverse learner needs, and foster higher engagement while prioritizing accessibility as a cornerstone. By harnessing advancements in interactive systems, data analytics, and natural language processing, these technologies deliver dynamic, tailored educational experiences that respond to individual and group-level requirements.

Intelligent Tutoring Systems (ITS) embody this transformative approach by facilitating highly personalized and adaptive learning processes. Leveraging sophisticated user modeling techniques, ITS adapt curriculum delivery based on the learner's performance, pace, and evolving preferences. These systems can identify knowledge gaps and offer targeted interventions such as practice exercises, explanatory videos, or scaffolded learning tasks [44]. Reinforcement learning paradigms further empower ITS, enabling systems to refine interventions over time by determining pedagogically effective strategies through iterative trial-and-error adaptations [17]. However, despite the educational potential of such systems, over-reliance on AI risks diminishing the indispensable role of human educators, raising important questions about maintaining the balance between automation and meaningful human oversight [25].

Alongside ITS, educational games and simulations serve as effective tools for enhancing learner engagement and comprehension. By embedding adaptive learning objectives into gamified and immersive environments, these applications harness multimodal feedback—such as real-time visual simulations and auditory cues—to facilitate active learning. Gamified systems successfully promote sustained cognitive and emotional engagement, dynamically adjusting their difficulty to preserve the learner's sense of flow and immersion [20]. Additionally, technologies incorporating affective computing mechanisms analyze and respond to students' emotional states, optimizing learning environments and mitigating frustration [97]. Nevertheless, a central challenge remains in balancing entertainment-driven mechanics with meaningful educational rigor, ensuring that engagement does not compromise the depth or quality of conceptual learning [98].

In collaborative learning settings, AI-enhanced tools assist both instructors and learners by offering real-time insights into classroom dynamics and individual progress. Interactive analytics dashboards enable educators to monitor students' learning trajectories and make customized adjustments to their instructional strategies [99]. For example, these dashboards facilitate the visualization of class-wide comprehension, enabling timely interventions and personalized feedback [44]. However, as the reliance on learning analytics grows, accompanying concerns about data privacy, ownership, and anonymity become more pressing, necessitating rigorous regulatory frameworks to protect student rights [23].

Ensuring accessibility for marginalized groups remains a critical focus area within AI-driven education. Technologies such as deep-learning algorithms designed to simplify complex texts and augmented reality applications tailored for students with cognitive disabilities aim to foster more inclusive learning environments [100]. However, implicit biases embedded in training datasets often limit the inclusivity of these tools, disproportionately affecting learners from underrepresented socioeconomic and cultural backgrounds [23]. Developing systems that deliver equitable experiences across diverse contexts requires greater commitment to representative datasets and participatory design principles [22].

Looking forward, emerging advancements in multimodal AI and explainable AI (XAI) hold great promise for refining educational tools, promoting transparent interactions, and building trust between learners, educators, and technology [54]. The future of intelligent education lies at the nexus of technical innovation and ethical, pedagogical, and cultural considerations, ensuring that all learners benefit equally from the transformative potential of AI-powered educational systems.

5.3 Human-Centered Healthcare and Therapeutic Technologies

Human-computer intelligent interaction (HCII) in healthcare represents a transformative shift in diagnostic accuracy, patient care personalization, and therapeutic interventions. Leveraging artificial intelligence (AI), these systems aim to minimize errors, enhance decision-making, and empower patients and clinicians with adaptive tools. Intelligent interfaces developed for healthcare enhance both the scope and precision of medical services, expanding accessibility, efficiency, and inclusivity.

AI-assisted diagnostic tools have demonstrated significant potential in recognizing patterns often imperceptible to human clinicians, such as in early disease detection. For example, machine learning-driven models incorporate multimodal data sources—text, imagery, and patient histories—to achieve remarkable accuracy in identifying conditions like cancer or cardiovascular disease [81]. Wearable devices integrated with real-time monitoring systems further emphasize this trend, enabling continuous data aggregation from biosensors to quantify health parameters and alert clinicians to anomalies [62].

Interactive systems have also been instrumental in advancing mental health care. Virtual health assistants supported by natural language processing (NLP) provide conversational interactions that mimic therapeutic dialogues, catering to patients coping with anxiety, depression, or cognitive impairments [62]. These systems integrate emotion recognition technologies, which adapt their interventions based on user affect. However, challenges such as user trust and overall system interpretability persist, as reliance on opaque algorithms may erode confidence in these solutions [101]. Addressing these challenges requires the broader adoption of explainability frameworks to ensure system transparency and user engagement [2].

Robotics in therapeutic applications introduces another dimension of HCII. Surgical robots, enhanced by shared control paradigms, enable precise and minimally invasive operations, reducing patient recovery periods. These systems often employ probabilistic models that account for surgeon input while ensuring operational safety [28]. Additionally, intelligent robots assist in physical rehabilitation by personalizing exercises and adjusting levels of support dynamically in real-time [30]. However, shared autonomy frameworks must continue to evolve to guarantee the alignment between human expectations and robotic capabilities [83].

Chronic disease management represents a promising area for HCII innovation, with AI-enhanced smart devices offering predictive modeling to optimize interventions. For instance, diabetes management systems leverage continuous glucose monitoring data to provide personalized recommendations, thereby improving disease control and reducing complications [31]. These systems exemplify the broader movement toward patient-specific healthcare, where adaptive technologies anticipate patient needs and enhance care delivery workflows.

Despite these advancements, HCII in healthcare must navigate critical challenges, such as ethical considerations around data privacy, integration with existing healthcare infrastructure, and mitigating inherent biases in AI-based solutions. Addressing these limitations while adhering to robust human-centric design principles will shape the trajectory of future innovation. Emerging interdisciplinary frameworks combining insights from cognitive science, HCI, and medical informatics may bridge existing gaps, enabling healthcare technologies that are ethically sound, transparent, and universally accessible.

5.4 Collaborative and Service-Oriented Human-Robot Interaction

Collaborative and service-oriented human-robot interaction (HRI) represents a pivotal domain within human-computer intelligent interaction, advancing robots from tools of automation to cooperative agents that dynamically adapt to human intentions and environmental contexts. These systems harness innovations in artificial intelligence, sensor technologies, and multimodal interaction frameworks to facilitate seamless collaboration across diverse industrial, domestic, and social applications, creating new opportunities for human-machine partnerships.

Central to collaborative HRI is the ability of robots to interpret and predict human intent through real-time multimodal perception and adaptive learning. For instance, in industrial environments, human-robot teams increasingly rely on integrated systems employing vision, speech, and gesture recognition to ensure safe, efficient task execution. Advances in intent recognition, such as probabilistic decision-making frameworks and active learning approaches, enhance robotic systems' capacity to infer human preferences even in scenarios marked by ambiguity or incomplete information [34]. This capability underscores the transformative potential for robots to act in alignment with dynamic human goals.

Intuitive communication is another cornerstone of collaborative HRI, with natural language processing (NLP) empowering robots to engage effectively with non-expert users. By combining deep learning models with rule-based algorithms, these systems achieve a balance between flexibility and interpretability, enabling seamless communication through speech and gestures [34], [102]. Such approaches improve usability and foster trust, particularly when paired with explainability paradigms that ensure users comprehend and anticipate robotic behavior [86].

In service-oriented domains, robots are increasingly deployed in unstructured and dynamic environments such as healthcare, disaster response, and public safety. Flexible autonomy paradigms, where control transitions between human operators and robots based on situational demands, empower these systems to navigate complex scenarios [24]. For instance, disaster response robots leverage multimodal sensor data to prioritize safety and optimize resource allocation while maintaining alignment with human objectives through shared mental models [65]. Such adaptability reinforces the role of robots as capable partners in high-pressure environments.

A persistent challenge in this domain centers on human trust calibration. Users often exhibit either over-reliance on or skepticism toward robotic systems, depending on perceived reliability. Studies demonstrate that incorporating behavioral explanations and interactive feedback mechanisms enhances trust by helping users better understand robotic capabilities and limitations, ultimately improving team performance [69], [103]. Addressing these dynamics remains critical to effective collaboration.

Emerging trends emphasize the importance of adaptive and ethically grounded HRI designs that incorporate considerations such as safety, fairness in task distribution, and data privacy. Seamless cross-platform interoperability also presents an ongoing challenge as robots are increasingly integrated into diverse and distributed operational contexts [104], [105].

Looking forward, the future of collaborative and service-oriented HRI lies in hybrid intelligence systems that integrate human oversight with robotic autonomy through real-time feedback loops and incremental learning. By advancing intent recognition, trust calibration, and ethical design principles, these systems are positioned to redefine teamwork paradigms, enabling robots to function as proactive, trustworthy collaborators in complex, human-centered environments.

5.5 Immersive and Creative Interactive Technologies

Immersive and creative interactive technologies are redefining the way users engage with digital environments by offering dynamic, multimodal experiences that integrate artificial intelligence (AI) to enhance creativity, immersion, and interaction. By leveraging advances in augmented reality

(AR), virtual reality (VR), and generative AI, these systems cater to diverse applications, including collaborative creative production, storytelling, and user-centric experiential design.

The integration of AI in AR and VR platforms facilitates personalized and context-aware interactions, enriching the user experience. For example, AI-driven augmented environments dynamically adapt to user behavior and intent, offering tailored content in educational or professional contexts [106]. This adaptivity is critical for maintaining engagement and enabling learning and productivity in immersive spaces. Moreover, in virtual reality, the use of deep neural networks has been pivotal in rendering complex, real-time environments while maintaining computational efficiency [68]. However, achieving seamless synchronization of multimodal inputs—such as visual, auditory, and haptic feedback—remains a technical and computational challenge requiring robust fusion architectures.

Generative AI is another transformative element in creative domains, exemplified by its application in collaborative digital art, music, and storytelling. Systems powered by generative models assist users in co-creating visual or auditory designs by generating novel content based on specific parameters or style prompts [76]. For instance, generative text-to-image frameworks allow artists to iteratively refine their work through user-guided controls, thus enabling intuitive co-creation workflows. These systems, however, raise concerns about intellectual property, shared agency, and over-reliance on automation, highlighting the need for transparent and responsibly designed interfaces [58].

Immersive social systems employing AI, such as conversational agents or multiplayer AR platforms, are also being deployed within therapeutic and well-being applications. These systems simulate social interactions to address mental health needs or promote user engagement in isolated environments [107]. Despite their potential, users often report anthropomorphic cues as manipulative or impersonal, necessitating careful consideration in conversational design and ethical safeguards [108].

Emerging trends include the development of collaborative creative systems, integrating user feedback and AI suggestions interactively. For instance, Deliberative AI systems utilize large language models (LLMs) to enable conversational negotiations between users and AI, fostering co-ownership of outputs in domains like film production or game design [109]. However, challenges persist in calibrating trust and balancing user control against AI autonomy. Studies underline the risk of over-reliance when AI-generated content appears flawless, emphasizing the importance of transparency and explainability [86].

Future research should aim to improve the scalability of multimodal interactions in resource-constrained environments, advance generalist AI models for adaptive content creation, and expand human-AI collaboration frameworks to accommodate diverse user preferences and cultural contexts. Additionally, interdisciplinary efforts are essential to address the ethical implications of immersive and creative technologies, ensuring their deployment is inclusive, accountable, and aligned with user interests. By fostering these developments, immersive and creative technologies have the potential to revolutionize human experiences and redefine the boundaries of human-computer intelligent interaction.

5.6 Intelligent Interaction in Smart Cities and Urban Spaces

Human-computer intelligent interaction technologies are pivotal in transforming the infrastructure and services of smart cities, fostering innovative approaches to urban development that prioritize sustainability, efficiency, and inclusivity. These technologies facilitate dynamic, real-time synergies between citizens, environments, and computational systems, enabling the creation of adaptive and user-centered urban ecosystems. By integrating multimodal artificial intelligence frameworks, IoT networks, and adaptive human-computer interfaces, smart cities can enhance decision-making processes and foster meaningful citizen engagement [94], [110].

One of the cornerstone applications of these interactive technologies in smart cities is AI-driven urban mobility systems. Predictive models and real-time sensor data optimize traffic management, reducing congestion and enhancing the flow of vehicles [111]. These systems also promote equity in transportation through personalized commuting experiences, integrating accessibility features tailored for individuals with disabilities [55]. However, computational constraints and the challenge of achieving equitable scalability in dynamic urban settings require continued innovation and investment.

IoT-enabled infrastructures further extend the capabilities of these technologies by enabling seamless interactions between human actions and urban utilities. From interactive pollution monitoring to smart energy grids and adaptive waste management, these systems employ multimodal sensors embedded across city landscapes to generate responsive urban environments. For instance, energy grids can dynamically adjust power distribution during peak hours, while conversational AI systems can notify citizens about waste disposal schedules [92], [110]. Nevertheless, interoperability across heterogeneous IoT platforms remains a significant technical hurdle requiring cohesive standardization efforts.

Augmented reality (AR) is also playing a transformative role in enhancing urban navigation and citizen engagement. AR maps enriched with real-time contextual data provide location-based services and tailored interactive assistance for individuals with sensory impairments [55]. AR tools further contribute to efficient crowd management, offering real-time visual insights during events or emergencies—a critical asset for disaster response in complex urban scenarios [112].

Additionally, interactive digital twin systems introduce new possibilities in urban planning through highly immersive simulations of city environments. By integrating multimodal data streams, these digital twins provide planners with actionable insights to predict the potential societal impact of infrastructure changes and policy decisions [113], [114]. Yet, limitations concerning the scalability and accuracy of current implementations underscore the importance of advancing robust multimodal learning techniques [115].

While these advancements unlock new opportunities, they also present ethical and practical challenges that necessitate thoughtful resolutions. Privacy concerns surrounding urban data collection, as well as the risks of algorithmic bias in decision-making systems, call for stringent regulatory frameworks and a commitment to ethical AI principles [92]. To ensure inclusivity and security, future research should embrace user-centered design approaches and employ privacy-preserving technologies, such as federated learning, for distributed and secure data processing [116].

Emerging trends in integrating large language and multimodal AI models within urban systems hold transformative potential for evolving smart cities into adaptive, human-centric ecosystems. However, ensuring sustainable, inclusive, and ethical practices remains paramount as these technologies continue to shape the cities of the future. Continuous strides in reliability, scalability, and responsible AI practices will be key to maintaining the human-centered ethos of smart city innovations.

6 Challenges, Evaluation, and Ethical Considerations

6.1 Usability, Accessibility, and Inclusivity Barriers

Designing human-computer intelligent interaction (HCI2) systems that are usable, accessible, and inclusive is an inherently multidimensional challenge. Despite significant advancements, these systems often struggle to accommodate the diverse abilities, preferences, cultural backgrounds, and contextual needs of global users. This subsection illuminates key barriers related to usability, accessibility, and inclusivity, while offering a rigorous analysis of current approaches and directions for overcoming these challenges.

Usability in HCI2 systems remains a critical concern. While machine learning models and advanced interaction mechanisms have ushered in more adaptive and intelligent interfaces, these systems frequently fail to prioritize user intuitiveness, demanding steep learning curves for non-expert users. For instance, the limited naturalness of interfaces based on modalities like voice often constrains their adoption [6]. Moreover, static design techniques fail to adapt to dynamic user contexts, limiting real-time usability, especially in high-stakes or evolving environments [16]. These inadequacies not only reduce adoption but also exacerbate digital inequities among user groups with varying digital literacy levels, a fundamental usability gap.

Accessibility barriers, meanwhile, stem from systems being disproportionately optimized for certain user groups, often reflecting the "WEIRD" (Western, Educated, Industrialized, Rich, and Democratic) user profile [22]. Although accessibility tools such as screen readers and gesture control exist, these features are often retrofitted rather than integrated during initial design, resulting in suboptimal user experiences for individuals with disabilities [4]. Advances in multimodal sensing, such as gaze and pinch interaction [117], promise to address these issues but remain in their infancy, with inconsistent implementation across platforms.

Inclusivity issues are further compounded by design biases embedded within technical systems. Algorithmic behavior that mirrors bias in training datasets perpetuates systemic exclusion, disproportionately affecting minority and marginalized populations [46]. Cultural and linguistic diversity also introduces challenges, with systems often failing to adapt to non-Western languages, dialects, and cultural norms for interaction. As evidenced, naturally interactive systems like conversational agents frequently struggle with multilingual adaptability, limiting their effectiveness in non-native environments [18]. Moreover, insufficient personalization in accommodating neurodiverse users remains a pervasive gap, as most systems fail to tailor interactions based on cognitive styles or sensory preferences [73].

Emerging trends incorporate participatory design approaches, actively involving diverse user groups throughout the design pipeline [46]. Interactive reinforcement learning systems have demonstrated potential in dynamically integrating user feedback to refine personalized and context-aware interfaces [17]. Innovative methodologies like designing hybrid systems, which combine user-centric contextual cues with multimodal computational adaptability, also show promise [118].

Future research must go beyond technical fixes, embedding inclusivity at the core of design processes. Incorporating strategies such as universal design principles alongside real-time multimodal feedback mechanisms can bridge usability gaps across diverse demographics. Moreover, advancing techniques in explainable AI (XAI) could foster transparency, particularly within underserved groups, enhancing trust and engagement [2]. With sustained efforts to eliminate biases, scale computational adaptivity, and align interaction paradigms with diverse user behaviors, human-computer intelligent interaction systems can evolve into truly inclusive and equitable technologies.

6.2 Explainability and Transparency in Intelligent Systems

Explainability and transparency are critical for fostering trust and understanding in intelligent systems, particularly as their decisions increasingly impact high-stakes and sensitive domains such as healthcare, education, and finance. However, achieving these properties is a multifaceted challenge attributable to the inherent complexity of modern AI and machine learning (ML) models. The opaque mechanisms underpinning many intelligent systems, such as deep learning architectures, often hinder users' ability to comprehend how decisions are made. This not only violates established human-computer interaction (HCI) usability principles but also deters system adoption by reducing user confidence and trust [119].

Efforts to enhance explainability typically fall into two categories: intrinsic explainability and post-hoc interpretability. Intrinsic approaches focus on creating models that are inherently transparent, such as decision trees and linear models, which offer interpretable decision-making processes. However, this often comes at the cost of reduced predictive accuracy compared to more complex models like deep neural networks [120]. Post-hoc interpretability, by contrast, applies external methods to elucidate the decisions of black-box systems. Techniques such as feature importance analysis, saliency maps, counterfactual explanations, and surrogate models aim to enhance user understanding without compromising performance [54], [59]. Nonetheless, post-hoc methods have limitations, as their explanations may lack fidelity and might not accurately represent the system's true decision-making process.

Tailoring explanations to meet the needs of diverse user groups introduces another significant challenge. Domain experts and lay users differ markedly in their informational and cognitive requirements. For instance, healthcare professionals may necessitate detailed, mechanistic transparency to validate diagnostic outputs, while patients might benefit more from concise, actionable explanations. Recognizing these differences has led to growing adoption of frameworks such as participatory design and value-sensitive design to actively involve stakeholders during system development. These methodologies align explanation strategies with the unique values, contexts, and expectations of users, promoting inclusivity and relevance in explainability initiatives [2], [46].

Recent studies further underscore the need to measure explainability's influence on user trust and decision-making outcomes through empirical research. Metrics such as trust calibration, task success rates, and user satisfaction have become standard for evaluating the effectiveness of various explanation methods. Findings suggest that tailoring the granularity and complexity of explanations to specific contextual needs can better balance transparency with usability [10]. Moreover, the integration of transparency, consistency, and accountability has been shown to significantly enhance trust levels, as these factors contribute to the reliability and predictability of intelligent systems [104].

Despite advances, achieving meaningful explainability in dynamic, multimodal systems remains an open challenge. Interactive systems, such as conversational AI or multimodal interfaces, pose unique complexities due to their reliance on multiple modalities and temporally interconnected interactions. Techniques like multimodal fusion interpretability and embedded, context-aware transparency show potential for addressing these issues, offering ways to provide explanations intuitively across diverse interaction channels [56].

Looking ahead, integrating explainability into iterative, user-centered prototyping workflows offers a pragmatic pathway to address explainability challenges early in development. For example, participatory design methodologies have demonstrated success in identifying user dissatisfaction and explanation gaps during iterative prototyping stages [121]. Addressing culturally embedded biases in explanation practices is also crucial for ensuring globally inclusive systems that resonate with diverse users [23]. By bridging explainability with user-centered design principles and rigorously validating its impact on trust, intelligent systems can become not only more transparent but also more comprehensible and aligned with ethical imperatives. In doing so, they can fulfill their potential to empower users across diverse contexts and promote equitable, trustworthy adoption on a global scale.

6.3 Evaluation Metrics and Methodologies

Evaluation metrics and methodologies in human-computer intelligent interaction (HCII) must address the multifaceted nature of these systems, encompassing technical performance, user experience, and contextual adaptability. This subsection examines the state-of-the-art approaches to evaluating HCII systems, emphasizing their strengths, limitations, and ongoing challenges.

A widely used category of evaluation metrics focuses on usability and user satisfaction. Quantitative measures such as task success rates, completion times, error rates, and interaction efficiency provide technical insights into system performance [19]. Surveys and subjective feedback analysis complement these measures, capturing perceived usability and satisfaction to reflect user-centric evaluation [14]. However, these methods often fail to capture real-world variability, necessitating a push toward more dynamic and context-sensitive evaluations [122].

Dynamic evaluation considers how HCII systems adapt to changing environments and user behaviors. For example, active experimentation frameworks allow the system to modify interaction strategies based on observed user responses, enhancing robustness while simultaneously measuring adaptability [30]. In human-robot collaboration, frameworks incorporating probabilistic models and game-theoretic approaches have additionally been employed to evaluate trust, shared control dynamics, and decision alignment [28], [35]. These methods provide deeper insights into dynamic user-system interactions but present challenges in computational cost and scalability.

Emerging frameworks advocate for human-centric evaluation models that integrate cognitive and socio-emotional indicators. Trust, empathy, and collaboration efficacy, for instance, have been measured through experimental studies that examine user perceptions of system transparency, human-likeness, and mutual influence in decision-making [31]. Similarly, frameworks like participatory or value-sensitive design emphasize iterative, user-involved evaluation to align system goals with human values [46].

While standardization efforts in evaluation metrics exist, such as benchmarks for cross-modal coherence in multimodal systems [81], there remain critical gaps. Benchmarks often lack applicability across diverse modalities, environments, or user demographics, and many frameworks struggle to account for blended technical and experiential dimensions. To address this, novel approaches, such as hybrid metrics combining neural network-based predictive modeling with user perception alignment, have shown promise [123].

Future research must focus on advancing generalizable frameworks for HCII evaluation. Standardized yet adaptive benchmarks capable of integrating diverse modalities and real-time user feedback are crucial. Furthermore, ethics-minded methodologies—quantifying fairness, inclusivity, and explainability—will play a pivotal role in fostering user trust and acceptance [2]. As HCII systems continue to grow in complexity, adopting holistic, user-informed evaluation approaches will ensure their alignment with human needs, equity, and societal values.

6.4 Ethical Implications of Intelligent Interaction

The ethical implications of intelligent interaction systems are both nuanced and far-reaching, encompassing algorithmic fairness, data privacy, anthropomorphism, and societal impacts. Operating at the intersection of advanced machine learning and human-centric design, these systems hold transformative potential but simultaneously raise critical concerns about their equitable deployment and broader consequences. Ensuring that intelligent interaction systems serve all users fairly, safeguard privacy, and minimize unintended harms is an ongoing challenge that necessitates continuous interdisciplinary scrutiny.

Algorithmic biases present a significant ethical concern within intelligent systems. Rooted in imbalances within training data and system design, these biases often disproportionately affect marginalized groups, exacerbating existing inequalities [10]. While fairness-aware algorithms have advanced, mitigation strategies frequently involve trade-offs, such as sacrificing accuracy or system scalability [124]. Additionally, fairness considerations vary widely across cultural, demographic, and socioeconomic contexts, highlighting the necessity of contextual sensitivity within algorithmic frameworks [46]. Effective mechanisms for evaluating fairness must span the entire lifecycle of intelligent systems, adapting to emergent biases that evolve during real-world, dynamic deployments [42].

Data privacy is another crucial facet of ethical intelligent interaction systems. To deliver personalization and adaptability, these systems rely heavily on extensive user data, exposing significant risks of data misuse, inferential privacy breaches, and exploitation [125]. Emerging privacy solutions, such as federated learning and differential privacy, offer promising pathways to address these challenges; however, their computational demands often render them impractical for certain real-world applications, particularly in resource-constrained multimodal systems [109]. Furthermore, obtaining informed user consent becomes increasingly complex as interaction modalities diversify, complicating users' understanding of what data is collected and how it is utilized [48].

Anthropomorphism introduces unique ethical dilemmas in intelligent systems. By emulating human-like behavior, conversational agents and embodied AI systems can inadvertently mislead or manipulate users, exploiting inherent cognitive biases in human perception [125]. For example, embedding emotional cues to enhance engagement risks fostering distrust when users recognize the deliberate intent to influence their behavior [126].

Beyond individual interactions, the societal impacts of intelligent systems require comprehensive attention. Although these systems promise wider accessibility, enhanced efficiency, and scalability, they risk amplifying digital divides and reinforcing disparities in technological access [11], [66]. Additionally, debates persist regarding the implications of intelligent automation for workforce dynamics, with concerns about the balance between technological augmentation and displacement raising questions of responsibility and equity [127].

Future advancements must prioritize sustainable, ethically grounded innovations that address these challenges holistically. Key directions include developing standardized fairness benchmarks, designing hybrid frameworks that integrate explainability with contextual adaptability, and fostering interdisciplinary research bridging HCI, cognitive science, and AI ethics [2], [85]. By actively engaging diverse stakeholders and conducting rigorous longitudinal evaluations, the field can ensure these systems enhance human potential while adhering to ethical principles. These considerations reinforce the broader themes of scalability, robustness, and adaptability, which are vital to ensuring the reliability and societal alignment of intelligent interactive systems.

6.5 Scalability and Robustness in Dynamic Environments

Scalability and robustness are critical considerations in designing human-computer intelligent interaction (HCII) systems for dynamic environments where user needs and system contexts continuously evolve. Achieving scalability mandates that the system processes and responds efficiently to varying loads while maintaining real-time interactivity, whereas robustness ensures that the system can sustain operational continuity in the face of incomplete or noisy data, environmental shifts, and unforeseen disruptions.

Dynamic environments demand real-time adaptability, requiring systems to process multimodal and context-sensitive inputs under high temporal constraints. Current advancements in scalable architectures leverage distributed systems and federated learning frameworks, which offer decentralized data processing and model updates while preserving scalability [21]. However, maintaining system efficiency in computationally constrained environments remains a challenge, particularly for multimodal systems dealing with high-dimensional data from diverse input streams [56]. Temporal modeling techniques, such as dynamic Bayesian networks and recurrent neural networks, have shown promise in handling evolving user interactions by capturing temporal dependencies, but their computational cost grows significantly with system complexity and scale [42].

Robustness under uncertainty is another pressing challenge in dynamic environments, necessitating HCII systems to function reliably amid incomplete inputs or noisy signals. Error-recovery mechanisms, such as reinforcement learning-based error correction and fault-tolerant design frameworks, have been effective in mitigating system failures [128]. Additionally, human-in-the-loop (HITL) systems provide an adaptive mechanism to refine model performance during unexpected disruptions by combining human expertise with real-time system learning [89]. However, as these systems rely heavily on human feedback, they introduce trust and reliance challenges, wherein user behavior may shift unpredictably [69].

Emerging approaches to scaling HCII system performance include the development of lightweight deep learning models, such as knowledge distillation, for efficient inference while preserving accuracy [40]. Furthermore, robust explainable AI (XAI) techniques enhance system reliability by providing contextual explanations that allow end-users to interpret and rectify misaligned outputs, ultimately improving systems' adaptability to varied scenarios [129]. Rigorous stress-testing protocols, encompassing simulated edge-case scenarios and real-time environmental variabilities, further ensure the scalability and robustness of deployed systems [85].

Future research must integrate bidirectional human-AI alignment, enabling mutual adaptation between systems and users to navigate dynamic uncertainty with greater efficacy [130]. Additionally, advancements in hybrid architectures combining neural-symbolic AI with rule-based systems could offer enhanced robustness to reasoning under diverse scenarios. Finally, cross-disciplinary evaluations incorporating human, computational, and contextual dimensions will be vital to extend the scalability and reliability of HCII systems in increasingly unpredictable use cases.

7 Conclusion

The survey presented an extensive exploration of the diverse dimensions of human-computer intelligent interaction (HCII), emphasizing its evolution, foundational principles, enabling technologies, applications, and challenges inherent in its implementation. The synthesis of this body of work underscores the transformative potential of HCII in fostering more cohesive, adaptive, and human-centered systems. Particularly, the integration of artificial intelligence (AI) into interactive systems has catalyzed the development of interfaces capable of dynamic learning, contextual awareness, and real-time personalization, representing a critical shift from traditional human-computer paradigms [7], [9].

The foundational analysis of multimodal interaction highlights its role in creating seamless user experiences by synthesizing diverse input modalities, such as speech, gestures, and gaze, into coherent interaction frameworks. Advanced computational architectures, including multimodal transformers and cross-modal learning methods, have further enriched this space but remain constrained by challenges like scalability, resource intensity, and alignment difficulties across diverse modalities [4], [6]. Additionally, theoretical models such as shared human-machine autonomy have helped redefine interaction dynamics, enhancing both trust and usability in hybrid systems [11], [127].

The analysis also underscores significant challenges, particularly around ethical concerns, transparency, and inclusivity—an area where biases in AI models and systemic inequities in design exacerbate marginalization. Researchers have called for integrating user-centered approaches that prioritize interpretability, such as explainable AI (XAI), while fostering inclusivity in diverse sociocultural and linguistic contexts [8], [22]. Trust in intelligent systems, pivotal for user acceptance, further highlights the importance of explainability matched with mechanisms that balance human oversight and system autonomy [10], [25].

Key technological advancements, such as brain-computer interfaces and extended reality (XR) systems, are paving innovative pathways for deeper human-computer collaboration. For instance, XR-AI frameworks and embodied sensor-based technologies hold promise in merging physical and virtual interactions, offering new tools to evaluate interaction fidelity and augment accessibility [74], [131]. However, significant questions remain regarding adaptability in real-world applications, especially under resource constraints or complex environments [64], [132].

Looking forward, the field must focus on reducing cognitive load through proactive, anticipatory systems that leverage AI to predict user needs without compromising transparency or ethical considerations. As systems transition from reactive designs to shared cognitive frameworks, adopting participatory design methodologies and hybrid socio-technical approaches will prove instrumental. The trajectory of HCII research hence lies in merging adaptive interaction technologies with a robust ethical and user-centric foundation, ensuring that these systems remain transparent, inclusive, and accessible across diverse domains and demographics [2], [48].

In summary, human-computer intelligent interaction stands as a cornerstone of advancing human-centric AI technologies. This synthesis confirms the progress made while emphasizing critical pathways to bridge existing conceptual and technological gaps. By addressing ethical challenges, investing in multimodal adaptability, and refining user-centered evaluation frameworks, HCII can aspire to create intelligent systems that not only adapt to but deeply enrich the human experience.

References

[1] M. Kaushik and R. Jain, "Natural user interfaces: Trend in virtual interaction," *ArXiv*, vol. abs/1405.0101, 2014. 1

[2] U. Ehsan and M. O. Riedl, "Human-centered explainable ai: Towards a reflective sociotechnical approach," *ArXiv*, vol. abs/2002.01092, 2020. 1, 2, 3, 9, 13, 15, 16, 17, 18

[3] H. R. Chennamma and X. Yuan, "A survey on eye-gaze tracking techniques," *ArXiv*, vol. abs/1312.6410, 2013. 1, 2, 4

[4] A. Carfí and F. Mastrogiovanni, "Gesture-based human–machine interaction: Taxonomy, problem definition, and analysis," *IEEE Transactions on Cybernetics*, vol. 53, pp. 497–513, 2021. 1, 4, 11, 15, 18

[5] I. Mackenzie, J. Read, and M. Horton, "Empirical research methods for human-computer interaction," *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023. 1

[6] L. Clark, P. R. Doyle, D. Garaialde, E. Gilmartin, S. Schlögl, J. Edlund, M. Aylett, J. P. Cabral, C. Munteanu, and B. R. Cowan, "The state of speech in hci: Trends, themes and challenges," *Interact. Comput.*, vol. 31, pp. 349–371, 2018. 1, 2, 6, 11, 15, 18

[7] J. Gao, S. A. Gebreegziabher, K. T. W. Choo, T. Li, S. Perrault, and T. W. Malone, "A taxonomy for human-llm interaction modes: An initial exploration," *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 2024. 1, 8, 18

[8] S. Mohseni, N. Zarei, and E. Ragan, "A multidisciplinary survey and framework for design and evaluation of explainable ai systems," *ACM Trans. Interact. Intell. Syst.*, vol. 11, pp. 24:1–24:45, 2018. 1, 4, 18

[9] M. Raees, I. Meijerink, I. Lykourentzou, V.-J. Khan, and K. Papangelis, "From explainable to interactive ai: A literature review on current trends in human-ai interaction," *ArXiv*, vol. abs/2405.15051, 2024. 1, 4, 18

[10] T. A. Bach, A. Khan, H. P. Hallock, G. Beltrao, and S. Sousa, "A systematic literature review of user trust in ai-enabled systems: An hci perspective," *International Journal of Human–Computer Interaction*, vol. 40, pp. 1251 – 1266, 2022. 1, 3, 4, 6, 7, 16, 17, 18

[11] D. Dellermann, A. Calma, N. Lipusch, T. Weber, S. Weigel, and P. Ebel, "The future of human-ai collaboration: A taxonomy of design knowledge for hybrid intelligence systems," in *Hawaii International Conference on System Sciences*, 2019, pp. 1–10. 1, 7, 17, 18

[12] W. Xu and Z. Gao, "Applying hcai in developing effective human-ai teaming: A perspective from human-ai joint cognitive systems," *Interactions*, vol. 31, pp. 32 – 37, 2023. 1

[13] S. Thuseethan and S. Kuhanesan, "Effective use of human computer interaction in digital academic supportive devices," *ArXiv*, vol. abs/1501.00529, 2014. 2

[14] Y. Lu, Y. Yang, Q. Zhao, C. Zhang, and T. J.-J. Li, "Ai assistance for ux: A literature review through human-centered ai," *ArXiv*, vol. abs/2402.06089, 2024. 2, 16

[15] K. Shah, K. R. Rathod, and S. J. Agravat, "A survey on human computer interaction mechanism using finger tracking," *ArXiv*, vol. abs/1402.0693, 2014. 2, 11

[16] X. B. Liu, J. N. Li, D. Kim, X. A. Chen, and R. Du, "Human i/o: Towards a unified approach to detecting situational impairments," *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2024. 2, 11, 15

[17] C. A. Cruz and T. Igarashi, "A survey on interactive reinforcement learning: Design principles and open challenges," *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, 2020. 2, 4, 5, 7, 9, 11, 12, 15

[18] P. R. Doyle, J. Edwards, O. Dumbleton, L. Clark, and B. R. Cowan, "Mapping perceptions of humanness in intelligent personal assistant interaction," *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*, 2019. 2, 15

[19] R. Michalski, J. Grobelny, and W. Karwowski, "The effects of graphical interface design characteristics on human-computer interaction task efficiency," *ArXiv*, vol. abs/1211.6712, 2006. 2, 16

[20] J. Frey, C. Mühl, F. Lotte, and M. Hachet, "Review of the use of electroencephalography as an evaluation method for human-computer interaction," in *International Conference on Physiological Computing Systems*, 2013, pp. 214–223. 2, 12

[21] L. Ponciano, F. Brasileiro, N. Andrade, and L. M. R. Sampaio, "Considering human aspects on strategies for designing and managing distributed human computation," *Journal of Internet Services and Applications*, vol. 5, 2014. 2, 7, 17

[22] K. Seaborn, G. Barbareschi, and S. Chandra, "Not only weird but "uncanny"? a systematic review of diversity in human–robot interaction research," *International Journal of Social Robotics*, pp. 1 – 30, 2023. 2, 7, 8, 11, 12, 15, 18

[23] U. Peters and M. Carman, "Cultural bias in explainable ai research: A systematic analysis," *ArXiv*, vol. abs/2403.05579, 2024. 2, 8, 12, 16

[24] A. Hussein and H. Abbass, "Mixed initiative systems for human-swarm interaction: Opportunities and challenges," *2018 2nd Annual Systems Modelling Conference (SMC)*, pp. 1–8, 2018. 2, 6, 13

[25] L. C. Siebert, M. Lupetti, E. Aizenberg, N. Beckers, A. Zgonnikov, H. Veluwenkamp, D. A. Abbink, E. Giaccardi, G. Houben, C. Jonker, J. van den Hoven, D. Forster, and R. Lagendijk, "Meaningful human control: actionable properties for ai system development," *AI and Ethics*, vol. 3, pp. 241–255, 2021. 2, 12, 18

[26] Q. Zhu, L. Chong, M. C. Yang, and J. Luo, "Reading users' minds from what they say: An investigation into llm-based empathic mental inference," *ArXiv*, vol. abs/2403.13301, 2024. 3, 8

[27] M. Axelsson, M. Spitale, and H. Gunes, ""oh, sorry, i think i interrupted you": Designing repair strategies for robotic longitudinal well-being coaching," *2024 19th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp. 13–22, 2024. 3

[28] S. Nikolaidis, J. Forlizzi, D. Hsu, J. Shah, and S. Srinivasa, "Mathematical models of adaptation in human-robot collaboration," *ArXiv*, vol. abs/1707.02586, 2017. 3, 9, 13, 16

[29] A. Broad, T. Murphey, and B. Argall, "Learning models for shared control of human-machine systems with unknown dynamics," *ArXiv*, vol. abs/1808.08268, 2017. 3, 5, 6, 9

[30] R. Pandya and C. Liu, "Safe and efficient exploration of human models during human-robot interaction," *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6708–6715, 2022. 3, 5, 9, 13, 16

[31] K. Akash, G. McMahon, T. Reid, and N. Jain, "Human trust-based feedback control: Dynamically varying automation transparency to optimize human-machine interactions," *IEEE Control Systems*, vol. 40, pp. 98–116, 2020. 3, 6, 9, 13, 16

[32] J. Zhou, R. Li, J. Tang, T. Tang, H. Li, W. Cui, and Y. Wu, "Understanding nonlinear collaboration between human and ai agents: A co-design framework for creative design," *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2024. 3

[33] M. J. Guzdial and M. O. Riedl, "An interaction framework for studying co-creative ai," *ArXiv*, vol. abs/1903.09709, 2019. 3

[34] N. Wilde, A. Blidaru, S. L. Smith, and D. Kulić, "Improving user specifications for robot behavior through active preference learning: Framework and evaluation," *The International Journal of Robotics Research*, vol. 39, pp. 651 – 667, 2019. 3, 4, 9, 13

[35] K. Darvish, E. Simetti, F. Mastrogiovanni, and G. Casalino, "A hierarchical architecture for human–robot cooperation processes," *IEEE Transactions on Robotics*, vol. 37, pp. 567–586, 2020. 3, 9, 16

[36] M. T. Llano, M. d'Inverno, M. Yee-King, J. Mccormack, A. Ilsar, A. Pease, and S. Colton, "Explainable computational creativity," *ArXiv*, vol. abs/2205.05682, 2022. 3

[37] A. Fuchs, A. Passarella, and M. Conti, "A cognitive framework for delegation between error-prone ai and human agents," *2022 IEEE International Conference on Smart Computing (SMARTCOMP)*, pp. 317–322, 2022. 3, 6

[38] K. M. Collins, I. Sucholutsky, U. Bhatt, K. Chandra, L. Wong, M. Lee, C. Zhang, Z.-X. Tan, M. Ho, V. K. Mansinghka, A. Weller, J. B. Tenenbaum, and T. L. Griffiths, "Building machines that learn and think with people," *Nature human behaviour*, vol. 8 10, pp. 1851–1863, 2024. 3

[39] H. Liu, V. Lai, and C. Tan, "Understanding the effect of out-of-distribution examples and interactive explanations on human-ai decision making," *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, pp. 1 – 45, 2021. 3

[40] Y. Rong, T. Leemann, T. trang Nguyen, L. Fiedler, P. Qian, V. Unhelkar, T. Seidel, G. Kasneci, and E. Kasneci, "Towards human-centered explainable ai: A survey of user studies for model explanations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, pp. 2104–2122, 2022. 3, 10, 17

[41] C. Rastogi, L. Leqi, K. Holstein, and H. Heidari, "A taxonomy of human and ml strengths in decision-making to investigate human-ml complementarity," *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 2022. 3

[42] V. Lai, C. Chen, Q. Liao, A. Smith-Renner, and C. Tan, "Towards a science of human-ai decision making: A survey of empirical studies," *ArXiv*, vol. abs/2112.11471, 2021. 3, 4, 17

[43] A. Holzinger, M. Plass, K. Holzinger, G. Crişan, C. Pintea, and V. Palade, "A glass-box interactive machine learning approach for solving np-hard problems with the human-in-the-loop," *ArXiv*, vol. abs/1708.01104, 2017. 4, 9

[44] N. Boukhelifa, A. Bezerianos, and E. Lutton, "Evaluation of interactive machine learning systems," in *Human and Machine Learning*, 2018, pp. 341–360. 4, 9, 12

[45] S. Zhang, J. Yu, X. Xu, C. Yin, Y. Lu, B. Yao, M. Tory, L. M. Padilla, J. Caterino, P. Zhang, and D. Wang, "Rethinking human-ai collaboration in complex medical decision making: A case study in sepsis diagnosis," *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2023. 4

[46] F. Delgado, S. Yang, M. A. Madaio, and Q. Yang, "Stakeholder participation in ai: Beyond "add diverse stakeholders and stir"," *ArXiv*, vol. abs/2111.01122, 2021. 4, 6, 12, 15, 16, 17

[47] Q. Liao, H. Subramonyam, J. Wang, and J. W. Vaughan, "Designerly understanding: Information needs for model transparency to support design ideation for ai-powered user experience," *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023. 4

[48] T. T. Khuat, D. Kedziora, and B. Gabrys, "The roles and modes of human interactions with automated machine learning systems," *ArXiv*, vol. abs/2205.04139, 2022. 4, 17, 18

[49] F. Lehmann and D. Buschek, "Examining autocompletion as a basic concept for interaction with generative ai," *i-com*, vol. 19, pp. 251 – 264, 2020. 4

[50] Z. Xiao, T. Wang, J. Wang, J. Cao, W. Zhang, B. Dai, D. Lin, and J. Pang, "Unified human-scene interaction via prompted chain-of-contacts," *ArXiv*, vol. abs/2309.07918, 2023. 4

[51] Y.-L. Li, S. Zhou, X. Huang, L. Xu, Z. Ma, H. Fang, Y. Wang, and C. Lu, "Transferable interactiveness knowledge for human-object interaction detection," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3580–3589, 2018. 4

[52] A. Sellen and E. Horvitz, "The rise of the ai co-pilot: Lessons for design from aviation and beyond," *Communications of the ACM*, vol. 67, pp. 18 – 23, 2023. 4

[53] J. Shi, R. Jain, H. Doh, R. Suzuki, and K. Ramani, "An hci-centric survey and taxonomy of human-generative-ai interactions," *ArXiv*, vol. abs/2310.07127, 2023. 5, 8

[54] Q. Liao, M. Research, C. Kush, R. Varshney, and K. R. Varshney, "Human-centered explainable ai (xai): From algorithms to user experiences," *ArXiv*, vol. abs/2110.10790, 2021. 5, 12, 16

[55] J. Xie, R. Yu, H. Zhang, S. Lee, S. M. Billah, and J. M. Carroll, "Emerging practices for large multimodal model (lmm) assistance for people with visual impairments: Implications for design," *ArXiv*, vol. abs/2407.08882, 2024. 5, 14

[56] A. Peña, I. Serna, A. Morales, J. Fierrez, A. Ortega, A. Herrarte, M. Alcántara, and J. Ortega-Garcia, "Human-centric multimodal machine learning: Recent advances and testbed on ai-based recruitment," *SN Computer Science*, vol. 4, 2023. 5, 7, 10, 16, 17

[57] C. Zhang, W. Lu, C. Ni, H. Wang, and J. Wu, "Enhanced user interaction in operating systems through machine learning language models," *ArXiv*, vol. abs/2403.00806, 2024. 5

[58] Y. Nakao, L. Strappelli, S. Stumpf, A. Naseer, D. Regoli, and G. D. Gamba, "Towards responsible ai: A design space exploration of human-centered artificial intelligence user interfaces to investigate fairness," *International Journal of Human–Computer Interaction*, vol. 39, pp. 1762 – 1788, 2022. 5, 7, 14

[59] H. Chen, C. Gómez, C.-M. Huang, and M. Unberath, "Explainable medical imaging ai needs human-centered design: guidelines and evidence from a systematic review," *NPJ Digital Medicine*, vol. 5, 2021. 5, 8, 16

[60] X. Huang, Z. Zhang, F. Guo, X. Wang, K. Chi, and K. Wu, "Research on older adults' interaction with e-health interface based on explainable artificial intelligence," *ArXiv*, vol. abs/2402.07915, 2024. 5, 8

[61] R. Kaufman, J. Costa, and E. Kimani, "Effects of multimodal explanations for autonomous driving on driving performance, cognitive load, expertise, confidence, and trust," *Scientific Reports*, vol. 14, 2024. 5

[62] F. de Arriba Pérez, S. García-Méndez, F. González-Castaño, and E. Costa-Montenegro, "Automatic detection of cognitive impairment in elderly people using an entertainment chatbot with natural language processing capabilities," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1 – 16, 2022. 6, 12

[63] Z. Durante, Q. Huang, N. Wake, R. Gong, J. Park, B. Sarkar, R. Taori, Y. Noda, D. Terzopoulos, Y. Choi, K. Ikeuchi, H. Vo, F.-F. Li, and J. Gao, "Agent ai: Surveying the horizons of multimodal interaction," *ArXiv*, vol. abs/2401.03568, 2024. 6, 9

[64] A. Madison, E. R. Novoseller, V. G. Goecks, B. T. Files, N. R. Waytowich, A. Yu, V. J. Lawhern, S. Thurman, C. Kelshaw, and K. McDowell, "Scalable interactive machine learning for future command and control," *2024 International Conference on Military Communication and Information Systems (ICMCIS)*, pp. 1–10, 2024. 6, 9, 18

[65] T. K. M. Vercauteren, M. Unberath, N. Padoy, and N. Navab, "Cai4cai: The rise of contextual artificial intelligence in computer-assisted interventions," *Proceedings of the Ieee. Institute of Electrical and Electronics Engineers*, vol. 108, pp. 198 – 214, 2019. 6, 13

[66] S. Zhu, T. Yu, T. Xu, H. Chen, S. Dustdar, S. Gigan, D. Gunduz, E. Hossain, Y. Jin, F. Lin, B. Liu, Z. Wan, J. Zhang, Z. Zhao, W. Zhu, Z. Chen, T. Durrani, H. Wang, J. Wu, T. Zhang, and Y. Pan, "Intelligent computing: The latest advances, challenges and future," *ArXiv*, vol. abs/2211.11281, 2022. 7, 17

[67] E. Schlicht, R. Lee, D. Wolpert, M. J. Kochenderfer, and B. D. Tracey, "Predicting the behavior of interacting humans by fusing data from multiple sources," *ArXiv*, vol. abs/1408.2053, 2012. 7

[68] L. Mathur, P. Liang, and L. philippe Morency, "Advancing social intelligence in ai agents: Technical challenges and open questions," *ArXiv*, vol. abs/2404.11023, 2024. 7, 14

[69] D. R. Honeycutt, M. Nourani, and E. Ragan, "Soliciting human-in-the-loop user feedback for interactive machine learning reduces user trust and impressions of model accuracy," *ArXiv*, vol. abs/2008.12735, 2020. 7, 13, 17

[70] A. Alkhatlan and J. Kalita, "Intelligent tutoring systems: A comprehensive historical survey with recent developments," *ArXiv*, vol. abs/1812.09628, 2018. 7

[71] C. Gao, W. Lei, X. He, M. de Rijke, and T. seng Chua, "Advances and challenges in conversational recommender systems: A survey," *AI Open*, vol. 2, pp. 100–126, 2021. 7

[72] Q. Motger, X. Franch, and J. Marco, "Software-based dialogue systems: Survey, taxonomy, and challenges," *ACM Computing Surveys*, vol. 55, pp. 1 – 42, 2021. 7

[73] F. Cabitza and A. Locoro, "Human-data interaction in healthcare," *Data Analytics in Medicine*, 2016. 7, 15

[74] A. Dahiya, A. M. Aroyo, K. Dautenhahn, and S. L. Smith, "A survey of multi-agent human-robot interaction systems," *Robotics Auton. Syst.*, vol. 161, p. 104335, 2022. 8, 18

[75] Q. Zheng, Y. Tang, Y. Liu, W. Liu, and Y. Huang, "Ux research on conversational human-ai interaction: A literature review of the acm digital library," *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022. 8

[76] J. D. Weisz, M. J. Muller, J. He, and S. Houde, "Toward general design principles for generative ai applications 130-144," in *IUI Workshops*, 2023, pp. 130–144. 8, 14

[77] A. Fadhil and G. Schiavo, "Designing for health chatbots," *ArXiv*, vol. abs/1902.09022, 2019. 8

[78] Y. Deng, L. Liao, Z. Zheng, G. H. Yang, and T.-S. Chua, "Towards human-centered proactive conversational agents," in *Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2024, pp. 807–818. 8

[79] N. Rajcic and J. McCormack, "Message ritual: A posthuman account of living with lamp," *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023. 8

[80] F. Semeraro, A. Griffiths, and A. Cangelosi, "Human-robot collaboration and machine learning: A systematic review of recent research," *Robotics Comput. Integr. Manuf.*, vol. 79, p. 102432, 2021. 8

[81] Y. Li, S. Bengio, and G. Bailly, "Predicting human performance in vertical menu selection using deep learning," *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 2018. 8, 12, 16

[82] T. L. Griffiths, J.-Q. Zhu, E. Grant, and R. T. McCoy, "Bayes in the age of intelligent machines," *ArXiv*, vol. abs/2311.10206, 2023. 9

[83] S. Matsumoto and L. Riek, "Shared control in human robot teaming: Toward context-aware communication," *ArXiv*, vol. abs/2203.10218, 2022. 9, 13

[84] M. C. Fontaine and S. Nikolaidis, "A quality diversity approach to automatically generating human-robot interaction scenarios in shared autonomy," *ArXiv*, vol. abs/2012.04283, 2020. 9

[85] S. Sreedharan, A. Kulkarni, and S. Kambhampati, "Explainable human–ai interaction: A planning perspective," *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 2022. 9, 10, 17

[86] K. Morrison, P. Spitzer, V. Turri, M. C. Feng, N. Kuhl, and A. Perer, "The impact of imperfect xai on human-ai decision-making," *Proceedings of the ACM on Human-Computer Interaction*, vol. 8, pp. 1 – 39, 2023. 9, 13, 14

[87] L. Ibrahim, S. Huang, L. Ahmad, and M. Anderljung, "Beyond static ai evaluations: advancing human interaction evaluations for llm harms and risks," *ArXiv*, vol. abs/2405.10632, 2024. 9

[88] C. P. Lee, "Design, development, and deployment of context-adaptive ai systems for enhanced user adoption," *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 2024. 9

[89] F. M. Zanzotto, "Human-in-the-loop artificial intelligence," *ArXiv*, vol. abs/1710.08191, 2017. 10, 17

[90] Y. Gao, T. Sun, L. Zhao, and S. R. Hong, "Aligning eyes between humans and deep neural network through interactive attention alignment," *Proceedings of the ACM on Human-Computer Interaction*, vol. 6, pp. 1 – 28, 2022. 10

[91] N. V. Volvoikar, "A survey on human action recognition," *International Journal for Research in Applied Science and Engineering Technology*, 2019. 10

[92] N. Rodis, C. Sardianos, G. Papadopoulos, P. I. Radoglou-Grammatikis, P. Sarigiannidis, and I. Varlamis, "Multimodal explainable artificial intelligence: A comprehensive review of methodological advances and future research directions," *ArXiv*, vol. abs/2306.05731, 2023. 10, 11, 14, 15

[93] M. J. Lyons, "Facial gesture interfaces for expression and communication," *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, vol. 1, pp. 598–603 vol.1, 2004. 11

[94] A. Gomaa, "Adaptive user-centered multimodal interaction towards reliable and trusted automotive interfaces," *Proceedings of the 2022 International Conference on Multimodal Interaction*, 2022. 11, 14

[95] Y. Wang, Y.-S. Lin, R. Huang, J. Wang, and S. Liu, "Enhancing user experience in large language models through human-centered design: Integrating theoretical insights with an experimental study to meet diverse software learning needs with a single document knowledge base," *ArXiv*, vol. abs/2405.11505, 2024. 11

[96] S. N. Gilani, D. Traum, A. Merla, E. Hee, Z. Walker, B. Manini, G. Gallagher, and L. Petitto, "Multimodal dialogue management for multiparty interaction with infants," *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, 2018. 11

[97] S. Benford, A. Løvlie, K. Ryding, P. Rajkowska, E. Bodiaj, D. P. Darzentas, H. R. Cameron, J. Spence, J. O. Egede, and B. Spanjevic, "Sensitive pictures: Emotional interpretation in the mu-

seum," *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022. 12

[98] K. Siu, A. Zook, and M. O. Riedl, "A framework for exploring and evaluating mechanics in human computation games," *Proceedings of the 12th International Conference on the Foundations of Digital Games*, 2017. 12

[99] R. Alfredo, V. Echeverría, Y. Jin, L. Yan, Z. Swiecki, D. Gašević, and R. M. Maldonado, "Human-centred learning analytics and ai in education: a systematic literature review," *Comput. Educ. Artif. Intell.*, vol. 6, p. 100215, 2023. 12

[100] B. Shneiderman, "Human-centered artificial intelligence: Reliable, safe & trustworthy," *International Journal of Human–Computer Interaction*, vol. 36, pp. 495 – 504, 2020. 12

[101] Z. R. Khavas, R. Ahmadzadeh, and P. Robinette, "Modeling trust in human-robot interaction: A survey," *ArXiv*, vol. abs/2011.04796, 2020. 13

[102] Z. Ding and J. Chan, "Mapping the design space of interactions in human-ai text co-creation tasks," *ArXiv*, vol. abs/2303.06430, 2023. 13

[103] Ángel Alexander Cabrera, A. Perer, and J. I. Hong, "Improving human-ai collaboration with descriptions of ai behavior," *Proceedings of the ACM on Human-Computer Interaction*, vol. 7, pp. 1 – 21, 2023. 13

[104] C. J. Smith, "Designing trustworthy ai: A human-machine teaming framework to guide development," *ArXiv*, vol. abs/1910.03515, 2019. 13, 16

[105] L. Ibrahim, L. Rocher, and A. Valdivia, "Characterizing and modeling harms from interactions with design patterns in ai interfaces," *ArXiv*, vol. abs/2404.11370, 2024. 13

[106] M. Chen, Y. Jiang, Y. Cao, and A. Y. Zomaya, "Creativebioman: A brain- and body-wearable, computing-based, creative gaming system," *IEEE Systems, Man, and Cybernetics Magazine*, vol. 6, pp. 14–22, 2019. 14

[107] T. Mildner, O. Cooney, A.-M. Meck, M. Bartl, G.-L. Savino, P. R. Doyle, D. Garaialde, L. Clark, J. Sloan, N. Wenig, R. Malaka, and J. Niess, "Listening to the voices: Describing ethical caveats of conversational user interfaces according to experts and frequent users," *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 2024. 14

[108] L. Alberts, U. Lyngs, and M. V. Kleek, "Computers as bad social actors: Dark patterns and anti-patterns in interfaces that act socially," *Proceedings of the ACM on Human-Computer Interaction*, vol. 8, pp. 1 – 25, 2023. 14

[109] S. Ma, Q. Chen, X. Wang, C. Zheng, Z. Peng, M. Yin, and X. Ma, "Towards human-ai deliberation: Design and evaluation of llm-empowered deliberative ai for ai-assisted decision-making," *ArXiv*, vol. abs/2403.16812, 2024. 14, 17

[110] L. Soenksen, Y. Ma, C. Zeng, L. Boussioux, K. V. Carballo, L. Na, H. Wiberg, M. L. Li, I. Fuentes, and D. Bertsimas, "Integrated multimodal artificial intelligence framework for healthcare applications," *NPJ Digital Medicine*, vol. 5, 2022. 14

[111] P. Jansen, M. Colley, and E. Rukzio, "A design space for human sensor and actuator focused in-vehicle interaction based on a systematic literature review," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, pp. 1 – 51, 2022. 14

[112] J. Cacace, A. Finzi, and V. Lippiello, "Multimodal interaction with multiple co-located drones in search and rescue missions," *ArXiv*, vol. abs/1605.07316, 2016. 14

[113] B. Zhuang, L. Liu, C. Shen, and I. Reid, "Towards context-aware interaction recognition for visual relationship detection," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 589–598, 2017. 14

[114] A. Srivastava, A. Rastogi, A. Rao, A. A. M. Shoeb, A. Abid, A. Fisch, A. R. Brown, A. Santoro, A. Gupta, A. Garriga-Alonso, A. Kluska, A. Lewkowycz, A. Agarwal, A. Power, A. Ray, A. Warstadt, A. W. Kocurek, A. Safaya, A. Tazarv, A. Xiang, A. Parrish, A. Nie, A. Hussain, A. Askell, A. Dsouza, A. Slone, A. Rahane, A. S. Iyer, A. Andreassen, A. Madotto, A. Santilli, A. Stuhlmuller, A. M. Dai, A. La, A. K. Lampinen, A. Zou, A. Jiang, A. Chen, A. Vuong, A. Gupta, A. Gottardi, A. Norelli, A. Venkatesh, A. Gholamidavoodi, A. Tabassum, A. Menezes, A. Kirubarajan, A. Mullokandov, A. Sabharwal, A. Herrick, A. Efrat, A. Erdem, A. Karakacs, B. R. Roberts, B. S. Loe, B. Zoph, B. Bojanowski, B. Ozyurt, B. Hedayatnia, B. Neyshabur, B. Inden, B. Stein, B. Ekmekci, B. Y. Lin, B. Howald, B. Orinion, C. Diao, C. Dour, C. Stinson, C. Argueta, C. F. Ram'irez, C. Singh, C. Rathkopf, C. Meng, C. Baral, C. Wu, C. Callison-Burch, C. Waites, C. Voigt, C. D. Manning, C. Potts, C. Ramirez, C. E. Rivera, C. Siro, C. Raffel, C. Ashcraft, C. Garbacea, D. Sileo, D. H. Garrette, D. Hendrycks, D. Kilman, D. Roth, D. Freeman, D. Khashabi, D. Levy, D. Gonz'alez, D. R. Perszyk, D. Hernandez, D. Chen, D. Ippolito, D. Gilboa, D. Dohan, D. Drakard, D. Jurgens, D. Datta, D. Ganguli, D. Emelin, D. Kleyko, D. Yuret, D. Chen, D. Tam, D. Hupkes, D. Misra, D. Buzan, D. C. Mollo, D. Yang, D.-H. Lee, D. Schrader, E. Shutova, E. D. Cubuk, E. Segal, E. Hagerman, E. Barnes, E. Donoway, E. Pavlick, E. Rodolà, E. Lam, E. Chu, E. Tang, E. Erdem, E. Chang, E. A. Chi, E. Dyer, E. Jerzak, E. Kim, E. E. Manyasi, E. Zheltonozhskii, F. Xia, F. Siar, F. Mart'inez-Plumed, F. Happ'e, F. Chollet, F. Rong, G. Mishra, G. I. Winata, G. de Melo, G. Kruszewski, G. Parascandolo, G. Mariani, G. X. Wang, G. Jaimovitch-L'opez, G. Betz, G. Gur-Ari, H. Galijasevic, H. Kim, H. Rashkin, H. Hajishirzi, H. Mehta, H. Bogar, H. Shevlin, H. Schutze, H. Yakura, H. Zhang, H. M. Wong, I. Ng, I. Noble, J. Jumelet, J. Geissinger, J. Kernion, J. Hilton, J. Lee, J. Fisac, J. B. Simon, J. Koppel, J. Zheng, J. Zou, J. Koco'n, J. Thompson, J. Wingfield, J. Kaplan, J. Radom, J. N. Sohl-Dickstein, J. Phang, J. Wei, J. Yosinski, J. Novikova, J. Bosscher, J. Marsh, J. Kim, J. Taal, J. Engel, J. O. Alabi, J. Xu, J. Song, J. Tang, J. W. Waweru, J. Burden, J. Miller, J. U. Balis, J. Batchelder, J. Berant, J. Frohberg, J. Rozen, J. Hernández-Orallo, J. Boudeman, J. Guerr, J. Jones, J. B. Tenenbaum, J. S. Rule, J. Chua, K. Kanclerz, K. Livescu, K. Krauth, K. Gopalakrishnan, K. Ignatyeva, K. Markert, K. D. Dhole, K. Gimpel, K. Omondi, K. Mathewson, K. Chiafullo, K. Shkaruta, K. Shridhar, K. McDonell, K. Richardson, L. Reynolds, L. Gao, L. Zhang, L. Dugan, L. Qin, L. Contreras-Ochando, L.-P. Morency, L. Moschella, L. Lam, L. Noble, L. Schmidt, L. He, L. O. Col'on, L. Metz, L. K. cSenel, M. Bosma, M. Sap, M. ter Hoeve, M. Farooqi, M. Faruqui, M. Mazeika, M. Baturan, M. Marelli, M. Maru, M. J. R. Quintana, M. Tolkiehn, M. Giulianelli, M. Lewis, M. Potthast, M. L. Leavitt, M. Hagen, M. Schubert, M. Baitemirova, M. Arnaud, M. McElrath, M. A. Yee, M. Cohen, M. Gu, M. Ivanitskiy, M. Starritt, M. Strube, M. Swkedrowski, M. Bevilacqua, M. Yasunaga, M. Kale, M. Cain, M. Xu, M. Suzgun, M. Walker, M. Tiwari, M. Bansal, M. Aminnaseri, M. Geva, M. Gheini, T. MukundVarma, N. Peng, N. A. Chi, N. Lee, N. G.-A. Krakover, N. Cameron, N. Roberts, N. Doiron, N. Martinez, N. Nangia, N. Deckers, N. Muennighoff, N. Keskar, N. Iyer, N. Constant, N. Fiedel, N. Wen, O. Zhang, O. Agha, O. Elbaghdadi, O. Levy, O. Evans, P. A. M. Casares, P. Doshi, P. Fung, P. Liang, P. Vicol, P. Alipoormolabashi, P. Liao, P. Liang, P. Chang, P. Eckersley, P. M. Htut, P. Hwang, P. Milkowski, P. Patil, P. Pezeshkpour, P. Oli, Q. Mei, Q. Lyu, Q. Chen, R. Banjade, R. E. Rudolph, R. Gabriel, R. Habacker, R. Risco, R. Milliere, R. Garg, R. Barnes, R. Saurous, R. Arakawa, R. Raymaekers, R. Frank, R. Sikand, R. Novak, R. Sitelew, R. L. Bras, R. Liu, R. Jacobs, R. Zhang, R. Salakhutdinov, R. Chi, R. Lee, R. Stovall, R. Teehan, R. Yang, S. Singh, S. Mohammad, S. Anand, S. Dillavou, S. Shleifer, S. Wiseman, S. Gruetter, S. R. Bowman, S. Schoenholz, S. Han, S. Kwatra, S. A. Rous, S. Ghazarian, S. Ghosh, S. Casey, S. Bischoff, S. Gehrmann, S. Schuster, S. Sadeghi, S. S. Hamdan, S. Zhou, S. Srivastava, S. Shi, S. Singh, S. Asaadi, S. Gu, S. Pachchigar, S. Toshniwal, S. Upadhyay, S. Debnath, S. Shakeri, S. Thormeyer, S. Melzi, S. Reddy, S. Makini, S.-H. Lee, S. B. Torene, S. Hatwar, S. Dehaene, S. Divic, S. Ermon, S. Biderman, S. Lin, S. Prasad, S. T. Piantadosi, S. M. Shieber, S. Misherghi, S. Kiritchenko, S. Mishra, T. Linzen, T. Schuster, T. Li, T. Yu, T. Ali, T. Hashimoto, T.-L. Wu, T. Desbordes, T. Rothschild, T. Phan, T. Wang, T. Nkinyili, T. Schick, T. Kornev, T. Tunduny, T. Gerstenberg, T. Chang, T. Neeraj, T. Khot, T. Shultz, U. Shaham, V. Misra, V. Demberg, V. Nyamai, V. Raunak, V. Ramasesh, V. U. Prabhu, V. Padmakumar, V. Srikumar, W. Fedus, W. Saunders, W. Zhang, W. Vossen, X. Ren, X. Tong, X. Zhao, X. Wu, X. Shen, Y. Yaghoobzadeh, Y. Lakretz, Y. Song, Y. Bahri, Y. Choi, Y. Yang, Y. Hao, Y. Chen, Y. Belinkov, Y. Hou, Y. Hou, Y. Bai, Z. Seid, Z. Zhao, Z. Wang, Z. J. Wang, Z. Wang, and Z. Wu, "Beyond the imitation game: Quantifying and extrapolating the capabilities of language models," *ArXiv*, vol. abs/2206.04615, 2022. 14

[115] P. Liang, A. Zadeh, and L.-P. Morency, "Foundations and recent trends in multimodal machine learning: Principles, challenges, and open questions," *ArXiv*, vol. abs/2209.03430, 2022. 14

[116] H. D. Vries, D. Bahdanau, and C. D. Manning, "Towards ecologically valid research on language user interfaces," *ArXiv*, vol.

abs/2007.14435, 2020. 15
[117] S. H. Vemprala, R. Bonatti, A. Bucker, and A. Kapoor, “Chatgpt for robotics: Design principles and model abilities,” *IEEE Access*, vol. 12, pp. 55 682–55 696, 2023. 15
[118] M. Mackeprang, C. Müller-Birn, and M. Stauss, “195 discovering the sweet spot of human—computer configurations: A case study in information extraction.” 15
[119] M. Eiband, D. Buschek, and H. Hussmann, “How to support users in understanding intelligent systems? structuring the discussion,” *Proceedings of the 26th International Conference on Intelligent User Interfaces*, 2020. 15
[120] S. T. Mueller, E. S. Veinott, R. Hoffman, G. Klein, L. Alam, T. Mamun, and W. Clancey, “Principles of explanation in human-ai systems,” *ArXiv*, vol. abs/2102.04972, 2021. 15
[121] K. Ahmad, M. Abdelrazek, C. Arora, A. A. Baniya, M. Bano, and J. C. Grundy, “Requirements engineering framework for human-centered artificial intelligence software systems,” *Appl. Soft Comput.*, vol. 143, p. 110455, 2023. 16
[122] J. Rezwana and M. Maher, “Understanding user perceptions, collaborative experience and user engagement in different human-ai interaction designs for co-creative systems,” *Proceedings of the 14th Conference on Creativity and Cognition*, 2022. 16
[123] A. Kangasrääsiö, K. Athukorala, A. Howes, J. Corander, S. Kaski, and A. Oulasvirta, “Inferring cognitive models from data using approximate bayesian computation,” *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2016. 16
[124] D. Leitão, P. Saleiro, M. A. T. Figueiredo, and P. Bizarro, “Human-ai collaboration in decision-making: Beyond learning to defer,” *ArXiv*, vol. abs/2206.13202, 2022. 17
[125] M. Jadeja and N. Varia, “Perspectives for evaluating conversational ai,” *ArXiv*, vol. abs/1709.04734, 2017. 17
[126] K. Vodrahalli, T. Gerstenberg, and J. Zou, “Do humans trust advice more if it comes from ai?: An analysis of human-ai interactions,” *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 2021. 17
[127] W. Xu, M. 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,” *International Journal of Human–Computer Interaction*, vol. 39, pp. 494 – 518, 2021. 17, 18
[128] A. Fridman, L. Ding, B. Jenik, and B. Reimer, “Arguing machines: Human supervision of black box ai systems that make life-critical decisions,” *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1335–1343, 2017. 17
[129] S. S. Y. Kim, E. A. Watkins, O. Russakovsky, R. C. Fong, and A. Monroy-Hernández, “"help me help the ai": Understanding how explainability can support human-ai interaction,” *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2022. 17
[130] H. Shen, T. Knearem, R. Ghosh, K. Alkiek, K. Krishna, Y. Liu, Z. Ma, S. Petridis, Y.-H. Peng, L. Qiwei, S. Rakshit, C. Si, Y. Xie, J. P. Bigham, F. Bentley, J. Chai, Z. Lipton, Q. Mei, R. Mihalcea, M. Terry, D. Yang, M. R. Morris, P. Resnick, and D. Jurgens, “Towards bidirectional human-ai alignment: A systematic review for clarifications, framework, and future directions,” *ArXiv*, vol. abs/2406.09264, 2024. 18
[131] C. Wienrich and M. Latoschik, “extended artificial intelligence: New prospects of human-ai interaction research,” in *Frontiers in Virtual Reality*, vol. 2, 2021. 18
[132] M. Pietikäinen and O. Silvén, “Challenges of artificial intelligence - from machine learning and computer vision to emotional intelligence,” *ArXiv*, vol. abs/2201.01466, 2022. 18