



Current status and trends of technology, methods, and applications of Human–Computer Intelligent Interaction (HCII): A bibliometric research

Zijie Ding¹ · Yingrui Ji^{2,3} · Yan Gan¹ · Yuwen Wang¹ · Yukun Xia¹

Received: 21 June 2023 / Revised: 10 September 2023 / Accepted: 29 December 2023 /

Published online: 30 January 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

This study delves into Human–Computer Intelligent Interaction (HCII), a burgeoning interdisciplinary field that builds upon traditional Human–Computer Interaction (HCI) by integrating advanced technologies like Natural Language Processing (NLP) and Machine Learning (ML). In this paper, we scrutinize 5,781 HCII papers published between 2000 and 2023, narrowing our focus to 803 most relevant articles to construct co-citation and interdisciplinary networks based on the CiteSpace Software. Our findings reveal that the publications of the United States and China are relatively high with 558 and 616 publications respectively. Furthermore, we found that machine learning and deep learning have emerged as the prevalent methodologies in HCII, which currently emphasizes multimodal emotion recognition, facial expression recognition, and NLP. We predict that HCII will be integrated into advanced applications such as neural-based interactive games and multi-sensory environments. In sum, our analysis underscores HCII’s role in advancing artificial intelligence, facilitating more intuitive and efficient human–computer interactions, and its prospective societal impact. We hope that our review and analysis may guide the efforts of researchers aiming to contribute to HCII and develop more powerful and intelligent methods, tools, and applications.

Keywords Human–Computer Intelligent Interaction · Human–Computer Interaction · Artificial intelligence · Bibliometric Research · CiteSpace

✉ Yan Gan
ganyan@hust.edu.cn

¹ School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

² Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100190, China

³ University of Chinese Academy of Sciences, Beijing 100190, China

1 Introduction

With the rise of artificial intelligence (AI), the field of Human–Computer Interaction (HCI) has undergone significant transformations, leading to more intuitive and efficient human–computer collaborations. Notably, Tesla’s Optimus Prime humanoid robot enhances human productivity and operational efficiency, while Apple’s Apple Vision Pro mixed-reality headset provides a multisensory, immersive user experience [1, 2]. The advancements in Natural Language Processing (NLP) and Computer Vision are instrumental in these developments, presenting both opportunities and challenges in HCI [3–5]. At the same time, a new paradigm known as Human–Computer Intelligent Interaction (HCII) has arisen, driven by the combined efforts of AI and HCI research [6]. Unlike traditional HCI, HCII focuses on the naturalization and humanization of interactions, particularly in complex and specialized application domains. The application of HCII, augmented by AI-centric methodologies, has demonstrated innovative solutions and feasibility across various sectors. These include healthcare [7–12], smart home systems [13, 14], transportation [15–21], intelligent manufacturing [22–25], educational frameworks [26, 27], and interactive gaming [28–30]. The evolution in HCI and the progress of HCII have captured the attention of interdisciplinary researchers, revealing the critical role of AI in shaping the future landscape of human–computer interactions. Recent advancements in Human–Computer Intelligent Interaction (HCII) have yielded a plethora of applications in diverse domains, including emotion and gesture recognition methods [33–41], as well as the integration of Augmented Reality (AR) in healthcare, education, and industrial settings [42–45]. These technologies will enable users to have more natural and immersive interactions with computers. Users will be able to operate computers or other digital devices using natural gestures, and they can also connect human thoughts to computers through brain–computer interfaces, facilitating a more direct and efficient mode of interaction. These technologies will significantly make our lives more convenient and intelligent. Despite these strides, a conflation often exists between HCII and its predecessor, Human–Computer Interaction (HCI), resulting in a vague understanding of their distinct characteristics among the general public and even within the academic community.

To bridge this gap and shed light on the evolutionary trajectory and future directions of HCII, this study employs CiteSpace [46] to conduct a comprehensive analysis spanning literature of two decades (2000–2023). The review and analysis are structured around the following three aspects: (1) Understanding the relationships and differences among HCII, AI, and HCI, particularly focusing on how AI-driven methodologies have transformed HCII from traditional HCI paradigms; (2) Mapping the geographical and institutional contributions to HCII, providing a nuanced understanding of research interests and academic engagements across nations and academic communities; (3) Predicting potential important areas for HCII applications and highlight emerging avenues for innovation. By addressing these aspects, this study aims to provide a rigorous academic framework that can guide and inform future efforts in HCII research and applications.

This study represents a novel endeavor in the Human–Computer Intelligent Interaction (HCII) research landscape, leveraging CiteSpace software for the analysis and visualization of literature. This innovative methodology enables an unprecedented, nuanced understanding of HCII’s evolving research themes, hotspots, and international scholarly collaborations. The analysis addresses existing challenges of correlation and clustering in conventional HCII research, thereby offering a robust analytical lens to capture the dynamism and trends shaping this rapidly growing field. The contributions of this study are three-fold:

(1) This paper rigorously defines HCII, elucidates its symbiotic relationship with Artificial Intelligence (AI) technologies, and outlines the architectural paradigms and technical methodologies underpinning human–computer intelligent interactions; (2) Utilizing the CiteSpace Software, this paper visualizes and scrutinizes HCII-related literature spanning two decades. The analysis yields critical insights into emergent research hotspots, thematic clusters, and longitudinal trends in the HCII domain; (3) This paper prognosticates future trajectories for HCII, identifying prospective intelligent interaction paradigms and methodologies, thereby informing researchers with strategic directions and conceptual scaffolding for future research efforts. In summary, this study aims to act as a foundational reference, providing both a thorough overview of current HCII research and a forward-thinking viewpoint to steer future academic and practical efforts.

The paper proceeds as follows. Section 2 provides a critical review of existing literature on HCII, Artificial Intelligence (AI) algorithms, and the modifications and applications that have evolved in HCII. Section 3 elaborates on the methodology employed in the study, focusing on the data screening processes, database creation, and the utilization of CiteSpace as the analytical tool. Section 4 presents the study's results, featuring analyses of literature growth trends, collaborative networks, co-citation structures, and disciplinary intersections within HCII research. Section 5 discusses the empirical findings, drawing connections to existing literature and theorizing on their implications for the field of HCII. Finally, Section 6 concludes the paper, offering recommendations for future research directions in HCII.

2 Background and related work

2.1 Intelligent human–computer interaction

Human–Computer Intelligent Interaction (HCII) represents an interdisciplinary confluence of Artificial Intelligence (AI) and Human–Computer Interaction (HCI) technologies [4]. At its core, HCII aims to enhance computational system performance through the integration of AI methodologies [47]. The overarching objective shared by both HCII and HCI is the optimization of the user experience [48]. AI technology encompasses a suite of algorithms and computational models designed to emulate human cognition and decision-making processes [49]. Prominent among these are machine learning (ML), natural language processing (NLP), and computer vision (CV). Conversely, HCI focuses on the design principles and implementation strategies that facilitate effective and efficient interaction between humans and computational systems [50]. These interactions may employ a myriad of modalities, including but not limited to, visual cues, auditory signals, haptic feedback, and gestural inputs [51, 52]. The symbiotic relationship between AI and HCI gives rise to the HCII domain, as illustrated in Fig. 1. Specifically, AI-driven techniques such as ML and NLP can be harnessed to refine HCI elements, thereby fostering smarter user interfaces, task automation, and personalized interaction experiences [53]. This synergy further catalyzes the development and application of avant-garde technologies, including autonomous vehicular systems.

Intelligent User Interfaces (IUI) represent a specialized facet of Human–Computer Intelligent Interaction (HCII). These interfaces serve as platforms not just for information exchange but also as adaptive, user-centric operational environments. Through advancements in smart HCI technologies, IUI has given rise to multimodal forms, revolutionizing the realm of user

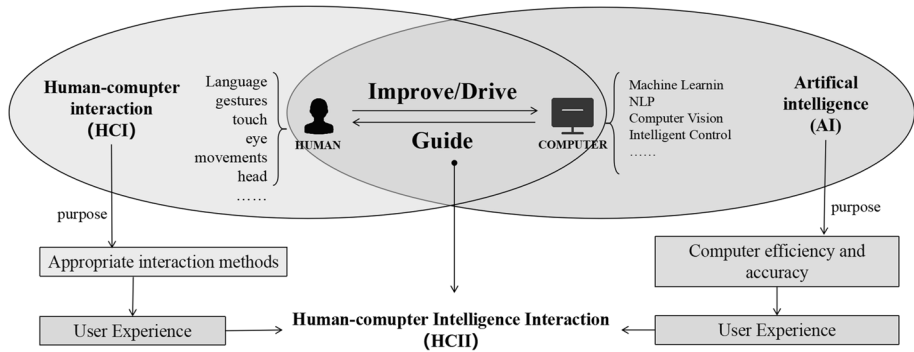


Fig. 1 The relationship between AI and HCI in human–computer intelligence interaction

interfaces [54]. These environments dynamically adjust layout, content, and interaction pathways based on user needs and behaviors, thus optimizing efficiency, efficacy, and naturalness [55–57]. Moreover, IUIs employ sophisticated modeling techniques, encompassing user, domain, task, conversation, and media models, to facilitate nuanced forms of interaction [58, 59].

Contrastingly, Human–Computer Dialogue Systems extend the scope of intelligent interactions by focusing on natural language-based communication [60]. Unlike HCII, which primarily aims for efficient and effective interactions, these systems deploy Natural Language Processing (NLP) technologies to simulate natural conversational dynamics. These dialogue systems find extensive application in sectors like voice-activated assistants [11], intelligent customer service [61], and chatbots [27].

Human–computer dialogue systems deploy a multi-tiered architectural framework to manage the complexity of information sensing, processing, decision-making, and control, ultimately culminating in application-specific functionalities [62] (Fig. 2). **Sensor Layer:** Constituting the foundational tier, the sensor layer captures a variety of human inputs, including vocal cues, visual data, and gestural signals, through specialized hardware such as microphones, cameras, and other sensors. **Signal Processing Layer:** As the subsequent layer, this module is tasked with the computational analysis of data amassed by the sensor layer. It employs algorithms for speech, image, and gesture recognition, serving as the bridge between raw data and actionable insights. **Intelligent Decision-Making Layer:** Leveraging the processed signals, this layer utilizes a range of machine learning, deep learning, and natural language processing algorithms to facilitate autonomous decision-making and control mechanisms. **Interactive Interface Layer:** Situated above the decision-making layer, this tier is responsible for presenting the processed information to the user. It accommodates various forms of user feedback through multiple interfaces, including graphical, vocal, and gestural pathways. **Application Layer:** Representing the apex of this architectural model, the application layer executes specific functionalities based on synthesized information and user interactions, thereby realizing the system’s ultimate objectives.

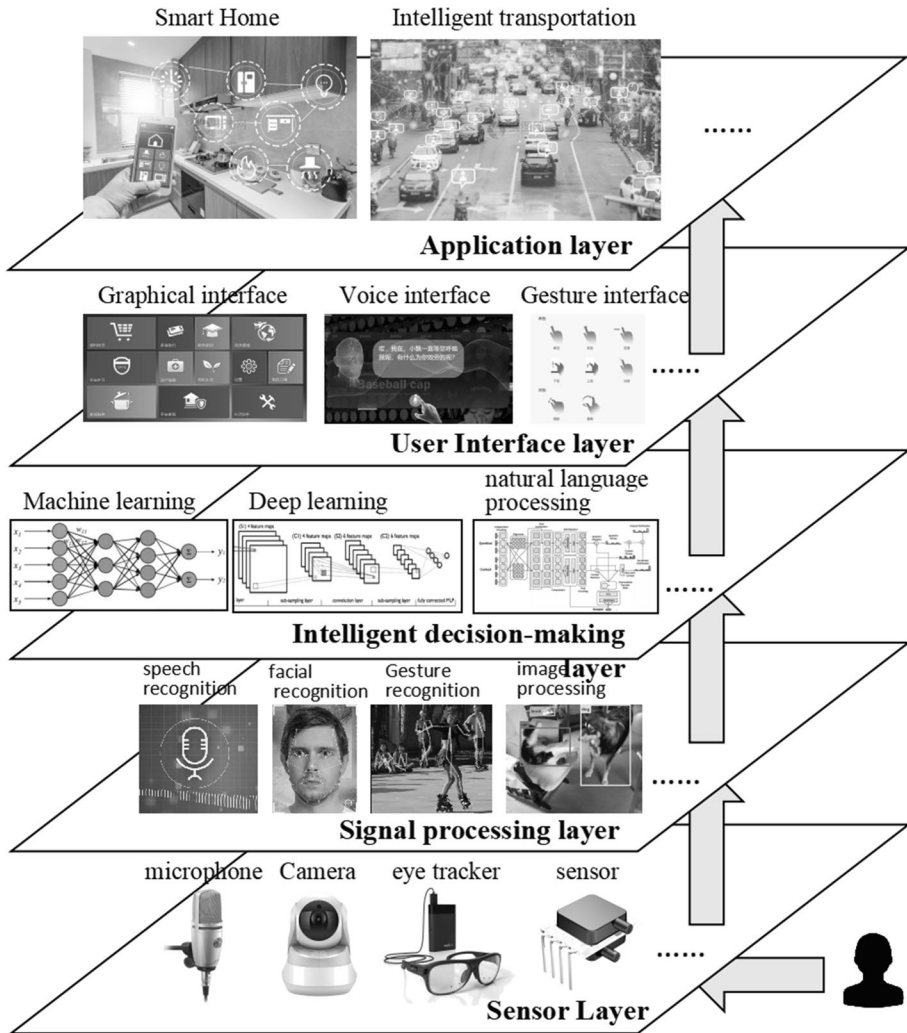


Fig. 2 HCI’s five levels of application architecture

2.2 AI algorithms in human–computer intelligent interaction

Machine Learning Algorithms (ML) [63] Encompassing supervised, unsupervised, and semi-supervised learning paradigms, these algorithms find applications in tasks such as classification, clustering, and predictive modeling.

Deep Learning Algorithms (DL) [64] Incorporating architectures like convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs), these algorithms are instrumental in image recognition and natural language processing tasks.

Decision tree algorithms These algorithms facilitate classification and predictive analytics by recursively partitioning feature spaces and constructing decision trees.

Support Vector Machines (SVM) Employed for classification and regression tasks, SVM algorithms map data into high-dimensional spaces to identify optimal hyperplanes for classification.

Clustering algorithms Algorithms such as K-means and hierarchical clustering serve to uncover the intrinsic structures within datasets by grouping them into clusters.

Reinforcement learning algorithms These algorithms enable agents to learn optimal decision-making strategies through a system of rewards and penalties.

2.3 Evolution and diversification of HCII: From HCI to multimodal interactions

Traditional HCI primarily emphasized visual interactions via interface operating systems to meet user requirements [70]. The subsequent advent of voice interaction technology marked a pivotal junction, fostering collaboration between HCI and AI disciplines [71]. As the field evolved into HCII, researchers have increasingly harnessed AI techniques to explore a plethora of interaction modalities. Firstly, HCII accentuates natural user interfaces, facilitating interactions through speech, gestures, and facial expressions rather than conventional devices like mice and keyboards [65, 66]. This naturalistic interface design aligns more closely with human behavior, thereby enriching user experience and interaction efficiency.

Secondly, HCII leverages AI-driven technologies capable of intelligent recognition and physiological signal analysis [51, 52], thereby achieving nuanced understanding of user intentions and requirements. This level of understanding enables the provision of highly personalized and intelligent services.

Moreover, HCII integrates multimodal interactions [58, 67], such as simultaneous voice and gesture recognition, amplifying the flexibility and efficacy of the interactive experience.

Lastly, emotional interaction capabilities have been incorporated [68, 69], allowing the system to recognize and respond to users' emotional states through vocal and facial cues, thereby delivering a more empathetic user experience. The outcomes have been transformative, diversifying HCII implementations into the following categories: Voice Interaction: Facilitates verbal communication between users and systems; Visual Interaction: Involves graphical or image-based interfaces; Gesture Interaction: Utilizes human gestures for non-verbal communication; Multimodal Interaction: Incorporates multiple interaction channels simultaneously, such as voice and gesture [58, 67]; Emotional Interaction: Employs sentiment analysis algorithms to interpret user emotions for more empathetic user experiences [68, 69].

In terms of cross-domain applications, HCII technologies are gaining traction across a wide range of sectors, including healthcare, smart homes, transportation, intelligent manufacturing, education, and interactive gaming. In these domains, HCII is principally employed for control, monitoring, administration, and service delivery. Core methodologies include: Recognition Technologies: Used for identifying voice, image, and gesture inputs; Machine Learning and Natural Language Processing: Employed for decision-making and language interpretation; Data Analytics: Utilized for extracting actionable insights

from large datasets; Virtual Reality Technologies: Employed for immersive user experiences; Sensor Technologies: Used for capturing a wide range of physical data; User Interface Design: Focuses on creating intuitive and user-friendly interfaces.

- (1) **Voice Interaction** Originating from spoken dialogue systems, voice interaction leverages technologies such as speech recognition, natural language comprehension, dialogue management, and speech synthesis [72]. The ubiquity and convenience of speech have propelled voice interaction into mainstream application technologies. Voice-controlled interfaces have found significant applications in vehicle control [13, 17], autonomous driving [73], and domestic settings [11], enhancing naturalness and precision in user interfaces [74]. Emotional cues in voice serve as invaluable data for emotion recognition algorithms [75, 76].
- (2) **Visual Interaction:** Visual interaction is integrally related to Intelligent User Interfaces (IUI), relying on cameras, sensors, and other hardware to facilitate human–machine visual communication. Innovations like virtual reality [29–31], augmented reality [30, 77], and digital twin technology [78–81] have expanded the scope of visual interaction. Its application areas include medical imaging [41], robot control [20, 21], and vehicle safety [14–16].
- (3) **Gesture Interaction:** Gesture interaction encompasses a variety of non-verbal cues such as movements and signs [51]. Due to the inherent complexity in gesture recognition [82], AI algorithms coupled with machine learning techniques are employed to enhance recognition accuracy. Gesture-based interactions are particularly relevant in robotics, somatosensory games, virtual reality, and sign language recognition [68, 83]. Research in this domain often focuses on pose estimation [84], motion capture [85, 86], and gesture classification.
- (4) **Multimodal and Emotional Interaction:** Multimodal interaction fuses multiple interaction modalities to offer a more natural and efficacious user experience. Emotional interactions often leverage multimodal recognition algorithms, utilizing facial and speech recognition to model emotional states [69, 87–90]. Extensive research is underway in innovative applications such as personality-based financial advisory systems [91], personalized smart clothing [92], interactive cultural heritage communities [93], and Alzheimer’s disease recognition through multimodal fusion [94].

2.4 Related work

Similar studies from the past ten years to this paper’s topic are [4, 95, 96]. First, Bumak et al. [95] identified and analyzed the research on HCII and Intelligent User Interface (IUI) between 2010 and 2021, concluding that deep learning and instance-based AI methods and algorithms are the most commonly used by researchers. Support Vector Machines (SVMs) are the most popular algorithms for recognising emotions, facial expressions, and gestures, whereas Convolutional Neural Networks (CNNs) are the most popular deep learning algorithms for recognition techniques. Quantitative studies have mapped the extant sensors, techniques, and algorithms used in HCII, but there are fewer quantitative studies and forecasts of future trends, which have not yet been scientifically summarised. The combination of CNN and Long Short-Term Memory Network (LSTM) can significantly enhance the accuracy and precision of action recognition. With the proliferation of interactive devices such as AR/VR, they predict that diverse sensors will be combined to store and control more complicated processes. They imply that smarter human–computer interaction may

be used in some public areas, and that virtual reality glasses will be used in entertainment, gaming, and industrial manufacturing, but they have not yet specified existing application scenarios with corresponding technical approaches. Karpov et al. [96] conducted a study on the future of cognitive interfaces and the most recent HCI technologies. Similar to [95], they simultaneously analyzed and reviewed HCII and IUI; however, the former focuses more on investigating the technology of human–computer interaction, whereas the latter focuses on the user interface. Examining the progression of user interfaces from command text to graphics to intelligent unimodal and multimodal user interfaces. It also analyses the development of multimodal interfaces between countries and derives some conclusions about HCI systems for people with disabilities and the trend towards "human-to-human" interaction.

In addition, there are a number of related and comparable review articles, such as the systematic review of HCII and AI in healthcare [97], the research progress between human–computer interaction and intelligent robots [98], and human–computer interaction and evaluation in AI conversational agents [99, 100]. There is no research that summarises the overall development of human–computer intelligent interaction and related application areas and popular trends. This paper seeks to comprehend the existing hot technologies and application areas in the field of human–computer intelligent interaction, to create a series of hot knowledge maps using bibliometric techniques, and to investigate the emerging methods and technologies of human–computer intelligent interaction as well as the future application scenarios.

3 Methodology

3.1 Database selection and search strategy

The Web of Science (WoS) Core Collection was elected as the primary database for this research, owing to its comprehensive and reliable citation indexing for scholarly literature across scientific and technical domains [63]. This database was favored for its high-quality content and its capability for detailed categorization of retrieved data, encompassing attributes such as title, author, institution, and publication type. The search query was formulated to capture articles pertinent to "Human–Computer Intelligent Interaction." Given the overlapping thematic scopes of "Artificial Intelligence" and "Human–Computer Interaction," these terms were included in the query. Further specificity was afforded by incorporating "Machine Learning" and "Deep Learning" into the search criteria. The final query string was thus constructed as follows: Initial search outcomes were refined by applying category-based filters, such as "Computer Science," "Electrical and Electronic Engineering," "Human Engineering," "Psychology," "Multidisciplinary Engineering," among others. Figure 3 shows the screening process of the literature in this paper. This step was implemented to exclude non-relevant literature in disparate fields like chemistry. This exercise resulted in an initial pool of 5,781 articles. Subsequently, articles appearing in journals not directly aligned with the thematic focus of HCII were excluded. Examples of such journals include but are not limited to "Lecture Notes in Computer Science," "Institute of Electrical Engineers Access," and "Sensors," yielding a final corpus of 803 papers. A meticulous manual screening process was conducted, wherein articles were evaluated based on their title, abstract, and keywords. The screening was narrowed to cover HCII-specific tasks such as data mining, signal processing, feature recognition, and recommendation models, within

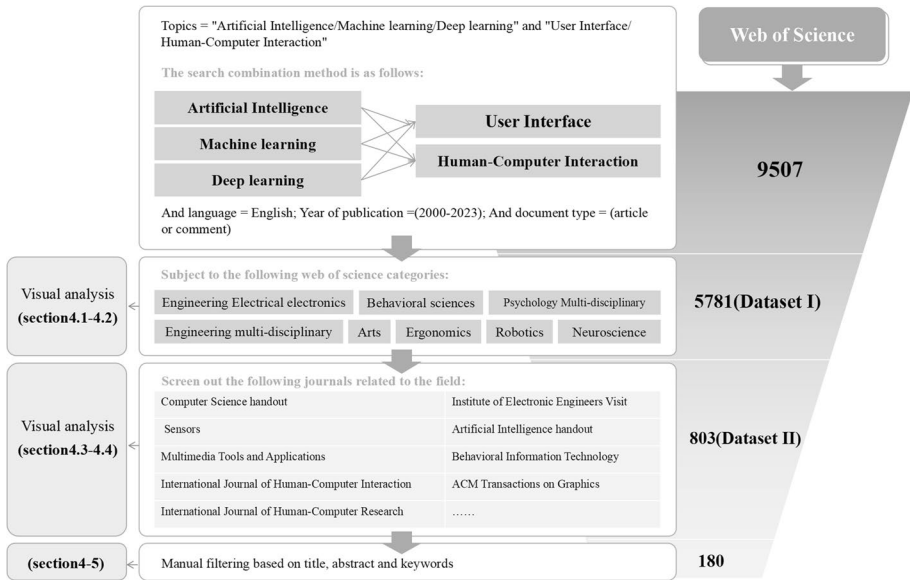


Fig. 3 Literature screening and utilization process

applicative domains like healthcare, smart homes, transportation, and education. A comprehensive analysis was subsequently carried out on a curated subset of 180 scholarly articles, aimed at encapsulating the existing paradigms in user-interface interactions and future directions in the field.

3.2 Bibliometric analysis and knowledge mapping

This study adopts a bibliometric approach combined with knowledge mapping to elucidate the current state, emergent trends, and future trajectories in the domain of Human-Computer Intelligent Interaction (HCII). Knowledge mapping is an interdisciplinary methodology designed to visualize, analyze, and interpret patterns in scientific literature, thereby contributing to the broader field of scientometrics [101, 102]. There are many visualisation tools that can be used as knowledge graphs, such as VOSviewer [103], SciMAT [104], etc., but CiteSpace is currently the more popular and powerful tool. We employ CiteSpace, a software application designed for the interactive and exploratory analysis of scientific fields. It is particularly useful for generating visual co-citation networks and conducting collaborative analyses across authors, institutions, and nations [105]. The software also facilitates co-occurrence analyses of terms, keywords, and subject categories, as well as co-citation analyses of referenced literature, authors, and journals [46]. CiteSpace offers Betweenness Centrality (BC) as a metric for identifying pivot nodes in a scientific network. The BC value is calculated based on the shortest paths connecting the node of interest to all other nodes in the network [106]. The time-zone graph feature in CiteSpace provides a temporal perspective on the dataset, enabling the identification of citation bursts and emergent research trends [106]. This tool has been widely applied in research fields such as sustainability and urban studies [107–109]. Figure 4 depicts the methodological process of using the CiteSpace tool. To achieve the research objectives, several parameters

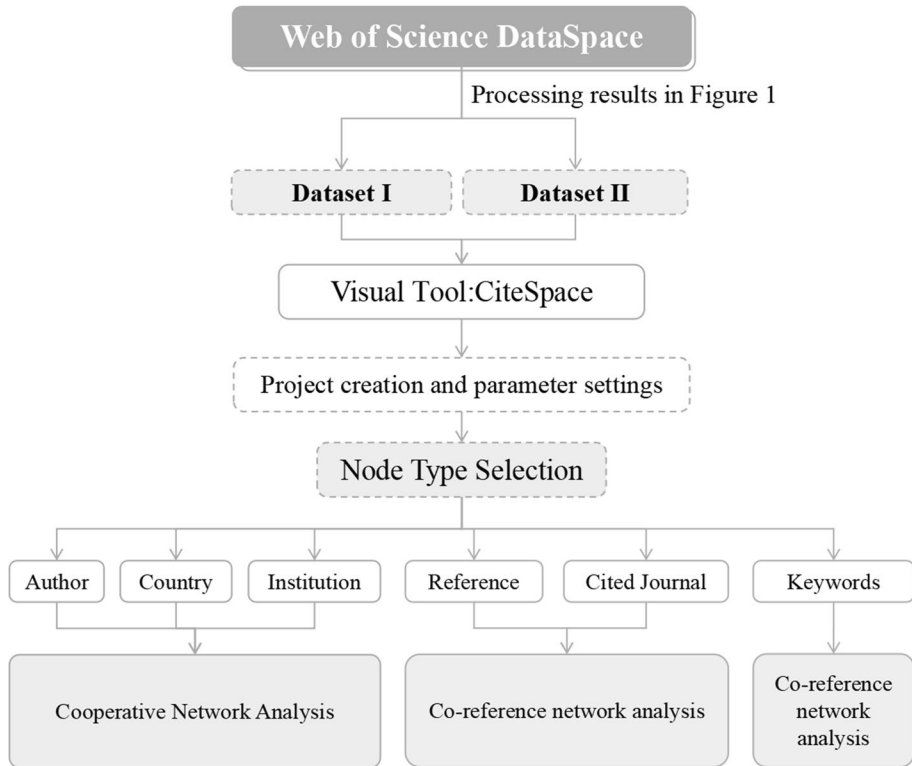


Fig. 4 CiteSpace visualization and analysis process

in CiteSpace — including time slices, node types, links, selection criteria, and pruning methods—were carefully configured during the analytical process [110]. The study uses CiteSpace to create various types of visual networks, including collaborative networks among countries, authors, and institutions; co-citation networks that highlight clusters of cited literature; and disciplinary category symbiosis networks that reveal the interconnections among key knowledge domains.

4 Results

4.1 Trends in literature growth

Figure 5 offers a comprehensive statistical analysis based on the initial screening of 5,781 articles. The scholarly output in the Human–Computer Intelligent Interaction (HCII) domain appears to have evolved through three distinct developmental phases: an exploratory phase from 2000 to 2013, an initial development phase from 2014 to 2016, and a rapid growth phase from 2017 to 2022. During the exploratory phase, a total of 872 papers were published, accounting for approximately 15% of the total literature, indicating the field was in its nascent stage. The initial development phase was relatively short-lived, signaling a swift uptick in research interest. The rapid growth phase,

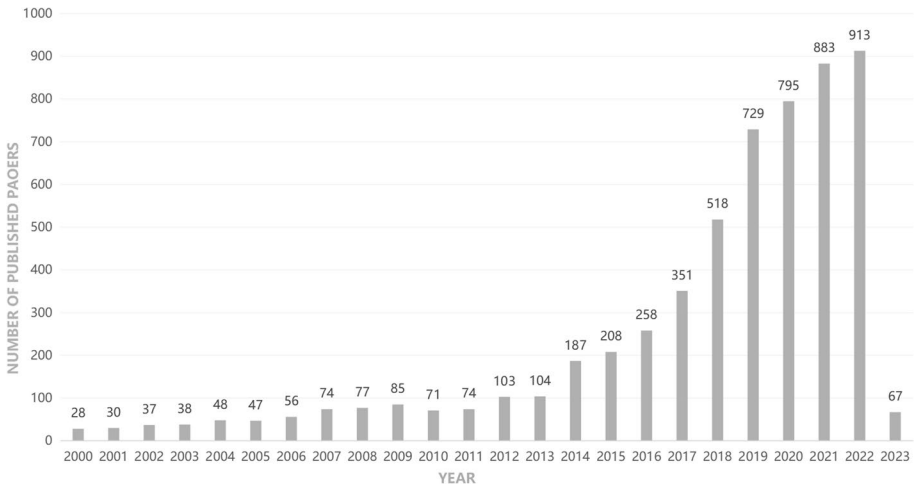


Fig. 5 Research and development trends in HCII domain data

stretching from 2017 to 2022, saw a substantial increase in applied research related to HCII. The average annual growth rate of the literature was around 100 articles during this period. Specifically, the years 2020, 2021, and 2022 witnessed the publication of approximately 800, 883, and 913 papers, respectively. Publications in this stage make up 72.4% of the entire dataset. Given the current trajectory, it is hypothesized that scholarly contributions to the HCII field will likely continue to proliferate at an accelerated rate.

4.2 Analysis of cooperation networks

A network analysis of author collaborations was undertaken, based on the initial screening of 5,781 articles, to identify patterns of cooperation among authors in the HCII domain.

4.2.1 Study author association analysis

Figure 6 presents a visual network graph generated using CiteSpace, where each node symbolizes an individual author. The graph reveals that 713 authors have engaged in HCII-related research, resulting in 410 collaborative instances. The density of this collaboration network is measured at 0.0016, suggesting that scholarly cooperation in this field is relatively limited. Most contributions appear to be independent efforts, without the formation of extensive, collaborative research teams.

Table 1 lists the top 10 prolific authors, ranked by their publication count. The most active author in this realm has contributed to nine relevant papers, and all authors in the top 10 have published at least four articles. This analysis underscores the current state of collaboration within the HCII field, indicating that although the field is growing, it has yet to develop extensive networks of academic cooperation.

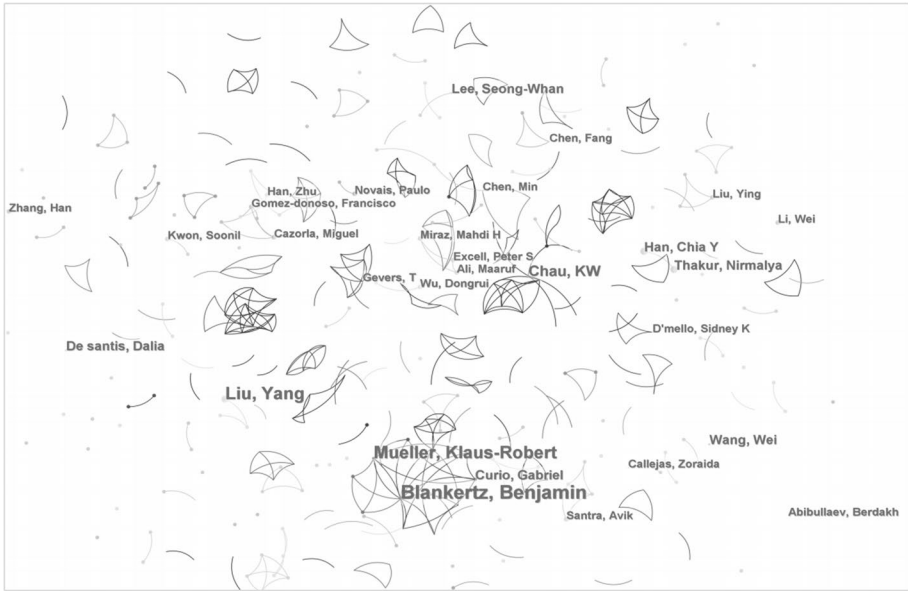


Fig. 6 Network diagram of research authors in the HCII domain of the WoS core database

Table 1 Top 10 research authors in the HCII domain of the WoS core database

Serial No	Author Name	Year of Debut	Number of Articles
1	Blankertz Benjamin	2006	9
2	Liu, Yang	2021	7
3	Mueller, Klaus-Robert	2007	7
4	Chau, KW	2001	5
5	Thakur, Nirmalya	2021	4
6	Wang, Wei	2006	4
7	Han, Chia Y	2020	4
8	Curio, Gabriel	2020	4
9	De santis, Dalia	2020	4
10	Lee, Seong-Whan	2019	4

4.2.2 Collaborative linkage analysis of research institutions

In this section, the focus shifts to the network of collaborations among research institutions engaged in HCII studies. Figure 7 provides a visual depiction of this network, where each node represents an institution. The size of each node is proportional to the institution’s publication count, while the links between nodes signify collaborative endeavors. An analysis reveals that 602 institutions have participated in HCII research, resulting in 466 instances of inter-institutional collaboration. The density of this

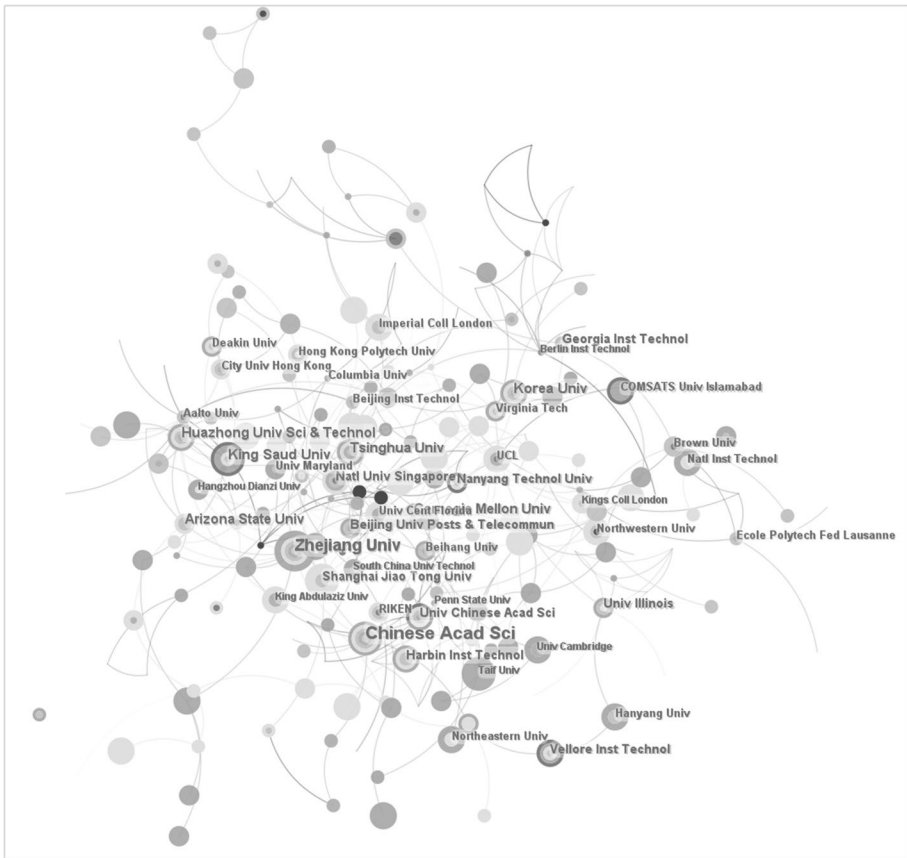


Fig. 7 Collaborative network of research institutions in the HCII area of the WoS core database

network is calculated to be 0.0026, which is notably higher than the author collaboration network, indicating tighter institutional linkages.

Table 2 lists the top 10 research institutions based on their publication output. Among them, 80 key research institutions have published at least five articles in the HCII domain. The Chinese Academy of Sciences leads the pack with 36 publications over the past two decades. It is followed by Korea University, Zhejiang University, King Saud University, Tsinghua University, Carnegie Mellon University, Huazhong University of Science and Technology, Arizona State University, Vellore Institute of Technology, and Shanghai Jiao Tong University, each contributing more than 10 articles.

The data suggests that while a multitude of institutions are active in HCII research, their efforts are relatively dispersed. Despite the high level of interest and attention from these institutions, there is room for enhancing collaboration and synergies. Currently, there are few institutions that have emerged as definitive leaders in the field.

In the last two years, the Chinese Academy of Sciences (CAS) has concentrated its efforts on several key areas within the realm of Human–Computer Intelligent Interaction (HCII). These focus areas include: CAS has been exploring the fusion of various interaction modalities to create a more natural and effective user experience. The institute

Table 2 Top 10 cooperating organizations in the HCII area of the WoS core database

Serial No	Name of Institution	Year of Debut	Number of Articles
1	Chinese Acad Sci	2005	36
2	Korea Univ	2017	21
3	Zhejiang Univ	2005	17
4	King Saud Univ	2020	16
5	Tsinghua Univ	2018	15
6	Carnegie Mellon Univ	2003	14
7	Huazhong Univ Sci & Technol	2006	13
8	Arizona State Univ	2008	13
9	Vellore Inst Technol	2020	12
10	Shanghai Jiao Tong Univ	2020	11

has been particularly active in this domain. Researchers have employed Convolutional Neural Networks (CNNs) to extract and classify implicit features from normalized facial images. Techniques such as facial cropping and rotation have been deployed to optimize CNN architectures, thereby enhancing the efficiency of facial expression recognition systems [111]. Tang Y et al. introduced a frequency neural network (FreNet) as a pioneering approach in frequency-based facial expression recognition [115], achieving superior performance [89, 114]. CAS has also delved into the creation of human models, particularly in the context of gesture detection [113] and 3D spatio-temporal skeleton rendering [7, 112]. Research at CAS has resulted in innovative medical applications based on deep learning models. These include: Skeleton-based human body recognition, facilitating 3D rendering of spatio-temporal skeletons [7]; Surgical robots controllable via eye movements [8, 9]; Pulse cumulative image mapping techniques using CNNs to accurately measure real heart rates [10, 11]. Promising results have been achieved using recurrent neural networks for the recognition and classification of EEG signals, contributing to the field of affective computing.

Overall, the extensive research portfolio of CAS manifests its leadership in the HCII domain, particularly in applying machine learning and deep learning techniques to tackle complex, real-world problems.

The second major contributor to the field of Human–Computer Intelligent Interaction (HCII) is a Korean university. Their research is largely application-driven, targeting the improvement of human–computer interaction and user interfaces in smart sensing devices and virtual systems. Below are some of the key focus areas: Virtual Reality Sports [116]: Development of immersive virtual reality experiences tailored for sports. Gaming Fitness [26]: Leveraging gaming elements to enhance physical fitness training. Interactive Sports Training Strategy [117]: Utilizing sensor data and machine learning algorithms to develop personalized and effective sports training strategies. AI Agent System for Multiple Vehicles [17]: Deployment of intelligent agents to manage fleets of vehicles. Real-time Bus Management System [18]: Utilizing data mining technologies for real-time management and optimization of bus services. Self-driving Car Takeover Performance [19]: Evaluating the efficiency and reliability of human takeover in self-driving cars through machine learning techniques. Facial Recognition [118]: Research into advanced facial recognition technologies for various medical applications. Gesture

Recognition [119]: Active investigation into gesture-based control and monitoring systems for healthcare applications.

Overall, the research emanating from this Korean university demonstrates a strong orientation towards applied HCII. The primary focus is on the recognition, extraction, and classification of human behaviors. This research is not only innovative but also finds practical applications in critical sectors like healthcare and transportation. The work contributes to the broader goal of making human–computer interaction more natural, efficient, and beneficial across various domains.

4.2.3 Global distribution of research contributions by country

The geographic distribution of research contributions serves as an indicator of a country's significance and influence in a specific field. Moreover, international collaboration is instrumental in fostering innovation and making strides in research. In this analysis, countries are represented as nodes, and metrics such as the number of publications and betweenness centrality are computed. Figure 8 presents a symbiotic network diagram that outlines the global distribution of HCII research contributions. The network revealed that: A total of 101 countries have engaged in HCII research ($N=101$). These countries have collaborated 374 times ($E=374$). The collaboration density is calculated to be 0.741, signifying a high level of international collaboration. China, the United States, and France emerged as the most interconnected nations in this domain, each contributing 7 articles. This attests to a vibrant academic environment in HCII within these countries. Table 3 extracts the top 15 nations based on their publication counts: China (616 publications) and the United States (558 publications) lead the pack with substantial contributions exceeding 500 publications each. Betweenness centrality values indicate that the U.S.(0.75) has a slightly higher influence compared to China (0.55) in the field. While Germany and France

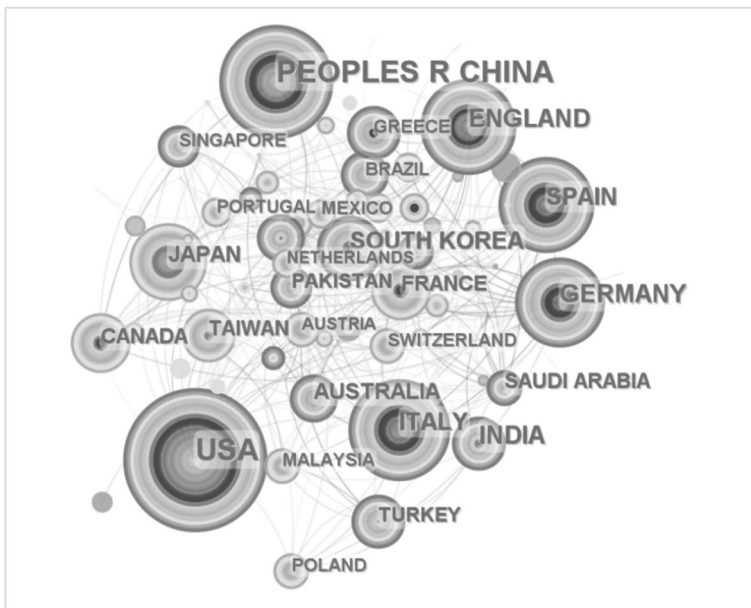


Fig. 8 Global distribution of research countries in the HCII domain of the WoS core database

Table 3 Top 15 countries in the HCII area of the WoS core database

Serial No	Country or Area	Year of Debut	Number of Articles	Centrality
1	PEOPLES R CHINA	2000	616	0.55
2	USA	2000	558	0.22
3	INDIA	2012	185	0.00
4	ENGLAND	2000	182	0.23
5	GERMANY	2000	179	0.44
6	SOUTH KOREA	2000	153	0.04
7	ITALY	2003	138	0.12
8	SPAIN	2004	135	0.11
9	AUSTRALIA	2000	96	0.19
10	JAPAN	2000	92	0.24
11	CANADA	2000	88	0.00
12	SAUDI ARABIA	2002	86	0.18
13	FRANCE	2014	84	0.49
14	TAIWAN	2000	72	0.26
15	PAKISTAN	2000	51	0.27

have fewer publications, their higher centrality scores indicate that their research contributions are of high quality and serve as pivotal nodes in the global HCII research network. In summary, the HCII field enjoys a rich tapestry of international contributions, with China being the most prolific contributor both in terms of publications and international collaborations. However, there is room for more active academic cooperation and integration to advance the field. From a global standpoint, the research output from the United States exerts a greater influence, whereas China's contributions are more abundant. For the future, there is a pressing need to encourage more diversified and intelligent development in HCII research across the globe.

4.3 Co-citation network analysis

Co-citation analysis offers a lens through which the academic landscape of a field can be examined in detail. Based on the 803 articles screened during the second phase, this section conducts both publication co-citation and literature co-citation analyses in the field of Human–Computer Intelligent Interaction (HCII).

4.3.1 Publication co-citation analysis

The co-citation of publications serves as a valuable metric for identifying influential work within a scientific domain. Academic publications constitute the building blocks of any research field, and analyzing them can guide both researchers and practitioners toward seminal work in the area. As illustrated in Fig. 9, nodes in this analysis represent cited journals. The network comprises: A total of 748 nodes ($N=748$), 1,939 links ($E=1,939$), A collaborative density bit of 0.0069. The size of each node reflects the journal's co-citation frequency. Journals with more than 60 co-citations are prominently displayed in the

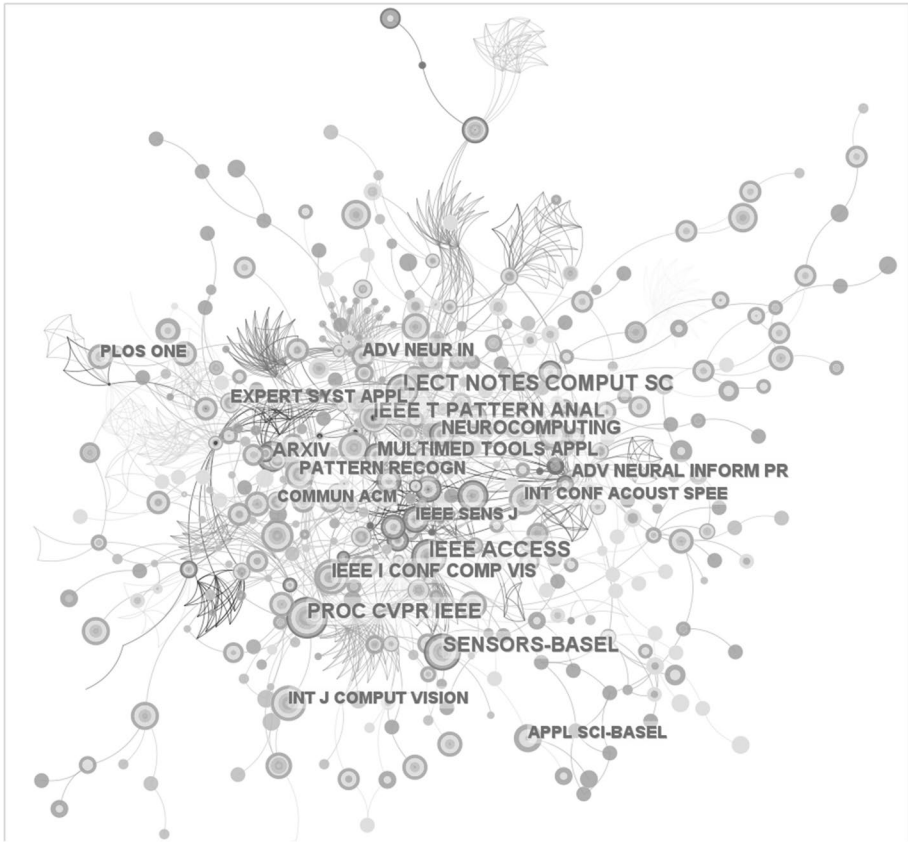


Fig. 9 Co-citation network map of publications in the HCII domain of the WoS core database

network. Table 4 describes the top 10 journals in terms of number of publications. Notably, "Lecture Notes in Computer Science" leads the pack with 226 publications, closely followed by "IEEE Access" with 222 citations and "Sensors" with 181 citations. The high

Table 4 Top 10 publications in the HCII area of the WoS core database

Serial No	Cited Journals	Year of Debut	Citations Count
1	LECT NOTES COMPUT SCI	2006	226
2	IEEE ACCESS	2018	222
3	SENSORS	2014	181
4	PROC CVPR IEEE	2016	177
5	IEEE T PATTERN ANAL	2001	136
6	ARXIV	2018	133
7	MULTIMED TOOLS APPL	2015	108
8	IEEE I CONF COMP VIS	2016	94
9	NEUROCOMPUTING	2015	94
10	EXPERT SYST APPL	2002	85

co-citation frequency of these core journals suggests that they publish research that is both in-depth and exhaustive. Their high co-citation count is indicative of their significant contributions to the field and their value as reliable sources of high-impact research (Table 5).

4.3.2 Literature co-citation analysis

Literature co-citation analysis serves as an invaluable tool for understanding the intellectual structure of a scientific field. By examining the co-citation relationships between academic articles, this analytical approach uncovers the seminal works that have shaped the field, identifies influential technologies, and elucidates current trends. The references within the HCII field are conceptualized as nodes to form a co-citation network, as illustrated in Fig. 10. The network exhibits the following characteristics: 706 nodes ($N=706$), 1,870 links ($E=1,870$), A collaborative density of 0.0075. Overall, the network reveals a robust pattern of citation relationships, highlighting an environment of interdisciplinary learning and cross-citation. The most frequently cited literature in the HCII field includes, as shown in fig. 5: He, Kaiming, et al.(2016) with 26 citations, Simonyan, K. with 17 citations, Vaswani, A. with 15 citations. These works predominantly focus on foundational artificial intelligence algorithms and architectures, especially those based on machine learning and deep learning. These algorithms are often deployed for tasks involving image and language recognition and classification. Specifically, they frequently employ convolutional neural networks as the basis for more advanced algorithms [120, 121]. The prominence of these works suggests that they have contributed significantly to the development and advancement of the HCII field. By identifying these cornerstone publications, we can gain a comprehensive understanding of the state-of-the-art technologies and methods that currently shape HCII.

4.4 Analysis of disciplinary symbiotic networks

4.4.1 Keyword co-occurrence analysis

Keywords serve as the intellectual core of an academic article, encapsulating the paper's primary focus, themes, methodologies, and research scope. They provide a snapshot of the research domain, methodologies, and specific topics covered. By

Table 5 Top 10 cited documents in the HCII domain of the WoS Core Database

No	Cited Literature	Count	Year	DOI
1	He, Kaiming et al	26	2016	DOI https://doi.org/10.1109/cvpr.2016.90
2	Simonyan K et al	17	2015	DOI https://doi.org/10.48550/arXiv.1409.1556
3	Vaswani A et al	15	2017	DOI https://doi.org/10.48550/arXiv.1706.03762
4	Krizhevsky A et al	14	2017	DOI https://doi.org/10.1145/3065386
5	Devlin, J et al	13	2019	DOI https://doi.org/10.18653/v1/N19-1423
6	Szegedy C et al	11	2015	DOI https://doi.org/10.1109/CVPR.2015.7298594
7	Lawhern VJ et al	10	2018	DOI https://doi.org/10.1088/1741-2552/aace8c
8	Ren SQ et al	10	2015	DOI https://doi.org/10.1109/TPAMI.2016.2577031
9	LeCun Y et al	9	2015	DOI https://doi.org/10.1038/nature14539
10	Huang G et al	9	2017	DOI https://doi.org/10.1109/CVPR.2017.243

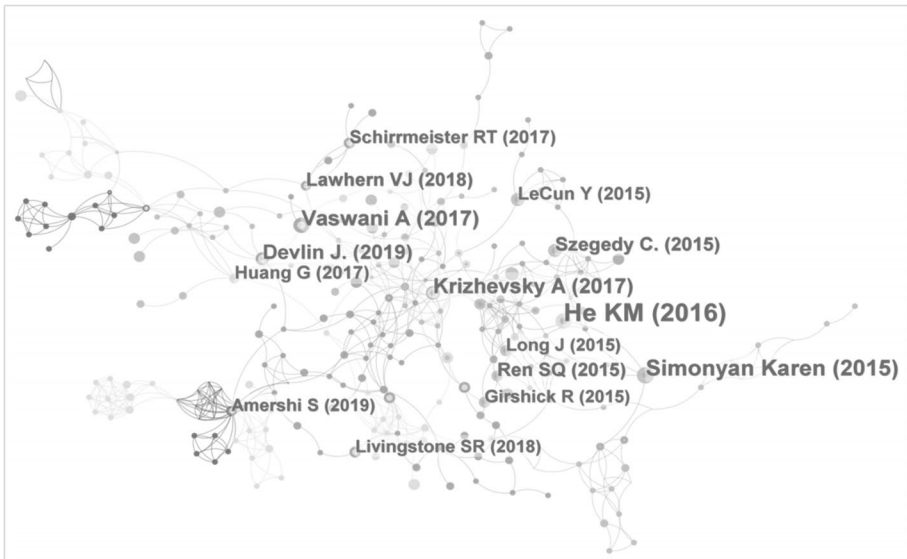


Fig. 10 Co-citation network of literature in the HCII domain of the WoS core database

analyzing keyword co-occurrence, we can construct a network graph that reflects the thematic concentration of the field. In Fig. 11, the size of each node represents the frequency of keyword occurrence. Nodes with purple edges indicate higher centrality values. The larger the purple circle, the more central the keyword is to the field. The keyword co-occurrence network consists of: 379 Keywords ($N = 379$), 850 Links ($E = 850$), A density of 0.0119.

This indicates a strong interrelation between the keywords, revealing a cohesive and robust correlation network for the HCII field. Table 6 lists the top 15 most frequently occurring keywords, their first year of appearance, their frequency, and their centrality measures. The most frequently occurring keywords are "Deep Learning" and "Machine Learning," with the latter having the highest mediational centrality. This suggests that machine learning serves as a foundational keyword within the field. Deep learning is commonly applied in areas such as multimodal human–computer interaction [4] and user interface evaluation [122]. The next most frequent keywords are: Human–Computer Interaction, Feature Extraction, Artificial Intelligence, Neural Network, Convolutional Neural Network, Model, System, Emotion Recognition, Gesture Recognition.

These keywords suggest that the primary focus of HCII research is on the "human" aspect of human–computer interaction. Specifically, there is a strong emphasis on recognizing various human senses, postures, and emotions. This is enabled by technologies such as artificial intelligence, and the most frequently used models aim to optimize the system for enhanced user interface usability and increased HCI effectiveness in a range of tasks. The keyword co-occurrence analysis provides valuable insights into the overarching themes and focal points of HCII research. The analysis reveals a strong emphasis on machine learning technologies, with a particular focus on deep learning, and highlights the role of AI in enhancing human–computer interactions.

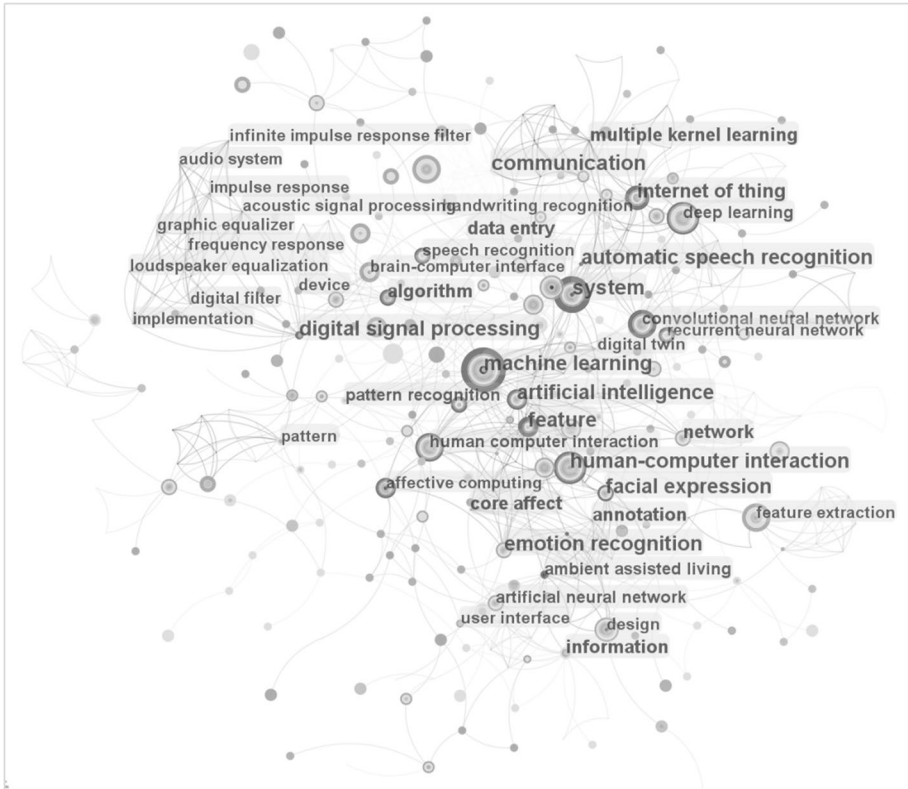


Fig. 11 Keyword co-occurrence network diagram for the HCII domain of the WoS core database

Table 6 Top 15 high-frequency keywords in the HCII domain of the WoS core database

Serial No	Keywords	Year of Debut	Frequency	Centrality
1	Deep learning	2015	127	0.06
2	Machine learning	2007	80	0.25
3	Human–computer interaction	2007	72	0.14
4	Feature Extraction	2017	54	0.07
5	Artificial Intelligence	2003	49	0.18
6	Neural Network	2001	43	0.00
7	Convolutional Neural Network	2016	41	0.11
8	Model	2003	38	0.04
9	System	2001	35	0.35
10	Emotion Recognition	2012	34	0.10
11	Gesture Recognition	2016	31	0.01
12	Recognition	2014	31	0.02
13	Classification	2017	30	0.08
14	Design	2014	21	0.09
15	Network	2004	20	0.09

4.4.2 Keyword clustering

Keyword clustering is an advanced technique used to group related keywords together, providing a lens through which to view various research hotspots within the domain. This kind of clustering can reveal the knowledge structure of a given field, offering insights into its core components and emergent trends. In this analysis, keywords were clustered using the Log-Likelihood Ratio (LLR) algorithm. The parameter for the "K clusters showing the largest in the cluster" was set to 14 to isolate the top 14 clusters within this research area. This resulted in the keyword clustering map for the HCII research domain shown in Fig. 12. Modularity Value = 0.791 > 0.3, Q = 0.791 > 0.3: Indicates that the clusters have a significant structure. Average Silhouette Value = 0.9056 > 0.5, S = 0.9056 > 0.5: Suggests that the clusters are well-defined and significant. The most prominent clusters identified were: Data Models (#0), Facial Expression Recognition (#1), Adaptation Models (#2). These clusters indicate that extensive research has been conducted in these specific areas.

The complete list of clusters, in order of prominence, includes: Data Models (#0), Facial Expression Recognition (#1), Adaptation Models (#2), Performance (#4), Brain-Computer Interface (#5), Artificial Intelligence (#6), Speech Emotion Recognition (#7), Brain Modeling (#8), Digital Signal Processing (#9), Human-Computer Interaction (#10), Action Recognition (#11), Natural Language Processing (#12), Feature Extraction (#13). The keyword clustering analysis reveals a well-defined and structured field

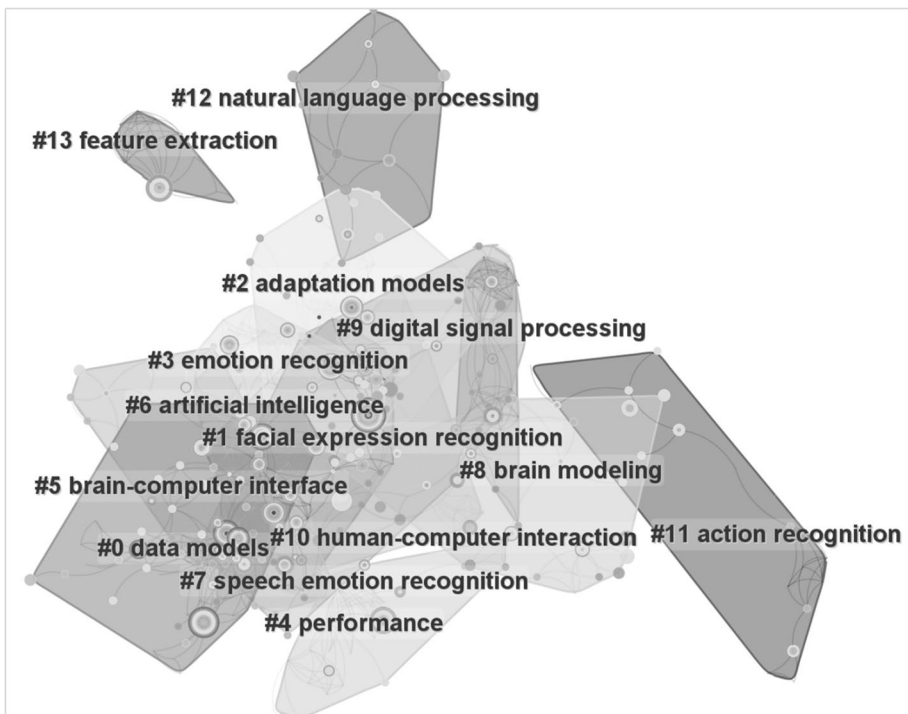


Fig. 12 Network diagram of keyword clustering mapping for HCII domains in the WoS core database

with several key areas of focus. This provides a comprehensive overview of the current state of HCII research, highlighting major themes and pointing to potential avenues for future exploration.

A preliminary analysis indicates that current research in the Human–Computer Intelligent Interaction (HCII) field predominantly lies at the intersection of three disciplines: Human Behavior (Cluster 4), Artificial Intelligence (Cluster 6), Human–Computer Interaction (HCI) (Cluster 10). The primary data source is human behavior, which is integrated with HCI and AI-related knowledge and technology to enhance the effectiveness and user experience of intelligent user interfaces (UI) [123]. The remaining clusters can be manually screened and reorganized into five main groups: Data Processing, Brain–Computer Interface, Multimodal Emotion Recognition, Action Recognition, Language Processing.

- (1) **Data Model Processing: Cluster 0 (Data Models):** This cluster focuses on the concept of data models, which are tools used in database design to abstract real-world scenarios for better data description [124]. Data models are the core foundation of any database system [125]. In the context of HCII, researchers predominantly use visual data, such as video and still images, or rely on sensor data to acquire target data [126]. Wearable sensors and computer vision technology are significant tools for acquiring human data models [127, 128]. Additionally, 3D data models of digitized products often require a depth sensor for data acquisition and processing [129]. **Cluster 2 (Adaptation Models):** This cluster has gained popularity across various fields recently. Adaptation models serve as one of the most effective ways to improve the performance of data models. They are often used for dataset training in webpage recommendation models [130]. By recording user behavior during webpage interactions, a fuzzy inference system is constructed to evaluate and continually optimize user experience [131]. Domain adaptation, when applied to complex and large image tasks with the same training dataset, significantly improves model performance [132, 133]. Within the HCII domain, the AI methods predominantly used are Deep Learning (DL) models, Convolutional Neural Networks (CNN), Long Short-Term Memory networks (LSTM), and Support Vector Machines (SVM). Researchers employ various network models to acquire datasets, which are then compared and analyzed experimentally. Based on these experiments, suitable models are selected, and optimizations are proposed to improve the precision or robustness of experimental results. In summary, the HCII field is a multidisciplinary area that has been making significant strides in integrating human behavior analytics, artificial intelligence, and HCI to create more intelligent, effective, and user-friendly interfaces.
- (2) **Brain–Computer Interface: Cluster 5 (Brain–Computer Interface):** This cluster is an emerging hot topic that bridges the HCII and biomedical fields. It primarily relies on EEG signals to control external devices, incorporating signal processing, classification techniques, and control theories to manage complex environments or execute fine motions [134]. Applications often target individuals with paralysis and include Deep Learning (DL) technology-based assistive rehabilitation devices with IoT modules [61, 135], as well as assistive mobility devices for disabled patients [136]. The technology's success has also fueled the growth of the neurogaming industry. In this context, EEG signals form datasets, Support Vector Machines (SVMs) and Long Short-Term Memory networks (LSTMs) train learning algorithms for in-game decision-making [27], and sensor devices host the game engine for user testing. Research in brain–computer interaction games primarily targets the enhancement of human–computer interaction

and user experience and is mainly applied to board games like mahjong [137] and backgammon [75].

- (3) **Multimodal Emotion Recognition: Cluster 3 (Emotion Recognition):** This cluster is a key research area in HCII. Emotion detection enables more natural computer interaction and the design of human-centered user interfaces and systems. The cluster encompasses Cluster 1 (Facial Expression Recognition) and Cluster 7 (Speech Emotion Recognition), both of which are hot topics. Facial expression recognition is particularly robust; it allows for direct assessment of user emotions. Researchers have enhanced the accuracy of such recognition by refining convolutional neural network (CNN) approaches [88, 89]. Speech emotion recognition, on the other hand, focuses on extracting emotional cues from discourse [76] and has shown improved classification capabilities with one-dimensional CNNs (1D CNNs). These technologies underscore the importance of speech emotion recognition in intelligent interaction and illustrate that most research in this area employs deep learning techniques to improve data model accuracy [90]. In line with the multimodal nature of emotion recognition models, existing frameworks like CNN and Gated Recurrent Units (GRU) are also being used to extract emotional features from videos, actions, and even brain activity, aiming to enhance multi-dimensional recognition accuracy [87, 138]. Overall, the HCII field is making strides in integrating advanced AI methodologies with human-centered approaches across various applications, from healthcare and rehabilitation to gaming and emotion recognition. These advances not only improve the user experience but also open new avenues for research and application in real-world scenarios.
- (4) **Action Recognition: Cluster 11 (Action Recognition):** This cluster has a broad range of applications, including human–computer interaction, healthcare, intelligent surveillance, autonomous driving, and virtual reality. The research within this cluster mainly targets human pose estimation, human action capture, and hand action recognition. Human pose estimation often begins with refining a dataset that captures various joints and key body parts, which can be applied in various studies [84]. There are a growing number of studies focusing on human motion capture and recognition, largely based on deep learning techniques. For example, local error convolutional neural network models [85] and hybrid deep learning-based activity recognition models like QWSA-HDLAR [86] offer significant insights into action recognition.
- (5) **Language Processing: Cluster 12 (Natural Language Processing):** NLP technologies have been successfully integrated into the HCII field, especially in dialogue systems across various sectors. These include human-like dialogue systems for industrial robots [22, 23], entertainment chatbots for the elderly to promote digital inclusion [32, 139], smart home systems with voice commands [12], and educational systems [24]. Additionally, there is a wealth of research in text mining and sentiment analysis. For example, semantic networks are being used for humor recognition in text [140], evaluations in VR games [28], and assessments of learning forum discussions [25]. Image caption generation is another growing area within this cluster [141].

The research in the HCII field is increasingly becoming interdisciplinary, merging advanced computational methods with a deep understanding of human behavior and needs. Whether it is through the lens of action recognition for healthcare and surveillance or natural language processing for more interactive and intuitive user interfaces, the advancements in HCII are setting the stage for a future where technology is not just a tool but an extension of human capability.

4.4.3 Keyword timeline analysis

The keyword timeline serves as a visual tool to track the development, duration, and interconnections among the hotspots in HCII research. By plotting the 14 clusters in a timeline view, as demonstrated in Fig. 13, several key insights emerge: **Interconnected Research:** The clusters within the HCII field are interconnected, reflecting an increasing trend in interdisciplinary studies. **Longevity and Impact:** Clusters such as #0 data models, #4 performance, and #7 speech emotion recognition have been around for a long time and have remained persistent in the field. For instance, data models have been pivotal since the early stages of HCII research, indicating their foundational role. **Emerging Focus Areas:** In recent years, areas like facial expression recognition, brain-computer interface, speech emotion recognition, brain modelling, and natural language processing have gained prominence. While clusters like #1 (facial expression recognition), #5 (brain-computer interface), and #8 (brain modelling) are relatively new, they show strong connections with other clusters, suggesting that they are extensions or innovations based on foundational research and current needs. **Leading Clusters:** The keywords 'sentiment recognition,' 'model,' 'design,' and 'user interface' frequently appear across the timeline, especially in cluster #3. This indicates the leading role and extensive, high-quality research value of cluster #3 in the HCII field. **Bridging Research:** Digital signal processing (cluster #9) may have appeared later in the timeline but shows an extended duration and links with other clusters, serving as a bridge in the overall research landscape. These insights not only offer a historical perspective but also help to identify enduring and emerging areas of interest, thereby providing a comprehensive understanding of the field's evolution. It is clear that as HCII research matures, it is becoming increasingly interdisciplinary, integrating more closely with emerging technologies and user needs.

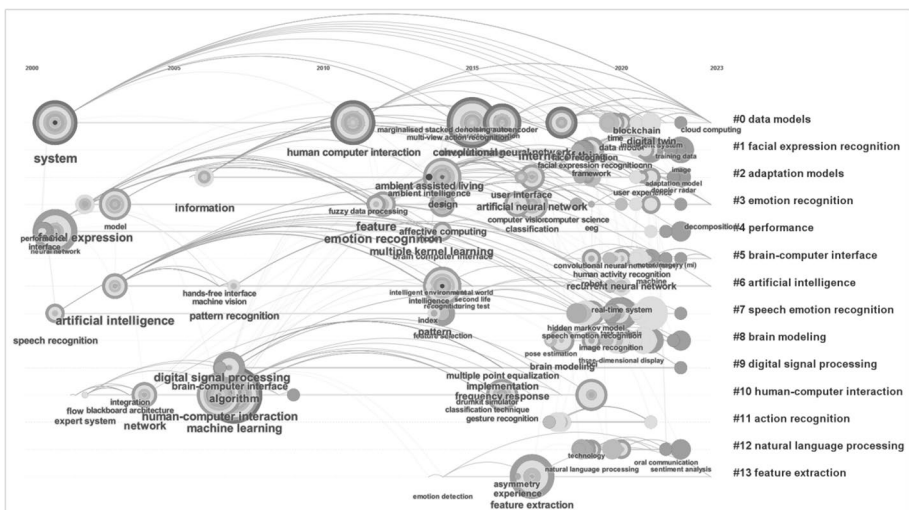


Fig. 13 WoS core database HCII domain keyword clustering timeline

5 Analysis and discussion of results

5.1 HCII application area hotspot time zone map

The time-zone diagram, depicted in Fig. 14, focuses on the four most prevalent application areas in HCII research: industrial manufacturing, healthcare, smart homes, and vehicles & transportation. The diagram serves multiple purposes: Identifying Common Hotspots: Keywords like "computational modelling," "programming," and "model" frequently appear across these sectors, signaling these as common areas of research focus. Unveiling Technical Methods: "ML/DL" (Machine Learning/Deep Learning) emerge as hot keywords, indicating that the future of HCII is increasingly intertwined with advancements in artificial intelligence. Computational modelling stands out as the most commonly used technique. Emotion Computing and Recognition: This has become a significant area of interest, especially in healthcare, smart homes, and industrial settings. The potential for emotion recognition technology to be integrated into a broader range of applications is high, aiming for increased accuracy and enhanced user experience. The Rise of IoT: In recent years, the Internet of Things (IoT) has gained significant attention, particularly in vehicles, transportation, and industrial applications. IoT technology combined with HCII can revolutionize various sectors: In Transportation: Smart parking, smart navigation, and traffic flow supervision are a few areas where this amalgamation can be beneficial. In Industry: Applications extend to smart warehousing, smart supply chains, and smart robotic arms, among others. In Smart Homes: Home appliance interconnection, home health, and security can be improved through IoT and HCII integration. The time-zone diagram analysis reveals that

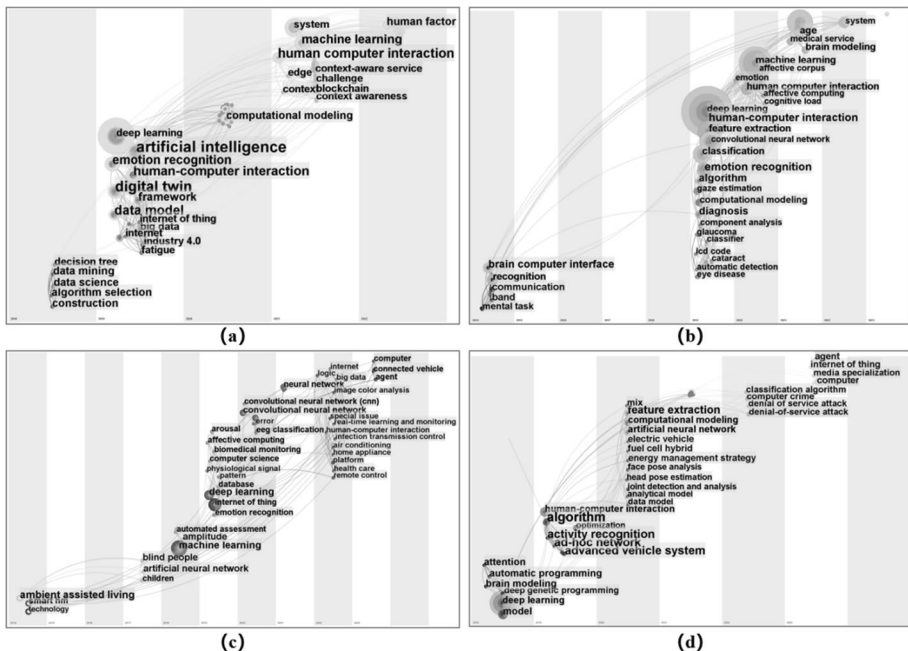


Fig. 14 Industrial manufacturing (a), healthcare (b), smart home (c), vehicles and transportation (d) with HCII domain keyword time zone map

HCII is moving towards a future deeply integrated with emerging technologies like artificial intelligence and IoT. Emotion recognition, once limited to specific contexts, is finding broader applications and is expected to evolve further. These developments present both opportunities and challenges for HCII research. On one hand, there is enormous potential for breakthroughs in improving human–computer interactions in various sectors. On the other, these advancements also demand a multidisciplinary approach to address complex challenges in data accuracy, user experience, and system integration.

In alignment with the Industry 4.0 paradigm, HCII in this sector is rapidly incorporating digital twin and IoT technologies. Decision tree algorithms, a classical AI technique, are also being applied to optimize manufacturing processes and decision-making. HCII is increasingly prevalent in healthcare, particularly in psychological diagnosis, psychotherapy, and emotional well-being. Brain-computer interface technology is being integrated into HCII and is expected to find applications in rehabilitation therapy, neuro-gaming, and telemedicine. HCII research in smart homes spans a broad and complex range, from foundational studies to practical applications, aimed at making homes more interactive, secure, and user-friendly. In this sector, the focus is primarily on human body recognition technologies, like head and facial gesture and action recognition. These technologies aim to improve the robustness and accuracy of HCII systems, leading to more intelligent and natural advanced vehicle systems. This also becomes a challenge for future HCII.

Archaeological site detection [142] HCII can facilitate the use of AI systems to assist archaeologists in identifying potential sites in aerial or satellite images. Semantic segmentation models can be used to draw accurate maps, enhancing operational efficiency.

Cultural heritage [143] Digital collections and 3D visualizations are becoming new user requirements. An intelligent framework, built on AI technology, can allow museum visitors to interact virtually with cultural artifacts.

Aerospace [144] HCII technology could be extremely useful in hazardous flying missions in complex high-altitude environments. AI technology can simulate spatial orientation obstacles and provide enhanced navigation functions. The applications of HCII are not limited to the above-mentioned fields. Future research could extend into space and deep-sea explorations, indicating the limitless potential and versatile applications of HCII. The field promises to continuously evolve, integrating with other emerging technologies to address increasingly complex human–computer interaction challenges across a myriad of scenarios.

5.2 Research hot spots and trends in HCII

Based on our analysis, the three main centers of contemporary research in Human–Computer Intelligent Interaction (HCII) are as follows: Multimodal and Emotion Recognition Technology: The future of HCII lies in developing more diverse databases and achieving higher accuracy and better user experiences. The objective is to make human–computer dialogues and collaborations more natural and akin to "human-to-human" interactions. New Interaction Media and Scenarios: Innovations in interaction mediums like VR glasses and smart medical care are on the horizon. These mediums will likely make interactions more efficient and versatile. Advancements in AI Algorithms: Natural language processing

has already become mainstream in AI, and the continuous improvements in deep learning algorithms are making computerized natural language more robust, accurate, and diverse.

Brain-computer interface development: The integration of neurology and HCII is poised to allow physically handicapped patients to interact with external devices through EEG signals. This would significantly enhance their quality of life and offer applications like multi-modal neuro-interactive games and intelligent rehabilitation devices. **Intelligent Human–Computer Interaction Systems:** Future systems will have more complex computing capacities. **Intelligent User Interfaces (UIs)** will offer a variety of sensory options and may even become de-physicalized, offering more natural and intuitive user experiences. **Emerging Applications and Industries:** HCII systems are expected to find applications in various emerging scenarios and industries. These could range from intelligent customer service in the financial sector, smart shopping carts in retail, intelligent control of urban facilities, to more realistic and engaging virtual reality interactions in the gaming industry. In summary, the field of HCII is rapidly evolving, with promising avenues for research and practical applications. The integration of advanced algorithms, new interaction mediums, and multi-disciplinary approaches are expected to revolutionize how humans interact with computers, thereby impacting various sectors from healthcare to industrial manufacturing and beyond.

6 Conclusion

This paper offers a comprehensive analysis of the evolution of Human–Computer Intelligent Interaction (HCII), elucidating its core concepts, definitions, and research hotspots. It affirms that HCII is a cross-discipline that synergizes Human–Computer Interaction (HCI) and Artificial Intelligence (AI) technologies. This synergy not only amplifies the capabilities of HCI through AI but also guides the development of AI technology itself, thereby producing significant research results across various application scenarios.

There are our main findings: **Interaction Modes:** The paper identifies and analyzes the five primary interaction modes in HCII—voice interaction, visual interaction, gesture interaction, multimodal interaction, and emotional interaction—contrasting them with traditional HCI to offer a roadmap for future technological advancements in HCII. **Research Hotspots:** Utilizing CiteSpace for bibliometric analysis, the paper pinpoints the key research areas in HCII over the past five years, including data processing, brain-computer interfaces, multimodal emotion recognition, action recognition, and language processing. **Future Trends:** The paper predicts broader application scenarios and greater research value for HCII in the future, particularly in the realms of multimodal interaction and emotion recognition. The development of more accurate algorithms and models, complex databases, and intelligent user interfaces is imminent. **Innovations like neural interaction games, multi-sensory interaction, and de-physicalized interfaces are on the horizon.** **Complex and Natural Interactions:** The use of advanced algorithms will likely make HCII systems more intuitive, allowing for more complex and natural interactions between humans and computers. **Multimodal Emotion Recognition:** Advances in this area will enable more nuanced and effective human–computer interactions, further narrowing the gap between human-to-human and human-to-computer communication. **Societal Contributions:** As HCII continues to evolve, it is set to offer more intelligent, intuitive, and efficient methods of interaction, thereby contributing significantly to the advancement of various sectors of human society. In summary, HCII is at a pivotal point in its development, with the potential for transformative impact on how humans interact with technology. As it continues to integrate

advancements from both HCI and AI, HCII is poised to revolutionize various sectors, from healthcare to industrial manufacturing, and beyond.

Funding This research is funded by the State Key Laboratory of Mechanical Systems and Vibrations of China (Grant No.MSV202014); the Fundamental Research Funds for the Central Universities of China (HUST: 2020kfyXJJS014); the Major Philosophy and Social Science Research Project for Hubei Province Colleges and Universities of China (Grant No.20ZD002).

Data availability My manuscript has data included as electronic supplementary material.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

References

1. Hirschberg K, Manning CD (2023) Advances in natural language processing. *Science* 349:261–266
2. Hussien RM, Al-Jubouri KQ, Gburi MA et al (1973) (2021) computer vision and image processing the challenges and opportunities for new technologies approach: A paper review. *J Phys: Conf Ser* 1:012002
3. Pantic M, Pentland A, Nijholt A, Huang TS (2007) Human computing and machine understanding of human behavior: a survey. In: Huang TS, Nijholt A, Pantic M, Pentland A (eds) *Artificial intelligence for human computing. Lecture notes in computer science*, vol 4451. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-72348-6_3
4. Lv Z, Poiesi F, Deep DQ et al (2022) Learning for intelligent human-computer interaction. *Appl Sci* 12:11457
5. Wang G, Li L, Xing S, Ding H (2018) Intelligent HMI in orthopedic navigation. *Adv Experiment Med Biol* 1093:207–224
6. Miao Y, Jiang Y, Peng L et al (2018) Telesurgery robot based on 5G tactile internet. *Mobile Netw Appl* 23:1645–1654
7. Li R, Fu H, Lo W, Chi Z, Song Z, Wen D (2019) Skeleton-based action recognition with key-segment descriptor and temporal step matrix model. *IEEE Access* 7:169782–169795. <https://doi.org/10.1109/ACCESS.2019.2954744>
8. Li P, Hou X, Duan X, Yip HM, Song G, Liu Y (2019) Appearance-based gaze estimator for natural interaction control of surgical robots. *IEEE Access* 7:25095–25110. <https://doi.org/10.1109/ACCESS.2019.2900424>
9. Li P, Hou X, Wei L, Song G, Duan X (2018) Efficient and low-cost deep-learning based gaze estimator for surgical robot control. In: 2018 IEEE international conference on real-time computing and robotics (RCAR). IEEE, Kandima, Maldives, pp 58–63. <https://doi.org/10.1109/RCAR.2018.8621810>
10. Senle Z, Rencheng S, Juan C, Yunfei Z, Xun C (2019) A feasibility study of a video-based heart rate estimation method with convolutional neural networks. In: 2019 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), pp 1–5. <https://doi.org/10.1109/CIVEMSA45640.2019.9071634>
11. Singh A, Kabra R, Kumar R, Lokanath MB, Gupta R, Shekhar SK (2021) On-device system for device directed speech detection for improving human computer interaction. *IEEE Access* 9:131758–131766. <https://doi.org/10.1109/ACCESS.2021.3114371>
12. Ioanna M, Alexandris C (2009) Verb processing in spoken commands for household security and appliances. In: Constantine S (ed) *Universal access in human-computer interaction. Intelligent and ubiquitous interaction environments*. Springer, Berlin Heidelberg, pp 92–99. https://doi.org/10.1007/978-3-642-02710-9_11
13. Hu W, Xiang L, Zehua L (2022) Research on auditory performance of vehicle voice interaction in different sound index. In: Kurosu M (ed) *Human-computer interaction. User experience and behavior*. Springer International Publishing, Cham, pp 61–69. https://doi.org/10.1007/978-3-031-05412-9_5

14. Waldron SM, Patrick J, Duggan GB, Banbury S, Howes A (2008) Designing information fusion for the encoding of visual–spatial information. *Ergonomics* 51(6):775–797. <https://doi.org/10.1080/00140130701811933>
15. Li Z, Li X, Zhang J et al (2021) Research on interactive experience design of peripheral visual interface of autonomous vehicle. In: Kurosu M (eds) *Human-computer interaction. Design and user experience case studies. HCII 2021. Lecture notes in computer science*, 12764. Springer
16. Hazoor A, Terrafino A, Di Stasi LL et al (2022) How to take speed decisions consistent with the available sight distance using an intelligent speed adaptation system. *Accid Anal Prev* Sep174:106758
17. Lee KM, Moon Y, Park I, Lee J-g (2023) Voice orientation of conversational interfaces in vehicles. *Behav Inf Technol* 1–12. <https://doi.org/10.1080/0144929X.2023.2166870>
18. Shanthi N, Sathishkumar VE, Upendra Babu K, Karthikeyan P, Rajendran S, Allayear SM (2022) Analysis on the bus arrival time prediction model for human-centric services using data mining techniques. *Comput Intell Neurosci* 2022:7094654. <https://doi.org/10.1155/2022/7094654>
19. Kim H, Kim W, Kim J et al (2022) Study on the take-over performance of level 3 autonomous vehicles based on subjective driving tendency questionnaires and machine learning methods. *ETRI J* 45:75–92
20. Zhu Z, Ye A, Wen F, Dong X, Yuan K, Zou W (2010) Visual servo control of intelligent wheelchair mounted robotic arm. In: 2010 8th World Congress on Intelligent Control and Automation, pp 6506–6511. <https://doi.org/10.1109/WCICA.2010.5554200>
21. Chen L, Haiwei Y, Liu P (2019) Intelligent robot arm: Vision-based dynamic measurement system for industrial applications. In: Haibin Y, Jinguo L, Liu Lianqing J, Zhaojie LY, Dalin Z (eds) *Intelligent robotics and applications*. Springer International Publishing, Cham, pp 120–130. https://doi.org/10.1007/978-3-030-27541-9_11
22. Chen L, Xiaochun Z, Dimitrios C, Hongji Y (2022) ToD4IR: A humanised task-oriented dialogue system for industrial robots. *IEEE Access* 10:91631–91649. <https://doi.org/10.1109/ACCESS.2022.3202554>
23. Chen L, Jinha P, Hahyeon K, Dimitrios C (2021) How can I help you? In: *An intelligent virtual assistant for industrial robots*. Association for Computing Machinery, New York. <https://doi.org/10.1145/3434074.3447163>
24. Wojciech K, Maciej M, Zurada Jacek M (2010) Intelligent E-learning systems for evaluation of user's knowledge and skills with efficient information processing. In: Rutkowski L, Rafa S, Tadeusiewicz R, Zadeh Lotfi A, Zurada Jacek M (eds) *Artificial intelligence and soft computing*. Springer, Berlin Heidelberg, pp 508–515. https://doi.org/10.1007/978-3-642-13232-2_62
25. Cheong Michelle LF, Chen Jean Y-C, Tian DB (2019) An intelligent platform with automatic assessment and engagement features for active online discussions. In: Wotawa F, Friedrich G, Pill I, Koitz-Hristov R, Ali M (eds) *Advances and trends in artificial intelligence. From theory to practice*. Springer International Publishing, Cham, pp 730–743. https://doi.org/10.1007/978-3-030-22999-3_62
26. Choi Y, Jeon H, Lee S et al (2022) Seamless-walk: Novel natural virtual reality locomotion method with a high-resolution tactile sensor. 2022 IEEE conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp 696–697
27. Amin M, Tubaishat A, Al-Obeidat F, Shah B, Karamat M (2022) Leveraging brain–computer interface for implementation of a bio-sensor controlled game for attention deficit people. *Comput Electr Eng* 102:108277. <https://doi.org/10.1016/j.compeleceng.2022.108277>
28. Gao Y, Anqi C, Susan C, Guangtao Z, Aimin H (2022) Analysis of emotional tendency and syntactic properties of VR game reviews. In: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp 648–649. <https://doi.org/10.1109/VRW55335.2022.00175>
29. Hu Z, Andreas B, Li S, Wang G (2021) FixationNet: Forecasting eye fixations in task-oriented virtual environments. *IEEE Trans Vis Comput Graph* 27(5):2681–2690. <https://doi.org/10.1109/TVCG.2021.3067779>
30. Kim J (2020) VIVR: Presence of Immersive Interaction for Visual Impairment Virtual Reality. *IEEE Access*, pp 196151–196159
31. Krepki R, Blankertz B, Curio G et al (2007) The Berlin Brain-Computer Interface (BCI) – towards a new communication channel for online control in gaming applications. *Multimed Tools Appl* 33:73–90
32. García-Méndez S, Arriba-Pérez D, Francisco G-C, Francisco J, Regueiro-Janeiro JA, Gil-Castiñeira F (2021) Entertainment Chatbot for the digital inclusion of elderly people without abstraction capabilities. *IEEE Access* 9:75878–75891. <https://doi.org/10.1109/ACCESS.2021.3080837>
33. Lee W, Son G (2023) Investigation of human state classification via EEG signals elicited by emotional audio-visual stimulation. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-023-16294-w>

34. Razzaq MA, Hussain J, Bang J (2023) Hybrid multimodal emotion recognition framework for UX evaluation using generalized mixture functions. *Sensors* 23:4373
35. Jo AH, Kwak KC (2023) Speech emotion recognition based on two-stream deep learning model using korean audio information. *Appl Sci* 13:2167
36. Aleisa HN, Alrowais FM, Negm N et al (2023) Henry gas solubility optimization with deep learning based facial emotion recognition for human computer Interface. *IEEE Access* 11:62233–62241
37. Gagliardi G, Alfeo AL, Cattrambone V, Diego C-R, Cimino Mario GCA, Valenza G (2023) Improving emotion recognition systems by exploiting the spatial information of EEG sensors. *IEEE Access* 11:39544–39554. <https://doi.org/10.1109/ACCESS.2023.3268233>
38. Eswaran KCA, Akshat P, Gayathri M (2023) Hand gesture recognition for human-computer interaction using computer vision. In: Kottursamy K, Bashir AK, Kose U, Annie U (eds) *Deep Sciences for Computing and Communications*. Springer Nature Switzerland, Cham, pp 77–90
39. Ansar H, Mudawi NA, Alotaibi SS et al (2023) Hand gesture recognition for characters understanding using convex Hull landmarks and geometric features. *IEEE Access* 11:82065–82078
40. Kothadiya DR, Bhatt CM, Rehman A, Alamri FS, Tanzila S (2023) SignExplainer: An explainable ai-enabled framework for sign language recognition with ensemble learning. *IEEE Access* 11:47410–47419. <https://doi.org/10.1109/ACCESS.2023.3274851>
41. Salman SA, Zakir A, Takahashi H (2023) Cascaded deep graphical convolutional neural network for 2D hand pose estimation. In: Salman SA, Zakir A, Takahashi H (eds) *Other conferences*. <https://api.semanticscholar.org/CorpusID:257799908>
42. Lyu Y, An P, Xiao Y, Zhang Z, Zhang H, Katsuragawa K, Zhao J (2023) Eggly: Designing mobile augmented reality neurofeedback training games for children with autism spectrum disorder. *Assoc Comput Machin* 7(2):1–29. <https://doi.org/10.1145/3596251>
43. Van Mechelen M, Smith RC, Schaper M-M, Tamashiro M, Bilstrup K-E, Lunding M, Petersen MG, Iversen OS (2023) Emerging technologies in K–12 education: A future HCI research agenda. *Assoc Comput Machine* 30(3):1073–0516. <https://doi.org/10.1145/3569897>
44. Ometto M (2022) An innovative approach to plant and process supervision. *Danieli Intelligent Plant. IFAC-PapersOnLine* 55(40):313–318. <https://doi.org/10.1016/j.ifacol.2023.01.091>
45. Matheus N, Joaquim J, João V, Regis K, Anderson M (2023) Exploring affordances for AR in laparoscopy. In: 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp 147–151. <https://doi.org/10.1109/VRW58643.2023.00037>
46. Rawat KS, Sood SK (2021) Knowledge mapping of computer applications in education using CiteSpace. *Comput Appl Eng Educ* 29:1324–1339
47. Grigsby Scott S (2018) Artificial intelligence for advanced human-machine symbiosis. In: Schmorror DD, Fidopiastis CM (eds) *Augmented cognition: intelligent technologies*. Springer International Publishing, Cham, pp 255–266. https://doi.org/10.1007/978-3-319-91470-1_22
48. Gomes CC, Preto S (2018) Artificial intelligence and interaction design for a positive emotional user experience. In: Karwowski W, Ahram T(eds) *Intelligent Human Systems Integration*. IHSI 2018. *Advances in intelligent systems and computing*, Springer
49. Zhang C, Lu Y (2021) Study on artificial intelligence: The state of the art and future prospects[J]. *J Ind Inf Integr* 2021(23-):23
50. Ahamed MM (2017) Analysis of human machine interaction design perspective-a comprehensive literature review. *Int J Contemp Comput Res* 1(1):31–42
51. Li X (2020) Human-robot interaction based on gesture and movement recognition. *Signal Process: Image Commun* 81:115686
52. Majaranta P, Riih a K-J, Aulikki H, Špakov O (2019) Eye movements and human-computer interaction. In: Klein C, Ettinger U (eds) *Eye movement research: an introduction to its scientific foundations and applications*. Springer International Publishing, Cham, pp 971–1015
53. Bi L, Pan C, Li J, Zhou J, Wang X, Cao S (2023) Discourse-based psychological intervention alleviates perioperative anxiety in patients with adolescent idiopathic scoliosis in China: A retrospective propensity score matching analysis. *BMC Musculoskelet Disord* 24(1):422. <https://doi.org/10.1186/s12891-023-06438-2>
54. Maybury M (1998) Intelligent user interfaces: An introduction. In: *Proceedings of the 4th international conference on intelligent user interfaces*. Association for Computing Machinery, New York, NY, pp 3–4. <https://doi.org/10.1145/291080.291081>
55. Jaimes A, Sebe N (2005) Multimodal human computer interaction: A survey. *Lect Notes Comput Sci* 3766:1
56. Qiu Y (2004) Evolution and trends of intelligent user interfaces. *Comput Sci*. <https://api.semanticscholar.org/CorpusID:63801008>

57. Zhao Y, Wen Z (2022) Interaction design system for artificial intelligence user interfaces based on UML extension mechanisms. IOS Press 2022. <https://doi.org/10.1155/2022/3534167>
58. Margiené A, Simona R (2019) Trends and challenges of multimodal user interfaces. In: 2019 Open Conference of Electrical, Electronic and Information Sciences (eStream), pp 1–5. <https://doi.org/10.1109/eStream.2019.8732156>
59. Maybury MT (1998) Intelligent user interfaces: an introduction. In: International conference on intelligent user interfaces. <https://api.semanticscholar.org/CorpusID:12602078>
60. Stavros M, Nikolaos B (2016) A survey on human machine dialogue systems. In: 2016 7th International conference on information, intelligence, systems & applications (IISA), pp 1–7. <https://doi.org/10.1109/IISA.2016.7785371>
61. Vinoj PG, Jacob S, Menon VG, Balasubramanian V, Piran J (2021) IoT-powered deep learning brain network for assisting quadriplegic people. *Comput Electr Eng* 92:107113. <https://doi.org/10.1016/j.compeleceng.2021.107113>
62. Peruzzini M, Grandi F, Pellicciari M (2017) Benchmarking of tools for user experience analysis in industry 4.0. *Procedia Manuf* 11:806–813
63. Mahesh B (2020) Machine learning algorithms-a review. *Int J Sci Res (IJSR)* 9:381–386
64. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444
65. Cantoni V, Cellario M, Porta M (2004) Perspectives and challenges in e-learning: towards natural interaction paradigms. *J Vis Lang Comput* 15(5):333–345
66. Changhoon O, Jungwoo S, Jinhan C et al (2018) You help but only with enough details: Understanding user experience of co-creation with artificial intelligence. *Assoc Comput Mach Pap* 649:1–13
67. Turk M (2005) multimodal human-computer interaction. In: Kisačanin B, Pavlović V, Huang TS (eds) Real-time vision for human-computer interaction. Springer, Boston, MA, pp 269–283. https://doi.org/10.1007/0-387-27890-7_16
68. Singh SK, Chaturvedi A (2022) A reliable and efficient machine learning pipeline for American sign language gesture recognition using EMG sensors. *Multimed Tools Appl* 82(15):23833–23871. <https://doi.org/10.1007/s11042-022-14117-y>
69. Zhang J, Qiu X, Li X, Huang Z, Mingqiu W, Yumin D, Daniele B (2021) Support vector machine weather prediction technology based on the improved quantum optimization algorithm, vol 2021, Hindawi Limited, London, GBR. <https://doi.org/10.1155/2021/6653659>
70. Giatsintov A, Kirill M, Pavel B (2023) Architecture of the graphics system for embedded real-time operating systems. *Tsinghua Sci Technol* 28(3):541–551. <https://doi.org/10.26599/TST.2022.9010028>
71. Lee S, Jeeyun O, Moon W-K (2022) Adopting voice assistants in online shopping: examining the role of social presence, performance risk, and machine heuristic. *Int J Hum-Comput Int* 39:2978–2992. <https://api.semanticscholar.org/CorpusID:250127863>
72. Johannes P (2005) Spoken dialogue technology: toward the conversational user interface by Michael F. McTear. *Comput Linguist* 31(3):403–416. <https://doi.org/10.1162/089120105774321136>
73. Du Y, Qin J, Zhang S, et al (2018) Voice user interface interaction design research based on user mental model in autonomous vehicle. In: Kurosu M (eds) Human-computer interaction. Interaction technologies. HCI 2018. Lecture notes in computer science. Springer
74. Koni YJ, Al-Absi MA, Saparmammedovich SA, Jae LH (2020) AI-based voice assistants technology comparison in term of conversational and response time. Springer-Verlag, Berlin, Heidelberg, pp 370–379. https://doi.org/10.1007/978-3-030-68452-5_39
75. Li M, Li F, Pan J et al (2021) The MindGomoku: An online P300 BCI game based on bayesian deep learning. *Sensors* 21:1613
76. Alnuaim AA, Mohammed Z, Aseel A, Chitra S, Atef HW, Hussam T, Kumar SP, Rajnish R, Vijay K (2022) Human-computer interaction with detection of speaker emotions using convolution neural networks, vol 2022. Hindawi Limited, London, GBR. <https://doi.org/10.1155/2022/7463091>
77. Charissis V, Falah J, Lagoo R et al (2021) Employing emerging technologies to develop and evaluate in-vehicle intelligent systems for driver support: Infotainment AR HUD case study. *Appl Sci* 11:1397
78. Fan Y, Yang J, Chen J, et al (2021) A digital-twin visualized architecture for flexible manufacturing system. *J Manuf Syst* 2021(60-):60
79. Wang T, Li J, Kong Z, Liu X, Snoussi H, Lv H (2021) Digital twin improved via visual question answering for vision-language interactive mode in human-machine collaboration. *J Manuf Syst* 58:261–269. <https://doi.org/10.1016/j.jmsy.2020.07.011>
80. Tian W, Jiakun L, Yingjun D et al (2021) Digital twin for human-machine interaction with convolutional neural network. *Int J Comput Integr Manuf* 34(7–8):888–897
81. Zhang Q, Wei Y, Liu Z et al (2023) A framework for service-oriented digital twin systems for discrete workshops and its practical case study. *Systems* 11:156

82. El OI, Benouini R, Zenkouar K et al (2022) RGB-D feature extraction method for hand gesture recognition based on a new fast and accurate multi-channel cartesian Jacobi moment invariants. *Multimed Tools Appl* 81:12725–12757
83. Miah ASM, Shin J, Hasan MAM et al (2022) BenSignNet: Bengali sign language alphabet recognition using concatenated segmentation and convolutional neural network. *Appl Sci* 12:3933
84. Munea TL, Jembre YZ, Weldegebriel HT et al (2020) The progress of human pose estimation: A survey and taxonomy of models applied in 2D human pose estimation. *IEEE Access*, pp 133330–133348
85. Xia H, Lei H, Yang J, Rahim K (2022) Human behavior recognition in outdoor sports based on the local error model and convolutional neural network, vol 2022. Hindawi Limited, London, GBR, pp 1687–5265. <https://doi.org/10.1155/2022/6988525>
86. Malibari AA, Alzahrani JS, Qahmash A (2022) Quantum water strider algorithm with hybrid-deep-learning-based activity recognition for human-computer interaction. *Appl Sci* 12:6848
87. Jia N, Zheng C, Sun W (2022) A multimodal emotion recognition model integrating speech, video and MoCAP. *Multimed Tools Appl* 81:32265–32286
88. Zhou Y, Feng Y, Zeng S, Pan B (2019) Facial expression recognition based on convolutional neural network. In: 2019 IEEE 10th International conference on software engineering and service science (ICSESS). IEEE, pp 410–413. <https://doi.org/10.1109/ICSESS47205.2019.9040730>
89. Yan T, Zhang Xingming H, Xiping WS, Haoxiang W (2021) Facial expression recognition using frequency neural network. *IEEE Tran Image Process* 30:444–457. <https://doi.org/10.1109/TIP.2020.3037467>
90. Bhavani R, Vijay MT, Kumar TR, Jonnadula N, Murali K, Harpreet K (2022) Deep learning techniques for speech emotion recognition. In: 2022 International conference on futuristic technologies (INCOFT), pp 1–5. <https://doi.org/10.1109/INCOFT55651.2022.10094534>
91. Lee CH, Yang HC, Su XQ (2022) A multimodal affective sensing model for constructing a personality-based financial advisor system. *Appl Sci* 12:10066
92. Haijuan D, Minglong L, Gengxin S (2022) Personalized smart clothing design based on multimodal visual data detection, vol 2022. Hindawi Limited, London, GBR. <https://doi.org/10.1155/2022/4440652>
93. Raptis GE, Kavvetsos G, Katsini C (2021) MuMIA: Multimodal interactions to better understand art contexts. *Appl Sci* 11:2695
94. Kumar PS, Singh SH, Shalendar B, Ravi J, Prasanna SRM (2022) Alzheimer’s dementia recognition using multimodal fusion of speech and text embeddings. In: Kim J-H, Madhusudan S, Javed K, Shanker TU, Marigankar S, Dhananjay S (eds) *Intelligent human computer interaction*. Springer International Publishing, Cham, pp 718–728. https://doi.org/10.1007/978-3-030-98404-5_64
95. Šumak B, Brdnik S, Pusnik M (2021) Sensors and artificial intelligence methods and algorithms for human–computer intelligent interaction: A systematic mapping study. *Sensors* 22. <https://api.semanticscholar.org/CorpusID:245441300>
96. Karpov AA, Yusupov RM (2018) Multimodal interfaces of human-computer interaction. *Her Russ Acad Sci* 88:67–74
97. Mobeen N, Muhammad Mansoor A, Eiad Y, Mazliham Mohd S (2021) A systematic review of human–computer interaction and explainable artificial intelligence in healthcare with artificial intelligence techniques. *IEEE Access* 9:153316–153348. <https://doi.org/10.1109/ACCESS.2021.3127881>
98. Diederich S, Brendel AB, Morana S et al (2022) (2022) on the design of and interaction with conversational agents: An organizing and assessing review of human-computer interaction research. *J Assoc Inf Syst* 1:23
99. Sheetal K, Shruti P, Jyoti C, Ketan K, Sashikala M, Ajith A (2022) AI-based conversational agents: a scoping review from technologies to future directions. *IEEE Access* 10:92337–92356. <https://doi.org/10.1109/ACCESS.2022.3201144>
100. Ren F, Bao Y (2020) A review on human-computer interaction and intelligent robots. *Int J Inf Technol Decis Mak* 19:5–47. <https://api.semanticscholar.org/CorpusID:213516319>
101. Vail EF III (1999) Knowledge mapping: getting started with knowledge management. *Inf Syst Manag* 16(4):16–23. <https://doi.org/10.1201/1078/43189.16.4.19990901/31199.3>
102. Lin S, Shen T, Guo W (2021) Evolution and emerging trends of kansei engineering: A visual analysis based on CiteSpace. *IEEE Access* 9:111181–111202
103. Eck NJ, Waltman L (2010) Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84:523–538
104. Cobo MJ, López-Herrera AG, Herrera-Viedma EE et al (2012) SciMAT: A new science mapping analysis software tool. *J Am Soc Inform Sci Technol* 63(8):1609–1630
105. Chaomei C (2018) visualizing and exploring scientific literature with citespace: an introduction. association for computing machinery. In: CHIR ’18, New York, NY, pp 369–370

106. Wei F, Grubestic TH, Bishop BW (2015) Exploring the gis knowledge domain using citespace. *Prof Geogr* 67:374–384
107. Zhong S, Chen R, Song F et al (2019) Knowledge mapping of carbon footprint research in a LCA perspective: A visual analysis using CiteSpace. *Processes* 7:818
108. Jiayi Y, Hong L (2022) Visualizing the knowledge domain in urban soundscape: A scientometric analysis based on CiteSpace. *Int J Environ Res Public Health* 19(21):13912. <https://www.mdpi.com/1660-4601/19/21/13912>
109. Chen Y, Wang Y, Zhou D (2021) Knowledge map of urban morphology and thermal comfort: A bibliometric analysis based on CiteSpace. *Buildings* 11:427
110. Chen C (2017) Science mapping: A systematic review of the literature. *J Data Inform Sci* 2:1–40
111. Li K, Jin Y, Akram MW et al (2020) Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy. *Vis Comput* 36:391–404
112. Wang J, Xinyu S (2019) Human motion modeling based on single action context. In: 2019 4th International conference on communication and information systems (ICCIS), pp 89–94. <https://doi.org/10.1109/ICCIS49662.2019.00022>
113. Xian Z, Zhan S, Jian G, Yanguo Z, Feng Z (2012) Real-time hand gesture detection and recognition by random forest. In: Maotai Z, Junpin S (eds) *Communications and information processing*. Springer, Berlin Heidelberg, pp 747–755. https://doi.org/10.1007/978-3-642-31968-6_89
114. Yong Z, Dong W, Hu B-G, Qiang J (2018) Weakly-supervised deep convolutional neural network learning for facial action unit intensity estimation. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, pp 2314–2323. <https://doi.org/10.1109/CVPR.2018.00246>
115. Guo Y, Liu W, Wei D, Chen Q (2019) Emotional recognition based on EEG signals comparing long-term and short-term memory with gated recurrent unit using batch normalization. <https://api.semanticscholar.org/CorpusID:181831183>
116. Afsar MM, Saqib S, Ghadi YY, Alsubhany SA, Jalal A, Park J (2022) Body worn sensors for health gaming and e-Learning in virtual reality. *Comput Mater Contin*. <https://api.semanticscholar.org/CorpusID:251164186>
117. Li X, Li Y (2022) Sports training strategies and interactive control methods based on neural network models. *Comput Intell Neurosci* 2022. <https://api.semanticscholar.org/CorpusID:247332344>
118. Ahn J, Nguyen TP, Kim Y-J, Kim T, Yoon J (2022) Automated analysis of three-dimensional CBCT images taken in natural head position that combines facial profile processing and multiple deep-learning models. *Comput Methods Programs Biomed* 226:107123. <https://doi.org/10.1016/j.cmpb.2022.107123>
119. Xi Y, Cho S, Fong S et al (2016) Gesture recognition method using sensing blocks. *Wireless Pers Commun* 91:1779–1797
120. Kaiming H, Xiangyu Z, Shaoqing R, Jian S (2016) Deep residual learning for image recognition. In: 2016 IEEE Conference on computer vision and pattern recognition (CVPR). IEEE, pp 770–778. <https://doi.org/10.1109/CVPR.2016.90>
121. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. *CoRR*:abs/1409.1556. <https://api.semanticscholar.org/CorpusID:14124313>
122. Keselj A, Milicevic M, Zubrinic K et al The application of deep learning for the evaluation of user interfaces. *Sensors* 22:9336
123. Khalil RA, Jones E, Babar MI et al (2019) Speech emotion recognition using deep learning techniques: A Review. *IEEE Access* 7:117327–117345
124. Bai J (2009) Panel data models with interactive fixed effects. *Econometrica* 77(4):1229–1279
125. Wu F (2016) Study on composition and development of the database management system. In: *Proceedings of the 2nd international conference on advances in mechanical engineering and industrial informatics (AMEII 2016)*. Atlantis Press, pp 159–163. <https://doi.org/10.2991/ameii-16.2016.33>
126. Ozcan T, Basturk A (2020) Static facial expression recognition using convolutional neural networks based on transfer learning and hyperparameter optimization. *Multimed Tools Appl* 79:26587–26604
127. Chen W, Yu C, Tu C et al (2020) A survey on hand pose estimation with wearable sensors and computer-vision-based methods. *Sensors* 20(4):1074
128. Lv T, Xiaojuan W, Lei J, Yabo X, Mei S (2020) A hybrid network based on dense connection and weighted feature aggregation for human activity recognition. *IEEE Access* 8:68320–68332. <https://doi.org/10.1109/ACCESS.2020.2986246>
129. Belhi A, Ahmed H, Alfaqheri T, Bouras A, Sadka AH, Fofouf S (2023) An integrated framework for the interaction and 3D visualization of cultural heritage. *Multimed Tools Appl* 1–29. <https://api.semanticscholar.org/CorpusID:255770349>

130. Echeverry-Correa JD, Martínez González B, Hernández R, Córdoba Herralde R de, Ferreiros López J (2014) Dynamic topic-based adaptation of language models: a comparison between different approaches. <https://api.semanticscholar.org/CorpusID:62300998>
131. Souza KES, Seruffo MCR, De Mello Harold D, Da Souza Daniel S, MBR VM, (2019) User experience evaluation using mouse tracking and artificial intelligence. *IEEE Access* 7:96506–96515. <https://doi.org/10.1109/ACCESS.2019.2927860>
132. Shubhajt B, Peter C, Faisal K, Rachel M, Michael S (2021) Learning 3D head pose from synthetic data: A semi-supervised approach. *IEEE Access* 9:37557–37573. <https://doi.org/10.1109/ACCESS.2021.3063884>
133. Yu-Wei C, Soo-Chang P (2022) Domain adaptation for underwater image enhancement via content and style separation. *IEEE Access* 10:90523–90534. <https://doi.org/10.1109/ACCESS.2022.3201555>
134. Padfield N, Camilleri K, Camilleri T (2022) A comprehensive review of endogenous EEG-based BCIs for dynamic device control. *Sensors* 22:5802
135. Jacob S, Mukil A, Menon Varun G, Manoj KB, Jhanjhi NZ, Vasaki P, Shynu PG, Venki B (2020) An adaptive and flexible brain energized full body exoskeleton with IoT edge for assisting the paralyzed patients. *IEEE Access* 8:100721–100731. <https://doi.org/10.1109/ACCESS.2020.2997727>
136. Tang J, Liu Y, Jiang J et al (2019) Toward brain-actuated Mobile platform. *Int J Hum-Comput Interact* 35(10):846–858
137. Xiaodi W, Qi Y, Zhu Xinyun X, Kedi ZJ, Jianmin Z, Yueming W (2022) A brain-controlled mahjong game with artificial intelligence augmentation. In: Lu F, Daniel P, Guangtao Z, Tao M, Ruiping W (eds) *Artificial intelligence*. Springer Nature Switzerland, Cham, pp 548–553. https://doi.org/10.1007/978-3-031-20503-3_47
138. Choi DY, Deok-Hwan K, Cheol SB (2020) Multimodal attention network for continuous-time emotion recognition using video and EEG signals. *IEEE Access* 8:203814–203826. <https://doi.org/10.1109/ACCESS.2020.3036877>
139. Pérez FD, García-Méndez S, González-Castaño FJ et al (2021) Evaluation of abstraction capabilities and detection of discomfort with a newscaster chatbot for entertaining elderly users. *Sensors* 21:5515
140. Xiong S, Wang R, Huang X (2022) Multidimensional latent semantic networks for text humor recognition. *Sensors* 22:5509
141. Wang H, Zhang Y, Yu X (2020) An overview of image caption generation methods. *Comput Intell Neurosci* 2020. <https://api.semanticscholar.org/CorpusID:210956524>
142. Casini L, Marchetti N, Montanucci A et al (2023) A human–AI collaboration workflow for archaeological sites detection. *Sci Rep* 13:8699
143. Narek M, Charles A-D, Alain P, Jean-Marc A, Didier S (2022) Human intelligent machine teaming in single pilot operation: a case study. In: Schmorow DD, Fidopiastis CM (eds) *Augmented cognition*. Springer International Publishing, Cham, pp 348–360. https://doi.org/10.1007/978-3-031-05457-0_27
144. Foucher J, Anne-Claire C, Le GK, Thomas R, Valérie J, Thomas D, Jérémie L, Marielle P-R, François D, Grunwald Arthur J, Jean-Christophe S, Bardy Benoît G (2022) Simulation and classification of spatial disorientation in a flight use-case using vestibular stimulation. *IEEE Access* 10:104242–104269. <https://doi.org/10.1109/ACCESS.2022.3210526>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.