IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 21 March 2025, accepted 27 April 2025, date of publication 5 May 2025, date of current version 13 May 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3566698

SURVEY

Embodied Conversational Agents in Extended Reality: A Systematic Review

FU-CHIA YANG, PEDRO ACEVEDO, SIQI GUO, MINSOO CHOI, AND CHRISTOS MOUSAS, (Member, IEEE)
Department of Computer Graphics Technology, Purdue University, West Lafayette, IN 47907, USA
Corresponding author: Christos Mousas (cmousas@purdue.edu)

ABSTRACT Embodied conversational agents (ECAs) that can interact with users in a human-like manner have demonstrated promising potential in various endeavors. With the ongoing advancement in extended reality (XR) and artificial intelligence (AI), ECAs are becoming increasingly sophisticated. Although previous reviews have predominantly focused on ECAs for non-XR applications, a growing number of research papers are exploring the capabilities of ECAs that utilize XR technologies. However, no prior systematic review has focused explicitly on XR ECAs, leading to a gap in understanding how ECAs are designed, implemented, and evaluated within immersive environments. Our work identified the gap between the existing reviews and the current trends in XR ECAs. We began with 1,717 related papers from January 2014 to June 2024. We narrowed down the selection to 23 papers using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework, which employed an iterative screening procedure and criteria defined by our research team. The resulting papers were analyzed and discussed in terms of the features of the ECA application, its design and implementation, and use cases. Our analysis highlights key trends in XR ECA design, including the dominance of VR-based implementations using head-mounted displays, the prevalence of human-like and female-presenting agents, the move from rule-based to neural-based conversational systems, and the primary use cases in training, therapy, and social interaction. We also summarize the evaluation methods employed across studies and discuss future research directions for developing more adaptive and human-like ECAs in XR environments.

INDEX TERMS Embodied conversational agents, extended reality, conversational interaction, use cases, evaluation methods.

I. INTRODUCTION

Conversational agents (CAs) are utilized in numerous applications throughout our daily lives. CAs often refer to text- or speech-based dialogue systems responding to users' natural language [1]. Moving beyond CAs to embodied conversational agents (ECAs), the significance of agent embodiment has influenced how humans perceive dialogue-based computer agents. Cassell [2] described ECAs as computer agents that converse in a manner similar to real humans and can produce both verbal and nonverbal communications. In 1996, Reeves and Nass [3] demonstrated that humans are polite to computers and treat them as social entities, even when they lack human-like appearances. Nowadays, with the advances in artificial intelligence (AI) and computer graphics, we can design digital applications with anthropomorphized interfaces, opening up more aspects to investigate within the realm of human-computer interaction. Apart from ECAs, other terms like virtual humans [4], intelligent virtual agents (IVAs), and socially interactive agents (SIAs) [5] emphasize digital characters with human modalities such as facial expressions, gestures, emotions, and social behaviors, are also widely adopted in the research field.

In recent years, the convergence of extended reality (XR) technologies and ECAs has paved the way for more immersive and interactive experiences that could transform how users engage with conversational agents in digital environments. However, previous reviews on ECAs primarily focused on adoptions in conventional applications, including smartphones, laptops, desktops, or other electronic

The associate editor coordinating the review of this manuscript and approving it for publication was Andrea F. Abate.

devices without virtual reality (VR), augmented reality (AR), or mixed reality (MR) functionalities [6], [7], [8]. As XR technologies become increasingly prevalent and transformative across various fields, we argue that it is time to comprehensively review XR ECAs.

To address the existing research gap, our review encompasses aspects such as current technologies, application features, attributes, use cases, and evaluation methods of XR ECAs. By providing a comprehensive review of the current state of XR ECAs, we can gain a deeper understanding of the advancements in technology and their future progression. We do so by answering the following research questions:

- **RQ1:** What are the XR technologies and devices used for ECAs?
- **RQ2:** What are the XR ECA application features regarding dialogue structures, conversational styles, back-end integrations, software platforms, and input/output modalities?
- **RQ3:** What are the XR ECA attributes regarding appearance, gender, representation, scale, mobility, and expressions?
- **RQ4:** What are the use cases for XR ECA applications?
- **RQ5:** What are the measurements, ratings, and qualitative methods researchers have used to evaluate XR ECA applications?

We conducted a systematic review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework [9]. The scope encompassed eight digital libraries—IEEE Xplore, ACM, ScienceDirect, Web of Science, SpringerLink, Wiley, JSTOR, and Taylor & Francis—and covered publications from 01/2014 to 06/2024. We categorized the filtered publications into relevant themes and topics to systematically and effectively analyze the results and answer the research questions.

The structure of this paper is as follows. We listed related work in Section II. We detailed our PRISMA systematic review procedures and screening criteria in Section III. We presented the resulting papers and categorizations in Section IV. We discussed our review results and answered our research questions in Section V. We mentioned the study's limitations in Section VI. Finally, we concluded our work and suggested future directions in Section VII.

II. RELATED WORK

A. BACKGROUND

At the beginning of 2000, Cassell et al. [10] provided one of the earliest comprehensive overviews of ECAs. They covered the fundamental elements of ECAs, such as interaction modalities, system designs, and applications. They defined ECAs as being able to recognize and generate verbal and nonverbal cues, perform turn-taking conversational interaction, signal discourse state indicators, and contribute propositions to the conversation [11]. Magnenat-Thalmann and Thalmann [12] investigated the history and evolution of virtual human technology. They examined the creation, application, and implications of digital human models in various fields, including medicine, education, commerce, and gaming. Their work explored technical advancements and forecasted future trends in the field of virtual human research. Ruttkay et al. [13] proposed a conceptual framework (see Figure 1) of ECA design aspects featuring three main categories: embodiment, mental capacities, and application interfaces. Embodiment is the visual appearance of the agent, output modalities, hand and body animation, and facial expressions. Mental capacities encompass the agent's social role, defined personality, emotional states, user model (e.g., adaptation to users and input modalities), and discourse capabilities, which often relate to the agent's system framework, such as the dialogue system or natural language processor and generator. The application interface encompasses the selection of display equipment, data transfer models, and agents' background knowledge processing, which also relates to the purpose of the ECA, whether it is for education, entertainment, companionship, or other use cases.

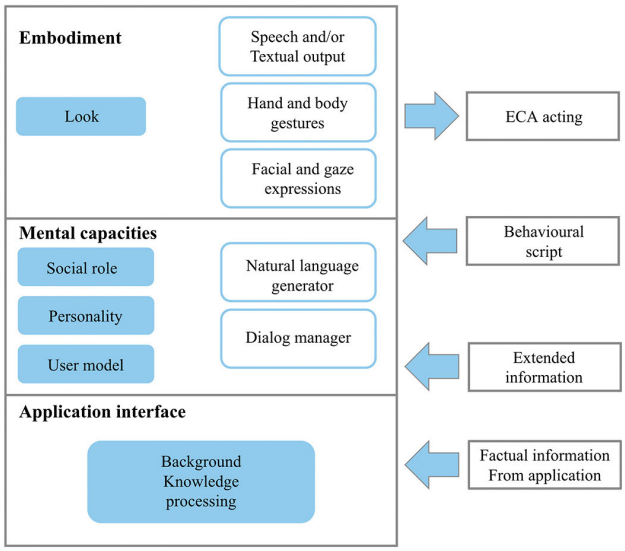


FIGURE 1. The conceptual framework for ECA design aspects as proposed by Ruttkay et al. [13].

As the mental capacities of ECA significantly impact user experience, past research has also shown interest in evaluating the connections between character animations, facial expressions, body postures, or gestures and human perceptions of agents' personalities [14], [15], [16]. Reviews that map agent personality and emotions to system usability and effectiveness also merit attention. Nijholt's [17] state-of-the-art report discussed humor modeling in ECA interaction protocols. Nijholt stated the benefit of incorporating humor into agents' responses, suggesting that humorous traits can make ECAs more relatable and compelling as social partners. Beale's and Creed's [18] structured review on the impact of agents' emotions on users' attitudes and perceptions provides researchers and developers with consolidated design guidelines for building ECAs in different endeavors. For example, in the games and entertainment domain, it was

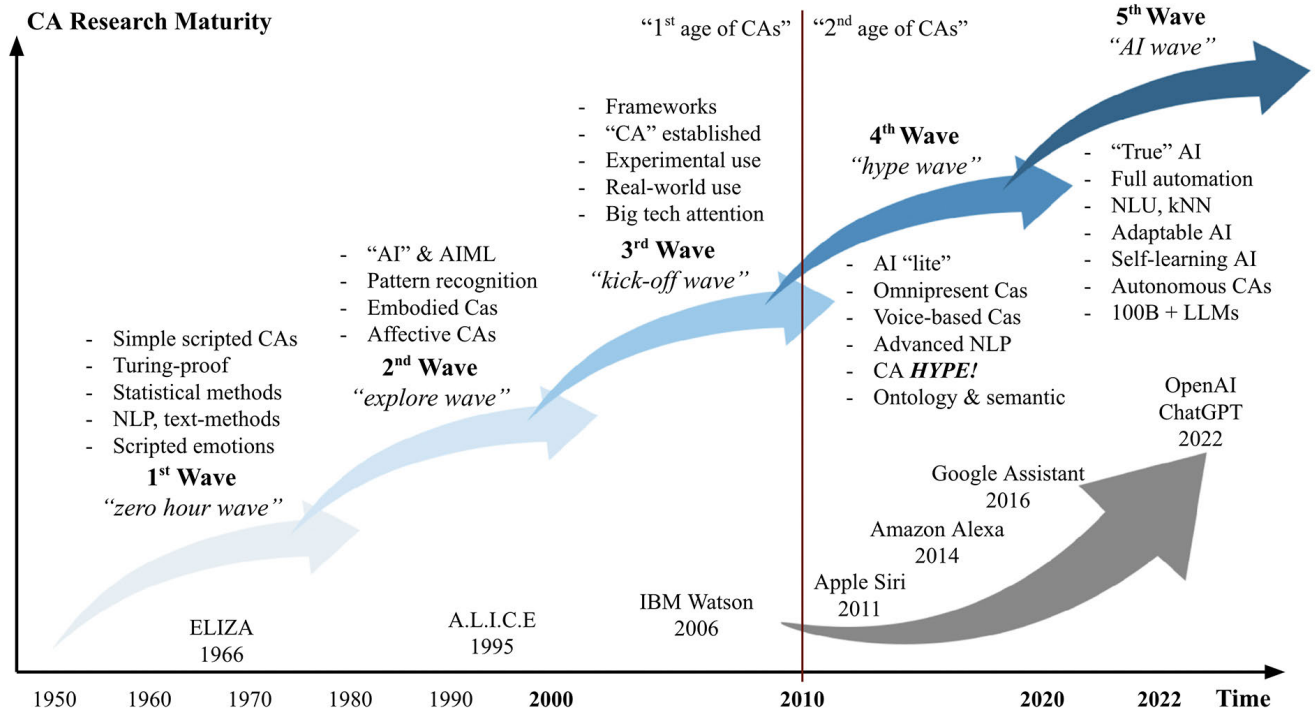


FIGURE 2. The five identified waves of CA evolutions as identified by Schobel et al. [8].

found that agents showing empathy through conversations can help reduce users' frustration and improve gameplay.

Past research has explored the multimodal interactions of ECAs [19], [20], [21], making conversational style and input/output modalities often the focus of review studies. André and Pelachaud [22] emphasized the importance of ECA design methodologies based on human dialogue observations. They distinguished conversational styles into TV-style, face-to-face, role-plays, and multi-party dialogue. TV-style refers to ECA applications where users watch as the agent speaks, which contradicts Cassell et al.'s [11] definition of ECA as having the ability to perform turn-taking human-agent conversational interaction. Role plays conversational style derived from TV commercials, where one salesman played the role of a buyer and interacted with another salesman, enabling viewers to easily get situated in the content by observing their interactions. We also find such an approach in games or social simulations, where multiple virtual agents were created to enhance human-agent interaction. Similar to TV-style, role-play dialogue does not require user input. Van Pinxteren et al. [6] reviewed previous studies to identify key attributes contributing to the human likeliness of CAs and proposed a research agenda for further exploration in this area. The authors identified two main classifications: modality and footing. Modality includes verbal behaviors, nonverbal behaviors, and appearance characteristics. Footing refers to the communication behaviors that enable agents to bond with users, including: human similarity, mimicking humans in general; individual similarity, mimicking individual users; and responsiveness, being sensitive and supportive to users' needs and others. Yousefi et al.'s [23] systematic review highlighted the potential of XR virtual characters in fostering prosocial behaviors through richer social cues. Their work identified past studies that focused on human-agent social interactions, such as perspective-taking, prosocial decision-making, and helping behaviors, providing insights into the state-of-the-art design of XR virtual characters' prosocial interaction attributes.

B. ECA SYSTEMS, EVALUATIONS, AND USE CASES

Another essential area to look at is the technological implementations. From traditional rule-based agents [24] and pattern recognition and markup language systems [25] to voice recognition and synthesis [26], ECAs have undergone continuous improvement in every aspect. Schobel et al. [8] observed that advancements in AI, machine learning (ML), natural language processing (NLP), generative AI (GAI), and large language models (LLMs) contribute to the development of chatbots, conversational agents, and other dialogue systems. They identified five waves of CA research, spanning past, present, and future evolution, including the zero-hour wave, the explore wave, the kick-off wave, the hype wave, and the AI wave (see Figure 2). Their work emphasized the progress in ML and AI technologies, highlighting technological tools such as OpenAI GPT[1] and

[1]https://openai.com/chatgpt/

BLOOM natural language understanding (NLU),[2] enabling more sophisticated dialogue systems capable of handling complex human-like conversations [8]. They also explored and documented the trend of keywords in CA publications, finding that ''embodied conversational agent,'' ''virtual agent,'' ''affective computing,'' and ''emotion'' showed a steady increase, with the most relevant research emphasizing trust, anthropomorphism, agent personality, and emotion expression. The data implied a trend of making CA more human-like to enhance interpersonal interaction.

As human-agent interactions are often complex and dynamic, depending on the various purposes and affordances of the system, developing a standardized evaluation model for all ECA applications poses a challenge. Weiss et al. [7] presented several assessment instruments and methods for evaluating multimodal ECAs, such as heuristic, model-based, experimental, and interaction parameter evaluation. Interaction parameter evaluation encompasses a wide range of log data in ECA research. For example, in turn-taking interactions, parameters can be outlined in a time-related manner (see Figure 3), comprising user response delay and action duration, as well as system response delay and action duration. Loveys et al.'s [27] systematic review highlighted past research with design features on ECA behavior, appearance, and language to strengthen the human-agent relationship and increase system effectiveness and user engagement. Several contributions were made to evaluating and understanding users' perceptions, allowing for adjustments to ECA characteristics accordingly.

Past reviews on ECAs that featured a specific endeavor were often found in the healthcare category. Provoost et al.'s [28] work evaluated ECA for clinical psychology in mood, anxiety, psychotic, autism spectrum, and substance use disorders. Ter Stal et al. [29] investigated ECAs in support of eHealth systems that lessen the burden on healthcare sectors. Kramer et al. [30] examined the relationship between ECA characteristics and their effectiveness in lifestyle coaching platforms. Reviews in the education domain were also conducted, with Khosrawi-Rad et al. [31] and Hobert and Meyer von Wolff [32] examining pedagogical conversational agents. Landim et al. [33] overlooked the development of CAs in e-commerce and analyzed existing CAs in computational classifications, including neural and rule-based approaches.

C. OUR CONTRIBUTIONS

Regarding virtual agents in XR, previous contributions focused on reviews of virtual humans, intelligent virtual agents, and embodied virtual agents, rather than XR ECAs. Hirzle et al.'s [34] review of AI usage in XR provided an overview of the varied interactions and functionalities that integrate AI technologies, including the creation of virtual agents and interactions with intelligent virtual assistants. Norouzi et al.'s [35] review on embodied agents in AR head-mounted display environments classified past research on several key dimensions, including agent embodiments, interaction technologies, application areas, and technological frameworks.

Despite the abundance of literature reviews on ECAs and the growing amount of research papers on virtual agents within XR environments, a notable gap remains at the intersection of these two fields. Based on our research, we did not find any comprehensive reviews that specifically explore ECAs within XR settings. Thus, bridging this gap could significantly advance our understanding of human-agent conversational interaction in immersive environments. Our systematic review was built on prior research and contributed as follows:

- provided an overview of immersive technologies and devices used for XR ECAs (**RQ1**);
- analyzed dialogue structures, conversational styles, backend integrations, software platforms, and input/output modalities in XR ECAs (**RQ2**);
- examined key attributes of XR ECAs, including appearance, gender, representation, scale, mobility, and expressions (**RQ3**);
- identified use cases of XR ECAs (**RQ4**); and
- reviewed user studies and evaluation methods for assessing XR ECAs (**RQ5**).

III. METHODOLOGY

A. OVERVIEW

We employed a structured literature review (see Figure 4) guided by the PRISMA framework [9] to investigate the intersection of ECAs and XR-based applications. The primary objective is to explore and capture existing publications in the field that meet our requirements. We included eight databases in our research: IEEE Xplore,[3] ACM Digital Library,[4] ScienceDirect,[5] Web of Science,[6] SpringerLink,[7] Wiley,[8] JSTOR,[9] and Taylor & Francis.[10] We retrieved 1,717 publications from these databases using our designed query and narrowed them down to 23 publications after screening. We then analyzed and answered our research questions with the remaining 23 papers. Detailed screening criteria and methods are documented in the following subsections.

B. IDENTIFICATION AND SEARCH STRATEGY

We collected publications within the past decade, from 2014 to 2024. The specific time range is from January 1st, 2014, to June 6th, 2024. We excluded book chapters, encyclopedia articles, and non-scholarly papers and included peer-reviewed publications, such as journal articles and conference papers, to increase credibility. We also excluded

[2]https://bigscience.huggingface.co/

[3]https://IEEE Xplore.ieee.org/
[4]https://dl.acm.org/
[5]https://www.sciencedirect.com/
[6]https://www.webofscience.com/wos/
[7]https://link.springer.com/
[8]https://www.wiley.com/
[9]https://www.jstor.org/
[10]https://taylorandfrancis.com/

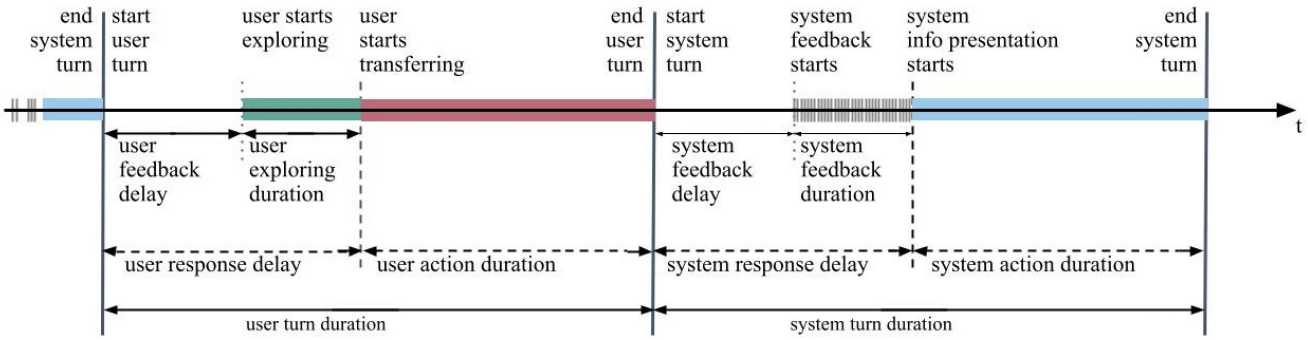


FIGURE 3. Time-related interaction parameters in a complete conversation exchange between the user and the system/agent as identified by Weiss et al. [7].

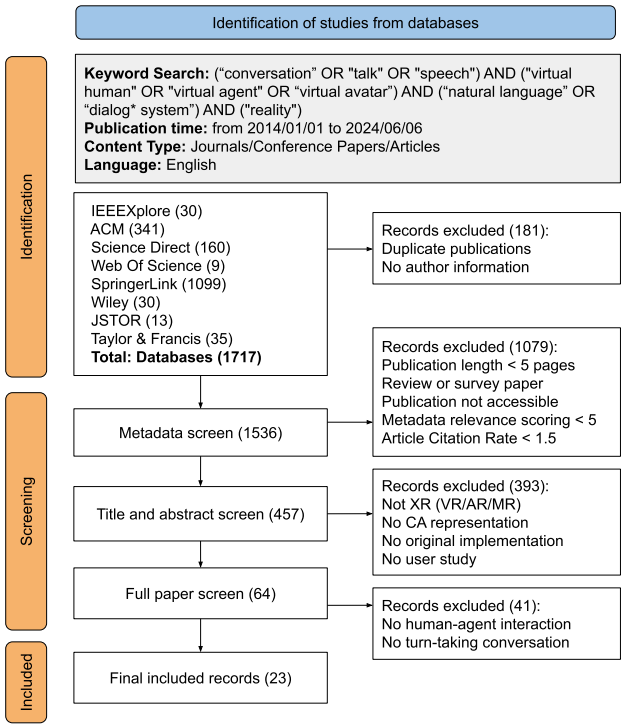


FIGURE 4. Our systematic review follows the PRISMA guidelines. This diagram demonstrates our identification, screening procedures, and final inclusion.

papers published in languages other than English to ensure consistency in analysis and avoid potential translation accuracy challenges. We defined keywords that best aligned with our review targets and objectives. Due to the search term and the number of Boolean operator restrictions in IEEE Xplore and ScienceDirect, we have set our search query to be applied consistently across all eight databases. We confined the search keyword as follows:

("conversation" OR "talk" OR "speech") AND ("virtual human" OR "virtual agent" OR "virtual avatar") AND ("natural language" OR "dialog* system") AND ("reality")

We exported the search results from the database website to BibTex or CSV format. Afterward, we employed Mendeley,[11] a reference management software, to eliminate duplicate publications and those lacking author details in consideration of research reliability and traceability. In total, 181 records were eliminated in this stage, leaving 1,536 publications for further screening.

C. INCLUSION AND EXCLUSION CRITERIA

Our screening procedure consisted of three main steps. First, in the metadata screening, we utilized Crossref[12] and Scopus APIs[13] to filter and exclude records based on our criteria. Second, during the title and abstract screening, we excluded unrelated publications based on their titles and abstracts. Third, during the full-text screening, we excluded publications after reading the full text of each paper. In the following sections, we detail the inclusion and exclusion criteria applied at each step.

1) METADATA SCREENING

a: PUBLICATION LENGTH

Although shorter papers may present modern technology improvements and demonstrations, they typically lack sufficient detail in user studies and result analyses. We set a threshold of five pages, which means that publications shorter than five pages were excluded.

b: DOCUMENT TYPE

We excluded surveys, reviews, and position papers as we aimed to focus on original research with technical development, user studies, and experimental results.

c: ACCESSIBILITY

We excluded records that could not be accessed due to paywalls, broken links, or missing files. We accessed all retrieved records through the Purdue University Libraries.[14]

[11] https://www.mendeley.com/
[12] https://www.crossref.org/documentation/retrieve-metadata/
[13] https://dev.elsevier.com/tecdoc_attribution_scopus.html
[14] https://lib.purdue.edu/

TABLE 1. The defined relevant keywords in our scoring system.

Aspects	Keywords
Conversational Interaction	embodied conversational agent, ECA, conversation agent, small talk, speech-based, voice-based, verbal, chatbot, communication, dialog, dialogue
Embodied Agent	virtual human, digital human, virtual avatar, virtual character, virtual agent, virtual assistant, agent, avatar, character, embodied agent, assistant, humanoid
XR	extended reality, virtual reality, augmented reality, mixed reality, XR, VR, AR, MR, virtual environment, virtual world, digital replica, digital twin, head-mounted device, HMD, immersive, 360
Series	VRST, CHI, ICVR, XR, ISMAR, SIGGRAPH, VR, EGVE

d: RELEVANCE SCORING

We formulated a keyword relevance scoring system that screens all metadata records. The scoring system comprised several aspects: conversation interaction, embodied agent, XR, and publication series type (see Table 1). As long as a defined keyword was identified in the metadata, a point was added to the total score of that entry. Records scoring below a threshold of five (< 5) were excluded, as these scores suggest lower relevance to the predefined search terms, indicating that the content might not be sufficiently focused on the key aspects of our topic.

e: ARTICLE CITATION RATE

We followed a similar approach from a prior systematic review [36] and measured the article citation rate (ACR). The ACR (see Equation 1) indirectly measures the study's impact and recognition in the field, as suggested by Hutchins [37]. The ACR was computed as:

$$ACR = \frac{numbers\ of\ cumulative\ citations}{(current\ year - publication\ year)}. \quad (1)$$

We used the Abstract Citations Count API[15] by Scopus to retrieve citation number information from the Scopus database. Considering that we conducted this study in June 2024, the current year variable was set to 2024.5, ensuring that publications in 2024 were also valid in the equation. Records with an ACR of less than 1.5 were excluded.

2) TITLE AND ABSTRACT SCREENING

Two of our research members performed this step. Both researchers read the titles and abstracts of 457 papers and decided whether to include or exclude each paper. If any disagreements arose, they resolved them through discussion or by consulting a third research member to reach a consensus. The screening exclusion criteria were the following:

- no XR technologies were involved;
- no conversational agent representation;
- no original implementation;
- no user study or preliminary study; and
- does not meet all prior metadata screening criteria.

[15]https://dev.elsevier.com/cited_by_scopus.html

3) FULL-TEXT SCREENING

The final full-text screening procedure narrowed the targeted publications from 64 to 23 papers. This step followed the exclusion criteria below:

- no human-agent interaction; and
- no turn-taking conversational interaction.

IV. RESULTS

In this section, we break down the 23 resulting papers from our PRISMA framework. Among the 23 papers, 13 were retrieved from SpringerLink (56.52%), six from ACM Digital Library (26.09%), two from ScienceDirect (8.70%), one from Web of Science (4.35%), and one from Taylor & Francis (4.35%) (see Figure 5). Four were published in 2024, three were published in 2019, 2020, and 2023, two were published in 2015, 2016, 2021, and 2022, and one in 2017 and 2018 (see Figure 6 for publication year trend). Among them, thirteen were journal articles (56.52%), and ten were conference papers (43.48%). We also collected information on the country in which the research was conducted based on the affiliations of the first authors. Among all selected papers, four were from the United States (17.39%), four

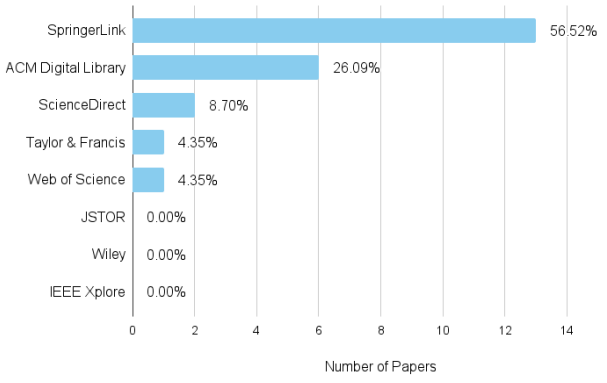


FIGURE 5. The number of papers and percentages of the contribution of each scientific database.

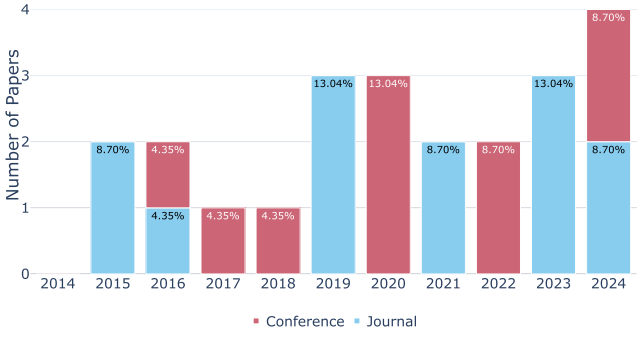


FIGURE 6. The number of publications trend per year in the included papers.

TABLE 2. The list of papers we included in our systematic review.

Paper	Publication	Type	Year	Country	Publisher
Pan et al. [38]	Frontiers in Robotics and AI	Journal	2015	United Kingdom	Frontiers Media SA
Heyselaar et al. [39]	Behavior Research Methods	Journal	2015	The Netherlands	Springer Science and Business Media LLC
Hartanto et al. [40]	Pervasive Computing Paradigms for Mental Health (MindCare)	Conf	2016	The Netherlands	Springer International Publishing
Saad et al. [41]	Multimedia Tools and Applications	Journal	2016	United Arab Emirates	Springer Science and Business Media LLC
Pejsa et al. [42]	Intelligent Virtual Agents (IVA)	Conf	2017	United States	Springer International Publishing
Ochs et al. [43]	6th International Conference on Human-Agent Interaction (HAI)	Conf	2018	France	Association for Computing Machinery
Herrero and Lorenzo [44]	Education and Information Technologies	Journal	2019	Spain	Springer Science and Business Media LLC
Ochs et al. [45]	Journal on Multimodal User Interfaces	Journal	2019	France	Springer Science and Business Media LLC
Slater et al. [46]	Scientific Reports - Nature	Journal	2019	Spain	Springer Science and Business Media LLC
Reinhardt et al. [47]	14th International Conference on Tangible, Embedded, and Embodied Interaction (TEI)	Conf	2020	Germany	Association for Computing Machinery
Guimarães et al. [48]	20th ACM International Conference on Intelligent Virtual Agents (IVA)	Conf	2020	Portugal	Association for Computing Machinery
Nguyen et al. [49]	Practical Aspects of Declarative Languages (PADL)	Conf	2020	United States	Springer International Publishing
Gorisse et al. [50]	Scientific Reports - Nature	Journal	2021	Spain	Springer Science and Business Media LLC
Souchet et al. [51]	Virtual Reality	Journal	2021	France	Springer Science and Business Media LLC
Hassan et al. [52]	2nd Workshop on Games Systems (GameSys)	Conf	2022	Norway	Association for Computing Machinery
Kato et al. [53]	Human Interface and the Management of Information: Applications in Complex Technological Environments (HCII)	Conf	2022	Japan	Springer International Publishing
Safadel et al. [54]	TechTrends	Journal	2023	United States	Springer Science and Business Media LLC
Zhu et al. [55]	International Journal of Human-Computer Interaction	Journal	2023	China	Taylor & Francis
Gan et al. [56]	Cognitive Systems Research	Journal	2023	China	Elsevier
Zhang et al. [57]	Conference on Human Factors in Computing Systems (CHI)	Conf	2024	New Zealand	Association for Computing Machinery
Wang et al. [58]	Conference on Human Factors in Computing Systems (CHI)	Conf	2024	China	Association for Computing Machinery
Spiegel et al. [59]	npj Digital Medicine - Nature	Journal	2024	United States	Springer Science and Business Media LLC
Llanes-Jurado et al. [60]	Expert Systems with Applications	Journal	2024	Spain	Elsevier

from Spain (17.39%), three from France (13.04%), three from China (13.04%), two from The Netherlands (8.70%), and the remaining were from the United Kingdom, United Arab Emirates, Germany, Portugal, Norway, Japan, and New Zealand (4.35%).

A. XR TECHNOLOGIES AND DEVICES

We identified the XR technologies and system devices adopted in the selected papers. XR technologies consisted of VR, AR, and MR. Within the 23 papers, system devices were classified into four categories: VR head-mounted displays (HMDs), cave automatic virtual environment (CAVE) systems, augmented reality (AR) smartphones, and mixed reality (MR) HMDs. Please see the results in Table 3. Moreover, Figure 7 illustrates the distribution of VR, AR, and MR applications by year and Figure 8 shows the distribution of each category.

B. APPLICATION FEATURES

We reported various features of XR ECA applications. These included dialogue structure, conversational style, backend integration, software platforms, and input/output modalities.

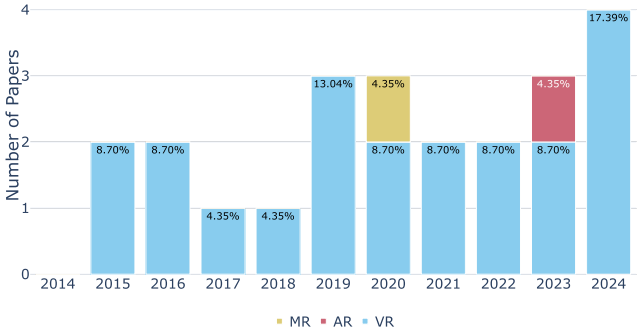


FIGURE 7. The distribution of XR technologies in the selected papers.

The dialogue structure encompassed several key features, including being task-oriented, semi-guided, and open-ended. The task-oriented dialogue facilitates goal-directed interactions, whereas open-ended structures provide users with dynamic exchanges with greater flexibility and agency. The conversational style was classified into two categories: one-on-one dialogue (face-to-face dialogue as defined by André and Pelachaud [22]) and multi-party dialogue. Following previous work [33], we classified backend integration into

TABLE 3. The list of XR technologies and devices used for our resulting papers (sorted by the publication years).

Paper	XR Technology	Device
Pan et al. [38]	VR	CAVE
Heyselaar et al. [39]	VR	NVIS nVisor SX60
Hartano et al. [40]	VR	HMD (not specified)
Saad et al. [41]	VR	Oculus Rift
Pejsa et al. [42]	VR	Oculus Rift CV1
Ochs et al. [43]	VR	Oculus Rift + CAVE
Herrero and Lorenzo [44]	VR	Oculus Rift
Ochs et al. [45]	VR	Oculus Rift + CAVE
Slater et al. [46]	VR	HTC Vive
Reinhardt et al. [47]	MR	Hololens 1
Guimarães et al. [48]	VR	HTC Vive
Nguyen et al. [49]	VR	Google Cardboard
Gorisse et al. [50]	VR	HTC Vive Pro
Souchet et al. [51]	VR	Samsung Gear VR 2
Hassan et al. [52]	VR	Oculus Quest 2
Kato et al. [53]	VR*	HTC Vive Pro
Safadel et al. [54]	VR	Oculus 1
Zhu et al. [55]	VR	HTC Vive Pro
Gan et al. [56]	AR	Smartphone
Zhang et al. [57]	VR	HP Reverb G2 headsets
Wang et al. [58]	VR	Oculus Quest 2
Spiegel et al. [59]	VR	Oculus Quest 2
Llanes-Jurado et al. [60]	VR	Semi-immersed projection/CAVE-like

*They refer to as XR cross reality.

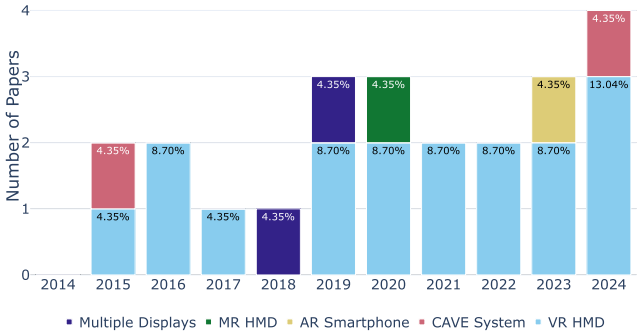


FIGURE 8. The distribution of device categories in the selected papers.

rule-based and neural-based. We listed out the software platforms utilized in each paper. Input/output modalities refer to the interaction methods between human subjects and XR ECA systems. Input modalities included voice, eye gaze, head movement, controller, touch, or mouse input. Output modalities spanned across pre-recorded or synthesized voice, visual output, or body movement. We documented each paper's input and output modalities based on the provided system descriptions. Please refer to Table 4 for a detailed description of the application features.

1) DIALOGUE STRUCTURE

Most reviewed studies employed task-oriented dialogue structures (17 papers, 73.91%). A task-oriented structure can vary in flexibility; for example, some task-oriented systems still allow for open-ended dialogues (three papers, 13.04%) [52], [56], [60]. In contrast, a smaller subset of studies (two papers, 8.70%) implemented supportive or therapeutic dialogue structures, focusing on empathy and emotional validation, particularly in mental health applications. Although often task-oriented, these structures incorporated elements of emotional support and flexibility to better suit users' needs. This demonstrates that dialogue structures are not strictly dichotomous, as task-oriented and open-ended characteristics can coexist depending on the interaction's design and objectives.

2) CONVERSATIONAL STYLE

The conversational style reflected how people communicated with XR ECAs. Specifically, we examined the number of participants in the conversational setting and identified two distinct conversational styles: one-on-one and multi-party. In the multi-party conversation scenario, participants were immersed in the virtual space with multiple virtual agents present simultaneously. Most conversational styles (see Figure 9) are one-on-one (19 papers, 82.61%), only three papers employed a multi-party conversational style (13.04%), and one paper (4.35%) employed both. Although researchers did not actively research the multi-party conversational style, they maintained a steady interest in it.

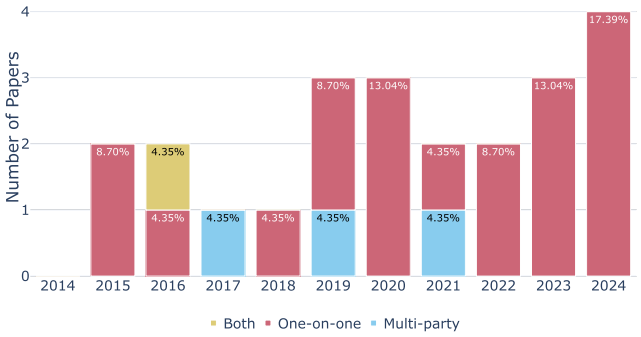


FIGURE 9. The distribution of the conversational style in the selected papers.

3) BACKEND INTEGRATION

The backend integration evolved from rule-based systems to more sophisticated neural-based models. Most reviewed studies (18 papers, 78.26%) utilized rule-based architectures, relying on pre-programmed scripts for predictable and controlled interactions. These systems were popular in contexts where consistency and structured guidance were necessary. However, we observed a significant shift in 2023 and 2024 with approximately 21.74% of the studies (five papers) adopting neural-based systems capable of generating contextually relevant and adaptive responses. This progression represented an enhancement of the XR ECA system's capabilities, emphasizing a more natural and human-like interaction. The transition toward neural-based frameworks also indicated a focus on refining user experiences through ECAs' real-time, dynamic, and adaptive communication (see Figure 10).

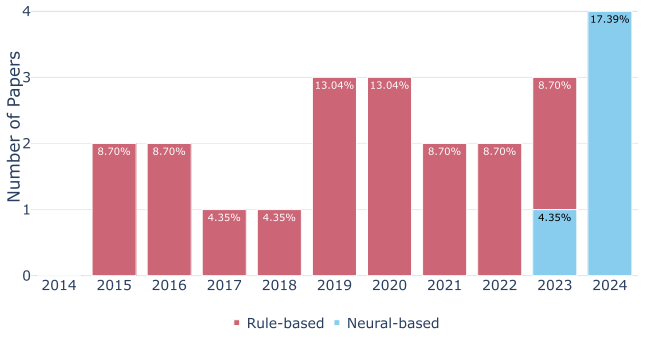


FIGURE 10. The distribution of backend integration in the selected papers.

4) SOFTWARE PLATFORM

The reviewed studies utilized a range of software platforms, with a notable dominance of game engines, such as Unity and Unreal Engine. Most studies (15 papers, 65.22%) employed Unity, utilizing its capabilities to build interactive extended reality applications. Zhang et al. [57] utilized Unreal Engine due to its advanced graphics capabilities. Meanwhile, a smaller subset of studies featured custom-developed solutions implemented to meet specialized interaction needs. For instance, Ochs et al. [45] and Ochs et al. [43] developed a 3D video playback player to support synchronized verbal and non-verbal cues in task-based interactions, seamlessly integrating it with a Unity-based system. Similarly, Nguyen et al. [49] introduced the VRASP, a platform that utilizes answer set programming (ASP) solvers and Web Speech API[16] for natural language processing to facilitate voice-based interactions.

The Memphis system developed by Hartanto et al. [40] stands out for its focus on VR-based social anxiety therapy, utilizing keyword recognition and speech detection for automated interactions. Another unique approach was observed by Saad et al. [41], who implemented the SitePal API[17] for multimodal interaction, combining voice recognition with pre-scripted responses. In contrast, some studies, such as Llanes-Jurado et al. [60], did not specify a particular software platform, while others, such as Pan et al. [38], employed the Platform Independent API for Virtual Characters (PIAVCA).[18] Spiegel et al. [59] processed conversations through a HIPAA-compliant server using OpenAI API,[19] GPT-4 model, for secure and adaptive therapeutic interactions.

5) INPUT/OUTPUT MODALITY

Voice-based input emerged as the dominant modality, having been implemented in 82.61% of the reviewed studies (19 papers), indicating its intuitive nature for seamless user interaction with the ECAs. This approach enhanced immersive dialogue experiences. A small portion of the studies (four papers, 17.39%) incorporated gaze tracking and movement-based inputs, thereby enriching the interaction with non-verbal communication. These inputs were particularly valuable in scenarios that required physical engagement or the incorporation of non-verbal cues. Specifically, 8.70% of the studies (two papers) explored touch and head movements as alternative modalities, indicating a trend toward creating multimodal virtual environments. The increasing use of diverse input methods suggested a growing emphasis on enhancing the interactive quality of ECAs.

The output modalities of the reviewed studies were diverse, with text-to-speech technology leading the way in 65.22% of cases (15 papers). This technology was crucial in delivering clear and lifelike responses, enhancing the believability of ECAs. Slightly over half of the reviewed studies (15 papers, 65.22%) included body movements and gestures, significantly contributing to more engaging and human-like interactions. Combining these output modalities addressed the importance of multi-sensory engagement, elevating user experience and the authenticity of human-agent interactions.

C. ATTRIBUTES OF ECAS

Researchers designed ECAs with various attributes based on the requirements of their XR applications. Therefore, we identified seven key attributes of XR ECAs: appearance, gender, representation, scale, mobility, and expressions. We categorized appearance into human (22 papers, 95.65%) and robotic forms (one paper, 4.35%). For human appearances (see Figure 11), we examined the gender of the ECAs and found three variations: male (three papers, 13.04%), female (12 papers, 52.17%), and male and female combination (seven papers, 30.43%). Representation referred to the format of the virtual character, including full-body (21 papers, 91.30%) and upper-body only (one paper, 4.35%), while scale indicated the character's size, composed of life-size (20 papers, 86.96%) and miniature (two papers, 8.70%). A life-sized agent appears at a scale comparable to an average human, whereas a miniature agent is significantly smaller than the human scale [61]. We determined whether the ECA is life-sized or miniature based on the figures provided in the papers. Mobility (see Figure 12) addressed whether ECAs were stationary (16 papers, 69.57%) or mobile (four papers, 17.39%). Lastly, expressions encompassed various non-verbal behaviors, including gestures (ten papers, 43.48%), lip-sync (nine papers, 39.13%), facial expressions (seven papers, 30.43%), eye gazes (seven papers, 30.43%), head orientation (two papers, 8.70%), and spatial orientation (one paper, 4.35%). We presented the findings related to these attributes in Table 5. Note that we used "NS" to indicate that a paper did not specify that information.

[16] https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API
[17] http://www.sitepal.com/
[18] https://github.com/marcogillies/Piavca
[19] https://openai.com/api/

TABLE 4. List of XR ECA application features of our resulting papers. NS denotes "not specified."

Paper	Dialogue Structure	Conversation Style	Backend Integration	Software Platform	Input Modality	Output Modality
Pan et al. [38]	Semi-guided interview	One-on-one	Rule-based	PIAVCA; XVR	Voice-based input	Pre-recorded verbal responses; body movements; gestures
Heyselaar et al. [39]	Syntactic priming in language interactions	One-on-one	Rule-based	Vizard (WorldViz)	Voice-based input	Pre-recorded verbal responses
Hartano et al. [40]	Task-oriented	One-on-one & multi-party	Rule-based	Memphis system	Voice-based input	Verbal responses; visual output
Saad et al. [41]	Task-oriented	One-on-one	Rule-based	SitePal API	Voice-based input	Text-to-speech; visual output
Pejsa et al. [42]	Multi-party	Multi-party	Rule-based	Unity	Voice-based input	Text-to-speech; body movement gaze shifts
Ochs et al. [43]	Task-oriented	One-on-one	Rule-based	Custom 3D player based on Unity	Voice-based input	Text-to-speech; body movement; gestures
Herrero and Lorenzo [44]	Task-oriented	Multi-party	Rule-based	Unity	Gaze tracking	Pre-recorded verbal responses; gestures
Ochs et al. [45]	Task-oriented	One-on-one	Rule-based	Custom 3D player based on Unity	Voice-based input	Text-to-speech; body movement; gestures
Slater et al. [46]	Self-conversation	One-on-one	Rule-based	Unity	Voice-based input	Pre-recorded responses; body movements
Reinhardt et al. [47]	Task-oriented	One-on-one	Rule-based	Unity	Voice-based input	Text-to-speech; body movement; gesture
Guimarães et al. [48]	Task-oriented	One-on-one	Rule-based	Unity	VR controllers or mouse input	Text-to-speech; gestures
Nguyen et al. [49]	Task-oriented	One-on-one	Rule-based	Custom VRASP; ASP solvers; Web Speech API	Voice-based input	Text-to-speech
Gorisse et al. [50]	Task-oriented	Multi-party	Rule-based	Unity	Voice-based and movement-based input	Pre-recorded responses; body movements
Souchet et al. [51]	Task-oriented	One-on-one	Rule-based	Manzalab; Unity	Head movement or mouse input	Pre-recorded responses; visual output
Hassan et al. [52]	Open-ended interview	One-on-one	Rule-based	Unity	Voice-based input	Text-to-speech; body movement; gesture
Kato et al. [53]	Task-oriented	One-on-one	Rule-based	Unity	Voice-based input	Text-to-speech
Safadel et al. [54]	Task-oriented	One-on-one	Rule-based	Unity	Voice-based input	Text-to-speech
Zhu et al. [55]	Task-oriented	One-on-one	Rule-based	Unity	VR controllers input	Text-to-speech; body movement
Gan et al. [56]	Task-oriented; open-ended conversation	One-on-one	Neural-based	Unity; Vuforia AR SDK	Voice-based and touch input	Text-to-speech; visual feedback; body movement; gestures
Zhang et al. [57]	Task-oriented	One-on-one	Neural-based	Unreal Engine	Voice-based input	Human-controlled body movements; verbal responses
Wang et al. [58]	Task-oriented	One-on-one	Neural-based	Unity	Voice-based input	Text-to-speech; visual feedback; body movement; gesture
Spiegel et al. [59]	Supportive; therapeutic conversation	One-on-one	Neural-based	HIPAA server	Voice-based input	Text-to-speech
Llanes-Jurado et al. [60]	Semi-guided; open-ended conversation	One-on-one	Neural-based	NS	Voice-based input	Text-to-speech; body movement

TABLE 5. The list of attributes of XR ECAs of our resulting papers. NS denotes "not specified."

Paper	Appearance	Gender	Representation	Scale	Mobility	Expressions
Pan et al. [38]	Human	Female	Full-body	Life-sized	Stationary (front)	Facial expression; gesture
Heyselaar et al. [39]	Human	Female	Full-body	Life-sized	Stationary (front)	Facial expression; lip-sync
Hartano et al. [40]	Human	Female and male	Full-body	Life-sized	Stationary (front)	NS
Saad et al. [41]	Human	Female	Upper-body	Miniature	Stationary (front)	NS
Pejsa et al. [42]	Human	Female and male	Full-body	Life-sized	Stationary (front)	Gaze; lip-sync; spatial orientation
Ochs et al. [43]	Human	Female	Full-body	Life-sized	Stationary (front)	Gaze; gesture
Herrero and Lorenzo [44]	Human	Female and male	Full-body	Life-sized	Mobile	Facial expression; gesture; lip-sync
Ochs et al. [45]	Human	Female	Full-body	Life-sized	Stationary (front)	Gaze; gesture
Slater et al. [46]	Human	Male	Full-body	Life-sized	Stationary (front)	Gesture
Reinhardt et al. [47]	Human	Female and male	Full-body	Life-sized	Stationary (front)	Gaze; lip-sync
Guimarães et al. [48]	Human	Male	Full-body	Life-sized	Stationary (front)	Gaze; gesture; lip-sync
Nguyen et al. [49]	Human	Male	Full-body	Life-sized	Stationary (front)	NS
Gorisse et al. [50]	Human	Female and male	Full-body	Life-sized	Mobile	Gaze; gesture; lip-sync
Souchet et al. [51]	Human	Female	Full-body	Life-sized	Stationary (front)	NS
Hassan et al. [52]	Human	Female	Full-body	Life-sized	Stationary (front)	Gaze; head orientation; lip-sync
Kato et al. [53]	Human	Female	Full-body	Life-sized	Stationary (front)	Facial expression
Safadel et al. [54]	Human	Female	Full-body	Life-sized	NS	Lip-sync
Zhu et al. [55]	Human	Female	Full-body	Life-sized	NS	Gesture
Gan et al. [56]	Human	Female	Full-body	Miniature	Stationary (front)	Facial expression; gesture
Zhang et al. [57]	Human	Female and male	Full-body	Life-sized	Mobile	NS
Wang et al. [58]	Human	Female	Full-body	Life-sized	Mobile	Facial expression; gesture
Spiegel et al. [59]	Robot	NS	Full-body	NS	NS	NS
Llanes-Jurado et al. [60]	Human	Female and male	Full-body	Life-sized	Stationary (front)	Facial expression; lip-sync

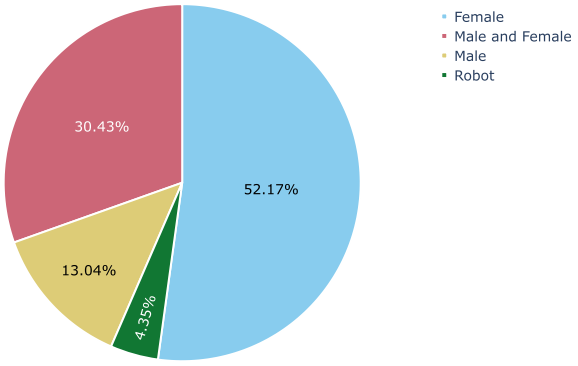


FIGURE 11. The appearance attribute distributions of ECAs in the examined papers.

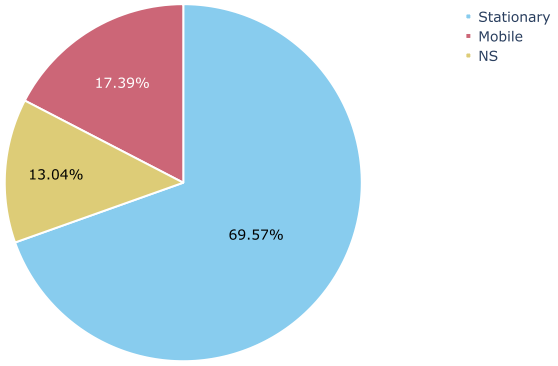


FIGURE 12. The mobility attribute distributions of ECAs in the examined papers.

D. USE CASES CATEGORIZATION

We categorized the papers according to their implementation use cases into seven main categories, including training/education (eight papers, 34.78%), social conversing (four papers, 17.39%), therapy/counseling (four papers, 17.39%), information retrieval/assistant (three papers, 13.04%), museum tour guide (two papers, 8.70%), and shopping/customer service (two papers, 8.70%). The training/education category encompasses papers exploring ECAs developed to improve skill acquisition, enhance learning experiences, or support instructors in immersive classroom settings or training environments. The social converse category covered research simulating or examining social interactions, daily dialogue, and communication using

XR ECAs. The therapy/counseling category included papers that showcase the adoption of XR ECAs in therapeutic, psychological, or psychiatric treatments, contributing to clinical settings. The information retrieval/assistant category provided papers that utilize ECAs as informational assistants in XR, focusing on helping users obtain information and efficiently cater to personalized tasks, such as weather forecasting or meeting scheduling. The museum/tour guide category contained research where ECAs were implemented as interactive agents guiding users through the immersive exhibitions and providing contextualized conversations related to the exhibit artifacts. The shopping/customer service category explored XR applications where ECAs assist with commercial activities such as VR shopping or addressing customer complaints. Figure 13 illustrates the distribution of use cases for the examined papers. We also summarized our results in Table 6.

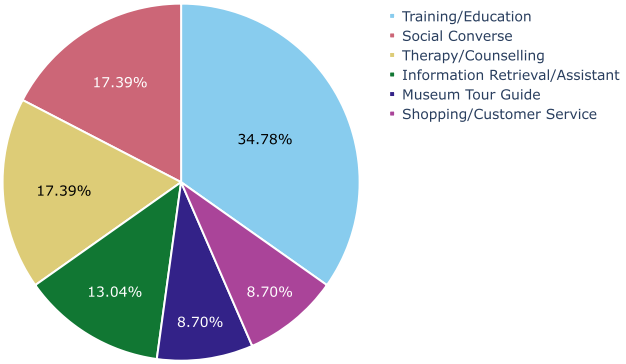


FIGURE 13. The distribution of selected papers based on embodied conversational agent use cases.

E. MEASUREMENTS, RATINGS, AND QUALITATIVE METHODS

It is essential to identify the instruments researchers use to collect data to assess the impact of XR ECAs. Therefore, we compiled a list of papers that employed both quantitative and qualitative methods. The quantitative methods consisted of questionnaires and measurements, focusing on the variables used in the questionnaires. Specifically, we classified these variables into four categories: system evaluation (seven papers, 30.43%), perception of XR ECAs (eight papers, 34.78%), user experience (18 papers, 78.26%), and psychological assessment (five papers, 21.74%). Regarding measurements, we identified ten papers (43.48%) that employed these methods. The researchers typically focused on conversation-related variables (seven papers, 30.43%), such as speaking time and the number of speech interactions. Motion capture (one paper, 4.35%) and heart rate (one paper, 4.35%) were also measured. For qualitative methods, we included interviews (eight papers, 34.78%), comments (four papers, 17.39%), and recorded audio (three papers, 13.04%) and video (three papers, 13.04%). We summarized our results in Table 7. Please note that we used "NS" to indicate cases where a paper did not provide the relevant information. Moreover, Figure 14 shows the distribution of different qualitative methods used in evaluating ECAs in the selected papers.

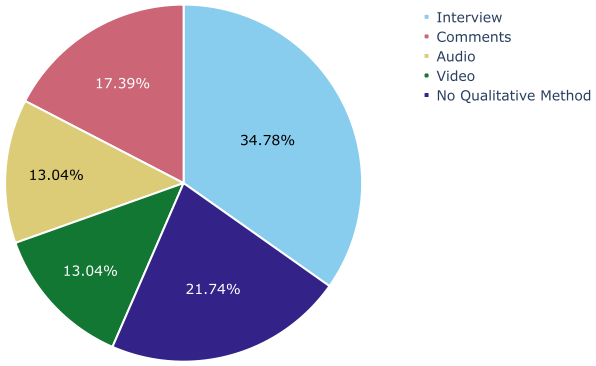


FIGURE 14. The distribution of the qualitative methods researchers used to evaluate their ECAs.

V. DISCUSSION

In the following sections, we discuss our findings and address the research questions.

A. RQ1: XR TECHNOLOGIES AND DEVICES

As seen in the distribution of XR technologies and devices (see Figure 7 and Figure 8), our selected papers highlighted a distinct preference for VR applications, predominantly focusing on HMDs and CAVE system projections. This trend has been significant, as it reflects the current focus of development within the field of XR ECA research. Building VR HMD applications has become increasingly accessible and cost-effective. Game engines have supported numerous toolkits, extensive libraries, and plugins that simplify complex tasks, such as 3D modeling, physics, and interaction design. They have created an environment where researchers can efficiently build prototypes and mockups for user testing and data collection within short development cycles [62]. This accessibility has enabled iterative design processes, allowing feedback to be quickly integrated and enhancing the user experience.

Moreover, VR HMDs have facilitated a controlled setting for experiments, ensuring consistent conditions for data collection across studies, which has been crucial for the validation and reproducibility of research findings. The technological ease and experimental rigor have made VR HMDs invaluable in human-agent interaction research. Conversely, CAVE systems have provided a different immersive experience through room-sized or 360-degree projections. Unlike HMDs, CAVE systems have not required users to wear headsets, thereby reducing barriers to participation and eliminating issues such as motion sickness. However, CAVE systems have generally been more costly and have required larger spaces dedicated to experiments, which has limited their accessibility compared to HMDs [63].

TABLE 6. The list of use case descriptions of our resulting papers.

Use Case Category	Paper	Use Case Description
Training/Education	Zhang et al. [57]	A VR cooking class scenario was created that allowed the instructor to switch between embodiments of two separate avatars, enabling simultaneous collaboration with two students. When the avatar was not in control of the instructor, a virtual agent took over the role of teaching.
	Hassan et al. [52]	The simulation incorporated an interview with a maltreated child to help train police and child protection service (CPS) workers. The system aimed to enhance their knowledge and skills in conversing with abused children.
	Guimarães et al. [48]	An immersive application that helped police practice interrogation conversations with suspects. The suspects were virtual intelligent agents who demonstrated verbal and nonverbal cues. The users needed to obtain as much information as possible without losing control of the interaction.
	Ochs et al. [43]	The system aimed to train doctors to break bad news to patients through conversational interaction.
	Ochs et al. [45]	The system provided a simulated scenario for doctors to break bad news to patients, as in [43].
	Nguyen et al. [49]	The proposed Virtual Reality Answer Set Programming (VRASP) application allowed students to learn programming in VR through conversing with an embodied agent in a virtual classroom. The agent answered students' questions and assisted in solving programming problems.
	Herrero and Lorenzo [44]	The VR application trained and improved the social and emotional skills of students with autism spectrum disorder (ASD) by simulating social scenes from virtual classrooms and playground setups.
	Souchet et al. [51]	The VR game recreated interview scenarios to let users practice responding to job interviews. Users selected the correct answer from the provided options to earn points in the game.
Social Conversation	Pan et al. [38]	The researchers asked participants to interview virtual agents who exhibited two distinct personalities: shy and confident. The study discussed how personality perceptions influenced social dynamics and responses in conversational contexts.
	Pejsa et al. [42]	The proposed model allowed virtual agents to signal conversational roles between speakers, addressees, and bystanders through eye gaze and spatial orientation. The use case focused on multi-party conversational interactions in a VR setting.
	Llanes-Jurado et al. [60]	The system provided users with a seamless conversation environment, allowing them to freely converse with an XR projection-based virtual human in dynamic, natural language.
	Heyselaar et al. [39]	They used a VR application to study language behaviors through a syntactic priming task, examining how humans adapt their sentence structure, or syntax, to match that of their conversing partner. The study aimed to prove that participants' language processing was comparable to interacting with virtual agents and human partners.
Therapy/Counselling	Hartanto et al. [40]	The home-based VR therapy involved virtual avatars to simulate social interactions, which helped patients reduce social phobia and anxiety.
	Spiegel et al. [59]	They developed an extended-reality artificial intelligence assistant (XAIA) platform that provided immersive mental health support via GPT-4-powered AI therapy agents in biophilic environments for individuals with mild-to-moderate anxiety or depression.
	Gorisse et al. [50]	The VR application allowed people to experience persecutory thoughts. They then observed their virtual body double engaging in regular social interactions, which reduced their anxiety. The study highlighted the potential of vicarious agency in influencing psychological states.
	Slater et al. [46]	The VR application enabled participants to alternate between embodying themselves and Sigmund Freud, allowing them to engage in self-dialogue. This method offered a potential approach for self-counseling, resulting in better psychological outcomes and a greater perception of change than interacting with a scripted character.
	Saad et al. [41]	They implemented four variants of virtual personal assistants that helped with email and calendar scheduling. The varying audio and visual immersion levels were designed to test whether higher visual and auditory immersion enhanced the user experience.
	Reinhardt et al. [47]	An AR HMD-based application with an ECA provided updated weather information. The study compared hyperrealistic and simply-designed humanoid characters.
	Safadel et al. [54]	A VR-based virtual librarian ECA provided information retrieval functionalities and library-related context inquiries. The study highlighted the potential of virtual librarians to enhance user experience in academic libraries.
Museum Tour Guide	Wang et al. [58]	The proposed VR museum tour guide system, powered by large language models, offered users multimodal interactions within a virtual museum. The study showed promising potential for engaging users in real-life adoption.
	Gan et al. [56]	A mobile-based AR museum touring application offered conversational interaction with a virtual guide that displayed emotional expression and animation. The study demonstrated that humanoid ECAs enhanced the believability of the agent and improved the tour experience.
Shopping/Customer Service	Zhu et al. [55]	The immersive VR shopping application in a pharmacy store incorporated a humanoid ECA with diverse output modalities, improving the virtual shopping experience.
	Kato et al. [53]	The developed XR Telexperience Portal connected the metaverse and real space with realistic, avatar-synthesized ECAs for communicating with customers.

B. RQ2: XR ECA APPLICATION FEATURES

Our analysis of dialogue structure showed that most studies favored a task-oriented dialogue structure designed to guide users toward specific goals or activities [44], [54]. These styles were goal-oriented, structured interactions where the virtual agent prompted users with questions or instructions, gathering task-relevant information to achieve a clear outcome. This approach was practical in scenarios

TABLE 7. The list of questionnaires, measurements, and qualitative methods reported in the resulting papers. NS denotes "not specified."

Page Title	Quantitative		Qualitative
	Questionnaires	Measurements	
Pan et al. [38]	Social anxiety; personality; presence	Waiting time	Comments; interviews
Heyselaar et al. [39]	Conflict; relationship	Response	NS
Hartano et al. [40]	Presence; anxiety	Heart rate	NS
Saad et al. [41]	Usability	NS	NS
Pejsa et al. [42]	Likeability; attractiveness; closeness; groupness; agent-behavior manipulation	Number of speaking turns; total speaking time	NS
Ochs et al. [43]	Presence; copresence	Motion captures	Audio; video
Herrero and Lorenzo [44]	Social and emotional reciprocity; non-verbal communication; inflexibility to changes; stereotypes and sensorial reactivity; performance in class	Number of interchanges in conversation; accumulated time of joint attention	Video
Ochs et al. [45]	Presence; copresence; perception of performance	NS	Audio; video
Slater et al. [46]	Depression; anxiety; perception of the problem; evaluation of changes; VR experience	NS	Interviews
Reinhardt et al. [47]	Attractiveness; usability	NS	Interviews
Guimarães et al. [48]	Social presence; usability	NS	NS
Nguyen et al. [49]	NS	Speech-to-query transcribing accuracy	Comments
Gorisse et al. [50]	Paranoid thought; paranoia; distress; anxiety; presence in virtual environment; body ownership; vicarious agency	NS	NS
Souchet et al. [51]	Quality of experience	Optometric measurements; game scores; response time	NS
Hassan et al. [52]	Quality of experience; responsiveness; flow; learning effect	NS	Comments
Kato et al. [53]	Correctness of mouth shape to speech voice; congruency of speech and facial expression; naturalness	NS	NS
Safadel et al. [54]	Technology acceptance; curiosity; intention to use	NS	NS
Zhu et al. [55]	Warmth; communication; trust; comfort; satisfaction	NS	Interviews
Gan et al. [56]	Sensory dimension; social dimension; affective dimension; behavioral dimension	Interaction duration; number of speech interaction	Interviews
Zhang et al. [57]	Usability; social presence	NS	Interviews
Wang et al. [58]	Understandability; focused attention; interest; novelty; engagement; usability; user comfort; presence; spatial awareness	NS	Comments; interviews
Spiegel et al. [59]	NS	NS	Interviews
Llanes-Jurado et al. [60]	Depression; anxiety; naturalness; realism; virtual character arousal and valence	Processing time; conversation duration; number of user's sentences	Audio

requiring clarity and precision, such as education and training, where efficiency and task completion were prioritized [64], [65], [66]. However, it often limits ECAs' flexibility to support more fluid and natural exchanges. Although less common, open-ended and therapeutic dialogue structures offered dynamic interactions that held promise for applications in mental health [59], [60]. Their ability to handle user-driven dialogue marked a step toward more engaging and context-sensitive ECAs, indicating a shift toward richer and more personalized experiences.

In terms of conversational style, we observed that most studies focused on a one-on-one conversation style. This trend might have resulted from the simplicity of conducting conditioned experiments under one-on-one human-agent interaction, as it was more straightforward for researchers to manage. Multi-party conversations require more complex verbal and nonverbal cues, as explored in Pejsa et al.'s [42] work. Nevertheless, with the ongoing advancement in immersive environments, we anticipate an increase in future XR studies that investigate multi-party conversational ECAs, as they can more closely mimic real-world social interactions and provide richer data in group settings.

Regarding rule-based and neural-based dialogue system implementations, we observed an increase in neural-based applications in 2023 and 2024. Rule-based systems functioned well for predefined tasks, but found it challenging to manage unstructured dialogue. Neural-based models, conversely, provided a significant leap in generating refined, context-aware responses. The results aligned with Schobel et al.'s [8] five waves of the evolution of conversational agents, where we are currently in the AI wave. With the advancement of LLMs, more complex and sophisticated dialogue systems have been developed and utilized. These models leveraged vast amounts of training data to generate responses that were not only contextually relevant but also adaptive to the conversation flow—something rule-based systems were not capable of. Moreover, neural-based dialogue systems provide a more dynamic conversational interaction with virtual agents, making them more realistic and engaging for users.

The analysis of software platforms revealed a strong preference for Unity, which was utilized in over half of the studies reviewed. This trend reflected Unity's popularity in virtual reality research due to its powerful toolset and

adaptability in handling complex 3D environments and AI-driven interactions. At the same time, a significant portion of the reviewed studies also showcased the development of custom solutions tailored to specific research needs. Custom-built platforms, such as the 3D video playback player developed by Ochs et al. [43] and the VRASP environment created by Nguyen et al. [49], demonstrated how bespoke systems could effectively address the limitations of generic engines, enabling more sophisticated control over agent behavior and user interactions. These platforms were particularly impactful in scenarios that required precise synchronization of verbal and nonverbal cues. Solutions like the Memphis system and the SitePal API have played a significant role in therapeutic and specialized applications, indicating a trend toward creating purpose-built systems closely aligned with the unique demands of mental health and social interaction scenarios. This focus on customized software suggested a growing need to tailor ECA designs to specific contexts, maximizing the effectiveness of user interactions. The contrast between using commercial game engines and creating bespoke platforms reflected a strategic approach in ECA development. Researchers have increasingly balanced the strengths of established software with the flexibility of custom-built solutions to enhance the realism, adaptability, and interactivity of virtual environments. This dual approach indicated the field's focus on refining user experiences through creative and practical software solutions.

In terms of input/output modality, voice input was preferred due to its natural alignment with real-time communication [49], [59]. However, gaze tracking and gesture-based inputs have been gaining importance, enabling more immersive and expressive user interactions [44], [50]. These multimodal input methods enhanced the interactivity of virtual environments by allowing more subtle and sophisticated exchanges, potentially enriching the user experience and broadening the applications of XR ECAs. Text-to-speech technology remained the most popular communication method, setting a high standard for vocal output [41]. When combined with facial expressions and gestures, these outputs elevated the realism of interactions, making ECAs more relatable and engaging [59]. The synchronization of speech with nonverbal cues significantly enhanced emotional resonance, which was vital for applications ranging from social simulations to therapeutic settings. This multimodal approach aligned well with creating ECAs that communicate effectively and connect with users on a more anthropomorphic level.

In summary, task-oriented dialogue structures remained the most prevalent, with open-ended and supportive approaches playing specialized roles in specific applications. The one-on-one conversational style was more dominant than multi-party conversation, featuring simpler turn-taking interactions. The transition from rule-based to neural-based dialogue systems signified a significant step toward more adaptive, contextually aware interactions, marking a significant evolution in virtual agent design. We observed a strong preference for Unity in the selected papers, while a growing trend toward custom-built platforms emerged to address specific research needs. Voice input remained the primary modality, complemented by gaze and gesture controls that enriched user engagement. Multimodal outputs, including text-to-speech, gestures, and facial animations, indicated a commitment to creating immersive and realistic virtual experiences. These advancements reflected ongoing efforts to develop ECAs that offered natural, engaging, and human-like interactions.

C. RQ3: XR ECA ATTRIBUTES

Most studies employed human-like virtual agents in XR as the embodiment method. For example, the ECA acted as a virtual patient in Ochs et al.'s [45] and Guimarães et al.'s [48] studies, which employed XR ECA as a suspect in interrogation conversations. Additionally, we found that most studies utilized female virtual agents in cases where the agents were supportive and assistive, such as guides [56], [58], information assistants [41], [54], or customer service agents [53], [55]. This finding aligns with Zimmerman et al.'s [67] finding, which suggests that people perceive female virtual agents as more supportive.

Regarding the representation and scale of XR ECAs, our results revealed that most studies used full-body and life-size virtual character 3D models. This reflects our findings on use cases, which indicated that researchers primarily employed XR ECAs in simulated scenarios. By using full-body, life-size models, researchers aimed to create immersive and lifelike experiences. Interestingly, all VR papers employed life-sized ECAs, while the AR paper employed a miniature XR ECA. Based on our findings in device categories and previous studies [68], [69], we argued that the miniature ECA was likely designed for smartphone applications.

In our analysis of the mobility of XR ECAs, we found that most XR ECAs were stationary rather than mobile. This design choice seemed to support conversations between humans and XR ECAs. Specifically, all stationary XR ECAs in our resulting papers were placed in front of users, facilitating interactions. On the contrary, for mobile XR ECAs, users sometimes had to approach the XR ECA to communicate with it. The user-centered interaction design helped explore the impacts of XR ECAs in terms of sensitive psychological topics, such as autism spectrum disorders [44] or paranoia [50].

Last, we observed a variety of expressions used by XR ECAs, including gestures, lip-sync, facial expressions, eye gaze, head orientation, and spatial orientation. Gesture and lip-sync were the most commonly employed expressions among the selected papers. Moreover, most papers used at least two expressions, highlighting the importance of thoughtfully designing XR ECAs.

D. RQ4: XR ECA USE CASES

From our analysis of the resulting papers, we observe that the most substantial application of XR ECAs was in training and education. XR ECAs were used to simulate real-world scenarios, enabling learners to practice without real-world consequences. Applications included medical students practicing diagnostic and communication skills with virtual patients [43], [45], child protection workers refining interview techniques [52], police officers enhancing interrogation strategies [48], autistic children developing social skills [44], users engaging in interview mockups [51], and interactive learning experiences in domains such as cooking [57] and programming [49]. These findings highlight the potential of XR ECAs in promoting experiential learning across various disciplines.

The second highest use of XR ECAs among the resulting papers was in social conversations, particularly in experimental settings where human interaction dynamics are studied. Applications included exploring personality perception in social dynamics [38], neurolinguistic behavior in language processing [39], and nonverbal cues in group conversations [42]. The integration of XR ECAs for social experiments has provided researchers with flexible and immersive environments to study complex human interactions that would be difficult or unethical to replicate in real life. Take Neyret et al.'s [70] study as an example, in which investigating victim perspectives in harassment scenarios would be ethically challenging to conduct with actual participants.

The therapy and counseling use case was also the second most common category. In these papers, XR ECAs offered users a nonjudgmental, readily accessible, and confidential support system. They were programmed to help users manage anxiety, depression, and other mental health issues [40], [50], [59]. Additionally, self-counseling through avatar switching leveraged XR ECAs to enable perspective-taking and self-reflection [46], a capability not possible in traditional counseling.

In the information retrieval and assistance category, we observed XR ECAs' ability to converse and assist through context-aware interactions. Applications included a holographic weather forecaster [47], a virtual personal assistant that optimizes email retrieval in various display formats [41], and virtual librarians that facilitate efficient knowledge navigation [54]. These use cases highlighted the potential of XR ECAs in enhancing efficiency, accessibility, and user engagement for digital assistance.

For use cases in museum tour guiding, XR ECAs demonstrated their abilities to enrich visitor experiences by offering personalized, multilingual guided tours. Applications included XR ECAs with LLM-powered navigation features [58] and AR guides with expressive humanoid avatars [56].

Similarly, XR ECAs in shopping and customer service showed the potential to transform retail experiences by providing product guidance, personalized recommendations, and virtual customer support. Applications such as a pharmacy shopping assistant [55] and a customer service agent with realistic facial expressions [53] suggested that XR ECAs could enhance convenience and engagement in virtual commerce.

E. RQ5: RATINGS, MEASUREMENTS, AND QUALITATIVE METHODS

Most of the studies in our resulting papers employed questionnaires to explore the impacts of XR ECAs and evaluate their XR system. Specifically, we identified four categories of questionnaires: system evaluation, perception of XR ECAs, user experiences, and psychological assessment. The system evaluation included numerous variables, such as usability [41], [47], [57], [58], technology acceptance [54], and learning effects [52]. Specifically, the System Usability Scale (SUS) [71] was employed to evaluate the usability of the developed applications. Regarding the perception of XR ECAs, researchers have investigated various variables, such as attractiveness [42], [47], [56], trust [55], and believability [48]. Most variables have been associated with positive perceptions and have revealed an interest in enhancing the user experience through XR ECAs. In the same vein, we found that most questionnaires about user experiences focused on aspects such as presence [36], [38], [39], [41], [52], [54], social presence [45], [46], and quality of experience [51], [52]. The Igroup Presence Questionnaire (IPQ) [72] has been primarily used to assess presence. Lastly, psychological assessments have mainly focused on anxiety or depression [38], [46], [50], [60]. Specifically, these assessments have incorporated various standardized questionnaires, including the Automatic Thoughts Questionnaire (ATQ) [73], the Depression Anxiety Stress Scales (DASS) [74], and the State-Trait Anxiety Inventory (STAI) [75].

Less than half of the selected papers collected quantitative measurements, focusing on conversation-related variables. Specifically, Gan et al. [56] measured interaction duration and the number of speech interactions, interpreting these variables as indicators of users' behavioral engagement. Also, Llanes-Jurado et al. [60] measured the level of dominance in a conversation by comparing the number of sentences between human participants and XR ECAs. Furthermore, researchers employed motion capture [43] to analyze nonverbal behaviors and heart rate [40] to assess participants' anxiety levels. Regarding qualitative methods, researchers conducted semi-structured interviews and collected comments and feedback at the end of participants' experiences. The authors provided questions about participants' perceptions of their interactions, the system's utility, the ECA's appearance, behavior, body language, and trustworthiness, as well as which aspects they preferred and why. The analysis included self-coded categories or thematic analysis for an in-depth understanding of participants' comments. Other analyses involved word frequency [45] and positive versus negative ratings [55].

F. IMPLICATIONS

1) TRENDS IN COMMON TECHNOLOGIES

The current trends in XR ECA applications demonstrate a distinct preference for XR technologies, particularly those utilizing VR HMDs and CAVE systems, which reflect their ability to provide a more immersive and situated environment. Game engines, such as Unity and Unreal Engine, are the dominant approach for developing such applications, offering a rich toolkit for 3D creation and interactive design. Additionally, the growing integration of LLMs has allowed more adaptive and dynamic conversational systems, shifting rule-based dialogue systems to neural-based ones. The increasing use of multimodal inputs, such as voice, gaze, and gestures, and outputs, such as text-to-speech, facial expressions, and spatial orientation, further enriches user engagement by creating richer and more responsive interactions.

2) GUIDELINES ON ECA DESIGN

Designing effective ECAs requires attention to several key factors, including the agent's persona, gender, and role. The gender of virtual agents often aligns with their functional roles, with a preference for female agents in supportive positions, such as guides or assistants [67]. Additionally, the choice of agent persona may also impact the ECA's trustworthiness. An agent's expertise should align with the application's use cases to avoid a sense of eeriness [76]. In terms of modalities, the design of ECAs should leverage non-verbal communication features, such as gestures, facial expressions, and gaze, which significantly contribute to creating life-like interactions and enhancing the overall experience.

3) GUIDELINES FOR CONDUCTING ECA-RELATED USER RESEARCH

To evaluate ECAs effectively, researchers should adopt a comprehensive approach to data collection and analysis. The most common measurements include usability scales, such as the SUS, and presence measures, like the IPQ. Studies often measure user perceptions of agent attractiveness [42], [47], [56]. Interviews and participant comments further help to uncover user experiences and provide context for the quantitative findings, ensuring a well-rounded evaluation of the ECA's impact.

VI. LIMITATIONS

Our study encountered several limitations that should be taken into consideration. While these limitations do not undermine the validity of our findings, they frame how the results should be interpreted and offer directions for future research. The selected 23 papers represented a wide diversity of XR ECA applications. Nevertheless, our selection criterion, while rigorous, may provide selection biases as our research limitation. One significant limitation was the scope of the literature reviewed. We primarily focused on peer-reviewed journals and conference papers. As a result, it may only capture part of the rapidly evolving landscape of ECA and XR research. Given the vast amount of available literature and the continuous emergence of new studies, some relevant and recent contributions, including book chapters, technical reports, dissertation theses, and industry reports, were excluded, even though they could have been influential to the field.

Another limitation pertained to the screening criteria employed. In our designed query, we utilized "reality" instead of all related terms such as "virtual reality," "augmented reality," and "extended reality," due to limitations on the number of Boolean operators in IEEE Xplore and ScienceDirect. We opted for "reality" as it broadly encompasses these terms. This limitation of our selection methodology may result in the omission of relevant studies. Additionally, we applied a stringent criterion to exclude publications with an ACR below 1.5. This approach has generally been considered reliable for filtering higher-quality, relevant research. However, it may have excluded pertinent studies within the domain of XR ECAs that were highly relevant to the review but had lower citation rates due to their niche focus areas [77], [78], [79]. Furthermore, the criterion requiring publications to exceed five pages was intended to ensure that the research provided sufficient detail in its implementation and experimental procedures. However, this may have inadvertently omitted shorter reports or conference abstracts that, while concise, contained valuable insights [80], [81].

Additional criteria, such as a focus on turn-taking conversational interactions or XR implementations without traditional screen displays, further narrowed the scope of the papers. A few excluded studies were designed in what André and Pelachaud [22] referred to as TV-style communication, where no turn-taking interactions were required between users and the agents. Most of these were in the domain of museum and virtual tour guides [82], [83], while a few were pedagogical agents [84], [85]. Others included avatar embodiment for social experiments [41], [70]. Although these studies provided insightful results regarding the implementation pipeline and design guidelines for communicative XR ECAs, the human-agent interaction model did not fit Cassell et al.'s [11] definition of ECAs. We excluded XR applications where agents were displayed on screen-based systems with multimodal sensory capabilities, such as a desktop setup with a Kinect sensor [86], [87] or an immersive CAVE system where the ECA was displayed on smaller screens [88]. Even though some consider these systems XR in a broad sense, we did not include them in our screening. We only included screen-based XR applications when they were implemented on mobile devices to avoid ambiguous filtering.

Among the selected papers, we did not find an emphasis on privacy restrictions or ethical considerations in the design of XR ECAs. Nevertheless, this should be an essential area for future exploration in the progression of XR ECAs. Moreover, hardware limitations such as motion sickness, restricted field

of view, and limited support for device mobility are also areas that hinder widespread adoption of XR ECAs and could be investigated in future research. In addition, we aimed to provide pointers to the existing source code of the selected papers as a foundation for future development. However, among the 23 selected papers, only two [49], [60] provided open-source code, which limits the reproducibility of the results and the potential for further development or validation by the research community.

Our methodology aimed to narrow the scope of the target research and provide insight into the current state of XR ECAs by focusing on high-impact studies that demonstrated the implementation of technology and user evaluation. However, these limitations restricted the review's comprehensiveness and may not have fully addressed all contributions that could further benefit the development and design of XR ECAs. In future research, these limitations should be considered to expand the inclusion criteria and cover a broader range of studies.

VII. CONCLUSION

This review highlighted a significant shift in focus from traditional CA applications to more immersive settings for ECAs. We identified a current research gap at the intersection of conversational human-agent interaction and embodied agents in XR. Our systematic review employed the PRISMA framework to analyze the results of XR technologies and devices, including application features, agent attributes, use cases, and evaluation methods. Looking ahead, as the capabilities of immersive and AI technologies continue to advance, more sophisticated and human-like ECAs will be introduced to enhance user engagement and effectiveness in various XR applications. Our work shed light on current trends and explorations of XR ECAs, offering insights that can guide future research in this dynamic field.

REFERENCES

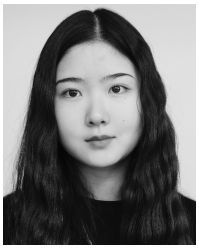
[1] J. Lester, K. Branting, and B. Mott, "Conversational agents," in *The Practical Handbook of Internet Computing*, 2004, pp. 220–240.
[2] J. Cassell, "Embodied conversational interface agents," *Commun. ACM*, vol. 43, no. 4, pp. 70–78, Apr. 2000.
[3] B. Reeves and C. Nass, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People*, vol. 10. Cambridge, U.K.: Cambridge Univ. Press, 1996, pp. 19–36.
[4] W. Swartout, J. Gratch, R. W. Hill, E. Hovy, S. Marsella, J. Rickel, and D. Traum, "Toward virtual humans," *AI Mag.*, vol. 27, no. 2, pp. 96–108, Jul. 2006.
[5] B. Lugrin, C. Pelachaud, and D. Traum, *The Handbook on Socially Interactive Agents: 20 Years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Interactivity, Platforms, Application*, vol. 2. New York, NY, USA: ACM, 2022.
[6] M. M. E. Van Pinxteren, M. Pluymaekers, and J. G. A. M. Lemmink, "Human-like communication in conversational agents: A literature review and research agenda," *J. Service Manage.*, vol. 31, no. 2, pp. 203–225, Jun. 2020.
[7] B. Weiss, I. Wechsung, C. Kühnel, and S. Möller, "Evaluating embodied conversational agents in multimodal interfaces," *Comput. Cognit. Sci.*, vol. 1, no. 1, pp. 1–21, Dec. 2015.
[8] S. Schöbel, A. Schmitt, D. Benner, M. Saqr, A. Janson, and J. M. Leimeister, "Charting the evolution and future of conversational agents: A research agenda along five waves and new frontiers," *Inf. Syst. Frontiers*, vol. 26, no. 2, pp. 729–754, Apr. 2024.
[9] A. Liberati, D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gøtzsche, J. P. Ioannidis, M. Clarke, P. J. Devereaux, J. Kleijnen, and D. Moher, "The prisma statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration," *Ann. Internal Med.*, vol. 151, no. 4, p. 65, 2009.
[10] J. Cassell, J. Sullivan, S. Prevost, and E. F. Churchill, *Embodied Conversational Agents*. Cambridge, MA, USA: MIT Press, 2000.
[11] J. Cassell, T. Bickmore, L. Campbell, H. Vilhjalmsson, and H. Yan, "Human conversation as a system framework: Designing embodied conversational agents," in *Embodied Conversational Agents*, 2000, pp. 29–63.
[12] N. Magnenat-Thalmann and D. Thalmann, "An overview of virtual humans," in *Handbook of Virtual Humans*, 2004, pp. 1–25.
[13] Z. Ruttkay, C. Dormann, and H. Noot, "Embodied conversational agents on a common ground," in *From Brows To Trust*, 2004.
[14] S. Castillo, P. Hahn, K. Legde, and D. W. Cunningham, "Personality analysis of embodied conversational agents," in *Proc. 18th Int. Conf. Intell. Virtual Agents*, Nov. 2018, pp. 227–232.
[15] P. Sajjadi, L. Hoffmann, P. Cimiano, and S. Kopp, "A personality-based emotional model for embodied conversational agents: Effects on perceived social presence and game experience of users," *Entertainment Comput.*, vol. 32, Dec. 2019, Art. no. 100313.
[16] M. Neff, Y. Wang, R. Abbott, and M. Walker, "Evaluating the effect of gesture and language on personality perception in conversational agents," in *Proc. Int. Conf. Intell. Virtual Agents*, Philadelphia, PA, USA. Springer, 2010, pp. 222–235.
[17] A. Nijholt, "Humor and embodied conversational agents," Tech. Rep., 2003.
[18] R. Beale and C. Creed, "Affective interaction: How emotional agents affect users," *Int. J. Hum.-Comput. Stud.*, vol. 67, no. 9, pp. 755–776, Sep. 2009.
[19] C. Pelachaud, "Multimodal expressive embodied conversational agents," in *Proc. 13th Annu. ACM Int. Conf. Multimedia*, Nov. 2005, pp. 683–689.
[20] J. Lee and S. Marsella, "Nonverbal behavior generator for embodied conversational agents," in *Proc. Int. Workshop Intell. Virtual Agents*, Cham, Switzerland. Springer, 2006, pp. 243–255.
[21] N. Novielli, F. de Rosis, and I. Mazzotta, "User attitude towards an embodied conversational agent: Effects of the interaction mode," *J. Pragmatics*, vol. 42, no. 9, pp. 2385–2397, Sep. 2010.
[22] E. André and C. Pelachaud, "Interacting with embodied conversational agents," in *Speech Technology: Theory and Applications*, 2010, pp. 123–149.
[23] M. Yousefi, S. E. Crowe, S. Hoermann, M. Sharifi, A. Romera,A. Shahi, and T. Piumsomboon, "Advancing prosociality in extended reality: Systematic review of the use of embodied virtual agents to trigger prosocial behaviour in extended reality," *Frontiers Virtual Reality*, vol. 5, May 2024, Art. no. 1386460.
[24] J. Weizenbaum, "ELIZA—A computer program for the study of natural language communication between man and machine," *Commun. ACM*, vol. 9, no. 1, pp. 36–45, Jan. 1966.
[25] R. S. Wallace, "The anatomy of ALICE," in *Parsing the Turing Test*. Cham, Switzerland: Springer, 2009.
[26] C. Rzepka, B. Berger, and T. Hess, "Voice assistant vs. Chatbot–examining the fit between conversational agents' interaction modalities and information search tasks," *Inf. Syst. Frontiers*, vol. 24, no. 3, pp. 839–856, Jun. 2022.
[27] K. Loveys, G. Sebaratnam, M. Sagar, and E. Broadbent, "The effect of design features on relationship quality with embodied conversational agents: A systematic review," *Int. J. Social Robot.*, vol. 12, no. 6, pp. 1293–1312, Dec. 2020.
[28] S. Provoost, H. M. Lau, J. Ruwaard, and H. Riper, "Embodied conversational agents in clinical psychology: A scoping review," *J. Med. Internet Res.*, vol. 19, no. 5, p. e151, May 2017.
[29] S. ter Stal, L. L. Kramer, M. Tabak, H. O. den Akker, and H. Hermens, "Design features of embodied conversational agents in eHealth: A literature review," *Int. J. Hum.-Comput. Stud.*, vol. 138, Jun. 2020, Art. no. 102409.
[30] L. L. Kramer, S. ter Stal, B. C. Mulder, E. de Vet, and L. van Velsen, "Developing embodied conversational agents for coaching people in a healthy lifestyle: Scoping review," *J. Med. Internet Res.*, vol. 22, no. 2, Feb. 2020, Art. no. e14058.
[31] B. Khosrawi-Rad, H. Rinn, R. Schlimbach, P. Gebbing, X. Yang, C. Lattemann, D. Markgraf, and S. Robra-Bissantz, "Conversational agents in education—A systematic literature review," Tech. Rep., 2022.

[32] S. Hobert and R. Meyer von Wolff, "Say hello to your new automated tutor—A structured literature review on pedagogical conversational agents," Tech. Rep., 2019.
[33] A. R. D. B. Landim, A. M. Pereira, T. Vieira, E. de B. Costa, J. A. B. Moura, V. Wanick, and E. Bazaki, "Chatbot design approaches for fashion e-commerce: An interdisciplinary review," *Int. J. Fashion Design, Technol. Educ.*, vol. 15, no. 2, pp. 200–210, May 2022.
[34] T. Hirzle, F. Müller, F. Draxler, M. Schmitz, P. Knierim, and K. Hornbæk, "When XR and AI meet—A scoping review on extended reality and artificial intelligence," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Apr. 2023, pp. 1–45.
[35] N. Norouzi, K. Kim, G. Bruder, A. Erickson, Z. Choudhary, Y. Li, and G. Welch, "A systematic literature review of embodied augmented reality agents in head-mounted display environments," in *Proc. Int. Conf. Artif. Reality Telexistence Eurograph. Symp. Virtual Environ.*, Jan. 2020, pp. 101–111.
[36] U. Radhakrishnan, K. Koumaditis, and F. Chinello, "A systematic review of immersive virtual reality for industrial skills training," *Behav. Inf. Technol.*, vol. 40, no. 12, pp. 1310–1339, Sep. 2021.
[37] B. I. Hutchins, X. Yuan, J. M. Anderson, and G. M. Santangelo, "Relative citation ratio (RCR): A new metric that uses citation rates to measure influence at the article level," *PLOS Biol.*, vol. 14, no. 9, Sep. 2016, Art. no. e1002541.
[38] X. Pan, M. Gillies, and M. Slater, "Virtual character personality influences participant attitudes and behavior—An interview with a virtual human character about her social anxiety," *Frontiers Robot. AI*, vol. 2, p. 1, Feb. 2015.
[39] E. Heyselaar, P. Hagoort, and K. Segaert, "In dialogue with an avatar, language behavior is identical to dialogue with a human partner," *Behav. Res. Methods*, vol. 49, no. 1, pp. 46–60, Feb. 2017.
[40] D. Hartanto, W. Brinkman, I. L. Kampmann, N. Morina, P. G. M. Emmelkamp, and M. A. Neerincx, "Home-based virtual reality exposure therapy with virtual health agent support," in *Proc. Int. Symp. Pervasive Comput. Paradigms Mental Health*, Sep. 2016, pp. 85–98.
[41] U. Saad, U. Afzal, A. El-Issawi, and M. Eid, "A model to measure QoE for virtual personal assistant," *Multimedia Tools Appl.*, vol. 76, no. 10, pp. 12517–12537, May 2017.
[42] T. Pejsa, M. Gleicher, and B. Mutlu, "Who, me? How virtual agents can shape conversational footing in virtual reality," in *Proc. Int. Conf. Intell. Virtual Agents*, Stockholm, Sweden. Cham, Switzerland: Springer, Aug. 2017, pp. 347–359.
[43] O. Magalie, J. Sameer, and B. Philippe, "Toward an automatic prediction of the sense of presence in virtual reality environment," in *Proc. 6th Int. Conf. Human-Agent Interact.*, Dec. 2018, pp. 161–166.
[44] J. F. Herrero and G. Lorenzo, "An immersive virtual reality educational intervention on people with autism spectrum disorders (ASD) for the development of communication skills and problem solving," *Educ. Inf. Technol.*, vol. 25, no. 3, pp. 1689–1722, May 2020.
[45] M. Ochs, D. Mestre, G. de Montcheuil, J.-M. Pergandi, J. Saubesty, E. Lombardo, D. Francon, and P. Blache, "Training doctors' social skills to break bad news: Evaluation of the impact of virtual environment displays on the sense of presence," *J. Multimodal User Interfaces*, vol. 13, no. 1, pp. 41–51, Mar. 2019.
[46] M. Slater, S. Neyret, T. Johnston, G. Iruretagoyena, M. Á. D. L. C. Crespo, M. Alabèrnia-Segura, B. Spanlang, and G. Feixas, "An experimental study of a virtual reality counselling paradigm using embodied self-dialogue," *Sci. Rep.*, vol. 9, no. 1, p. 10903, Jul. 2019.
[47] J. Reinhardt, L. Hillen, and K. Wolf, "Embedding conversational agents into AR: Invisible or with a realistic human body?" in *Proc. 14th Int. Conf. Tangible, Embedded, Embodied Interact.*, Feb. 2020, pp. 299–310.
[48] M. Guimarães, R. Prada, P. A. Santos, J. Dias, A. Jhala, and S. Mascarenhas, "The impact of virtual reality in the social presence of a virtual agent," in *Proc. 20th ACM Int. Conf. Intell. Virtual Agents*, Oct. 2020, pp. 1–8.
[49] V. T. Nguyen, Y. Zhang, K. Jung, W. Xing, and T. Dang, "VRASP: A virtual reality environment for learning answer set programming," in *Proc. Int. Symp. Practical Aspects Declarative Lang.*, New Orleans, LA, USA, Jan. 2020, pp. 82–91.
[50] G. Gorisse, G. Senel, D. Banakou, A. Beacco, R. Oliva, D. Freeman, and M. Slater, "Self-observation of a virtual body-double engaged in social interaction reduces persecutory thoughts," *Sci. Rep.*, vol. 11, no. 1, p. 23923, Dec. 2021.
[51] A. D. Souchet, S. Philippe, A. Lévéque, F. Ober, and L. Leroy, "Short- and long-term learning of job interview with a serious game in virtual reality: Influence of eyestrain, stereoscopy, and apparatus," *Virtual Reality*, vol. 26, no. 2, pp. 583–600, Jun. 2022.
[52] S. Z. Hassan, P. Salehi, R. K. Røed, P. Halvorsen, G. A. Baugerud, M. S. Johnson, P. Lison, M. Riegler, M. E. Lamb, C. Griwodz, and S. S. Sabet, "Towards an AI-driven talking avatar in virtual reality for investigative interviews of children," in *Proc. 2nd Workshop Games Syst.*, Jun. 2022, pp. 9–15.
[53] R. Kato, Y. Kikuchi, V. Yem, and Y. Ikei, "Reality avatar for customer conversation in the metaverse," in *Proc. Int. Conf. Human-Comput. Interact.*, Jan. 2022, pp. 131–145.
[54] P. Safadel, S. N. Hwang, and J. M. Perrin, "User acceptance of a virtual librarian chatbot: An implementation method using IBM watson natural language processing in virtual immersive environment," *TechTrends*, vol. 67, no. 6, pp. 891–902, Nov. 2023.
[55] S. Zhu, W. Hu, W. Li, and Y. Dong, "Virtual agents in immersive virtual reality environments: Impact of humanoid avatars and output modalities on shopping experience," *Int. J. Human–Comput. Interact.*, vol. 40, no. 19, pp. 5771–5793, Oct. 2024.
[56] Q. Gan, Z. Liu, T. Liu, Y. Zhao, and Y. Chai, "Design and user experience analysis of AR intelligent virtual agents on smartphones," *Cognit. Syst. Res.*, vol. 78, pp. 33–47, Mar. 2023.
[57] J. Zhang, B. Han, Z. Dong, R. Wen, G. A. Lee, S. Hoermann, W. Zhang, and T. Piumsomboon, "Virtual triplets: A mixed modal synchronous and asynchronous collaboration with human-agent interaction in virtual reality," in *Proc. Extended Abstr. CHI Conf. Human Factors Comput. Syst.*, May 2024, pp. 1–8.
[58] Z. Wang, L.-P. Yuan, L. Wang, B. Jiang, and W. Zeng, "VirtuWander: Enhancing multi-modal interaction for virtual tour guidance through large language models," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2024, pp. 1–20.
[59] B. M. R. Spiegel, O. Liran, A. Clark, J. S. Samaan, C. Khalil, R. Chernoff, K. Reddy, and M. Mehra, "Feasibility of combining spatial computing and AI for mental health support in anxiety and depression," *NPJ Digit. Med.*, vol. 7, no. 1, p. 22, Jan. 2024.
[60] J. Llanes-Jurado, L. Gómez-Zaragozá, M. E. Minissi, M. Alcañiz, and J. Marín-Morales, "Developing conversational virtual humans for social emotion elicitation based on large language models," *Expert Syst. Appl.*, vol. 246, Jul. 2024, Art. no. 123261.
[61] I. Wang, J. Smith, and J. Ruiz, "Exploring virtual agents for augmented reality," in *Proc. CHI Conf. Human Factors Comput. Syst.*, May 2019, pp. 1–12.
[62] J. Q. Coburn, I. Freeman, and J. L. Salmon, "A review of the capabilities of current low-cost virtual reality technology and its potential to enhance the design process," *J. Comput. Inf. Sci. Eng.*, vol. 17, no. 3, Sep. 2017, Art. no. 031013.
[63] T. Combe, J.-R. Chardonnet, F. Merienne, and J. Ovtcharova, "CAVE and HMD: Distance perception comparative study," *Virtual Reality*, vol. 27, no. 3, pp. 2003–2013, Sep. 2023.
[64] K.-C. Li, M. Chang, and K.-H. Wu, "Developing a task-based dialogue system for English language learning," *Educ. Sci.*, vol. 10, no. 11, p. 306, Oct. 2020.
[65] A. Rese and P. Tränkner, "Perceived conversational ability of task-based chatbots – which conversational elements influence the success of text-based dialogues?" *Int. J. Inf. Manage.*, vol. 74, Feb. 2024, Art. no. 102699.
[66] T. E. Kim and A. Lipani, "A multi-task based neural model to simulate users in goal oriented dialogue systems," in *Proc. 45th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2022, pp. 2115–2119.
[67] J. Zimmerman, E. Ayoob, J. Forlizzi, and M. McQuaid, "Putting a face on embodied interface agents," Tech. Rep., 2005.
[68] H. Jeong, J. H. Yoo, and H. Song, "Virtual agents with augmented reality in digital healthcare," in *Proc. 13th Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2022, pp. 2016–2021.
[69] Q. Gan, Z. Liu, T. Liu, and Y. Chai, "An indoor evacuation guidance system with an AR virtual agent," *Proc. Comput. Sci.*, vol. 213, pp. 636–642, Jan. 2022.
[70] S. Neyret, X. Navarro, A. Beacco, R. Oliva, P. Bourdin, J. Valenzuela, I. Barberia, and M. Slater, "An embodied perspective as a victim of sexual harassment in virtual reality reduces action conformity in a later milgram obedience scenario," *Sci. Rep.*, vol. 10, no. 1, p. 6207, Apr. 2020.
[71] J. Brooke, "SUS-A quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, no. 194, pp. 4–7, 1996.

[72] T. W. Schubert, "The sense of presence in virtual environments: A three-component scale measuring spatial presence, involvement, and realness," *Zeitschrift Für Medienpsychologie*, vol. 15, no. 2, pp. 69–71, Apr. 2003.
[73] S. D. Hollon and P. C. Kendall, "Cognitive self-statements in depression: Development of an automatic thoughts questionnaire," *Cognit. Therapy Res.*, vol. 4, no. 4, pp. 383–395, Dec. 1980.
[74] P. F. Lovibond and S. H. Lovibond, "The structure of negative emotional states: Comparison of the depression anxiety stress scales (DASS) with the beck depression and anxiety inventories," *Behav. Res. Therapy*, vol. 33, no. 3, pp. 335–343, Mar. 1995.
[75] C. D. Spielberger, F. Gonzalez-Reigosa, A. Martinez-Urrutia, L. F. Natalicio, and D. S. Natalicio, "The state-trait anxiety inventory," *Revista Interamer. De Psicologia/Interamer. J. Psychol.*, vol. 5, no. 3, pp. 1–15, 1971.
[76] F.-C. Yang, K. Duque, and C. Mousas, "The effects of depth of knowledge of a virtual agent," *IEEE Trans. Vis. Comput. Graph.*, vol. 30, no. 11, pp. 7140–7151, Nov. 2024.
[77] J. Zhu, R. Kumaran, C. Xu, and T. Höllerer, "Free-form conversation with human and symbolic avatars in mixed reality," in *Proc. IEEE Int. Symp. Mixed Augmented Reality (ISMAR)*, Oct. 2023, pp. 751–760.
[78] X.-D. Jhan, S.-K. Wong, E. Ebrahimi, Y. Lai, W.-C. Huang, and S. V. Babu, "Effects of small talk with a crowd of virtual humans on users' emotional and behavioral responses," *IEEE Trans. Vis. Comput. Graph.*, vol. 28, no. 11, pp. 3767–3777, Nov. 2022.
[79] M. Ma, S. Coward, and C. Walker, "Interact: A mixed reality virtual survivor for holocaust testimonies," in *Proc. Annu. Meeting Austral. Special Interest Group Comput. Human Interact.*, Dec. 2015, pp. 250–254.
[80] A. Hartholt, E. Fast, A. Reilly, W. Whitcup, M. Liewer, and S. Mozgai, "Ubiquitous virtual humans: A multi-platform framework for embodied AI agents in XR," in *Proc. IEEE Int. Conf. Artif. Intell. Virtual Reality (AIVR)*, Dec. 2019, pp. 308–3084.
[81] N. Bouali, E. Nygren, S. S. Oyelere, J. Suhonen, and V. Cavalli-Sforza, "Imikode: A VR game to introduce OOP concepts," in *Proc. 19th Koli Calling Int. Conf. Comput. Educ. Res.*, Nov. 2019, pp. 1–2.
[82] R. Hammady, M. Ma, C. Strathern, and M. Mohamad, "Design and development of a spatial mixed reality touring guide to the Egyptian museum," *Multimedia Tools Appl.*, vol. 79, nos. 5–6, pp. 3465–3494, Feb. 2020.
[83] A. Bönsch, D. Hashem, J. Ehret, and T. W. Kuhlen, "Being guided or having exploratory freedom: User preferences of a virtual agent's behavior in a museum," in *Proc. 21st ACM Int. Conf. Intell. Virtual Agents*, Sep. 2021, pp. 33–40.
[84] F.-C. Yang, C. Mousas, and N. Adamo, "Holographic sign language avatar interpreter: A user interaction study in a mixed reality classroom," *Comput. Animation Virtual Worlds*, vol. 33, nos. 3–4, p. 2082, Jun. 2022.
[85] D. Peeters, "Bilingual switching between languages and listeners: Insights from immersive virtual reality," *Cognition*, vol. 195, Feb. 2020, Art. no. 104107.
[86] A. Robb, C. White, A. Cordar, A. Wendling, S. Lampotang, and B. Lok, "A comparison of speaking up behavior during conflict with real and virtual humans," *Comput. Hum. Behav.*, vol. 52, pp. 12–21, Nov. 2015.
[87] S. L. Burke, T. Bresnahan, T. Li, K. Epnere, A. Rizzo, M. Partin, R. M. Ahlness, and M. Trimmer, "Using virtual interactive training agents (ViTA) with adults with autism and other developmental disabilities," *J. Autism Develop. Disorders*, vol. 48, no. 3, pp. 905–912, Mar. 2018.
[88] H. Hofmann, V. Tobisch, U. Ehrlich, and A. Berton, "Evaluation of speech-based HMI concepts for information exchange tasks: A driving simulator study," *Comput. Speech Lang.*, vol. 33, no. 1, pp. 109–135, Sep. 2015.

FU-CHIA YANG received the B.S. degree in computer science from The Chinese University of Hong Kong and the M.S. degree in computer graphics technology from Purdue University, where she is currently pursuing the Ph.D. degree with the Department of Computer Graphics Technology. Her research interests include conversational agents in extended reality, virtual humans, games, and human–computer interaction.

PEDRO ACEVEDO received the B.S. and M.S. degrees in software engineering from the Universidad del Norte, Colombia, in 2020 and 2021, respectively. He is currently pursuing the Ph.D. degree with the Department of Computer Graphics Technology, Purdue University. His research interests include virtual reality, educational technology, human–computer interaction, graphics, procedural content generation, and game development.



SIQI GUO received the B.S. and M.S. degrees in animation and computer graphics technology from Purdue University, West Lafayette, IN, USA, where she is currently pursuing the Ph.D. degree with the Department of Computer Graphics Technology, where she specializes in virtual reality and human–computer interaction, with a particular focus on virtual humans.

MINSOO CHOI received the B.E. and M.S. degrees in computer engineering from Hongik University, Seoul, Republic of Korea. He is currently pursuing the Ph.D. degree with the Department of Computer Graphics Technology, Purdue University. His research interests include virtual reality, intelligent virtual agents, human–computer interaction, and computer games.

CHRISTOS MOUSAS (Member, IEEE) received the five-year integrated master's (B.Sc. and M.Sc.) degrees in audiovisual science and art from the Department of Audio and Visual Arts, Ionian University, in 2009, and the M.Sc. degree in multimedia applications and virtual environments and the Ph.D. degree in informatics from the Department of Informatics (School of Engineering and Informatics), University of Sussex, in 2011 and 2014, respectively. From 2015 to 2016, he was a Postdoctoral Researcher with the Department of Computer Science, Dartmouth College. He is currently an Associate Professor and the Director of the Virtual Reality Laboratory, Department of Computer Graphics Technology, and a Core Faculty Member of the Applied AI Research Center, Purdue University. His research interests include virtual reality, virtual humans, intelligent virtual agents, computer graphics and animation, and human–computer interaction. He is a member of ACM. He has received the Best Paper Award from IEEE ISMAR, in 2024, ACM TiiS, in 2023, and ACM SIGGRAPH VRCAI, in 2022, as well as an Honorable Mention Award (top 5%) from ACM CHI, in 2022, and ACM CHI PLAY, in 2021. He serves as Associate Editor for the *Computer Animation and Virtual Worlds* and *Frontiers in Virtual Reality*, as well as on the organizing and program committees of numerous conferences in the fields of virtual reality, computer graphics/animation, and human–computer interaction.