

Chatbots integrated in the metaverse: a conceptual model based on interaction ritual chain theory

Yan Yan¹ and Mengjuan Fan²

¹ Faculty of Data Science, City University of Macau, Macau, China

² Culture, Media, Tourism and Sports Bureau of Jiangmen Municipality, Jiangmen, China

ABSTRACT

Chatbots are increasingly prevalent as conversational agents in the metaverse, enabled by commercial potential and artificial intelligence (AI) advancements that facilitate natural human-computer interaction (HCI). However, integrating chatbots as virtual community members remains challenging due to deficiencies in conversational and social intelligence. Current metaverse research prioritizes foundational infrastructure while neglecting commonalities between chatbots and community members. This gap is evident in the absence of standardized models for chatbots' social capabilities. To address this, we propose a conceptual model of chatbots' social capabilities based on the interaction ritual chain (IRC) theory, derived from a systematic analysis of 1,273 articles. Our qualitative methodology synthesizes 83 references, all of which dealt with the topic of chatbots and had direct or potential implications for the metaverse. The model addresses core challenges through four elements—virtual coexistence, dynamic constraint, shared intention, and positive incentive—providing solutions for chatbot integration. Finally, we examine social capability interrelationships and suggest future research directions.

Subjects Human-Computer Interaction, Artificial Intelligence, Emerging Technologies, Natural Language and Speech, Network Science and Online Social Networks

Keywords Chatbot, Metaverse, Interaction ritual chains

Submitted 4 April 2025
Accepted 1 September 2025
Published 29 September 2025

Corresponding author
Yan Yan,
d22092100108@cityu.edu.mo

Academic editor
Othman Soufan

Additional Information and
Declarations can be found on
page 27

DOI 10.7717/peerj-cs.3235

© Copyright
2025 Yan and Fan

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

INTRODUCTION

Chatbots and metaverse

Chatbots, which are computer programs employing artificial intelligence (AI) to interpret customer inquiries and automate conversational responses, have gained renewed prominence following ChatGPT's emergence. This resurgence is driven by two developments. First, large language models (LLMs) have transformed work, leisure, and social interactions (*Zan et al., 2023*), making chatbots indispensable in the metaverse. Users increasingly rely on natural language interfaces for human-computer interaction (HCI), navigating virtual spaces as search engines guide web exploration. Second, technological advances enhance chatbots' human-like qualities. Digital-human systems enable 3D real-time rendering with motion capture, while augmented reality/virtual reality (AR/VR) and haptic technologies provide visual-tactile immersion during interactions. Consequently, chatbots have evolved into integral virtual community components. For this survey, we define chatbots as non-embodied social agents facilitating natural HCI in the metaverse, supporting both general and task-specific community activities.

The metaverse has evolved significantly in recent years, developing interoperable, immersive virtual communities accessible through user avatars (Xu et al., 2023). Current iterations feature diverse chatbot applications. In multiplayer online games, chatbots enhance player engagement by improving non-player characters (NPCs)' narrative contributions to storylines. Platforms like Roblox—a user-generated 3D community with over 80 million daily active users—enable chatbot creation via Google or OpenAI application programming interface (API). At its 2023 Developers Conference, Roblox announced conversational AI assistants to streamline experience creation. Chatbots serving as digital sales avatars also enhance personalized shopping on virtual commerce platforms, exemplified by Amazon's 'Rufus' virtual shopping assistant. Additional applications span education (Kuhail et al., 2023), healthcare (Tam et al., 2023), and beyond.

While chatbots show promise in the metaverse, their development remains nascent. The metaverse accommodate massive concurrent users, chatbots, and entities cohesively. For instance, 3D multiplayer online games frequently host over 100 entities per session. The metaverse, as conceptualized, necessitates seamless social dynamics among multiple agents that mirror the structural characteristics of real-world social systems. Current chatbots predominantly operate in one-on-one interactions, struggling to maintain consistent, continuous dialogues. Effective chatbots require real-time metaverse feedback to assess conversation states, yet their deep involvement in virtual community management remains limited industrially.

Research question

This study examines the challenges encountered by chatbots operating as members of virtual communities within the metaverse and explores potential solutions. As an emerging domain with diverse demands, the metaverse has garnered growing scholarly interest. While some conceptualize it as a virtual mapping of the real world, the inherent complexity of real-world social systems renders the development of accurate mathematical models or technical indicators for short-term goals a daunting task. Consequently, establishing a standardized framework to guide chatbot design poses significant challenges. These challenges are further exacerbated by the fact that, unlike shared technological infrastructures, chatbots inherently possess social attributes within the metaverse. Understanding the complexities of user-chatbot interactions thus requires an interdisciplinary approach. Moreover, the integration of chatbots into the metaverse remains in its infancy, necessitating further forward-looking technological exploration.

Employing an analytical framework rooted in the interaction ritual chain (IRC) theory, this research evaluates the social capabilities of chatbots. The IRC theory offers a robust framework for analyzing the interaction mechanisms among participants in virtual communities. Its core components—bodily co-presence, barriers to outsiders, mutual focus of attention, and shared mood—interact to form rituals, which in turn generate group solidarity, individual emotional energy, symbols of social relationships, and moral norms.

Table 1 Comparison of key focuses: related chatbot-specific surveys vs. this study.

Reference	A key focus of the survey	How our survey differs
<i>Huang, Zhu & Gao (2020)</i>	Addressing three core challenges (semantics, consistency, interactiveness) in open-domain chatbots and summarizing supporting techniques.	Our survey focuses on chatbots in the metaverse, exploring the social capabilities that chatbots should possess to become members of virtual communities, and organizes existing technical achievements from an interdisciplinary perspective.
<i>Wang et al. (2024a)</i>	Comprehensive survey on LLM-based agents, covering construction, multi-field applications, evaluation, and key challenges.	
<i>Xi et al. (2025)</i>	Exploring LLM-based agents' philosophical origins, general frameworks, and future directions like artificial general intelligence.	
<i>Ng & Zhang (2025)</i>	Systematic review of trust in AI chatbots, including conceptualization, influencing factors, and outcomes.	Rather than focusing on how to use chatbots, we pay more attention to how to implement chatbots in the metaverse so that they can establish stable interactive relationships with human users.
<i>Kurniawan et al. (2024)</i>	Evaluating AI chatbots for chronic illness management, focusing on user acceptance and health outcomes.	
<i>Yusuf, Money & Daylamani-Zad (2025)</i>	Analyzing pedagogical AI agents in higher education, developing frameworks for educational applications.	
<i>Senadheera et al. (2025)</i>	Reviewing chatbot adoption in local governments, including purposes, benefits, and implementation frameworks.	

This study aims to enhance researchers' and designers' understanding of chatbots' social capabilities and identify solutions to existing challenges. The establishment of such standards could facilitate a clearer grasp of objectives, challenges, and enabling technologies, thereby potentially boosting user trust in chatbots.

Related works

Recent years have witnessed significant efforts to enhance chatbots for user experiences, accompanied by several related surveys. [Table 1](#) summarizes these surveys and contextualizes the value added by our work.

[Huang, Zhu & Gao \(2020\)](#) established foundational insights by addressing three core challenges in developing open-domain chatbots—semantics, consistency, and interactivity—and summarizing key techniques (retrieval-based, generation-based, and hybrid methods) that underpin such systems. Building on this, recent surveys have expanded focus to LLM-based agents: [Wang et al. \(2024a\)](#) offered a comprehensive overview of LLM-based autonomous agents, covering their construction (encompassing profiling, memory, planning, and action modules), applications across social science, natural science, and engineering, evaluation methods, and key challenges (e.g., role-playing capability and hallucination); similarly, [Xi et al. \(2025\)](#) examined these agents from philosophical origins to a general framework (brain, perception, action), their applications in single-agent, multi-agent, and human-agent scenarios, and agent societies, and highlighted synergies between LLM and agent research, evaluation dimensions, and future directions (e.g., paths to artificial general intelligence and Agent as a Service).

Beyond technical advancements, cross-disciplinary research has explored how chatbots foster user relationships across various contexts. [Ng & Zhang \(2025\)](#) conducted a systematic review of trust in AI chatbots, identifying user-, machine-, interaction-, social-, and context-related factors influencing trust, with outcomes spanning affective, relational, behavioral, cognitive, and psychological domains, and emphasizing the need for longitudinal studies to understand trust dynamics. In healthcare, [Kurniawan et al. \(2024\)](#) reviewed randomized controlled trials on AI-powered chatbots for chronic illness management, noting high user acceptance and partial effectiveness in improving health outcomes. In education, [Yusuf, Money & Daylamani-Zad \(2025\)](#) analyzed 92 studies on pedagogical AI conversational agents in higher education, developing a framework that categorizes their pedagogical applications and technological functions, and identifies opportunities to enhance assessment and personalization. In public administration, [Senadheera et al. \(2025\)](#) reviewed chatbot adoption in local governments, identifying common purposes (e.g., information provisioning), benefits (e.g., enhanced engagement), risks (e.g., ethical concerns), and influencing factors (e.g., perceived humanness), and proposing a framework to inform implementation. Collectively, these studies demonstrate the evolving role of chatbots as versatile tools, with their impact shaped by technical advancements, trust dynamics, and context-specific needs across disciplines.

Although the technical aspects of implementing the metaverse have been extensively reviewed in the literature, there is a lack of a unified framework to guide the integration of chatbots into the metaverse as virtual community members. [Park & Kim \(2022\)](#) outlined the components and approaches necessary for implementing the metaverse, including hardware, software, content, user interaction, implementation, and applications. [Tang et al. \(2023b\)](#) focused on delineating the roadmap for the metaverse concerning communication and networking in sixth-generation (6G) wireless systems, integrating technologies such as intelligent sensing, digital twins (DT), space-air-ground-sea integrated networks, multi-access edge computing, and blockchain into 6G. [Xu et al. \(2023\)](#) emphasized the convergence of AI and blockchain, which can facilitate service delivery for an edge-enabled metaverse. Additionally, [Wang et al. \(2023b\)](#) discussed privacy invasions, security breaches, and proposed research directions for future metaverse systems. These surveys primarily review and discuss research centered on key implementation technologies and applications. Our work, by contrast, aims to direct designers' attention to chatbots, focusing on their integration into the metaverse to advance virtual community development.

Next, the research is structured into five sections: background and methodology, result 1: conversational intelligence, result 2: social intelligence, discussions, and conclusions. The background and methodology section provides a detailed explanation of the systematic analysis inspired by the IRC theory, focusing on the interconnected components between chatbots and virtual community members in the metaverse. The review process employed a retrieval strategy that ensures comprehensive and fair procedures, thereby avoiding any imbalanced assessment that might favor particular interpretations or perspectives. The result sections offer insights into the conversational and social intelligence of chatbots, as well as their challenges and prospective solutions.

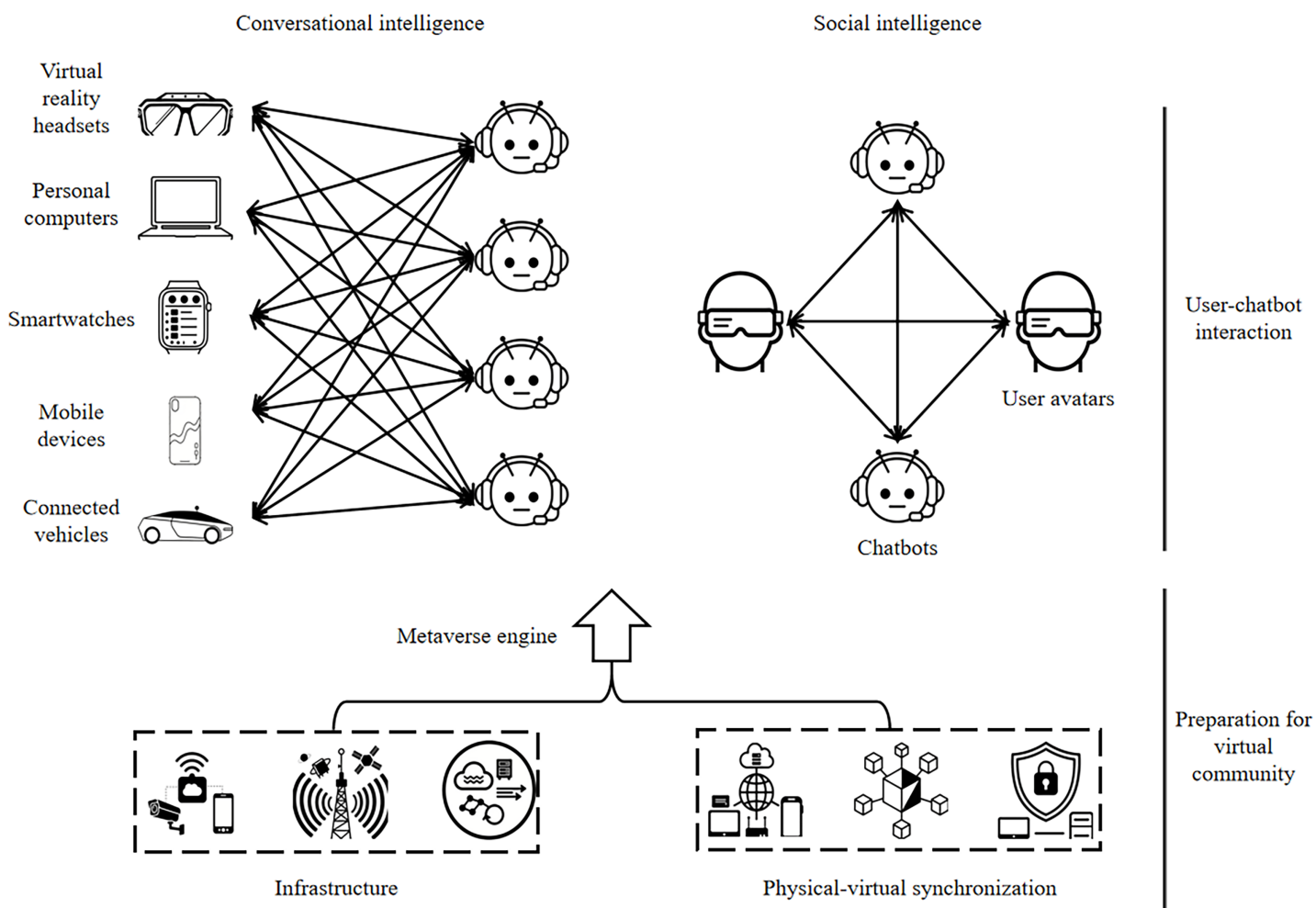


Figure 1 Chatbot integration in the metaverse. The foundational infrastructure enabling this integration is represented by icons corresponding to sensing, communication, storage, Internet, blockchain, and cybersecurity technologies. Positioned centrally, chatbots leverage conversational intelligence to interact with users *via* diverse devices, including VR headsets, personal computers, smartwatches, mobile devices, and connected vehicles. Additionally, their interactions with digital avatars exemplify social intelligence, underscoring the role of chatbots in facilitating and reinforcing virtual community construction within the metaverse. [Full-size DOI: 10.7717/peerj-cs.3235/fig-1](https://doi.org/10.7717/peerj-cs.3235/fig-1)

These findings enable designers to make informed decisions by selecting the appropriate subset of social capabilities and inspire further investigations by researchers. The discussions section elaborates on the significance of these findings, emphasizing the importance of social capability interrelations and future directions in chatbot design. Finally, the research concludes with a section on conclusions.

BACKGROUND AND METHODOLOGY

Chatbots integrated in the metaverse

The metaverse comprises virtual communities built on advanced Internet technologies that enable shared activities and emotional engagement. As shown in [Fig. 1](#), virtual community construction involves two stages: preparation and user-chatbot interaction. Key infrastructure supports the metaverse, including: parallel sensing, ubiquitous

cognition, scalable storage, and real-time analysis. These elements sustain its dynamic interactivity. Physical-virtual synchronization relies on foundational technologies: AR/VR (immersive experiences), integrated sensing-communication (seamless interaction), DT (virtual replication), AI (intelligent automation), and online social networks (social connectivity). Collectively, these components empower member-driven protocols for cohesive, responsive virtual communities.

Furthermore, these communities emerge through effective user-chatbot interactions, generating group cohesion, individual emotional engagement, shared social symbols, and collective ethical standards. Despite differing objectives and characteristics between chatbots and community members, fundamental commonalities establish robust bonds and mutual benefits. Successfully leveraging chatbots for virtual community construction hinges critically on user willingness to form stable IRCs with chatbots.

Chatbot intelligence critically influences both experience design and user engagement. Consequently, an interdisciplinary perspective—widely applied in virtual community research—can elucidate interaction-relevant intelligence types and augment chatbot capabilities. Designers must consider how chatbot behavior affects user interactions (including metaverse integrations), user experience, and participation levels. While HCI offers valuable development insights (*e.g.*, manipulation techniques, interface design, and behavioral patterns), chatbot design complexities inherently transcend traditional HCI boundaries.

Chatbots and virtual community members: interconnected elements

Conversation

Conversation serves as a foundational component for chatbots and virtual community members. Guided by five dialogic principles—mutuality, propinquity, empathy, risk, and commitment—these interactions establish frameworks for high-level conversation ([Kent & Taylor, 2002](#); [Jiang et al., 2022](#); [Kent & Taylor, 2021](#)). High-level conversations feature: risk-taking, altruistic information exchange, turn-taking respect, inclusive ethical deliberation, rhetorical sensitivity to others' needs, and participant-adapted communication. This distinguishes them from basic bidirectional exchanges. Since the Turing test ([Turing, 2009](#)), AI conversational systems have evolved substantially. [Nißen et al. \(2022\)](#) classifies chatbot engagement into three categories: *Ad-hoc* Supporters, Temporary Advisors, and Persistent Companions. Designed for longitudinal relationships, persistent companions demonstrate: autonomous adaptation, advanced personalization, anthropomorphic design, socio-emotional behaviors, and socially oriented communication. Their experiential learning and model optimization intrinsically satisfy high-level conversational criteria, especially inclusiveness.

The primary objective of conversational intelligence is maintaining persistent user engagement. This necessitates chatbots with social capabilities that enhance human-agent co-presence while implementing dynamic constraints to mitigate external interference during interactions. User-centric design prioritizes three core functionalities in chatbots: initiative support, responsibility management, and communication assistance, enabling

seamless the metaverse access. Beyond content delivery, high-level conversational chatbots must exhibit behavioral inclusiveness through multimodal information fusion, knowledge orchestration, and cross-domain adaptation.

Sociability

Beyond providing reliable and personalized conversational intelligence, user-centric chatbot services require sophisticated social intelligence. The social intelligence are fundamentally collaborative in virtual communities. Within the metaverse, members function as autonomous agents that interact to achieve shared or individual goals. These agents may include software programs, intelligent devices, or human users. Metaverses constitute complex cyber-physical-social systems that integrate virtual and real-world socio-economic ecosystems ([Zhang et al., 2018](#); [Wang et al., 2022](#)), with chatbots functioning as critical sociotechnical interfaces. Their deployment in university platforms exemplifies how personalized, immediate support capabilities enhance user experience ([Xie et al., 2023](#)). Chatbots' socially assistive roles—such as supporting isolated populations, neurodiverse individuals, and psychologically vulnerable groups ([De Gennaro, Krumhuber & Lucas, 2020](#); [Valtolina & Marchionna, 2021](#); [Xygykou et al., 2024](#); [Ali et al., 2020](#); [Zhou et al., 2020](#))—demonstrate significant societal value. Thus, integrating social intelligence into chatbots requires IRC-aligned designs to foster shared emotional states and identity formation. This integration involves three key actions: evaluating chatbots as ritual participants ([Sun, Zhan & Such, 2024](#)); mitigating the limitations of LLMs; and implementing cooperative multi-agent reinforcement learning (MARL) systems, where chatbots evolve through environmental interaction to sustain mutual benefits.

Applying the IRC theory to virtual communities provides a critical framework for analyzing evolving social intelligence. Human-agent interaction has fundamentally transformed community ownership and interaction paradigms ([Qiqi, 2023](#); [Simons, 2019](#)), necessitating evaluations that go beyond content analysis to include longitudinal engagement, emotional resonance, and sensory effects—key dimensions for enhancing sociability in the metaverse. To understand the social intelligence involved in facilitating user-chatbot interactions and advancing chatbot development, we examine the IRC theory. This framework helps elucidate the key components that foster the emergence of shared emotions and behaviors during virtual social interactions. [Table 2](#) presents a comparative analysis of the elements in our chatbot model and the IRC theory.

Survey methodology

The number of review articles published in computer science-focused journals has remained relatively stable, a trend that reflects the enduring impact of such research. Our systematic review aims to synthesize and analyze evidence related to research questions about the social capabilities of chatbots. This study follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to enhance the transparency and rigor of the literature review, as shown in [Fig. 2](#).

Table 2 Mapping from four elements of IRC to that of chatbots. The first column provides an overview of chatbots' social capabilities, which include conversational intelligence and social intelligence. The subsequent columns delineate the IRC and chatbot models, respectively. Each row contrasts methods from both domains that aim to achieve related short-term goals. For instance, bodily co-presence in the IRC theory emphasizes monitoring participants for interaction, while the chatbot model focuses on gathering user information for effective state synchronization.

Capability	IRC	Chatbots	Objective
Conversational intelligence	Bodily co-presence	Virtual coexistence	Chatbots and users can be in the same situation to influence each other.
	Barriers to outsiders	Dynamic constraint	Chatbots can determine the division of member groups to avoid outsiders' disturbances.
Social intelligence	Mutual focus of attention	Shared intention	Chatbots can collaborate with users to ensure that the group's common focus remains condensed.
	Shared mood	Positive incentive	Chatbots can form a system with other intelligent agents, often with long-lasting effects.

Eligibility criteria and search strategy

To identify the most relevant databases, preliminary searches were conducted across multiple platforms, with relevance evaluated based on the number of retrieved articles and their alignment with the research questions.

An extensive review of existing publications on “chatbots integrated in the metaverse” revealed a scarcity of literature directly addressing the research question, largely attributable to the topic's novelty. Consequently, this study initially identified empirical studies and prior research potentially relevant to answering the research question. Several databases and search engines were employed, including ScienceDirect, Taylor & Francis Online, Wiley Online Library, IEEE Xplore, Springer, Google Scholar, and arXiv. Google Scholar and arXiv served as supplementary resources due to their coverage of interdisciplinary research and emerging technologies in recent years.

Filters applied during literature searching included: scientific articles, English language, and no initial date restriction. However, this review focused on articles published between 2018 and 2025 (directly or indirectly related to chatbot technologies and the metaverse), as elaborated in “Reports excluded, reason 1”.

Search keywords included “chatbot and metaverse,” “chatbot in metaverse,” “chatbot role in metaverse,” “chatbot applications in metaverse,” and “chatbot challenges for metaverse.” Given the extensive body of research on chatbots in computer science, additional keywords such as “conversational agent,” “conversational model,” “dialogue system,” and “social agent” were incorporated to capture relevant literature. Notably, metaverse development is closely intertwined with advancements in AI; accordingly, research hotspots including “large language model,” “human-agent interaction,” and “multi-agent system” were included. Both chatbots and the metaverse have emerged as interdisciplinary research hotspots; however, this review specifically focuses on AI and HCI within computer science. Publications outside this scope were excluded during initial screening. The initial search yielded 1,273 articles.

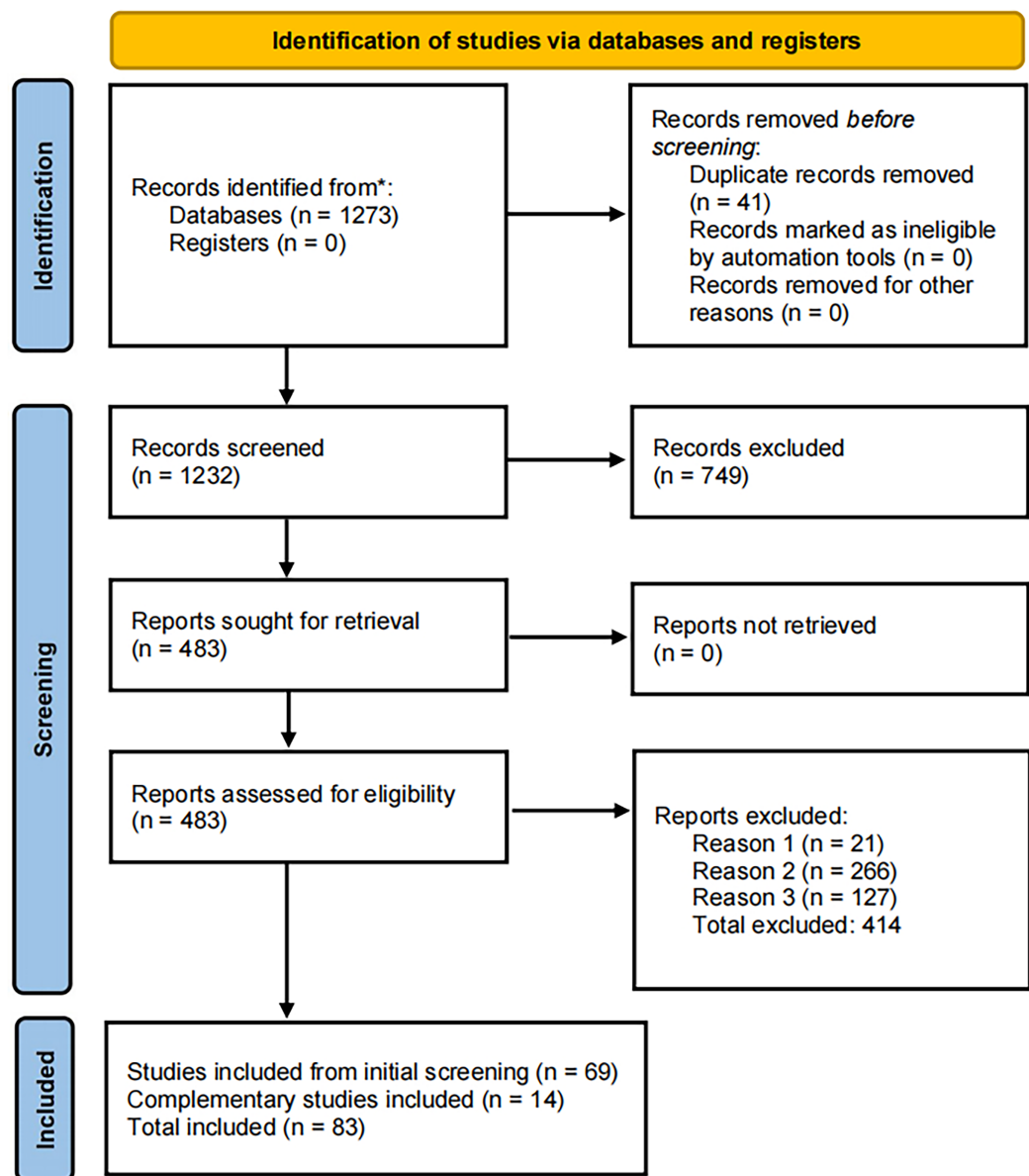


Figure 2 PRISMA flow chart process of article selection. Full-size [DOI: 10.7717/peerj-cs.3235/fig-2](https://doi.org/10.7717/peerj-cs.3235/fig-2)

Screening

Article selection was guided by keywords and abstracts, with each manuscript assessed for relevance to the research question and academic impact. The screening process adhered to PRISMA guidelines.

Duplicate removal: A total of 41 duplicate articles were identified and removed using reference management software (e.g., Zotero), resulting in 1,232 unique articles.

Title and abstract screening: A total of 749 articles were excluded after screening titles, abstracts, and keywords for misalignment with core topics (metaverse, chatbot, natural

language processing (NLP), multi-agent systems, virtual world), leaving 483 articles for full-text assessment.

- Reports excluded, reason 1: ($n = 21$) articles published before 2018. Although no initial date filter was applied (“anytime”), subsequent analysis determined that studies included in the systematic literature review should be restricted to the preceding 8 years (2018–2025). This timeframe aligns with the emergence of modern chatbot frameworks (e.g., transformer-based seq-to-seq models post-2018) and the growing prominence of the metaverse, as noted by [Dwivedi et al. \(2022\)](#), who highlight increased scholarly attention to the metaverse after 2020—including Facebook’s rebranding to Meta in 2021.
- Reports excluded, reason 2: ($n = 266$) irrelevance to the research topic. These included articles mentioning “metaverse” but outside computer science (e.g., pedagogy) or addressing the topic without providing substantive insights (e.g., metaverse infrastructure).
- Reports excluded, reason 3: ($n = 127$) misalignment with the research objective. These studies focused primarily on hardware-software integration in robotic systems rather than chatbot programming for virtual community development.

Supplemental searching: Conducted in response to peer reviewer feedback indicating omissions of literature on technical security aspects of deep machine learning and federated learning—both prevalent in chatbot training and deployment, particularly in metaverse applications. No additional duplicates were identified in supplemental searches.

Selection

- Reports included, reason 1: ($n = 16$) articles aligned with the topic, explicitly addressing “chatbot,” “metaverse,” or “HCI,” or focusing on relevant technologies (e.g., AR/VR, NLP, HCI) despite lacking explicit mention of core terms.
- Reports included, reason 2: ($n = 53$) articles on chatbots not centered on the metaverse but applicable to virtual community research.
- Reports included, reason 3: ($n = 14$) articles from supplemental searches, focusing on deep machine learning, federated learning, and technical security.

Final inclusion: For synthesis, 83 articles were included and analyzed through the lens of the chatbot conceptual model, as shown in [Fig. 3](#).

RESULT 1: CONVERSATIONAL INTELLIGENCE

Virtual coexistence

Virtual coexistence aims to intelligently analyze social dynamics and generate contextually appropriate responses—mirroring IRC’s bodily co-presence principle—to enhance interaction success through comprehensive user profiling. Face-to-face engagement and communal spaces are essential for boosting participative comprehension and interactivity. However, technological constraints severely limit implementation: unlike conventional voice/text dialogues, metaverse chatbots must engage diverse entities in open-domain

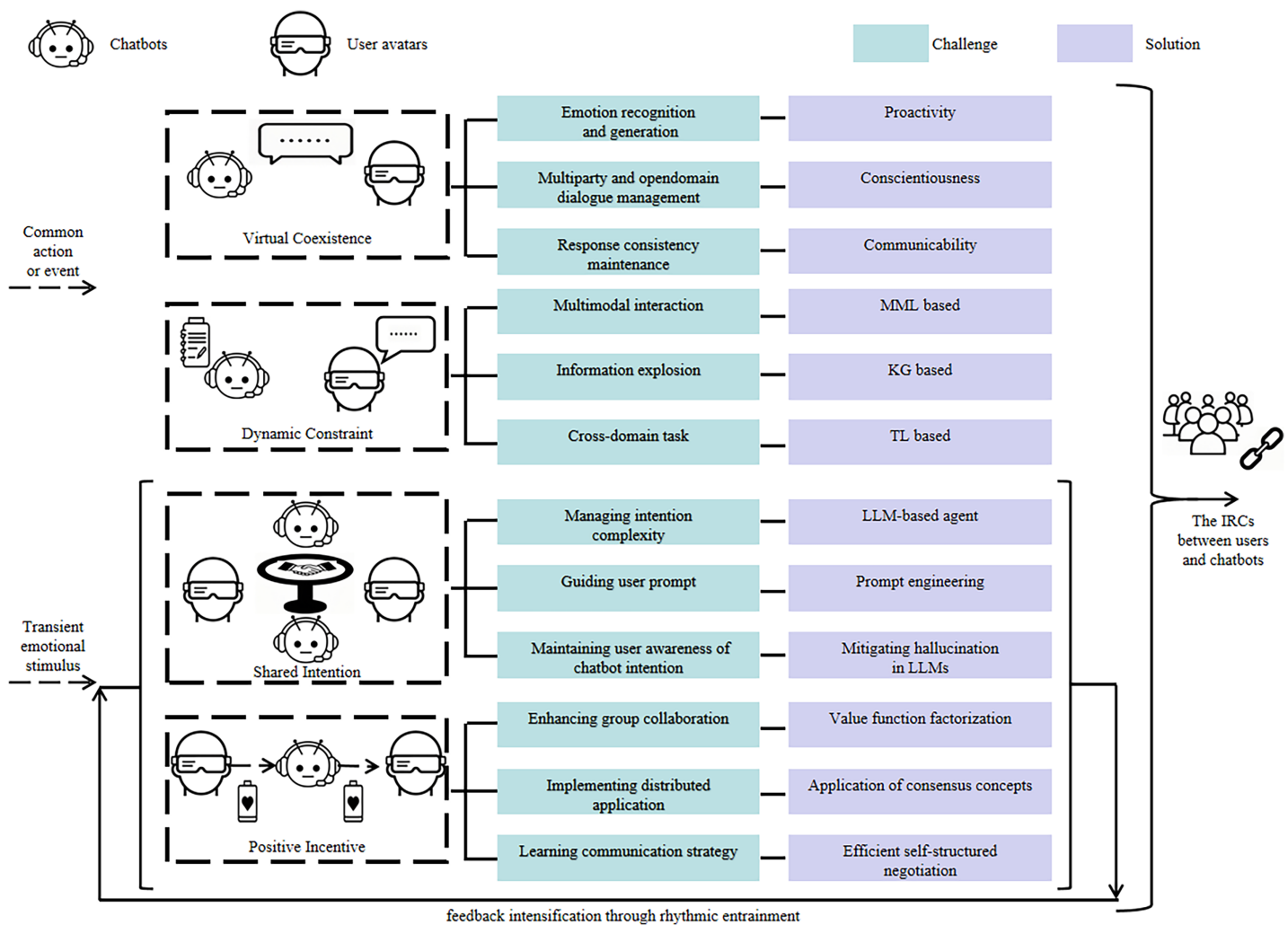


Figure 3 The model of interaction ritual between users and chatbots. Common action or event, as well as transient emotional stimulus, drive the initiation of interactive rituals. The four social abilities of chatbots facilitate the development of these interactive rituals in the metaverse. The interaction between users and chatbots in virtual coexistence, combined with dynamic constraints, provides optimal conditions for users to share intentions and transmit positive incentives. This, in turn, generates moments of intersubjectivity that produce the outcomes of IRCs between users and chatbots. [Full-size DOI: 10.7717/peerj-cs.3235/fig-3](https://doi.org/10.7717/peerj-cs.3235/fig-3)

contexts and decode complex social presence indicators (e.g., similarity, familiarity, attraction) from multimodal inputs. These requirements exceed current conversational intelligence capabilities, presenting fundamental implementation barriers.

Challenges of virtual coexistence

[C1] Emotion recognition and generation: Ineffective emotion recognition and generation jeopardize virtual coexistence success. Current techniques face difficulties in contextual modeling, capturing conversational dynamics, and representing speaker states. Chatbots with deficient emotion recognition produce generic responses, analogous to humans offering non-committal replies when uncertain. Furthermore, traditional neural language models struggle to generate logically coherent and emotionally resonant personalized

responses. Simply mirroring user sentiments without providing actionable guidance causes message neglect and impairs emotion generation capabilities.

[C2] Multiparty and open-domain dialogue management: Chatbots must sustain efficient goal-oriented interactions in multiparty open-domain dialogues. A restricted response repertoire may indicate inattentiveness to user inputs. Distinguishing nuanced variations across similar dialogue compositions remains an unresolved challenge. Ensuring structured conversational flow while masking multiparty constraints requires extensive training on addressee-annotated datasets. Users frequently evaluate chatbot membership suitability based on response diversity and generalization capability. Although multi-turn interactions are essential for complex goal attainment, they increase context coherence fragility.

[C3] Response consistency maintenance: Inconsistent chatbot responses can cause user misinterpretations and impair communicability. Unlike traditional software interfaces, chatbots articulate functionality and interaction principles through multi-turn dialogues. Their semantic analysis engines—which convert user utterances into executable commands—are critical components. However, multilingual models show reduced reliability for semantic analysis. While natural language effectively defines tasks and describes diverse entities, current chatbots face challenges integrating vision-language models, particularly when interpreting complex visual scenarios. Ensuring fairness and inclusion is essential for communicability; neglecting these principles impedes common ground establishment, potentially destabilizing virtual coexistence.

Solutions of virtual coexistence

[S1] Proactivity: Proactivity enables chatbots to actively engage users and enhance conversational flow through initiating dialogues, proposing topics, supplementing information, and asking follow-up questions. For emotion modeling, [Lei et al. \(2023\)](#) introduced emotional cognitive loss and hybrid continuous attributive networks to reduce speaker modeling overfitting, improving cross-scenario performance. [Ranasinghe et al. \(2022\)](#) developed an NLP-acoustic framework for detecting emotional transitions in telehealth. [Sun, Li & Tao \(2022\)](#) proposed a syntax-constrained bidirectional framework integrating emotion/topic keywords to generate contextually coherent responses.

Chatbots also must balance comfort provision with informational content, particularly for distressed users. [Deng et al.'s \(2023\)](#) knowledge-enhanced framework enables proactive mental health dialogues. [Tu et al.'s \(2022\)](#) strategy-aware model infers emotional states to deploy multi-tactic support. Addressing prior limitations, [Cheng et al.'s \(2022\)](#) multi-turn system detects nuanced emotional cues and employs A*-inspired lookahead heuristics for optimal strategy selection.

[S2] Conscientiousness: Conscientiousness enables chatbots to achieve effective virtual coexistence in the metaverse. This attribute allows comprehensive contextual understanding of utterances and autonomous recovery from operational failures. Unlike dyadic dialogues, multiparty conversations require responses predicated on recipient role prediction. [Liu et al.'s \(2023e\)](#) pseudo-variational hierarchical model identifies subtle semantic variations without prior knowledge, synthesizing variables optimized *via*

distribution consistency objectives. [Lv et al.'s \(2023b\)](#) hierarchical duality learning emulates human cognition in inferring utterances from logical structures. To address limited multiparty pretraining data, [Li & Zhao \(2023\)](#) proposed expectation maximization for iterative addressee annotation and response optimization.

In open-domain dialogues, chatbots achieve generalization through many-to-many mappings where single contexts elicit diverse responses. [Zhao et al. \(2023\)](#) introduced robust metrics for one-to-many evaluation. For multiparty contextual coherence, [Lv et al.'s \(2023a\)](#) dialogue path-sampling projects conversations onto extended Brownian Bridges, sampling latent variables to construct coherent paths that enrich training data.

[S3] Communicability: Communicability constitutes the third critical dimension for chatbots, particularly vital in the metaverse's multilingual environments. Chatbots must perform robust multilingual semantic analysis to accommodate users who frequently switch languages. [Held et al.'s \(2023\)](#) doubly aligned multilingual parser significantly advances zero-shot capabilities while reducing model parameters, enhancing cross-linguistic communication.

[Liu et al.'s \(2023f\)](#) InternGPT equips chatbots with planning/reasoning capacities through interactive visualization. This framework enables comprehension of non-verbal commands (e.g., gestures toward on-screen content) during vision-centric tasks, bridging visual-linguistic gaps to boost communication efficiency.

[Sicilia & Alikhani \(2023\)](#) formalized fairness in text generation *via* computational learning theories. Their work clarifies anthropomorphic learning-fairness relationships, demonstrating that algorithms can produce equitable text from biased data under fairness-preserving conditions.

In virtual communities, active chatbots may number in the hundreds, while the devices they serve can scale to millions. Each chatbot deployment requires distinct adaptations for both devices and users. Federated learning—a distributed paradigm that trains global models using client data without sharing raw data—faces two critical limitations. First, research reveals significant privacy risks, as malicious servers can reverse-engineer model gradients to compromise data ([Mothukuri et al., 2021](#)). Second, global models often underperform for individual clients when local data exhibit heterogeneity in quantity, quality, or distribution. To address these challenges, [Zhang et al. \(2024\)](#) proposed a privacy-preserving, personalized framework that partitions gradient computation between servers and local clients, thereby optimizing models for specific contexts. Complementarily, [Xu et al. \(2022\)](#) introduced a method to mitigate accuracy degradation from irregular user participation, ensuring training outcomes primarily reflect high-quality data contributions.

Dynamic constraint

Dynamic constraint employs computational methods that leverage virtual coexistence data to formulate conversation strategies aligned with virtual community structures, reflecting the IRC theory's Barriers to Outsiders principle. These strategies explicitly differentiate participants from non-participants by utilizing communal space information to reinforce interaction rituals, where accurate identification of active participants is

essential. This capability enables chatbots to distinguish and segregate disengaged members during social interactions. Dynamic constraint perceives multidimensional social boundaries—including temporal and spatial dimensions—to ensure information relevance for community members. Consequently, chatbots initiate context-sensitive interaction patterns that facilitate seamless integration into target virtual communities. This section delves into potential solutions across various research domains, including multimodal machine learning (MML), knowledge graph (KG), and transfer learning (TL).

Challenges of dynamic constraint

[C1] Multimodal interaction: The influx of multimodal messages poses a significant challenge for chatbots in the metaverse, particularly in restricting unauthorized participants in real-time interactions. User–chatbot interactions involve a complex interplay of various modalities such as speech, non-verbal sounds, visuals, and text. Chatbots face two main challenges in processing this diverse information. The first is synchronizing the different modalities within a message, known as the fusion problem. The second is identifying the implicit content often conveyed through multimodal messages, known as the recognition problem. For instance, face-to-face interactions often involve asynchronous and emotionally charged multimodal messages.

[C2] Information explosion: Dynamic constraint necessitates managing diverse information, which may compromise technical performance. Effective chatbots require strategies to distill relevant content from extensive data—a core aspect of dynamic constraint. Knowledge management system development remains complex due to intricate representation and reasoning requirements. As virtual community knowledge dynamically evolves with metaverse applications, chatbots must rapidly acquire and generalize new knowledge. Addressing information overload through concise abstractions and scalable frameworks presents significant technical challenges.

[C3] Cross-domain task: Predicting user behavior for chatbots implementing dynamic constraints in cross-domain tasks presents significant complexity. Chatbots must first navigate socio-contextual factors and task variations across domains before applying constraints. Traditional training methods assume consistent input requirements and social contexts within single domains—an assumption often invalid when tasks exhibit ambiguity or indeterminacy. Consequently, cross-domain training data accessibility becomes critical. Enhancing chatbot robustness for cross-domain tasks is fundamental to successful dynamic constraint implementation.

Solutions of dynamic constraint

[S1] MML-based solution: These approaches enable chatbots to comprehend and generate video-grounded dialogues. [Wang et al. \(2023c\)](#) established benchmarks for video-based dialogue understanding and generation, incorporating multimodal characteristics like scene/topic transitions. [Zheng et al. \(2023a\)](#) proposed a contrastive visual-textual transformation method for sign language recognition, leveraging variational autoencoders for contextual pretraining and contrastive alignment for cross-modal consistency.

Simultaneously, [Zhang et al.'s \(2023a\)](#) quality-aware multimodal fusion framework optimizes information utilization while mitigating low-quality data impacts, and [Lv et al.'s \(2021\)](#) progressive modality reinforcement addresses misaligned streams through iterative modality-hub exchanges.

For multimodal emotion recognition in multiparty dialogues, visual cues are critical: [Zheng et al.'s \(2023b\)](#) two-stage framework extracts utterance-aligned facial expressions *via* face recognition/clustering/matching and enhances recognition *via* frame-level emotion distributions; [Liu et al.'s \(2023a\)](#) brain-machine coupled approach trains on visual-EEG data for cross-domain facial emotion analysis; while [Rasendrasoa et al.'s \(2022\)](#) knowledge-aware network employs multi-head attention on dialogue history and participant knowledge for joint emotion/dialogue act recognition.

[S2] KG-based solution: KGs serve as collective repositories within organizations, enabling knowledge representation, accumulation, curation, and propagation. Chatbots leverage KGs to refine dialogue generation: [Tang et al.'s \(2023a\)](#) framework integrates language models with graph operations through multi-stage feature fusion to enhance response informativeness. Furthermore, KG embeddings effectively capture uncertainty and domain-specific knowledge. [Yao et al.'s \(2023c\)](#) AnKGE framework augments KG embeddings with self-supervised analogical inference, generating analogical object embeddings from pre-trained elements. While neural link predictors identify missing triples in incomplete graphs, they face computational intensity and negation-query limitations. [Wang, Chen & Grau \(2023\)](#) addresses these challenges, and [Cui et al.'s \(2023\)](#) lifelong model transfers knowledge through masked autoencoders and embedding regularization without re-embedding. Alternatively, [Liu et al.'s \(2023c\)](#) IterDE employs knowledge distillation to compress KG models, improving storage/inference efficiency while outperforming baselines in link prediction.

[S3] TL-based solution: Chatbots can acquire new tasks by drawing on prior experience. Invariance learning techniques help distinguish spurious features from invariant ones within inputs. [Tan et al. \(2023\)](#) have proposed an environment-agnostic invariance learning method requiring neither environmental indices nor auxiliary data. Domain generation offers another effective approach, though model efficacy depends on distributional alignment between generated and target domains. [Dai et al. \(2023\)](#) introduced moderate distributional exploration for domain generalization, identifying common factors across training domains *via* uncertainty subsets with shared semantics.

Zero-shot domain adaptation challenges chatbot deployment in unseen domains without supervised data. Parameter-efficient TL addresses this limitation. [Aksu, Kan & Chen \(2023\)](#) developed the Prompter method, which uses target domain slot descriptions to generate dynamic prefixes interacting with self-attention keys/values per layer, incorporating conversational semantics and co-occurrence frequency. Previous work prioritized data/model-level augmentation for generalization but inadequately isolated sample semantics, constraining zero-shot performance. [Wang et al. \(2023a\)](#) proposed a divide-conquer-combine strategy that disentangles observed data semantics *via* mixture-of-experts, enhancing robustness and performance.

Chatbot training employs diverse methods to process multimodal data (images, text, numerical inputs). Formulating cross-modal constraints remains challenging, as user behaviors—varying across temporal sequences and conversational contexts—are critically important for open-domain dialogue generation. Within federated learning paradigms, ethical concerns arise regarding the use of data. Privacy risks persist despite raw data localization: uploaded gradients/parameters may reveal local information, enabling models to infer client data distributions. [Nasr, Shokri & Houmansadr \(2019\)](#) demonstrated this *via* membership inference attacks exploiting stochastic gradient descent vulnerabilities. Fairness is equally critical, requiring mitigation of biases against disadvantaged groups (e.g., ethnic minorities, women, the elderly). Such biases originate from both clients ([Jang, Zheng & Wang, 2021](#)) and servers ([Nishio & Yonetani, 2019](#)). Violations of privacy or fairness are ethically unacceptable. [Chen et al. \(2023a\)](#) surveyed interactions between these principles, revealing their interdependence. Systematic co-optimization prevents sacrificing one for the other.

RESULT 2: SOCIAL INTELLIGENCE

Shared intention

Shared intention constitutes the foundational pillar of social intelligence, aligning with the principle of mutual focus of attention in the IRC theory. Chatbots' social roles extend beyond dyadic user interactions, particularly in open-world environments. They are engineered to maintain granular chronicles of social exchanges through natural language, iteratively synthesize these experiences into higher-level cognitive abstractions, and dynamically retrieve such knowledge to optimize behavioral strategies. LLMs prove indispensable in augmenting chatbots' capabilities, transcending the limitations of conventional question-answering systems or recommendation algorithms. However, the inherently complex and expansive nature of shared intention presents fundamental challenges: LLMs' architectural constraints impede chatbots' capacity to manage the continuous evolution of multi-agent intentions, resolve emergent conflicts, track longitudinal event trajectories, and navigate intricate social dynamics arising between users and autonomous agents within collaborative ecosystems.

Challenges of shared intention

[C1] Managing intention complexity: Designing advanced chatbots capable of multi-entity collaboration for complex intention planning, memory enhancement, and action execution remains an open challenge. These chatbots require adaptability and optimization, with core functionalities including: task decomposition, multi-timescale memory management, tool invocation, and multi-agent communication. Although LLMs demonstrate strong general capabilities, many chatbots rely exclusively on short-text inputs and LLMs' native planning/acting functions. This limitation causes stochastic outputs and interaction inconsistencies—a critical challenge requiring further investigation.

[C2] Guiding user prompt: Prompt design critically shapes model outputs ([Lu et al., 2022](#)). Analogously, LLMs function as chatbot brains where prompts serve as

commands—minor variations yield substantially divergent results. Prompts encompass syntactic elements (e.g., length, blanks, example ordering) and semantic components (e.g., phrasing, example selection, instructions). Prompt quality directly determines LLM task performance, with optimal prompts typically requiring human crafting. When novice users unaware of chatbot limitations issue capability-testing prompts, unmet expectations cause frustration. Consequently, prompt fragility significantly compromises chatbot intention realization.

[C3] Maintaining user awareness of chatbot intention: Chatbots must provide accurate responses for complex intentions. Hallucination—where generated content is illogical or unfaithful to source material—remains prevalent (Filippova, 2020). LLMs frequently produce hallucinations containing misleading information, often undetectable due to textual fluency. Lee et al. (2022) evaluated these using entity-based metrics and textual entailment. Given the emphasis on user-centric responses and alignment with user directives in LLMs, hallucinations in natural language generation tasks can be divided into factuality hallucination and factual fabrication (Huang et al., 2025).

Solutions of shared intention

[S1] LLM-based agent: Recent advances in LLM-based agents enable users to define virtual agent attributes, conduct dynamic dialogues, and receive automated feedback. Hasan et al. (2023) introduced SAPIEN—an advanced platform supporting multilingual communication (13 languages), emotional expression through facial/vocal cues, and non-verbal interaction. Significant progress has also been made in automating complex problem-solving. Hong et al. (2024) proposed MetaGPT, a meta-programming framework that integrates human workflows into multi-agent systems. Using an assembly-line approach, MetaGPT delegates roles and decomposes complex tasks into manageable subtasks to enhance agent collaboration.

An LLM-based agent is characterized by the integration of LLMs as the cognitive core, supplemented by several essential components:

- **Planning:** Language models typically solve general problems across diverse tasks. However, their token-level left-to-right decision-making limits performance on tasks requiring exploration, Strategic lookahead, or critical initial decisions. Wei et al. (2022) demonstrated that Chain-of-Thought prompting—providing exemplar demonstrations—enhances LLM reasoning. Yao et al. (2023a) extended this approach with the Tree of Thoughts framework, which examines coherent text units (“thoughts”) as intermediary problem-solving steps. This enables models to evaluate multiple reasoning paths and strategically select next actions. Wang et al. (2024c) proposed Solo Performance Prompting, where a single LLM emulates distinct personas through multi-turn self-collaboration. This method improves complex problem-solving by dynamically adapting to task inputs.
- **Memory:** The rise of high-performance conversational agents like ChatGPT stems from advances in LLMs. Such agents must retain key dialogue information to generate contextually relevant responses. However, limited user control over memory

management often harms conversational coherence. To tackle this, [Huang et al. \(2023\)](#) proposed Memory Sandbox—an interactive system for transparent user oversight of conversational memory. This framework modularizes memories as data objects, enabling viewing, editing, summarization, and sharing. Meanwhile, [Chen et al. \(2023b\)](#) utilized LLMs’ narrative generation to synthesize artificial memories, converting stories into actionable sequences for guiding agents in hypothetical scenarios. Traditional reinforcement learning methods struggle with LLM-based agents due to costly fine-tuning and sample inefficiency. [Shinn et al. \(2023\)](#) introduced Reflexion, a framework that improves language agents *via* linguistic feedback rather than weight updates. Reflexion agents analyze task feedback (e.g., success/failure signals), produce self-critiques, and store them in episodic memory. This iterative refinement achieves 91% pass@1 accuracy on HumanEval, surpassing GPT-4’s 80% (previous state-of-the-art).

- **Action:** While LLMs possess strong generative capabilities, they often struggle with tasks requiring complex planning and external tool usage. To address this, [Liu et al. \(2023b\)](#) leveraged a language-only GPT-4 to generate multimodal instruction-following datasets combining text and images. By refining these instructions, they developed a large vision-language assistant—an end-to-end multimodal model integrating a vision encoder with LLMs for general-purpose visual and linguistic understanding. [Yao et al. \(2023b\)](#) investigated LLM-based agents’ reasoning and action capabilities. Their work focused on using LLMs to generate structured reasoning traces alongside task-specific actions. These traces help the model infer, monitor, and adjust action plans, while actions facilitate interaction with external tools to gather additional information. Further enhancing task planning and tool utilization, [Kong et al. \(2024\)](#) proposed a framework with three key components: API retriever, demo selector, and LLM finetuner.

[S2] Prompt engineering: To address prompt brittleness, [Zhou et al. \(2023\)](#) proposed an automated prompt engineering system that generates and optimizes instructions. By treating instructions as trainable “programs,” their method iteratively refines candidate prompts using LLM-generated suggestions. Experiments demonstrated that this approach enhances few-shot learning performance by improving standard in-context learning prompts. It also effectively identifies high-quality zero-shot chain-of-thought prompts, leading to more accurate and informative model outputs. Complementing this work, [Raman et al. \(2023\)](#) introduced model-tuning *via* prompts, a lightweight adaptation method for downstream tasks. The method appends tunable prompt templates to inputs and generates predictions *via* text infilling or completion. When integrated with adversarial training, model-tuning *via* prompts not only preserves performance on clean data but also significantly improves robustness against adversarial attacks.

[S3] Mitigating hallucination in LLMs: A strategy to diminish hallucinations involves Retrieval Augmentation, a method that anchors the model’s input to external knowledge sources. [Ge et al. \(2023\)](#) augmented the generalization capabilities of language models with mixture-of-memory augmentation, a process that retrieves augmentation documents from multiple information sources, referred to as ‘external memories’, and permits the addition

of new memory during inference. [Liu et al. \(2023d\)](#) developed a hierarchical transformer retriever, specialized in dialogue domain data, for personalized retrieval. They proposed a context-aware prefix encoder to integrate retrieved information effectively. Furthermore, [Zhang et al. \(2023b\)](#) introduced a knowledge merging framework oriented towards compatibility, aiming to enhance open-domain question answering systems. This framework posits that answers corroborated by multiple sources are more reliable.

Metaverse applications require diverse model architectures to support heterogeneous training tasks. Consequently, chatbots must integrate multiple features—each associated with unique pre-training tasks and activation conditions. This architectural complexity exacerbates intellectual property risks, particularly model-stealing attacks. For instance, [Carlini et al. \(2024\)](#) successfully extracted all projection matrices from OpenAI's Ada and Babbage models for less than \$20, proving the feasibility of creating functional replicas. [Hui et al. \(2024\)](#) further demonstrated prompt theft through adversarial chatbot outputs. To counter these threats, [Zhou et al. \(2024\)](#) proposed an inversion-based defense using neural decoders to analyze suspicious model outputs for legitimacy verification. In parallel, [Li et al. \(2023b\)](#) developed generative API watermarking techniques to deter remote imitation attacks.

Positive incentive

Positive incentives align with the principle of shared mood in the IRC theory, as both generate 'emotional energy'—a pleasurable sense of group affiliation that motivates repeated ritual participation ([Collins, 2005](#)). In user-chatbot interactions, these incentives function as 'batteries' that store and release interactive energy. They enable experience recall and initiate subsequent interaction rituals, forming interconnected chains. This mechanism addresses inherent complexity from dynamic environments and combinatorial challenges in chatbot operations. The approach prioritizes training models for proficient group collaboration through methodologies like value function factorization, consensus-building techniques, and structured negotiation protocols.

Challenges of positive incentive

[C1] Enhancing group collaboration: A key challenge in multi-agent systems lies in learning value functions that effectively promote collaborative behavior. Traditional approaches involve either optimizing a joint reward function or reformulating the problem as a single-agent system. However, traditional reinforcement learning methods often fail to reach global optima in large state-action spaces, primarily due to two factors: (1) agents' inability to accurately assess their individual contributions to collective rewards, leading to diminished motivation, and (2) suboptimal policies from certain agents that degrade overall performance and hinder optimal policy discovery. Value function factorization has emerged as a promising solution, enabling precise attribution of each agent's contribution to the global reward. This approach maintains the theoretical advantages of centralized training while facilitating decentralized execution, effectively addressing the credit assignment problem in cooperative multi-agent environments.

[C2] Implementing distributed application: Chatbots must achieve an optimal solution by reaching consensus with their peers. With an increasing number of chatbots, the information they manage could surpass their processing capacity. Furthermore, in the metaverse, centralizing local data onto a central server presents challenges due to energy constraints, privacy issues, and hardware failures. The issue becomes how to facilitate communication among local agents and adjust the network topology to enable them to collaborate effectively and learn the optimal policy. The consensus approach entails structuring agents within a sparse network, allowing each to communicate only with a limited number of neighboring agents. This setup encourages agents to work toward an optimal solution that aligns with their neighbors' outcomes.

[C3] Learning communication strategy: The aforementioned challenge presupposes that chatbots strive for a consensus that yields the optimal policy. However, the specific information necessary for an agent to learn this policy is not always evident. Efficient negotiation through the exchange of critical messages allows agents to discern the content, timing, and recipients of their communications. This process also fosters the development of negotiation skills within the virtual community, independent of user-chatbot interactions.

Solutions of positive incentive

[S1] Value function factorization: Recent advances in multi-agent value decomposition have significantly enhanced cooperative learning in complex environments. [Sunehag et al. \(2018\)](#) pioneered value decomposition networks to quantify individual agent contributions to collective rewards by factorizing the team value function into agent-specific components based on local state-action histories. Building on this, [Rashid et al. \(2020\)](#) introduced a monotonicity constraint in the mixing network to preserve the relationship between centralized and decentralized value functions, while incorporating global state information and non-linear transformations to improve solution quality. To address representation limitations and approximation errors, [Shen et al. \(2022\)](#) proposed residual functions for optimal joint policy learning. [Mei, Zhou & Lan \(2023\)](#) advanced this further by formulating optimal projection of mixing functions as a regret minimization problem over state-action value weights. In metaverse environments characterized by partial observability and policy diversity, action sequencing affects representable function classes and introduces training estimation errors. [Zhou, Lan & Aggarwal \(2022\)](#) developed a PAC-assisted framework using counterfactual predictions and a novel loss function for value decomposition. For risk-sensitive scenarios, [Shen et al. \(2023\)](#) established the risk-sensitive individual-global-max principle, aligning decentralized risk assessments with centralized policy. [Yuan et al. \(2022a\)](#) introduced task decomposition to streamline learning in complex environments through concentrated coordination.

[S2] Application of consensus concepts: In multi-agent systems, two primary scenarios apply when agents adhere to distinct policies. In the first, agents possess knowledge solely of their states and rewards. Conversely, in the second, a global state is visible to all, with each agent receiving a unique local reward. [Cassano, Yuan & Sayed \(2020\)](#) introduced the fast diffusion for policy evaluation algorithm, merging off-policy learning with eligibility

traces and linear function approximation, enabling the evaluation of both individual and team policies within a finite dataset. This framework permits agents to exchange information with immediate neighbors; however, the lack of message aggregation may result in suboptimal policy learning. To mitigate this, [Guan et al. \(2025\)](#) proposed multi-agent communication *via* self-supervised information aggregation, developing a permutation-invariant message encoder to amalgamate common information into concise representations, thereby refining local policy. [Wang et al. \(2024b\)](#) presented the multi-agent goal imagination framework, which employs self-supervised learning to optimize representation by reconstructing and forecasting future states, thus aligning agents toward a shared objective.

In the design of consensus-reaching processes, the number of discussion rounds and the degree of agent harmony are pivotal for measuring efficiency ([Hassani et al., 2022](#)). To minimize communication overhead, [Zhang & Zavlanos \(2024\)](#) introduced a distributed zero-order policy optimization method, enabling agents to compute the local policy gradients necessary for policy function updates. [Lu et al. \(2021\)](#) conceptualized the consensus-reaching process as a distributed constrained Markov decision process involving networked agents. They proposed a decentralized policy gradient approach that optimizes policies based on this model over a network.

[S3] Efficient self-structured negotiation: In the study of communication strategies, [Mordatch & Abbeel \(2018\)](#) explored a joint reward scenario where agents are aware of each other's positions and messages, sharing their policies and action spaces and utilizing symbols from a predefined vocabulary to denote their actions. However, the indiscriminate sharing of parameters can impede exploration and diminish performance. To counteract this, [Yang et al. \(2022\)](#) introduced an innovative framework that learns dynamic subtask allocation using clustering agents with analogous skill sets. [Li et al. \(2023a\)](#) proposed a hybrid framework, integrating an evolutionary algorithm with MARL to enhance coordination. Evolutionary Algorithm drives population evolution through crossover and mutation, offering a variety of experiences to MARL, which in turn refines its policies for further optimization.

In virtual communities with fluctuating numbers of chatbots, uniform message sharing or targeted communication learning may face scalability challenges in MARL communication. To overcome these issues, [Jiang & Lu \(2018\)](#) developed the attention-based targeted observation communication algorithm, which incorporates an attention mechanism to decide when to integrate information from other agents. Agents establish communication groups, maintaining a majority of consistent collaborators when they opt to communicate. Additionally, [Yuan et al. \(2022b\)](#) introduced the multi-agent incentive communication framework, enabling agents to craft incentive-laden messages that can sway the value functions of other agents, fostering effective collaboration. [Guo, Shi & Fan, 2023](#) proposed the transformer-based email mechanism, a framework where agents engage in local communication, sending messages exclusively to observed agents, rather than modeling the entire agent population.

Chatbot failures risk collapsing entire virtual communities, making comprehensive pre-release testing across all relevant device types critically important. As chatbot scale

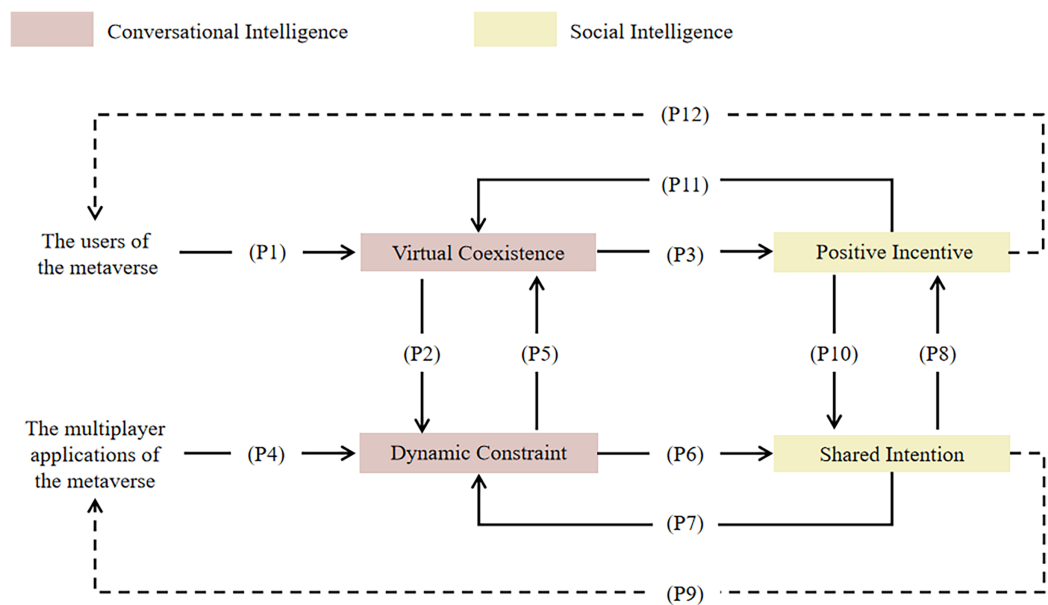


Figure 4 Interrelationships among the elements. The interaction among four pivotal elements within chatbots' conversational and social intelligence in the metaverse is crucial. These elements comprise virtual coexistence, dynamic constraint, shared intention, and positive incentive. The influence of virtual coexistence and dynamic constraint is primarily driven by immediate user interactions and the interplay of multiplayer applications. Conversely, shared intention and positive incentive contribute to significant long-term influences on user engagement and the collaborative efficacy of various applications. The interplay of these elements, both in the short-term, significantly shapes the user experience within the metaverse.

Full-size [DOI: 10.7717/peerj-cs.3235/fig-4](https://doi.org/10.7717/peerj-cs.3235/fig-4)

increases, training data demands escalate, compelling practitioners to automate and outsource data curation for state-of-the-art performance. This absence of human supervision in data collection introduces security vulnerabilities: manipulated training data can control and degrade downstream model behaviors (Goldblum et al., 2023). For instance, adversarial samples significantly contribute to failures. Zhu et al. (2024) demonstrated this vulnerability by deploying multi-level adversarial textual attacks against LLM prompts, revealing contemporary models' fragility. Zhou et al. (2022) reframed adversarial analysis through the lens of advanced persistent threats, modeling attacks as interconnected systems where strategically similar attack clusters form components. This systemic view enables identification of cross-stage defensive relationships, facilitating integrated frameworks that combine multiple defenses to enhance chatbot robustness.

DISCUSSIONS

Interrelationships among chatbots' social capabilities

In the preceding sections, we classified the social capabilities of chatbots into four key elements and examined the interdependent objectives that either promote or supplement one another. This section introduces a theoretical framework that outlines these interconnections. The aim is to elucidate the inherent relationships within potential

solutions, highlighting the complexity of chatbot development for virtual communities within the metaverse. As depicted in Fig. 4, the boxes represent the key elements, with different colors signifying their respective capabilities. Solid lines represent ten short-term influences, whereas dotted lines denote two long-term influences. The notion of endowing chatbots with social capabilities is not one of replicating human traits. However, as chatbots advance in conversational and social intelligence, they become more adept at forming IRCs with users.

- **Virtual coexistence:** Our chatbot model demonstrates that virtual coexistence enhances the interaction between chatbots and users (P1), acknowledging the complexity of managing conversational status in multi-participant and open-domain dialogues. Proactive conversation, grounded in contextual understanding, enhances chatbot proactivity in establishing dynamic constraint (P2), thereby demonstrating attentiveness to participants. This element also bolsters positive incentive (P3), as more communicative chatbots encourage greater user engagement in user–chatbot interactions.
- **Dynamic constraint:** Dynamic constraint is rooted in the multiplayer environment of the metaverse (P4), addressing challenges arising from multimodal interactions, information overload, and cross-domain tasks. Advanced multimodal information processing, robust knowledge management, and cross-domain adaptability strengthen the efficacy of dynamic constraint. This element parallels virtual coexistence (P5) by ensuring dialogue consistency through the derivation and recall of information from previous interactions. Furthermore, dynamic constraint reflects shared intention (P6), as chatbots can recall common actions or events across sessions.
- **Shared intention:** Shared intention impacts dynamic constraint (P7) by facilitating the exclusion of external participants, thereby enhancing group cohesion. For instance, chatbots can identify potential interlopers through improved language comprehension and generation capabilities, leveraging large language models (LLMs). Shared intention also supports positive incentive (P8), as it enables chatbots to autonomously plan and communicate, effectively applying policies from a reserved pool of positive initiative. This element can refine the multiplayer experience in the metaverse (P9), with chatbots serving as tireless work assistants, contingent on addressing issues like prompt brittleness and hallucination in future research.
- **Positive incentive:** Positive incentive influences shared intention (P10), as it can be used as a strategy to punish requests diverging from shared intention by reducing rewards while preserving interactive continuity. Positive incentive also influences virtual coexistence (P11), as it can help chatbots avoid multi-role conflicts and reduce the negative effects of dynamic environments. Positive incentive can make users and chatbots more interactive (P12) by providing the capability to respond effectively to group collaboration.

Future research directions

Research on chatbots, particularly in the metaverse, addresses a critical priority, as their integration remains an emerging area of study. The challenges identified for chatbots align with those observed in human-human interaction. However, the metaverse introduces unique interactive environments and technologies that shape the development of chatbots. While bodily co-presence (*e.g.*, face-to-face encounters) fosters intense and energizing interactions, some scholars argue that low-intensity, technology-mediated interactions can also be socially effective due to their accessibility and frequency (*Johannessen & Collins, 2024*). Thus, we maintain an optimistic outlook on chatbots as members of virtual communities in the metaverse. This study outlines prospective research directions, emphasizing key areas for chatbot integration in the metaverse.

- Enhancing multimodal socio-emotional intelligence. Significant improvements are needed in chatbots' ability to perceive, interpret, and express socio-emotional cues through multimodal integration (text, speech, facial expressions, and environmental context). Advancements in emotion recognition and generation systems—such as graph-based dialogue state representations and diffusion models—are essential for synchronized, context-aware emotional expressions. Strengthening Theory-of-Mind capabilities through mental state graphs can help track user personas, beliefs, and intentions more effectively. Additionally, aligning responses with diverse social norms and cultural nuances ensures inclusive interactions in global virtual communities.
- Establishing effective human-chatbot teaming capabilities. Metaverse chatbots must function as effective “co-participants” in interaction rituals, requiring mutual predictability, directability, and common ground maintenance. Transparent decision-making processes (*e.g.*, policy gradient methods) can enhance user trust, while dynamic autonomy adjustments allow real-time customization (*e.g.*, modifying response styles in virtual education). Maintaining common ground ensures alignment with evolving user contexts, and adaptive goal negotiation frameworks enable flexible collaboration in scenarios like virtual project management.
- Building robust technical foundations and ensuring ecosystem integration. Overcoming technical limitations is crucial for scalable, safe deployment in the metaverse. Key challenges include improving interpretability in deep neural dialogue models and mitigating hallucinations *via* hybrid neural-symbolic approaches. Addressing data scarcity through LLM-driven augmentation is vital for niche domains. Explicit safety constraints (*e.g.*, signaling limitations) prevent coordination failures in high-stakes interactions. Evaluation should extend beyond accuracy to user satisfaction and task success. Standardized, schema-driven design principles are needed for cross-platform compatibility in applications ranging from customer service to collaborative problem-solving.
- Advancing scalable multi-agent collaboration frameworks. Future research should prioritize MARL frameworks to manage the metaverse's dynamic, large-scale interactions. Techniques such as mean-field MARL can effectively model aggregated

user behavior in densely populated environments (e.g., social hubs), while Q-function factorization enables the decomposition of complex joint tasks, facilitating seamless collaboration among chatbots, virtual assistants, and NPCs in activities like virtual events or co-creation projects. Additionally, integrating MARL with game theory, social choice theory, and distributed optimization could enhance interaction models, improve strategic behavior handling, and optimize coordination mechanisms for complex multi-agent tasks.

Limitations

This study has several limitations. First, as the survey focused exclusively on disembodied, text-based chatbots, literature concerning embodied robots and speech-based conversational agents was excluded. We acknowledge that incorporating these modalities can enhance chatbots' social capabilities, particularly for traits significantly influenced by physical representation, tone, and accent (e.g., identity, politeness, and reasoning). However, embodiment and multimodal interaction introduce additional complexities (e.g., speech synthesis or facial expression simulation) that lie beyond this study's scope and require more advanced technological capabilities. Furthermore, this exclusion may have resulted in the under-representation of specific research domains. For instance, a substantial body of work within human-robot interaction, particularly in the humanoid robotics discipline (e.g., [Tong & Haotian Liu \(2024\)](#), [Rawal & Stock-Homburg \(2022\)](#)), investigates facial emotion expression and is therefore not represented here. Similarly, research within design engineering (e.g., [Spatola & Wudarczyk \(2021\)](#), [Lawrence et al. \(2025\)](#)) has extensively examined specific social attributes and behaviors of robots, focusing primarily on social rather than functional aspects of interaction, and is consequently underrepresented in this study.

Second, because the concept of social capability lacks consensus within the literature and chatbots have been studied across diverse domains, their application within specific subfields may have been overlooked. To mitigate this limitation, we incorporated multiple synonyms into our search strategy and leveraged Google Scholar and arXiv for its broad interdisciplinary coverage.

Finally, the conceptual model of social capabilities was derived through a qualitative coding process informed by theories such as IRC. Like any qualitative method, this approach relies on researcher interpretation of interactive phenomena. To address this limitation, the researchers discussed social capabilities and categories in expert forums until consensus was reached. Both the framework and the relationships among social capabilities were based on findings from the surveyed studies.

CONCLUSIONS

Although chatbots demonstrate significant efficacy in handling metaverse-related tasks, their social capabilities—essential for fostering virtual communities—remain understudied. We propose an analytical framework, grounded in the IRC theory, to standardize the evaluation of chatbots' social capabilities. By aligning these capabilities

with the core components of the IRC theory, our framework facilitates a systematic assessment of their conversational and social intelligence. We further identify and elaborate on four key elements—virtual coexistence, dynamic constraint, shared intention, and positive incentive—to highlight challenges and propose potential solutions.

Proactive chatbots play a pivotal role in virtual coexistence by recognizing emotions, engaging in supportive dialogue, and inferring user intentions to provide context-aware services. They must handle both single-turn and multi-turn interactions seamlessly, foster trust through anticipatory responses, and ensure communicability—especially in initial interactions—to guide users effectively. Future research could draw on metaverse literature to develop intelligent interactive systems capable of minimizing user misunderstandings. The metaverse generates vast amounts of multimodal data characterized by high variety, velocity, and veracity, presenting challenges for dynamic constraint management. MML could enhance emotion recognition and foster a sense of belonging, while KGs may serve as dynamic knowledge bases for social network representation and complex reasoning. TL could enable chatbots to generalize across constrained domains without requiring in-domain data. Despite their autonomy, LLM-based chatbots exhibit performance gaps compared to avatars, particularly in long-term reasoning, decision-making, and instruction adherence—issues that warrant further investigation. Successful interaction rituals require chatbots to provide positive incentives through multi-agent mechanisms: Value function factorization quantifies individual contributions to global rewards. Consensus idea application limits communication to sparsely connected networks. Efficient self-structured negotiation enables collaborative learning and strategy adaptation.

This study also presented the interdependencies among these elements and proposed future research directions for developing chatbot-enhanced metaverse environments. Our analysis revealed that: (1) virtual coexistence improves user-chatbot interactions, enforces dynamic constraints through proactive dialogue, and strengthens positive incentives for engagement; (2) dynamic constraints, inherent in multiplayer metaverse settings, maintain conversational consistency while reflecting shared intentions through common action recall; (3) shared intentions reinforce group cohesion through dynamic constraints, enable autonomous planning *via* positive incentives, and optimize multiplayer experiences; and (4) positive incentives regulate shared intentions by addressing divergent requests, affect virtual coexistence to avoid multi-role conflicts, and enhance interactivity in group collaboration. Key areas for future investigation include: developing advanced multimodal socio-emotional intelligence, establishing effective human-chatbot teaming capabilities, building robust technical foundations and ensuring ecosystem integration, and advancing scalable multi-agent collaboration frameworks. This study provides both theoretical and practical contributions for optimizing chatbot efficacy within the metaverse.

ACKNOWLEDGEMENTS

The authors would like to thank all academic advisors of the Faculty of Data Science at the City University of Macau for their insightful comments.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This research is supported by NSFC-FDCT under its Joint Scientific Research Project Fund (Grant No. 0051/2022/AFJ), China & Macau S.A.R. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

NSFC-FDCT under its Joint Scientific Research Project Fund: 0051/2022/AFJ.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Yan Yan conceived and designed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Mengjuan Fan performed the experiments, performed the computation work, prepared figures and/or tables, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

This article is a literature review.

REFERENCES

- Aksu T, Kan M-Y, Chen N. 2023.** Prompter: zero-shot adaptive prefixes for dialogue state tracking domain adaptation. In: Rogers A, Boyd-Graber J, Okazaki N, eds. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg: Association for Computational Linguistics, 4588–4603.
- Ali MR, Razavi SZ, Langevin R, Al Mamun A, Kane B, Rawassizadeh R, Schubert LK, Hoque E. 2020.** A virtual conversational agent for teens with autism spectrum disorder: experimental results and design lessons. In: *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents, IVA '20*. New York, NY, USA: Association for Computing Machinery.
- Carlini N, Paleka D, Dvijotham KD, Steinke T, Hayase J, Cooper AF, Lee K, Jagielski M, Nasr M, Conmy A, Wallace E, Rolnick D, Tramèr F. 2024.** Stealing part of a production language model. In: *Proceedings of the 41st International Conference on Machine Learning, ICML'24*. JMLR.org.
- Cassano L, Yuan K, Sayed AH. 2020.** Multiagent fully decentralized value function learning with linear convergence rates. *IEEE Transactions on Automatic Control* **66**(4):1497–1512 DOI [10.1109/tac.2020.2995814](https://doi.org/10.1109/tac.2020.2995814).
- Chen Z, Zhou E, Eaton K, Peng X, Riedl M. 2023b.** Ambient adventures: teaching ChatGPT on developing complex stories. ArXiv DOI [10.48550/arXiv.2308.01734](https://doi.org/10.48550/arXiv.2308.01734).
- Chen H, Zhu T, Zhang T, Zhou W, Yu PS. 2023a.** Privacy and fairness in federated learning: on the perspective of tradeoff. *ACM Computing Surveys* **56**(2):1–37 DOI [10.1145/3606017](https://doi.org/10.1145/3606017).

- Cheng Y, Liu W, Li W, Wang J, Zhao R, Liu B, Liang X, Zheng Y. 2022.** Improving multi-turn emotional support dialogue generation with lookahead strategy planning. In: *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 3014–3026.
- Collins R. 2005.** *Interaction ritual chains*. Princeton: Princeton University Press.
- Cui Y, Wang Y, Sun Z, Liu W, Jiang Y, Han K, Hu W. 2023.** Lifelong embedding learning and transfer for growing knowledge graphs. *Proceedings of the AAAI Conference on Artificial Intelligence* 37(4):4217–4224 DOI 10.1609/aaai.v37i4.25539.
- Dai R, Zhang Y, Fang Z, Han B, Tian X. 2023.** Moderately distributional exploration for domain generalization. In: *Proceedings of the 40th International Conference on Machine Learning, ICML'23*. JMLR.org.
- De Gennaro M, Krumhuber EG, Lucas G. 2020.** Effectiveness of an empathic chatbot in combating adverse effects of social exclusion on mood. *Frontiers in Psychology* 10:3061 DOI 10.3389/fpsyg.2019.03061.
- Deng Y, Zhang W, Yuan Y, Lam W. 2023.** Knowledge-enhanced mixed-initiative dialogue system for emotional support conversations. In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Toronto, Canada: Association for Computational Linguistics.
- Dwivedi YK, Hughes L, Baabdullah AM, Ribeiro-Navarrete S, Giannakis M, Al-Debei MM, Dennehy D, Metri B, Buhalis D, Cheung CMK, Conboy K, Doyle R, Dubey R, Dutot V, Felix R, Goyal DP, Gustafsson A, Hinsch C, Jebabli I, Janssen M, Kim Y-G, Kim J, Koos S, Kreps D, Kshetri N, Kumar V, Ooi K-B, Papagiannidis S, Pappas IO, Polyviou A, Park S-M, Pandey N, Queiroz MM, Raman R, Rauschnabel PA, Shirish A, Sigala M, Spanaki K, Wei-Han Tan G, Tiwari MK, Viglia G, Wamba SF. 2022.** Metaverse beyond the hype: multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management* 66(3):102542 DOI 10.1016/j.ijinfomgt.2022.102542.
- Filippova K. 2020.** Controlled hallucinations: learning to generate faithfully from noisy data. In: *Findings of the Association for Computational Linguistics: EMNLP 2020*. Online: Association for Computational Linguistics, 864–870.
- Ge S, Xiong C, Rosset C, Overwijk A, Han J, Bennett P. 2023.** Augmenting zero-shot dense retrievers with plug-in mixture-of-memories. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Singapore: Association for Computational Linguistics, 1796–1812.
- Goldblum M, Tsipras D, Xie C, Chen X, Schwarzschild A, Song D, Madry A, Li B, Goldstein T. 2023.** Dataset security for machine learning: data poisoning, backdoor attacks, and defenses. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45(2):1563–1580 DOI 10.1109/tpami.2022.3162397.
- Guan C, Chen F, Yuan L, Zhang Z, Yu Y. 2025.** Efficient communication via self-supervised information aggregation for online and offline multiagent reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems* 36(5):9044–9056 DOI 10.1109/tnnls.2024.3420791.
- Guo X, Shi D, Fan W. 2023.** Scalable communication for multi-agent reinforcement learning via transformer-based email mechanism. In: *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 126–134.
- Hasan M, Ozel C, Potter S, Hoque E. 2023.** SAPIEN: affective virtual agents powered by large language models. In: *2023 11th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)*. Piscataway: IEEE.

- Hassani H, Razavi-Far R, Saif M, Herrera-Viedma E. 2022. Reinforcement learning-based feedback and weight-adjustment mechanisms for consensus reaching in group decision making. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 53(4):2456–2468 DOI 10.1109/tsmc.2022.3214221.
- Held W, Hidey C, Liu F, Zhu E, Goel R, Yang D, Shah R. 2023. DAMP: doubly aligned multilingual parser for task-oriented dialogue. In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Strousburg: Association for Computational Linguistics, 3586–3604.
- Hong S, Zhuge M, Chen J, Zheng X, Cheng Y, Wang J, Zhang C, Wang Z, Yau SKS, Lin Z, Zhou L, Ran C, Xiao L, Wu C, Schmidhuber J. 2024. MetaGPT: meta programming for a multi-agent collaborative framework. In: *The Twelfth International Conference on Learning Representations*.
- Huang Z, Gutierrez S, Kamana H, MacNeil S. 2023. Memory sandbox: transparent and interactive memory management for conversational agents. In: *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, UIST '23 Adjunct*. New York, NY, USA: Association for Computing Machinery.
- Huang L, Yu W, Ma W, Zhong W, Feng Z, Wang H, Chen Q, Peng W, Feng X, Qin B, Liu T. 2025. A survey on hallucination in large language models: principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems* 43(2):1–55 DOI 10.1145/3703155.
- Huang M, Zhu X, Gao J. 2020. Challenges in building intelligent open-domain dialog systems. *ACM Transactions on Information Systems* 38(3):1–32 DOI 10.1145/3383123.
- Hui B, Yuan H, Gong N, Burlina P, Cao Y. 2024. Pleak: prompt leaking attacks against large language model applications. In: *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security, CCS '24*. New York, NY, USA: Association for Computing Machinery, 3600–3614.
- Jang T, Zheng F, Wang X. 2021. Constructing a fair classifier with generated fair data. *Proceedings of the AAAI Conference on Artificial Intelligence* 35(9):7908–7916 DOI 10.1609/aaai.v35i9.16965.
- Jiang H, Cheng Y, Yang J, Gao S. 2022. AI-powered chatbot communication with customers: dialogic interactions, satisfaction, engagement, and customer behavior. *Computers in Human Behavior* 134(3):107329 DOI 10.1016/j.chb.2022.107329.
- Jiang J, Lu Z. 2018. Learning attentional communication for multi-agent cooperation. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS'18*. Red Hook: Curran Associates Inc., 7265–7275.
- Johannessen LE, Collins R. 2024. The future of interaction rituals: an interview with Randall Collins. *Theory and Society* 53(6):1491–1504 DOI 10.1007/s11186-024-09587-y.
- Kent ML, Taylor M. 2002. Toward a dialogic theory of public relations. *Public Relations Review* 28(1):21–37 DOI 10.1016/s0363-8111(02)00108-x.
- Kent ML, Taylor M. 2021. Fostering dialogic engagement: toward an architecture of social media for social change. *Social Media+ Society* 7(1):2056305120984462 DOI 10.1177/2056305120984462.
- Kong Y, Ruan J, Chen Y, Zhang B, Bao T, Shiwei S, Qing dG, Hu X, Mao H, Li Z, Zeng X, Zhao R, Wang X. 2024. TPTU-v2: boosting task planning and tool usage of large language model-based agents in real-world industry systems. In: Derroncourt F, Preoțiu-Pietro D, Shimorina A, eds. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*. Miami, Florida: US. Association for Computational Linguistics, 371–385.

- Kuhail MA, Alturki N, Alramlawi S, Alhejori K. 2023.** Interacting with educational chatbots: a systematic review. *Education and Information Technologies* **28**(1):973–1018 DOI [10.1007/s10639-022-11177-3](https://doi.org/10.1007/s10639-022-11177-3).
- Kurniawan MH, Handiyani H, Nuraini T, Hariyati RTS, Sutrisno S. 2024.** A systematic review of artificial intelligence-powered (AI-powered) chatbot intervention for managing chronic illness. *Annals of Medicine* **56**(1):2302980 DOI [10.1080/07853890.2024.2302980](https://doi.org/10.1080/07853890.2024.2302980).
- Lawrence S, Jouaiti M, Hoey J, Nehaniv CL, Dautenhahn K. 2025.** The role of social norms in human-robot interaction: a systematic review. *Journal of Human-Robot Interaction* **14**(3):1–44 DOI [10.1145/3722120](https://doi.org/10.1145/3722120).
- Lee N, Ping W, Xu P, Patwary M, Fung PN, Shoeybi M, Catanzaro B. 2022.** Factuality enhanced language models for open-ended text generation. In: Koyejo S, Mohamed S, Agarwal A, Belgrave D, Cho K, Oh A, eds. *Advances in Neural Information Processing Systems*. Vol. 35. Red Hook, New York: Curran Associates, Inc., 34586–34599.
- Lei S, Wang X, Dong G, Li J, Liu Y. 2023.** Watch the speakers: a hybrid continuous attribution network for emotion recognition in conversation with emotion disentanglement. In: *2023 IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI)*, 881–888.
- Li P, Hao J, Tang H, Zheng Y, Fu X. 2023a.** RACE: improve multi-agent reinforcement learning with representation asymmetry and collaborative evolution. In: *Proceedings of the 40th International Conference on Machine Learning, ICML’23*. JMLR.org.
- Li Z, Wang C, Wang S, Gao C. 2023b.** Protecting intellectual property of large language model-based code generation apis via watermarks. In: *Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security, CCS ’23*. New York, NY, USA: Association for Computing Machinery, 2336–2350.
- Li Y, Zhao H. 2023.** EM pre-training for multi-party dialogue response generation. In: Rogers A, Boyd-Graber J, Okazaki N, eds. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg: Association for Computational Linguistics, 92–103.
- Liu S, Cho H, Freedman M, Ma X, May J. 2023d.** RECAP: retrieval-enhanced context-aware prefix encoder for personalized dialogue response generation. In: Rogers A, Boyd-Graber J, Okazaki N, eds. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg: Association for Computational Linguistics, 8404–8419.
- Liu D, Dai W, Zhang H, Jin X, Cao J, Kong W. 2023a.** Brain-machine coupled learning method for facial emotion recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **45**(9):10703–10717 DOI [10.1109/tpami.2023.3257846](https://doi.org/10.1109/tpami.2023.3257846).
- Liu Y, Feng S, Wang D, Zhang Y, Schütze H. 2023e.** PVGRU: generating diverse and relevant dialogue responses via pseudo-variational mechanism. In: Rogers A, Boyd-Graber J, Okazaki N, eds. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg: Association for Computational Linguistics, 3295–3310.
- Liu Z, He Y, Wang W, Wang W, Wang Y, Chen S, Zhang Q, Lai Z, Yang Y, Li Q, Yu J, Li K, Chen Z, Yang X, Zhu X, Wang Y, Wang L, Luo P, Dai J, Qiao Y. 2023f.** InternGPT: solving vision-centric tasks by interacting with chatgpt beyond language. ArXiv DOI [10.48550/arXiv.2305.05662](https://doi.org/10.48550/arXiv.2305.05662).
- Liu H, Li C, Wu Q, Lee YJ. 2023b.** Visual instruction tuning. In: *Proceedings of the 37th International Conference on Neural Information Processing Systems, NIPS ’23*. Red Hook: Curran Associates Inc.

- Liu J, Wang P, Shang Z, Wu C. 2023c.** IterDE: an iterative knowledge distillation framework for knowledge graph embeddings. In: *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence, AAAI'23/IAAI'23/EAAI'23*. AAAI Press.
- Lu Y, Bartolo M, Moore A, Riedel S, Stenetorp P. 2022.** Fantastically ordered prompts and where to find them: overcoming few-shot prompt order sensitivity. In: Muresan S, Nakov P, Villavicencio A, eds. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg: Association for Computational Linguistics, 8086–8098.
- Lu S, Zhang K, Chen T, Basar T, Horesh L. 2021.** Decentralized policy gradient descent ascent for safe multi-agent reinforcement learning. *Proceedings of the AAAI Conference on Artificial Intelligence* **35(10)**:8767–8775 DOI [10.1609/aaai.v35i10.17062](https://doi.org/10.1609/aaai.v35i10.17062).
- Lv F, Chen X, Huang Y, Duan L, Lin G. 2021.** Progressive modality reinforcement for human multimodal emotion recognition from unaligned multimodal sequences. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 2554–2562.
- Lv A, Li J, Chen Y, Xing G, Zhang J, Yan R. 2023a.** DialoGPS: dialogue path sampling in continuous semantic space for data augmentation in multi-turn conversations. In: Rogers A, Boyd-Graber J, Okazaki N, eds. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Piscataway: IEEE, 1267–1280.
- Lv A, Li J, Xie S, Yan R. 2023b.** Envisioning future from the past: hierarchical duality learning for multi-turn dialogue generation. In: Rogers A, Boyd-Graber J, Okazaki N, eds. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg: Association for Computational Linguistics, 7382–7394.
- Mei Y, Zhou H, Lan T. 2023.** ReMIX: regret minimization for monotonic value function factorization in multiagent reinforcement learning. ArXiv DOI [10.48550/arXiv.2302.05593](https://doi.org/10.48550/arXiv.2302.05593).
- Mordatch I, Abbeel P. 2018.** Emergence of grounded compositional language in multi-agent populations. In: *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence, AAAI'18/IAAI'18/EAAI'18*. AAAI Press.
- Mothukuri V, Parizi RM, Pouriyeh S, Huang Y, Dehghantanha A, Srivastava G. 2021.** A survey on security and privacy of federated learning. *Future Generation Computer Systems* **115(4)**:619–640 DOI [10.1016/j.future.2020.10.007](https://doi.org/10.1016/j.future.2020.10.007).
- Nasr M, Shokri R, Houmansadr A. 2019.** Comprehensive privacy analysis of deep learning: passive and active white-box inference attacks against centralized and federated learning. In: *2019 IEEE Symposium on Security and Privacy (SP)*, 739–753.
- Ng SWT, Zhang R. 2025.** Trust in AI chatbots: a systematic review. *Telematics and Informatics* **97(12)**:102240 DOI [10.1016/j.tele.2025.102240](https://doi.org/10.1016/j.tele.2025.102240).
- Nishio T, Yonetani R. 2019.** Client selection for federated learning with heterogeneous resources in mobile edge. In: *ICC, 2019—2019 IEEE International Conference on Communications (ICC)*. Piscataway: IEEE.
- Nißen M, Selimi D, Janssen A, Cardona DR, Breitner MH, Kowatsch T, von Wangenheim F. 2022.** See you soon again, chatbot? a design taxonomy to characterize user-chatbot relationships with different time horizons. *Computers in Human Behavior* **127(2)**:107043 DOI [10.1016/j.chb.2021.107043](https://doi.org/10.1016/j.chb.2021.107043).

- Park S-M, Kim Y-G. 2022.** A metaverse: taxonomy, components, applications, and open challenges. *IEEE Access* **10**:4209–4251 DOI [10.1109/access.2021.3140175](https://doi.org/10.1109/access.2021.3140175).
- Qiqi W. 2023.** A study of factors influencing user stickiness behavior in online learning communities—based on the theory of interactive ritual chains. *The Frontiers of Society, Science and Technology* **5**(4):50–56 DOI [10.25236/fsst.2023.050409](https://doi.org/10.25236/fsst.2023.050409).
- Raman M, Maini P, Kolter J, Lipton Z, Pruthi D. 2023.** Model-tuning via prompts makes NLP models adversarially robust. In: Bouamor H, Pino J, Bali K, eds. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg: Association for Computational Linguistics, 9266–9286.
- Ranasinghe S, Gamage G, Moraliyage H, Mills N, McCaffrey N, Bucholc J, Lane K, Cahill A, White V, De Silva D. 2022.** An artificial intelligence framework for the detection of emotion transitions in telehealth services. In: *2022 15th International Conference on Human System Interaction (HSI)*.
- Rasendrasoa S, Pauchet A, Saunier J, Adam S. 2022.** Real-time multimodal emotion recognition in conversation for multi-party interactions. In: *Proceedings of the 2022 International Conference on Multimodal Interaction, ICMI '22*. New York, NY, USA: Association for Computing Machinery, 395–403.
- Rashid T, Samvelyan M, De Witt CS, Farquhar G, Foerster J, Whiteson S. 2020.** Monotonic value function factorisation for deep multi-agent reinforcement learning. *Journal of Machine Learning Research* **21**(1):7234–7284.
- Rawal N, Stock-Homburg RM. 2022.** Facial emotion expressions in human-robot interaction: a survey. *International Journal of Social Robotics* **14**(7):1583–1604 DOI [10.1007/s12369-022-00867-0](https://doi.org/10.1007/s12369-022-00867-0).
- Senadheera S, Yigitcanlar T, Desouza KC, Mossberger K, Corchado J, Mehmood R, Li RYM, Cheong PH. 2025.** Understanding chatbot adoption in local governments: a review and framework. *Journal of Urban Technology* **2025** **32**(3):35–69 DOI [10.1080/10630732.2023.2297665](https://doi.org/10.1080/10630732.2023.2297665).
- Shen S, Ma C, Li C, Liu W, Fu Y, Mei S, Liu X, Wang C. 2023.** RiskQ: risk-sensitive multi-agent reinforcement learning value factorization. In: *Proceedings of the 37th International Conference on Neural Information Processing Systems, NIPS '23*. Red Hook: Curran Associates Inc.
- Shen S, Qiu M, Liu J, Liu W, Fu Y, Liu X, Wang C. 2022.** ResQ: a residual Q function-based approach for multi-agent reinforcement learning value factorization. In: *Proceedings of the 36th International Conference on Neural Information Processing Systems, NIPS '22*. Red Hook: Curran Associates Inc.
- Shinn N, Cassano F, Gopinath A, Narasimhan K, Yao S. 2023.** Reflexion: language agents with verbal reinforcement learning. In: *Proceedings of the 37th International Conference on Neural Information Processing Systems, NIPS '23*. Red Hook: Curran Associates Inc.
- Sicilia A, Alikhani M. 2023.** Learning to generate equitable text in dialogue from biased training data. In: Rogers A, Boyd-Graber J, Okazaki N, eds. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg: Association for Computational Linguistics, 2898–2917.
- Simons I. 2019.** Events and online interaction: the construction of hybrid event communities. *Leisure Studies* **38**(2):145–159 DOI [10.1080/02614367.2018.1553994](https://doi.org/10.1080/02614367.2018.1553994).
- Spatola N, Wudarczyk OA. 2021.** Implicit attitudes towards robots predict explicit attitudes, semantic distance between robots and humans, anthropomorphism, and prosocial behavior: from attitudes to human robot interaction. *International Journal of Social Robotics* **13**(5):1149–1159 DOI [10.1007/s12369-020-00701-5](https://doi.org/10.1007/s12369-020-00701-5).

- Sun X, Li J, Tao J. 2022.** Emotional conversation generation orientated syntactically constrained bidirectional-asynchronous framework. *IEEE Transactions on Affective Computing* **13**(1):187–198 DOI [10.1109/taffc.2019.2923619](https://doi.org/10.1109/taffc.2019.2923619).
- Sun G, Zhan X, Such J. 2024.** Building better AI agents: a provocation on the utilisation of persona in LLM-based conversational agents. In: *Proceedings of the 6th ACM Conference on Conversational User Interfaces, CUI '24*. New York, NY, USA: Association for Computing Machinery.
- Sunehag P, Lever G, Gruslys A, Czarnecki WM, Zambaldi V, Jaderberg M, Lanctot M, Sonnerat N, Leibo JZ, Tuyls K, Graepel T. 2018.** Value-decomposition networks for cooperative multi-agent learning based on team reward. In: *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, AAMAS '18*. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2085–2087.
- Tam W, Huynh T, Tang A, Luong S, Khatri Y, Zhou W. 2023.** Nursing education in the age of artificial intelligence powered chatbots (AI-chatbots): are we ready yet? *Nurse Education Today* **129**(8):105917 DOI [10.1016/j.nedt.2023.105917](https://doi.org/10.1016/j.nedt.2023.105917).
- Tan X, Lin Y, Zhu S, Qu C, Qiu X, Xu Y, Cui P, Qi Y. 2023.** Provably invariant learning without domain information. In: *Proceedings of the 40th International Conference on Machine Learning, ICML'23*. JMLR.org.
- Tang F, Chen X, Zhao M, Kato N. 2023b.** The roadmap of communication and networking in 6G for the metaverse. *IEEE Wireless Communications* **30**(4):72–81 DOI [10.1109/mwc.019.2100721](https://doi.org/10.1109/mwc.019.2100721).
- Tang C, Zhang H, Loakman T, Lin C, Guerin F. 2023a.** Enhancing dialogue generation via dynamic graph knowledge aggregation. In: Rogers A, Boyd-Graber J, Okazaki N, eds. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg: Association for Computational Linguistics, 4604–4616.
- Tong Y, Haotian Liu ZZ. 2024.** Advancements in humanoid robots: a comprehensive review and future prospects. *IEEE/CAA Journal of Automatica Sinica* **11**(JAS-2023-1103):301 DOI [10.1109/jas.2023.124140](https://doi.org/10.1109/jas.2023.124140).
- Tu Q, Li Y, Cui J, Wang B, Wen J-R, Yan R. 2022.** MISC: a mixed strategy-aware model integrating COMET for emotional support conversation. In: Muresan S, Nakov P, Villavicencio A, eds. *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg: Association for Computational Linguistics, 308–319.
- Turing AM. 2009.** *Computing machinery and intelligence*. Cham: Springer.
- Valtolina S, Marchionna M. 2021.** Design of a chatbot to assist the elderly. In: *End-User Development: 8th International Symposium, IS-EUD 2021, Virtual Event, July 6–8, 2021, Proceedings*. Berlin, Heidelberg: Springer-Verlag, 153–168.
- Wang D, Chen Y, Grau BC. 2023.** Efficient embeddings of logical variables for query answering over incomplete knowledge graphs. In: *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence, AAAI'23/IAAI'23/EAAI'23*. AAAI Press.
- Wang Q, Ding L, Cao Y, Zhan Y, Lin Z, Wang S, Tao D, Guo L. 2023a.** Divide, conquer, and combine: mixture of semantic-independent experts for zero-shot dialogue state tracking. In: Rogers A, Boyd-Graber J, Okazaki N, eds. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg: Association for Computational Linguistics, 2048–2061.

- Wang L, Ma C, Feng X, Zhang Z, Yang H, Zhang J, Chen Z, Tang J, Chen X, Lin Y, Zhao WX, Wei Z, Wen J. 2024a. A survey on large language model based autonomous agents. *Frontiers of Computer Science* 18(6):186345 DOI 10.1007/s11704-024-40231-1.
- Wang Z, Mao S, Wu W, Ge T, Wei F, Ji H. 2024c. Unleashing the emergent cognitive synergy in large language models: a task-solving agent through multi-persona self-collaboration. In: Duh K, Gomez H, Bethard S, eds. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. Stroudsburg: Association for Computational Linguistics, 257–279.
- Wang F-Y, Qin R, Wang X, Hu B. 2022. Metasocieties in metaverse: metaeconomics and metamanagement for metaenterprises and metacities. *IEEE Transactions on Computational Social Systems* 9(1):2–7 DOI 10.1109/tcss.2022.3145165.
- Wang Y, Su Z, Zhang N, Xing R, Liu D, Luan TH, Shen X. 2023b. A survey on metaverse: fundamentals, security, and privacy. *IEEE Communications Surveys & Tutorials* 25(1):319–352 DOI 10.1109/COMST.2022.3202047.
- Wang Y, Zheng Z, Zhao X, Li J, Wang Y, Zhao D. 2023c. VSTAR: a video-grounded dialogue dataset for situated semantic understanding with scene and topic transitions. In: Rogers A, Boyd-Graber J, Okazaki N, eds. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Toronto, Canada: Association for Computational Linguistics, 5036–5048.
- Wang L, Zhu K, Zhu F, Yao X, Zhang S, Ye D, Fu H, Fu Q, Yang W. 2024b. Reaching consensus in cooperative multi-agent reinforcement learning with goal imagination. ArXiv DOI 10.48550/arXiv.2403.03172.
- Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, Chi EH, Le QV, Zhou D. 2022. Chain-of-thought prompting elicits reasoning in large language models. In: *Proceedings of the 36th International Conference on Neural Information Processing Systems, NIPS '22*. Red Hook: Curran Associates Inc.
- Xi Z, Chen W, Guo X, He W, Ding Y, Hong B, Zhang M, Wang J, Jin S, Zhou E, Zheng R, Fan X, Wang X, Xiong L, Zhou Y, Wang W, Jiang C, Zou Y, Liu X, Yin Z, Dou S, Weng R, Qin W, Zheng Y, Qiu X, Huang X, Zhang Q, Gui T. 2025. The rise and potential of large language model based agents: a survey. *Science China Information Sciences* 68(2):121101 DOI 10.1007/s11432-024-4222-0.
- Xie Q, Lu W, Zhang Q, Zhang L, Zhu T, Wang J. 2023. Chatbot integration for metaverse—a university platform prototype. In: *2023 IEEE International Conference on Omni-layer Intelligent Systems (COINS)*. Piscataway: IEEE.
- Xu G, Li H, Zhang Y, Xu S, Ning J, Deng RH. 2022. Privacy-preserving federated deep learning with irregular users. *IEEE Transactions on Dependable and Secure Computing* 19(2):1364–1381 DOI 10.1109/tdsc.2020.3005909.
- Xu M, Ng WC, Lim WYB, Kang J, Xiong Z, Niyato D, Yang Q, Shen X, Miao C. 2023. A full dive into realizing the edge-enabled metaverse: visions, enabling technologies, and challenges. *IEEE Communications Surveys & Tutorials* 25(1):656–700 DOI 10.1109/comst.2022.3221119.
- Xygkou A, Siriaraya P, She W-J, Covaci A, Ang CS. 2024. Can I be more social with a chatbot?": social connectedness through interactions of autistic adults with a conversational virtual human. *International Journal of Human-Computer Interaction* 40(24):8937–8954 DOI 10.1080/10447318.2023.2292880.
- Yang M, Zhao J, Hu X, Zhou W, Zhu J, Li H. 2022. LDSA: learning dynamic subtask assignment in cooperative multi-agent reinforcement learning. In: *Proceedings of the 36th International*

- Conference on Neural Information Processing Systems, NIPS '22*. Red Hook: Curran Associates Inc.
- Yao S, Yu D, Zhao J, Shafran I, Griffiths TL, Cao Y, Narasimhan K. 2023a.** Tree of thoughts: deliberate problem solving with large language models. In: *Proceedings of the 37th International Conference on Neural Information Processing Systems, NIPS '23*. Red Hook: Curran Associates Inc.
- Yao Z, Zhang W, Chen M, Huang Y, Yang Y, Chen H. 2023c.** Analogical inference enhanced knowledge graph embedding. In: *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence, AAAI'23/IAAI'23/EAAI'23*. AAAI Press.
- Yao S, Zhao J, Yu D, Du N, Shafran I, Narasimhan K, Cao Y. 2023b.** React: synergizing reasoning and acting in language models. In: *International Conference on Learning Representations (ICLR)*.
- Yuan L, Wang C, Wang J, Zhang F, Chen F, Guan C, Zhang Z, Zhang C, Yu Y. 2022a.** Multi-agent concentrative coordination with decentralized task representation. In: *Proceedings of the 31st International Joint Conference on Artificial Intelligence, IJCAI 2022*. International Joint Conferences on Artificial Intelligence, 599–605.
- Yuan L, Wang J, Zhang F, Wang C, Zhang Z, Yu Y, Zhang C. 2022b.** Multi-agent incentive communication via decentralized teammate modeling. *Proceedings of the AAAI Conference on Artificial Intelligence* **36(9)**:9466–9474 DOI [10.1609/aaai.v36i9.21179](https://doi.org/10.1609/aaai.v36i9.21179).
- Yusuf H, Money A, Daylamani-Zad D. 2025.** Pedagogical AI conversational agents in higher education: a conceptual framework and survey of the state of the art. *Educational Technology Research and Development* **73(2)**:815–874 DOI [10.1007/s11423-025-10447-4](https://doi.org/10.1007/s11423-025-10447-4).
- Zan D, Chen B, Zhang F, Lu D, Wu B, Guan B, Yongji W, Lou J-G. 2023.** Large language models meet NL2Code: a survey. In: Rogers A, Boyd-Graber J, Okazaki N, eds. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg: Association for Computational Linguistics, 7443–7464.
- Zhang Y, Khalifa M, Logeswaran L, Lee M, Lee H, Wang L. 2023b.** Merging generated and retrieved knowledge for open-domain QA. In: Bouamor H, Pino J, Bali K, eds. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Stroudsburg: Association for Computational Linguistics, 4710–4728.
- Zhang G, Liu B, Zhu T, Ding M, Zhou W. 2024.** PPFed: a privacy-preserving and personalized federated learning framework. *IEEE Internet of Things Journal* **11(11)**:19380–19393 DOI [10.1109/jiot.2024.3360153](https://doi.org/10.1109/jiot.2024.3360153).
- Zhang JJ, Wang F-Y, Wang X, Xiong G, Zhu F, Lv Y, Hou J, Han S, Yuan Y, Lu Q, Lee Y. 2018.** Cyber-physical-social systems: the state of the art and perspectives. *IEEE Transactions on Computational Social Systems* **5(3)**:829–840 DOI [10.1109/tcss.2018.2861224](https://doi.org/10.1109/tcss.2018.2861224).
- Zhang Q, Wu H, Zhang C, Hu Q, Fu H, Zhou JT, Peng X. 2023a.** Provable dynamic fusion for low-quality multimodal data. In: *Proceedings of the 40th International Conference on Machine Learning, ICML'23*. JMLR.org.
- Zhang Y, Zavlanos MM. 2024.** Cooperative multiagent reinforcement learning with partial observations. *IEEE Transactions on Automatic Control* **69(2)**:968–981 DOI [10.1109/tac.2023.3288025](https://doi.org/10.1109/tac.2023.3288025).
- Zhao K, Yang B, Lin C, Rong W, Villavicencio A, Cui X. 2023.** Evaluating open-domain dialogues in latent space with next sentence prediction and mutual information. In: Rogers A, Boyd-Graber J, Okazaki N, eds. *Proceedings of the 61st Annual Meeting of the Association for*

Computational Linguistics (Volume 1: Long Papers). Stroudsburg: Association for Computational Linguistics, 562–574.

- Zheng J, Wang Y, Tan C, Li S, Wang G, Xia J, Chen Y, Li SZ. 2023a.** CVT-SLR: contrastive visual-textual transformation for sign language recognition with variational alignment. In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 23141–23150.
- Zheng W, Yu J, Xia R, Wang S. 2023b.** A facial expression-aware multimodal multi-task learning framework for emotion recognition in multi-party conversations. In: Rogers A, Boyd-Graber J, Okazaki N, eds. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Stroudsburg: Association for Computational Linguistics, 15445–15459.
- Zhou L, Gao J, Li D, Shum H-Y. 2020.** The design and implementation of xiaoice, an empathetic social chatbot. *Computational Linguistics* **46**(1):53–93 DOI [10.1162/coli_a_00368](https://doi.org/10.1162/coli_a_00368).
- Zhou H, Lan T, Aggarwal V. 2022.** PAC: assisted value factorisation with counterfactual predictions in multi-agent reinforcement learning. In: *Proceedings of the 36th International Conference on Neural Information Processing Systems, NIPS '22*. Red Hook: Curran Associates Inc.
- Zhou S, Liu C, Ye D, Zhu T, Zhou W, Yu PS. 2022.** Adversarial attacks and defenses in deep learning: from a perspective of cybersecurity. *ACM Computing Surveys* **55**(8):1–39 DOI [10.1145/3547330](https://doi.org/10.1145/3547330).
- Zhou Y, Muresanu AI, Han Z, Paster K, Pitis S, Chan H, Ba J. 2023.** Large language models are human-level prompt engineers. In: *The Eleventh International Conference on Learning Representations*.
- Zhou S, Zhu T, Ye D, Zhou W, Zhao W. 2024.** Inversion-guided defense: detecting model stealing attacks by output inverting. *IEEE Transactions on Information Forensics and Security* **19**:4130–4145 DOI [10.1109/tifs.2024.3376190](https://doi.org/10.1109/tifs.2024.3376190).
- Zhu K, Wang J, Zhou J, Wang Z, Chen H, Wang Y, Yang L, Ye W, Zhang Y, Gong N, Xie X. 2024.** PromptRobust: towards evaluating the robustness of large language models on adversarial prompts. In: *Proceedings of the 1st ACM Workshop on Large AI Systems and Models with Privacy and Safety Analysis, LAMPS '24*. New York, NY, USA: Association for Computing Machinery, 57–68.