The Five Ws of Multi-Agent Communication: Who Talks to Whom, When, What, and Why - A Survey from MARL to Emergent Language and LLMs

Jingdi Chen[1], Hanqing Yang[1], Zongjun Liu[1], and Carlee Joe-Wong[1]

[1]Affiliation not available

September 11, 2025

The Five Ws of Multi-Agent Communication: Who Talks to Whom, When, What, and Why
A Survey from MARL to Emergent Language and LLMs

Jingdi Chen* *jingdic@arizona.edu*
University of Arizona
Tucson, AZ 85719

Hanqing Yang *hanqing3@andrew.cmu.edu*
Carnegie Mellon University
Pittsburgh, PA 15213

Zongjun Liu *msnliu@arizona.edu*
University of Arizona
Tucson, AZ 85719

Carlee Joe-Wong *cjoewong@andrew.cmu.edu*
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

Multi-agent sequential decision-making underlies many real-world systems, from autonomous vehicles and robotics to collaborative AI assistants. In dynamic and partially observable environments, effective communication is crucial for reducing uncertainty and enabling coordination. While research in multi-agent communication (MA-Comm) spans diverse methods and paradigms, its central challenges can often be understood through the guiding lens of the *Five Ws of communication*: who talks to whom, when to speak, what to convey, and why communication is beneficial. These questions provide an intuitive thread across different approaches, even when not used as explicit section divisions. Progress in this field has been rapid. Within the *Multi-Agent Reinforcement Learning (MARL)* framework, early work emphasized static, hand-designed protocols, while later approaches introduced trainable, end-to-end communication models optimized with deep learning. This shift sparked interest in *emergent language*, where agents develop symbolic or structured messaging strategies through interaction. More recently, *large language models (LLMs)* have opened new possibilities, enabling natural language as a medium for reasoning, planning, and collaboration in more open-ended environments. Despite this momentum, there is still no dedicated survey that brings together these different lines of work. Most existing reviews focus narrowly on MARL, without fully addressing how communication is evolving from simple message passing to symbolic reasoning and language use. This paper aims to fill that gap. We provide a structured survey of MA-Comm, spanning traditional MARL approaches and emergent language studies. In light of growing interest in agentic and embodied AI, we also examine how LLMs are reshaping communication in both MARL contexts and broader multi-agent ecosystems. By using the *Five Ws* as a conceptual lens, our goal is to clarify the landscape, highlight key trends, and provide a foundation for future research at the intersection of communication, coordination, and learning in multi-agent systems.

*This work was done during Jingdi Chen's postdoctoral research at Carnegie Mellon University.

Contents

1 Introduction

Multi-agent decision-making plays a critical role in a wide range of real-world applications, including robotics Gu et al. (2016); Zhang et al. (2015), such as navigation Candido & Hutchinson (2011) and manipulation Pajarinen & Kyrki (2017); autonomous systems Talpaert et al. (2019), including autonomous driving Wang et al. (2019a); and planning under uncertainty Wang et al. (2019b); Cheng et al. (2018). In these settings, multiple agents must act independently while interacting with both the environment and one another. Depending on the task, agents may cooperate to achieve shared objectives, compete over limited resources, or engage in mixed-motive interactions that involve both collaboration and competition. A central challenge in these systems is uncertainty, which can arise from multiple sources: limited or partial observability of the global state or of other agents' intentions, as well as inherent stochasticity in environment dynamics. Effective multi-agent systems must therefore address problems of coordination, decentralized control, and decision-making under uncertainty, often in real time and with limited communication.

To address these challenges, reinforcement learning (RL) has been widely explored as a framework for training agents to make sequential decisions through trial and error. Deep Reinforcement Learning (DRL), in particular, has been extensively studied in both simulated Mnih et al. (2013); Lillicrap et al. (2015) and real-world Gu et al. (2016); Zhang et al. (2015); Qureshi et al. (2017); Meng et al. (2019) scenarios. While many DRL algorithms assume fully observable environments modeled as Markov Decision Processes (MDPs) Lample & Chaplot (2016); Schulman et al. (2015); Pipattanasomporn et al. (2009); Li et al. (2002), real-world multi-agent systems often involve Partially Observable Markov Decision Processes (POMDPs), where each agent has only a limited view of the system state. In such cases, inter-agent communication can help mitigate partial observability by allowing agents to exchange information and build a more complete representation of the environment Chen et al. (2024b). As a result, an important research direction in MARL focuses on designing effective communication strategies to improve decision-making in complex environments.

Effective communication is a core capability in multi-agent systems, particularly in scenarios requiring coordination, negotiation, or competition among autonomous agents. In MARL, agents must make decisions based on local observations while interacting with others in partially observable and often stochastic environments. Communication enables agents to exchange task-relevant information, such as observations, goals, intentions, or strategies, which helps them align actions, resolve uncertainty, and coordinate more effectively. While this is critical in fully cooperative settings like multi-robot collaboration Li et al. (2002) or smart grid control Pipattanasomporn et al. (2009), communication also plays a nuanced but important role in competitive and mixed-motive scenarios. For example, in adversarial games like StarCraft Samvelyan et al. (2019) or Dota Berner et al. (2019), agents may engage in strategic signaling, misinformation, or negotiation to manipulate opponents or form temporary alliances. In such cases, communication becomes not just a tool for cooperation, but also a mechanism for influencing, deceiving, or adapting to others' behaviors in strategic interactions.

To move beyond manually designed protocols, recent research has focused on learned communication, where agents develop communication strategies through interaction. A widely used approach involves enabling gradient flow between agents during training, allowing them to optimize both message content and timing based on task performance. This line of work offers a scalable and flexible alternative to hand-crafted rules, enabling agents to discover efficient and adaptive messaging schemes tailored to their goals and environmental dynamics—whether cooperative, competitive, or somewhere in between.

While MARL has made substantial strides in learning communication protocols within reinforcement learning frameworks, alternative approaches for enabling agent cooperation have increasingly drawn attention. One such direction is emergent language, which studies how agents can develop structured communication protocols through repeated interaction. This line of work provides insights into how meaningful communication can emerge organically in cooperative settings, without requiring pre-defined languages or protocols Lazaridou & Baroni (2020); Li et al. (2022). More recently, large language models (LLMs) have opened new possibilities for multi-agent coordination. With their strong reasoning abilities, vast prior knowledge, and capacity to generalize across tasks, LLMs offer a flexible and powerful alternative to traditional MARL agents. LLM-based systems can communicate using natural language, infer shared goals, and adapt to new environments with minimal additional training Yang et al. (2025). These complementary approaches—emergent language and LLM-based agents—highlight the growing need to broaden the scope of multi-agent communication beyond conventional MARL frameworks.

Communication in MARL Research on multi-agent communication has evolved significantly, mirroring broader advances in artificial intelligence and multi-agent systems. Initial efforts in this area focused on enabling communication in multi-agent reinforcement learning to enhance coordination under partial observability. Early studies addressed foundational questions such as *who should communicate* and *what information should be shared* Paulos et al. (2019); Lowe et al. (2017); Foerster et al. (2016a); Sukhbaatar et al. (2016); Jiang & Lu (2018); Das et al. (2018); Rangwala & Williams (2020). Many of these early approaches employed *centralized training with decentralized execution*, where a centralized coordinator facilitates communication during training while agents act independently during execution. For example, *CommNet* Sukhbaatar et al. (2016) and *IC3Net* Singh et al. (2018) aggregate hidden states across agents to produce shared representations. Later methods introduced more flexible mechanisms, such as attention-based message routing in *TarMAC* Das et al. (2018) and graph-structured message propagation in *DICG* Li et al. (2020), allowing agents to dynamically select communication partners. While these models often rely on centralized components during training, they still support decentralized execution, preserving scalability and autonomy in deployment. While these neural architectures enabled learned communication, they often *lacked interpretability*, as deep neural networks generate complex messages that are difficult to analyze Brown et al. (2020); LeCun et al. (2015). Moreover, most approaches assumed agents could exchange continuous, real-valued messages, ignoring the practical constraints of real-world communication networks, which typically rely on discrete and bandwidth-limited transmissions Foerster et al. (2016a); Lowe et al. (2017); Mordatch & Abbeel (2018); Freed et al. (2020b;a). These limitations motivated research into *emergent discrete communication*, where agents develop structured, interpretable protocols for exchanging information.

Communication in Emergent Language Learned communication in emergent language has been a promising avenue for improving interpretability and structure in multi-agent systems. To move beyond opaque continuous message spaces, early research explored how agents could develop discrete communication protocols through interaction—using one-hot messages Chaabouni et al. (2019); Kottur et al. (2017); Havrylov & Titov (2017); Lazaridou et al. (2016); Lee et al. (2017), binary signals Foerster et al. (2016a), or other constrained message formats Eccles et al. (2019). These efforts made agent communication more interpretable to humans, but often revealed coordination challenges, such as zero-shot failures where independently trained agents could not understand each other's learned protocols Hu et al. (2020b). To address this, researchers have begun to design environments and training methods that encourage more robust, generalizable, and human-aligned communication Bullard et al. (2021; 2020). While many of these efforts originate in the context of MARL, their implications extend to multi-agent decision-making more broadly, including settings beyond traditional reinforcement learning. For example, some works have aimed to align emergent communication with natural language Lee et al. (2019); Lowe et al. (2020), while others integrate pre-trained language models to introduce linguistic priors and promote structured, interpretable coordination Lazaridou et al. (2020); Tucker et al. (2021). These developments reflect a growing trend toward bridging learned communication protocols with human language to support not only efficient coordination in MARL, but also broader agent collaboration and human-AI interaction.

Communication in Large Language Models The rise of Large Language Models has introduced a new paradigm for multi-agent communication, expanding beyond traditional MARL frameworks. Unlike communication in MARL, where agents develop protocols from scratch through reinforcement learning, LLM-based agents leverage pre-trained linguistic knowledge to engage in structured, natural language-based exchanges and decision-making Du et al. (2023); Liang et al. (2023); Wang et al. (2023c). This is distinct from emergent language research, where the focus is on how agents develop communication protocols in task-driven settings without any pre-existing linguistic knowledge. Recent work has proposed a range of communication architectures for LLM-based multi-agent systems, including direct messaging, chain-of-thought interactions, hierarchical structures, and graph-based exchanges Zhang et al. (2023c); Du et al. (2023); Qian et al. (2023); Hong et al. (2023); Holt et al.; Wu et al. (2023b); Jiang et al. (2023); Chan et al. (2023); Qian et al. (2024b); Zhuge et al. (2024b). These frameworks enable agents to *collaborate, debate, plan, and reason* across diverse environments while offering scalability and adaptability across tasks. A key distinction is that LLM-based communication builds on existing natural language capabilities, whereas emergent language approaches study how communication can arise from scratch through learning. Nevertheless, insights from emergent language research can inform LLM-based systems by guiding the design of more interpretable, compositional, and generalizable communication protocols. Bridging these lines of work offers a promising direction toward hybrid communication frameworks that integrate learned behaviors, structured emergent signals, and the expressiveness of pre-trained language models.

The Need for a Dedicated Survey Despite substantial progress, most existing work on multi-agent communication is discussed as a subsection within broader MARL surveys Wong et al. (2023); Gronauer & Diepold (2022); Oroojlooy & Hajinezhad (2023); Nguyen et al. (2020); Hernandez-Leal et al. (2019); Zaïem & Bennequin (2019). Given the increasing complexity and diversity of approaches—from MARL-based communication and EL to LLM-driven coordination—there is a growing need for a *dedicated and structured review of MA-Comm*. This survey aims to formalize the field, highlight key research challenges, and provide a roadmap for future research at the intersection of reinforcement learning, natural language processing, and multi-agent cooperation. The main contributions of this paper are as follows:

- **A Novel Survey on Multi-Agent Communication:** This paper is the first to provide a dedicated survey on Multi-Agent Communication beyond the scope of traditional MARL-based methods. By integrating insights from EL research and recent advances in LLMs, we offer a broader perspective on agent communication for decision-making in multi-agent systems.

- **A Three-Dimensional Review Framework:** We introduce a structured framework for analyzing MA-Comm research, categorizing existing studies into: (1) MARL with Communication, (2) Emergent Language in Multi-Agent Systems, and (3) Multi-Agent Approaches with Large Language Models. This framework enables a systematic comparison across different paradigms, bridging past and recent developments in multi-agent communication.

- **Comprehensive Analysis of Key Works:** Within each research dimension, we systematically categorize studies and highlight key contributions. We compare and contrast representative methodologies, distilling essential paradigms and trends that have shaped the evolution of MA-Comm.

- **Future Research Directions and Open Challenges:** We identify critical gaps in current MA-Comm research and propose a roadmap for future developments. Key challenges include improving communication efficiency, interpretability, and scalability, as well as exploring novel techniques to enhance multi-agent cooperation in real-world applications.

In the order of MARL with Communication, Emergent Language, and LLM-based Multi-Agent Systems, we structure the paper as follows. Section 2 reviews existing surveys on MARL, emergent communication, and LLM-based agents, identifying the need for a unified perspective on multi-agent communication (MA-Comm). Section 3, Section 4, and Section 5 then present the three core dimensions of MA-Comm: MARL with learned communication protocols, emergent language systems, and LLM-powered multi-agent communication, respectively. For each dimension, we first introduce essential background and key modeling assumptions, followed by a categorization framework and a survey of representative works. Each section highlights the role of communication in enabling coordination, adaptability, and reasoning under different paradigms. Finally, in Section 6, we provide an in-depth discussion of challenges, limitations, and future research directions, drawing connections across the three communication paradigms and identifying open questions at the intersection of MARL, emergent language, and large language models.

2 Related Work - Surveying the Three Dimensions of Multi-Agent Communication

Prior efforts to survey communication in multi-agent systems have evolved along several lines. In this work, we focus on three prominent and interrelated dimensions: communication in MARL, emergent language systems, and LLM-driven communication frameworks. While not exhaustive, these dimensions capture a wide spectrum of recent research exploring how agents exchange information to support coordination, learning, and reasoning. Existing MARL surveys often address communication only indirectly, emphasizing policy learning while overlooking communication-specific challenges. Surveys on emergent communication focus on language development and interpretability but are typically siloed from broader multi-agent learning discussions. More recent reviews on LLM-based agents emphasize architectural workflows and coordination capabilities, yet often remain disconnected from prior frameworks. Our survey aims to bridge these perspectives through a unifying framework (Fig. 1) that supports comparative analysis and highlights shared trends, challenges, and open questions across these three influential threads of multi-agent communication research.

2.1 Surveys on Communication in MARL

Previous surveys on MARL have primarily provided broad overviews of the field, only occasionally addressing communication as a secondary or supporting topic. Hernandez-Leal et al. Hernandez-Leal et al. (2019) reviewed various MARL methodologies, emphasizing algorithmic frameworks and their critiques, with limited exploration of agent communication strategies. Similarly, Nguyen et al. Nguyen et al. (2020) presented a comprehensive analysis of MARL, highlighting challenges and general solutions but allocating minimal attention to the specifics of communication mechanisms.

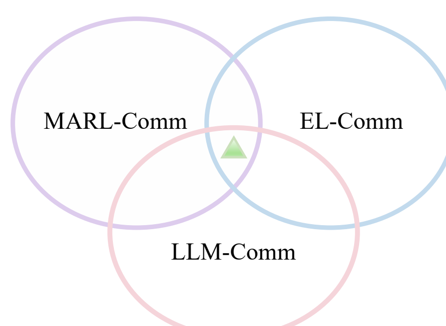


Figure 1: Three core dimensions of multi-agent communication surveyed in this work: MARL-based, emergent language, and LLM-driven communication.

Gronauer and Diepold Gronauer & Diepold (2022) conducted an extensive survey focusing predominantly on deep reinforcement learning techniques and their multi-agent extensions. They discussed communication briefly, primarily in the context of agent coordination tasks, without providing an in-depth or structured exploration. Wong et al. Wong et al. (2023) also outlined various

challenges and future directions in multi-agent systems, briefly
touching upon communication but primarily maintaining a focus on general reinforcement learning issues rather than systematically reviewing communication techniques.

Recent surveys like that of Oroojlooy and Hajinezhad Oroojlooy & Hajinezhad (2023) continue this trend, examining cooperation mechanisms extensively while only superficially addressing communication methodologies, and Beikmohammadi Beikmohammadi (2024) presents a systematic review of 16 MARL communication studies from 2016–2021, focusing on their methodological rigor and identifying gaps such as a lack of standardized environments and consideration of real-world communication constraints. Even the works specifically addressing communication, such as Zaiem et al. Zaïem & Bennequin (2019) and Zhu et al. Zhu et al. (2024) have limitations, either narrowly focusing on MARL-specific methods or not incorporating recent advancements involving emergent languages and large language models (LLMs).

In contrast, our work uniquely addresses the significant gap in the literature by providing the first dedicated, structured survey explicitly focusing on Communication. Our contributions differ significantly from prior surveys in several ways. First, we establish a novel three-dimensional review framework that integrates MARL-based communication approaches, emergent language paradigms, and recent advancements in LLM-driven agent communication. Secondly, we systematically categorize and comprehensively analyze key works across these dimensions, highlighting essential trends, similarities, and differences. Finally, we explicitly outline open challenges and suggest clear research directions to enhance the efficiency, interpretability, scalability, and applicability of MA-Comm in real-world multi-agent systems.

2.2 Surveys on Emergent Communication Protocols in Multi-Agent Systems

Research on emergent language in multi-agent systems has gained traction as a promising pathway toward interpretable and grounded communication. While Boldt and Mortensen Boldt & Mortensen (2024) provide a broad survey covering emergent communication across machine learning, natural language processing, and cognitive science, their review spans beyond the scope of decision-making. In contrast, our focus centers specifically on how emergent language facilitates coordination and policy learning in sequential decision-making settings. They emphasize emergent communication's potential for explainability and human-agent interaction, while also categorizing practical and theoretical use cases. Peters et al. Peters et al. (2024) present a large-scale review of 181 papers, proposing a taxonomy for discrete emergent language. They systematically analyze evaluation metrics and identify challenges in language comparability, compositionality, and semantic alignment. Their contribution is methodological, aiming to standardize how emergent language is assessed. Wolff et al. Wolff et al. (2024) address the limitations of traditional reference games and introduce more realistic, situated environments for emergent language learning. Their work highlights the emergence of bidirectional and sparse communication strategies that arise naturally under coordination pressure in complex MARL settings. Other focused surveys and domain-specific efforts include Chafii et al.Chafii et al. (2023), which explores EC-MARL in future 6G wireless networks, and Gupta et al.Gupta et al. (2020), who analyze emergent communication grounded in graph-based multi-agent topologies. While valuable, these works are contextually bounded and do not offer a general-purpose, integrative perspective.

Prior surveys on emergent language have primarily focused on the linguistic and structural properties of communication protocols, such as syntax, semantics, compositionally, and their resemblance to human language. These studies are often motivated by cognitive science and language understanding, aiming to investigate how structured language can emerge in artificial agents through interaction, without necessarily grounding that communication in task-driven decision-making. In contrast, Boldt and Mortensen's recent survey Boldt & Mortensen (2024) is among the few that explicitly address emergent communication in the context of multi-agent coordination, i.e., how structured language can emerge through agent interactions when multiple agents interact and communicate with each other. This highlights its role not just as a linguistic phenomenon but as a mechanism for enabling effective joint decision-making.

The inclusion of EL in this survey is motivated by the pioneering work of Mordatch & Abbeel (2018), which first established a connection between emergent communication and MARL methods through MADDPG Lowe et al. (2017). This work demonstrated how grounded, compositional language can emerge among agents as a mechanism to complete tasks and achieve goals in multi-agent environments. The symbolic and discrete nature of the communication protocols developed in this setting subsequently inspired a range of

MARL communication methods, such as Havrylov & Titov (2017); Chen et al. (2024b), which adopted discrete messaging to reduce communication overhead and enhance interpretability. Summarizing EL research in the context of multi-agent communication thus offers an interpretable framework for agent collaboration and promotes the development of human-understandable communication protocols, contributing toward the broader goal of human-centered AI.

Therefore, our survey complements and extends these efforts by incorporating EL as one of the three central dimensions of multi-agent communication. Unlike prior reviews that emphasize emergent communication in isolation, we provide a unified framework that links MARL communication, emergent language, and LLM-driven communication strategies. This broader integration facilitates comparisons, highlights cross-cutting challenges, and sets the stage for designing generalizable and human-aligned communication protocols in future multi-agent systems.

2.3 Integrating LLMs into Multi-Agent Communication: Recent Surveys and Trends

Recent surveys have extensively explored the integration of LLMs into multi-agent systems (MAS), particularly focusing on enhancing inter-agent communication, coordination, and collaboration. Wang et al. Wang et al. (2024a) provide a comprehensive review of LLM-based autonomous agents, categorizing core capabilities such as communication, planning, memory, and tool use. Li et al. Li et al. (2024d) offer a workflow-centric perspective, outlining key infrastructure components including profile modeling, self-action reasoning, perception, interaction, and learning in LLM-based MAS.

Guo et al. Guo et al. (2024b) survey the progress and challenges of large language model-based multi-agents, covering aspects such as reasoning, cooperation, and agent modularity. In the context of dialogue systems, Yi et al. Yi et al. (2024) and Guan et al. Guan et al. (2025) analyze recent advances and evaluation strategies for LLM-based multi-turn conversations, highlighting communication bottlenecks and coordination failures. Hu et al. Hu et al. (2024b) focuses on LLM-based agents in game environments, discussing planning, interaction dynamics, and decision-making complexity. While these surveys highlight the growing role of LLMs in multi-agent decision-making, they largely treat communication as a secondary concern rather than a primary focus, leaving open the need for a more targeted synthesis of communication strategies in LLM-powered multi-agent systems.

Several surveys have examined collaboration and coordination among LLM agents more directly. Tran et al. Tran et al. (2025a) review mechanisms for multi-agent collaboration using LLMs, while Sun et al. Sun et al. (2025b) survey coordination strategies across diverse application domains. Sun et al. Sun et al. (2024a) specifically explore how LLMs are being integrated with MARL, summarizing current progress and proposing future research directions. Aratchige and Ilmini Aratchige & Ilmini (2025) further examine the technological challenges in building effective LLM-based multi-agent systems, addressing issues like communication reliability, shared memory, and uncertainty handling.

Complementary to these broad surveys, several evaluation studies have investigated the fundamental limitations of LLM-based multi-agent cooperation. Mosquera et al. Mosquera et al. (2024) empirically assess cooperative behaviors in LLM-augmented agents using the Melting Pot benchmark, while Cemri et al. Cemri et al. (2025) provide an in-depth analysis of why multi-agent LLM systems often fail, identifying critical weaknesses in alignment, communication consistency, and robustness. Moreover, Liu et al. Liu et al. (2025a) present a broader vision of 'foundation agents', exploring brain-inspired architectures, evolutionary collaboration mechanisms, and the pursuit of safe, scalable multi-agent intelligence.

In contrast to existing surveys that primarily focus on architectural capabilities, evaluation benchmarks, or emerging challenges in LLM-based multi-agent systems, our work frames LLM-enabled communication as one of three central paradigms within a broader taxonomy of MA-Comm. By connecting LLM-driven interaction with traditional MARL communication protocols and EL systems, we provide an integrative perspective that spans diverse research communities. While MA-Comm addresses a broad range of coordination and decision-making objectives beyond human involvement, the incorporation of LLMs introduces new opportunities for more natural, interpretable, and potentially human-aligned communication. This broader lens helps uncover shared challenges and promising directions for advancing communication strategies across agent types and application domains.

Table 1: Comparison of Surveys Across MA-Comm Dimensions: our survey uniquely bridges three domains, MARL-Comm, EL, and LLM-based MAS, offering a unified framework that highlights cross-cutting trends, systematic methodologies, and open challenges for building scalable, interpretable, and human-centric multi-agent communication systems.

Survey	**Focus Area**	**Communication Focused**	**LLM Focused**	**Emergent Language**	**Systematic Framework**
Hernandez-Leal et al. (2019)	MARL	✗	✗	✗	✗
Nguyen et al. (2020)	MARL	✗	✗	✗	✗
Gronauer & Diepold (2022)	MARL	▲	✗	✗	✗
Wong et al. (2023)	MARL	▲	✗	✗	✗
Oroojlooy & Hajinezhad (2023)	MARL	▲	✗	✗	✗
Beikmohammadi (2024)	MARL-COMM	✓	✗	✗	✓
Zaïem & Bennequin (2019)	MARL-COMM	✓	✗	✗	✓
Zhu et al. (2024)	MARL-COMM	✓	✗	✗	✓
Boldt & Mortensen (2024)	Emergent	▲	✗	✓	✓
Peters et al. (2024)	Emergent	✓	✗	✓	✓
Wolff et al. (2024)	Emergent	✓	✗	✓	✓
Chafii et al. (2023)	Emergent (Wireless)	▲	✗	✓	✗
Gupta et al. (2020)	Emergent (Graphs)	▲	✗	✓	✓
Wang et al. (2024a)	LLM	▲	✓	✗	✓
Li et al. (2024d)	LLM	✓	✓	✗	✓
Guo et al. (2024b)	LLM	▲	✓	✗	✓
Yi et al. (2024)	LLM (Dialogue)	▲	✓	✗	✓
Guan et al. (2025)	LLM (Dialogue Eval.)	▲	✓	✗	✓
Hu et al. (2024b)	LLM (Games)	▲	✓	✗	✓
Tran et al. (2025a)	LLM (Collaboration)	✓	✓	✗	✓
Sun et al. (2025b)	LLM (Coordination)	▲	✓	✗	✓
Sun et al. (2024a)	LLM (MARL)	▲	✓	✗	✓
Mosquera et al. (2024)	LLM (Evaluation)	✓	✓	✗	✗
Cemri et al. (2025)	LLM (Evaluation)	✓	✓	✗	✗
Liu et al. (2025a)	LLM (Future)	▲	✓	✗	✓
Aratchige & Ilmini (2025)	LLM (Technological Aspects)	✓	✓	✗	✓
Our Survey	**All Above**	✓	✓	✓	✓

Table 1 provides a structured comparison of major surveys across the key dimensions of MA-Comm. While earlier works focus narrowly on either MARL communication, emergent language, or LLM-based systems, none comprehensively address all three paradigms together. In contrast, our survey uniquely bridges these domains, offering a unified framework that highlights cross-cutting trends, systematic methodologies, and open challenges for building scalable, interpretable, and human-centric multi-agent communication systems.

3 Communication in Multi-Agent Reinforcement Learning

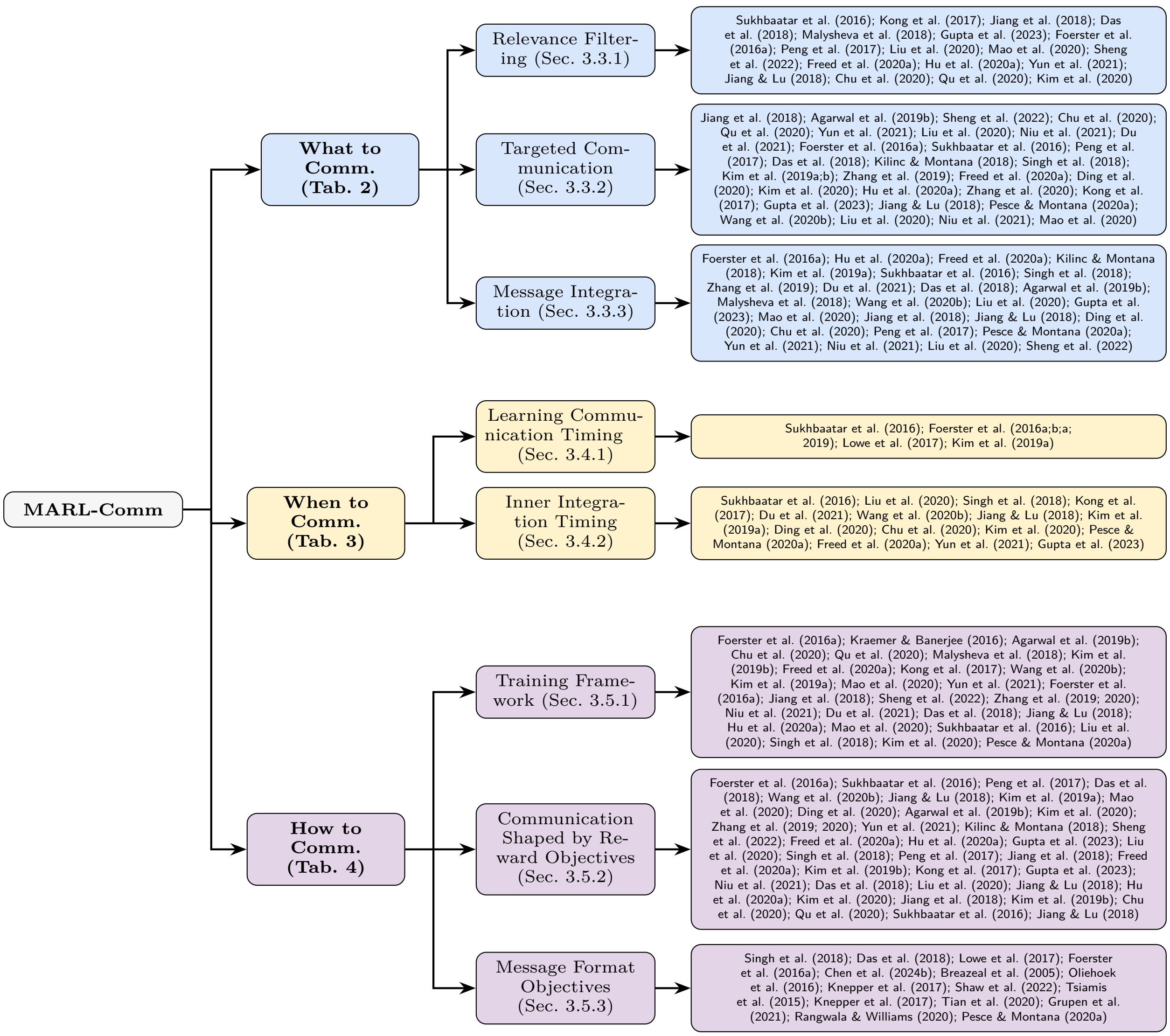


Figure 2: MARL-Comm Agents Taxonomy

Communication in Multi-Agent Reinforcement Learning (MARL-Comm) is vital for enabling agents to coordinate, share knowledge, and act collectively in dynamic environments. However, effective communication is not trivial—it must contend with constraints such as partial observability, limited bandwidth, and the need for decentralized execution. In this section, We begin with background information (Sec. 3.1), summarizing foundational concepts, and identifying challenges faced in real-world MARL systems. An overview of the related research is categorized and summarized in Sec. 3.2. To clarify the design space, we introduce a taxonomy of MARL-Comm research, organized along three key dimensions: **what to communicate** (Sec. 3.3), **when to communicate** (Sec. 3.4), and **how to communicate** (Sec. 3.5). As illustrated in Fig. 2 and Fig. 3,

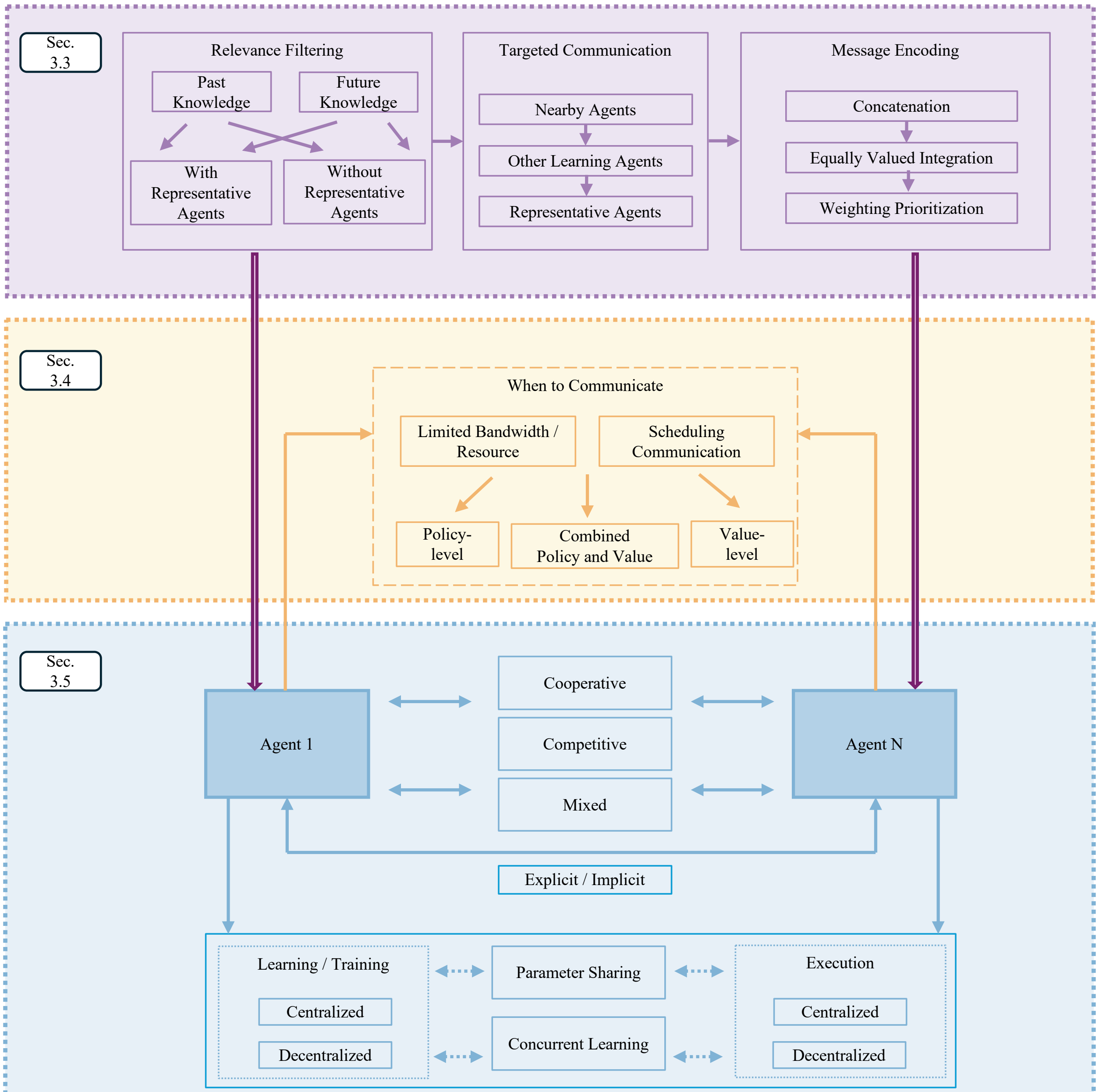


Figure 3: Organization of the key dimensions of communication design in Sec. 3: what and whom to communicate with (Sec. 3.3), when to communicate under resource constraints (Sec. 3.4), and how communication is shaped by interaction scenarios and learning architectures (Sec. 3.5). It highlights how various communication mechanisms are conditioned on task structure, bandwidth limitations, agent relationships, and training/execution paradigms.

each dimension reflects distinct algorithmic challenges, from selecting and structuring relevant messages, to deciding the timing of communication, to integrating learned protocols within the training framework. This organization helps unify the field's diverse approaches and provides a conceptual foundation for evaluating and extending MARL communication strategies.

Throughout this section, we provide illustrative examples, key algorithms, and evaluation tasks to bridge theory and real-world applications. All content adheres to the foundational Markov Decision Process (MDP) framework, ensuring consistent theoretical grounding and supporting a wide range of MARL-Comm research directions. Summarized tables at the end of each subsection serve as a reference for researchers designing future communication-centric MARL algorithms. This structure aims to provide a comprehensive reference for researchers exploring communication strategies in MARL. All discussions are contextualized within the foundational Markov Decision Process (Puterman (2014)) framework, with $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{T}, r, \rho_0(s), \gamma)$ serving as the base environment. Here, $\mathcal{S}$ represents the state space, $\mathcal{A}$ the action space, $\mathcal{T} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ the transition function, $\rho_0 : \mathcal{S} \rightarrow [0, 1]$ the initial state distribution, $r : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ the reward function, and $\gamma \in (0, 1]$ the discount factor, ensuring a consistent theoretical basis for the exploration of MARL-Comm strategies.

3.1 Background on MARL-Comm

Reinforcement learning trains models to make sequential decisions in uncertain environments by maximizing cumulative rewards through trial and error Sutton & Barto (2018). While it is traditionally designed for single agents, many reinforcement learning applications involve multiple agents, making MARL a more appropriate framework Busoniu et al. (2008). MARL studies decision-making among agents in shared environments, with applications in cyber-physical systems Adler & Blue (2002); Wang et al. (2016), communication networks Cortes et al. (2004); Choi et al. (2009), and social science Castelfranchi (2001); Leibo et al. (2017). MARL problems can be fully cooperative, fully competitive, or mixed, depending on the agents' objectives. Multi-agent communication is a fundamental aspect of MARL, enabling agents to share information, coordinate strategies, and achieve collective objectives efficiently. In MARL, agents operate in a shared environment, and their joint actions influence the state transitions and rewards. Communication allows agents to overcome partial observability, align their policies, and avoid conflicts or redundant actions.

Formal Framework Decentralized Partially Observable Markov Decision Processes (Dec-POMDPs) Bernstein et al. (2002) are a multi- agent extension of a partially observable Markov decision process, which models cooperative MARL, where agents lack complete information about the environment and have only local observations. Communication is crucial for strategy coordination in this scenario. A Dec-POMDP with communication is often modeled as a tuple Chen et al. (2024b) $D = \langle S, A, P, \Omega, O, I, n, R, \gamma, f \rangle$, where $I = \{1, 2, \ldots, n\}$ is a set of n agents, S is the joint *state* space and $A = A_1 \times A_2 \times \cdots \times A_n$ is the joint *action* space. Here $\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in A$ denotes the joint action of all agents and A_i is the individual action space of agent i. The joint action influences the environment, which transitions to a new state and provides a global reward r. $P(\boldsymbol{s}'|\boldsymbol{s}, \boldsymbol{a}) : S \times A \times S \rightarrow [0, 1]$ is the *state transition function*. Ω is the *observation* space. $O(\boldsymbol{s}, i) : S \times I \rightarrow \Omega$ is a function that maps from the joint state space to distributions of observations for each agent i. $R(\boldsymbol{s}, \boldsymbol{a}) : S \times A \rightarrow \mathbb{R}$ is the *reward function* in terms of state s and joint action $\boldsymbol{a}$, and γ is the discount factor. Communication can be modeled as an additional action space, where agents send and receive messages m_i to share information. At each time step, agent i generates a message $m_i = f_i(o_i, h_i)$ based on its local observation $o_i \in O$ and internal state $h_i \in H$, where $f_i : O \times H \rightarrow M$ is a message-generation function that maps observations and internal states to a message space M, often implemented as a neural network. Each agent i receives messages $\{m_j\}_{j \in \mathcal{N}_i}$ from a (possibly dynamic) subset of other agents $\mathcal{N}_i \subseteq \{1, \ldots, N\} \setminus \{i\}$, and aggregates them into a context vector $c_i = g_i(\{m_j\}_{j \in \mathcal{N}_i})$, where g_i is an aggregation function such as averaging, concatenation, or attention. The agent's policy π_i uses the aggregated context c_i along with its local observation o_i to select an action $a_i \sim \pi_i(o_i, c_i)$. This framework highlights the dual role of communication: (1) enabling agents to share local observations and (2) facilitating coordination by aligning their policies based on shared information.

Based on this background, we provide a review of communication-based MARL algorithms in the following subsections, with a particular focus on works whose primary novelty lies in the design of communication strategies. One category of such algorithms aims to enhance cooperative multi-agent learning by improving the efficiency and relevance of the communication process. These works either introduce mechanisms such as gating and selective message filtering to reduce redundant communication or leverage advanced techniques like attention mechanisms and graph-based communication protocols to ensure meaningful message exchange. Another category focuses on extended MARL setups, such as hierarchical or task-specific communication, utilizing the idea that communication messages can serve as embeddings or representations of tasks or roles within the multi-agent system.

3.2 An overview on MARL-Comm

Communication plays a crucial role in MARL by enabling agents to share information, coordinate effectively, and tackle complex tasks in partially observable environments. Early foundational works, such as CommNet Sukhbaatar et al. (2016), DIAL, and RIAL Foerster et al. (2016a), pioneered communication learning for cooperative agents. CommNet uses a shared neural network to process local observations and aggregates messages via mean pooling for decentralized decision-making. While effective in small-scale settings, it struggles with scalability and communication delays. DIAL and RIAL introduced more explicit communication protocols, where agents exchange discrete or continuous messages and jointly learn both their policies and communication strategies through reinforcement learning. These approaches, while innovative, require fully connected communication networks, leading to high bandwidth demands as the number of agents increases.

To address the scalability and efficiency challenges, subsequent research has focused on adaptive and selective communication mechanisms. For instance, IC3Net Singh et al. (2018) employs gating mechanisms to allow agents to dynamically decide whether to communicate based on their current states, reducing redundant messages. TarMAC Das et al. (2018) introduces multi-headed attention to improve the relevance of communicated information, enabling selective sharing among agents. Similarly, ATOC Jiang & Lu (2018) forms dynamic communication groups based on agent proximity and employs bi-directional LSTMs to aggregate messages within groups. More advanced methods, such as SARNet Rangwala & Williams (2020), incorporate structured reasoning and attention mechanisms to assess the relevance of received messages and past memories, enabling more nuanced decision-making. These methods collectively explore key questions such as *when*, *what*, and *with whom* agents should communicate, offering significant performance improvements in MARL tasks.

Recent work also explores communication constraints, such as limited bandwidth and noisy channels. For example, SchedNet Kim et al. (2019a) selects only a subset of agents to communicate, optimizing within bandwidth limits, while IMAC Wang et al. (2020b) regularizes message entropy to reduce communication overhead. These approaches emphasize the trade-offs between communication efficiency and performance, paving the way for practical MARL systems in real-world scenarios. To further advance MARL-Comm, new methods must address challenges such as scalability, interpretability, and performance guarantees under realistic constraints Chen et al. (2024b). Future work should explore integrating sparse, discrete communication protocols with theoretical guarantees, such as regret minimization, to enhance robustness and applicability in large-scale environments Chen et al. (2024b).

Based on the foundational advancements in MARL-Comm, we provide a review of communication-based MARL algorithms in the following subsections, with a particular focus on works whose primary novelty lies in designing efficient and scalable communication strategies. One category of such algorithms seeks to enhance MARL-Comm by introducing mechanisms like attention or gating to optimize when, what, and with whom agents should communicate. These approaches aim to reduce redundant communication, improve message relevance, and address practical constraints such as limited bandwidth or noisy channels.

Key Challenges Despite its potential, multi-agent communication in MARL faces several core challenges: **(1). Scalability in Large State and Agent Spaces:** In large-scale environments, each agent typically observes only a small portion of the global state. Agents often have limited local observations, making it difficult to infer the global state or the intentions of other agents. Communication is essential for enabling coordination under these conditions, but exchanging high-dimensional or unfiltered observations among many agents can be inefficient and computationally prohibitive Cao et al. (2013). The challenge lies in designing scalable protocols that selectively share information while preserving task-relevant signals. **(2). Communication Overhead:** As the number of agents increases, communication complexity grows rapidly, leading to significant overhead. Techniques such as message filtering, attention mechanisms, and hierarchical structures aim to mitigate this. For instance, CommNet Sukhbaatar et al. (2016) uses a shared centralized architecture to enable continuous communication, but it does not scale well with the number of agents. MADDPG Lowe et al. (2017) introduces a centralized critic and decentralized actors, where agents generate communication as part of their actions. However, its policy network often overfits to specific agent configurations, limiting generalizability in large-scale scenarios. **(3). Non-Stationarity:** In MARL, each agent operates in a non-stationary environment because other agents are concurrently learning and updating their policies. This dynamic context makes it difficult to learn stable communication strategies, as the meaning and utility of messages may shift over time. **Credit Assignment:** In cooperative settings, determining the contribution of each agent to the collective reward is non-trivial, yet essential for effective learning. Without proper credit assignment, agents cannot accurately update their policies based on individual impact, which can lead to suboptimal or unstable learning dynamics. Communication can help agents better coordinate their actions, but it also introduces additional dependencies between agents' behaviors and messages, making it more complex to attribute success or failure to specific contributions. **(5). Interpretability:** Learned communication protocols in MARL are often embedded in latent continuous spaces, making them hard to interpret. Unlike hand-crafted protocols with explicit semantics, learned messages lack transparency, making it difficult to analyze their meaning, influence on decision-making, or alignment with human reasoning. **(6). Lack of Theoretical Guarantees:** Many MARL-Comm methods rely on heuristic mechanisms for learning continuous or discrete communication Foerster et al. (2016a); Singh et al. (2018); Jiang & Lu (2018); Das et al. (2018); Rangwala & Williams (2020); Havrylov & Titov (2017); Mordatch & Abbeel (2018). While some approaches model discrete, human-like messages, they often lack formal guarantees such as policy convergence or regret minimization, limiting their theoretical robustness and practical reliability.

Key Methodologies Several approaches have been proposed to address these challenges, which can be broadly categorized as: **(1). Centralized Communication**, such as CommNet (Sukhbaatar et al. (2016)), uses a shared neural network to aggregate messages from all agents and compute a global context. The message generation and aggregation steps are formalized as: $m_i = f_i(o_i, h_i); \quad c_i = \frac{1}{N-1}\sum_{j\neq i} m_j$, where c_i is the averaged message from all other agents. While this approach simplifies coordination, it introduces a single point of failure and may not scale well to large systems. **(2). Decentralized Communication**, such as TarMAC (Das et al. (2018)), enables agents to communicate directly with each other using attention mechanisms. The message aggregation step is formalized as: $c_i = \sum_{j\neq i} \alpha_{ij} m_j, \quad \alpha_{ij} = \text{softmax}(u_i^T W u_j)$, where α_{ij} is the attention weight between agents i and j, u_i and u_j are learned embeddings, and W is a learned weight matrix. This approach is more scalable but requires robust protocols to handle message delays and redundancies. **(3). Learning to Communicate**, such as IC3Net (Singh et al. (2018)), focuses on end-to-end learning of communication protocols. The policy update step is formalized as: $a_i \sim \pi_i(o_i, c_i), \quad c_i = g_i(\{m_j\}_{j\neq i})$, where g_i is a learned aggregation function. This approach allows agents to develop communication strategies that are tailored to the specific task and environment. **(4). Emergent Communication**, such as in Mordatch & Abbeel (2018); Lazaridou et al. (2016), studies how agents can develop their emergent communication protocols through interaction, without explicit supervision. The message generation step is formalized as: $m_i = f_i(o_i, h_i), \quad f_i$ is learned through reinforcement learning. This approach is often studied in the context of cooperative games or language learning tasks. We will use a separate section to introduce this line of work in Sec. 4. **(5). LLM-Guided Communication:** Approaches such as Li et al. leverage Large Language Models (LLMs) to guide agent communication toward more interpretable, natural language-like structures. The message generation process is formalized as $m_i = f_i(o_i, h_i, \text{LLM})$, where f_i is a function conditioned not only on the agent's observation o_i and internal state h_i, but also informed by a pre-trained LLM. This integration allows agents to generate structured, human-understandable messages while preserving or even enhancing task performance. LLM-guided communication has shown promise in tasks like collaborative planning and ad-hoc teamwork, leading to improved generalization and faster communication emergence. We provide a detailed discussion of this research direction in Sec. 5. **(6). Return Gap Minimization**, such as in Chen et al. (2024b), quantifies the gap between the optimal expected average return of an ideal policy with full observability, i.e., $\pi^* = [\pi_i^*(a_i|o_1, \ldots, o_n), \forall i]$ and the optimal expected average return of a communication-enabled, partially-observable policy, i.e., $\pi = [\pi_i(a_i|o_i, \boldsymbol{m_{-i}}), \forall i]$. This result enables such works to recast multi-agent communication into a novel online clustering problem over the local observations at each agent, with messages as cluster labels and the upper bound on the return gap as clustering loss.

Review Scope and Structure To systematically analyze advancements in MARL-Comm, we categorize existing research based on three fundamental questions that communication algorithms seek to address: **(1) What to communicate?**, **(2) When to communicate?**, and **(3) How to communicate?**. These questions encapsulate the core challenges in designing efficient, scalable, and interpretable communication protocols for multi-agent systems. First, we examine **information selection and compression**, which focuses on determining what information should be shared among agents to maximize coordination while minimizing redundancy and communication overhead. Next, we explore **communication timing and adaptation**, which studies when agents should exchange information, balancing the trade-off between frequent updates and computational efficiency. Finally, we analyze **protocol design and coordination mechanisms**, investigating how different communication structures—whether centralized, decentralized, explicit, or implicit, which will affect performance and scalability in MARL. By structuring our review around these three dimensions, we provide a comprehensive overview of existing methodologies, highlight key advancements, and identify open research challenges in multi-agent communication in MARL.

3.3 Answering 'What to Communicate?' - Information Selection and Compression

One of the main challenges in MARL-Comm is determining what information should be shared among agents to maximize coordination while minimizing redundant or unnecessary communication. Previous surveys have treated *what* to communicate and *who* to communicate with as separate topics, despite their significant overlap in both research focus and technical solutions. In reality, these aspects are often studied jointly Das et al. (2018); Rangwala & Williams (2020), with many techniques addressing both simultaneously. Therefore, in this section, we present a unified discussion and categorization of these two topics and categorize it as 'what to communicate'. We identify three key challenges in this area: (1). **Relevance Filtering:** Agents should selectively communicate only the most critical information to enhance coordination and decision-making. (2). **Avoiding Redundancy:** Redundant information should be minimized, ensuring that agents do not transmit data that can be inferred independently. (3). **Compression and Encoding:** Efficient representation and encoding methods are necessary to reduce communication bandwidth while preserving essential information.

3.3.1 Relevance Filtering of Communication Messages

After establishing communication links among agents through a communication policy, agents must decide which pieces of information should be shared. Under the common assumption of partial observability, local observations become critical for coordination. Additionally, agents can use their historical experience, intended actions, or future plans to generate more informative messages.

Encoding Past Knowledge In this category, agents utilize existing knowledge (e.g., past observations or actions) to facilitate communication, and most recent works use an encoding of this knowledge as messages. In particular, a model from the RNN (recurrent neural networks), e.g., LSTM (long-short term memory) and GRU (gated recurrent unit)) family is often employed as the encoding function, which is able to selectively forget and store historical observations Sukhbaatar et al. (2016); Kong et al. (2017); Jiang et al. (2018); Das et al. (2018); Malysheva et al. (2018); Kilinc & Montana (2018); Singh et al. (2018); Pesce & Montana (2020a); Zhang et al. (2019); Wang et al. (2020b); Liu et al. (2020); Mao et al. (2020); Sheng et al. (2022); Freed et al. (2020a); Hu et al. (2020a); Zhang et al. (2020); Gupta et al. (2023); Niu et al. (2021); Du et al. (2021); Yun et al. (2021), or action-observation histories Foerster et al. (2016a); Peng et al. (2017). Nevertheless, if there is a representative agent, messages will be generated and transformed from agents to the representative agent, and then from the representative agent to agents. Therefore, we differentiate recent works on using existing knowledge as messages into the following two cases.

- **Through Representative Agents:** When using a representative agent, communicated messages will be generated by two steps. Firstly, local observations can be encoded Sukhbaatar et al. (2016); Kong et al. (2017); Jiang et al. (2018); Das et al. (2018); Malysheva et al. (2018) or directly sent Das et al. (2018); Malysheva et al. (2018) to the representative agent. Then, the representative agent who gathers local (encoded) observations can generate a single new message to all agents Gupta et al. (2023), or individualised messages to each agent Sukhbaatar et al. (2016); Kong et al. (2017); Jiang et al. (2018); Das et al. (2018); Malysheva et al. (2018). Both ways produce a message containing global information and agents do not need to make any efforts to consider how to combine messages.

- **Without Representative Agents:** Without using a representative agent, messages are directly sent to each agent. DIAL and RIAL Foerster et al. (2016a) use an encoding of past observations and actions, and local observation as messages. BiCNet Peng et al. (2017) is not only fed into the local view of each agent but also a global observation of the environment. Other works directly communicate observations Liu et al. (2020), or use simple feed-forward networks Liu et al. (2020); Mao et al. (2020); Sheng et al. (2022); Freed et al. (2020a), MLP (Multi-Layer Perceptron) Freed et al. (2020a); Hu et al. (2020a), autoencoder Freed et al. (2020a), CNN (Convolutional Neural Network) Liu et al. (2020); Mao et al. (2020), RNN (Recurrent Neural Network) Sukhbaatar et al. (2016); Kong et al. (2017); Jiang et al. (2018); Das et al. (2018); Malysheva et al. (2018), or GNN (Graph Neural Network) Liu et al. (2020); Freed et al. (2020a) are commonly used to encode local observations into compact representations for downstream communication and decision-making.

  In addition, agents could communicate more specific information, for example, in GAXNet Yun et al. (2021), agents coordinate their local attention weights used for combining hidden states from neighboring agents.

Encoding Future Knowledge We use 'Future Knowledge' to refer to either intended actions Jiang & Lu (2018), a policy fingerprint (i.e., current action probabilities in a particular state) Chu et al. (2020); Qu et al. (2020), or future plans Kim et al. (2020), which can be generated by simulating a model of environment dynamics Kim et al. (2020). As intentions and plans are state-related, recent works usually encode intentions together with local observations to generate more relevant messages.

3.3.2 Avoiding Redundancy via Targeted Communication

A key feature of human communication is the ability to engage in targeted interactions. Instead of broadcasting messages to all agents, as explored in previous work CommNet Sukhbaatar et al. (2016), RIAL and DIAL Foerster et al. (2016a), and IC3Net Singh et al. (2018), directing specific messages to intended recipients can enhance collaboration in complex environments. For instance, in a team of search-and-rescue robots, a message like 'smoke is coming from the kitchen' is relevant to a firefighter but irrelevant to a bomb defuser. In this section, we focused on how agents learn **whom to communicate** to while performing cooperative tasks in partially observable environments. The communicatee type defines which agents are potential recipients of messages in a MARL-Comm system. In the literature, communicatee types can be categorized based on whether agents communicate directly with one another or through a representative agent.

Communicating with Nearby Agents In many MARL systems, communication is only allowed between nearby agents. Nearby agents are defined in various ways, such as agents that are observable Yun et al. (2021), agents within a certain distance Jiang et al. (2018); Agarwal et al. (2019b); Sheng et al. (2022), or agents that are neighbors on a graph Chu et al. (2020); Qu et al. (2020). Neighboring agents can also emerge during learning instead of being predefined, as in GA-Comm Liu et al. (2020), MAGIC Niu et al. (2021), and FlowComm Du et al. (2021), which explicitly learn a graph structure among agents. However, in GA-Comm and MAGIC, a central unit (e.g., GNN) learns the graph structure and coordinates messages, so the agents do not communicate directly and are instead managed through a representative agent. Dynamic Graph Networks (DGN) Jiang et al. (2018) restrict communication to the three closest neighbors based on a predefined distance metric. This locality-based communication is particularly important in dynamic environments, where agents' positions and states change frequently, requiring timely and context-specific message exchanges to maintain effective coordination and avoid communication bottlenecks. Similarly, the Agent-Entity Graph Agarwal et al. (2019b) also uses distance to measure proximity, allowing communication between agents as long as they are close to each other. MAGNet-SA-GS-MG Malysheva et al. (2018) uses attention mechanisms to optimize communication by selecting the most relevant messages, with a pretrained graph limiting communication to neighboring agents. Learning Structured Communication (LSC) Sheng et al. (2022) enables agents within a cluster radius to decide whether to become a leader agent, with all non-leader agents communicating with the leader in that cluster. This structured communication protocol improves both communication efficiency and interpretability. NeurComm Chu et al. (2020) and Intention Propagation (IP) Qu et al. (2020) restrict communication to neighbors on a preset graph structure during learning, facilitating coordination by sharing decision-making policies and intentions (i.e., action probabilities) with neighboring agents. FlowComm Du et al. (2021) dynamically manages communication using message-passing techniques in a flow-based system, optimizing the routing of information based on environmental flow, while also learning a graph structure to identify neighboring agents. Lastly, GAXNet Yun et al. (2021) uses attention mechanisms to combine hidden states from observable neighboring agents, allowing for selective message passing that helps agents focus on critical information while filtering out irrelevant data.

Communicating with Other Learning Agents When nearby agents are not explicitly defined (e.g., via proximity-based graphs), communication typically occurs among all learning agents or through mechanisms that selectively determine recipients. Several methods support communication in this scenario by either broadcasting messages or learning whom to communicate with. DIAL (Differentiable Inter-Agent Learning) Foerster et al. (2016a) enables differentiable communication between agents, allowing them to optimize message content jointly via backpropagation. RIAL (Reinforced Inter-Agent Learning) Foerster et al. (2016a) instead uses reinforcement learning to select messages that maximize collective rewards, assuming a fixed set of recipients. CommNet Sukhbaatar et al. (2016) facilitates communication by averaging hidden states across all agents, implicitly assuming full communication among teammates. BiCNet (Bidirectionally Coordinated Network) Peng et al. (2017) applies RNNs to support recurrent communication of local and global observations, again relying on shared message exchange among all agents. TarMAC (Targeted Multi-Agent Communication) Das et al. (2018) introduces learned attention weights that determine *whom* to send messages to at each step, enabling targeted communication and reducing unnecessary message passing. MADDPG-M (Multi-Agent DDPG with Messages) Kilinc & Montana (2018) extends the MADDPG framework with continuous message exchange but does not explicitly control recipient selection, assuming messages are accessible to all agents. IC3Net (Implicit Communication in Collective Intelligence) Singh et al. (2018) enhances communication efficiency by allowing agents to learn a gating mechanism that determines *when* to communicate, reducing bandwidth usage. SchedNet Kim et al. (2019a) implements a scheduler that selects which agents can communicate in each timestep, effectively learning both *when* and *who* communicates under constrained bandwidth. DCC-MD Kim et al. (2019b) scales message exchange by dynamically dropping messages based on importance, which indirectly controls the recipient set. VBC (Value-Based Communication) Zhang et al. (2019) uses value function information to prioritize which state information to share, helping agents coordinate by highlighting key states without broadcasting all messages. Diff Discrete Freed et al. (2020a) and I2C (Intentional Information Communication) Ding et al. (2020) reduce communication load by encouraging sparse or goal-relevant communication. I2C, in particular, guides agents to share only information aligned with their current intentions, which can implicitly filter recipients. Information Sharing (IS) Kim et al. (2020) focuses on transmitting internal state representations to synchronize behavior in cooperative tasks. ETCNet (Event-Triggered Communication Network) Hu et al. (2020a) triggers communication only upon detecting significant environmental changes, thus reducing redundant message broadcasts. Finally, Variable-length Coding Freed et al. (2020a) and TMC (Temporal Message Compression) Zhang et al. (2020) compress historical information into concise messages, improving efficiency but still assuming full communication unless paired with selective mechanisms.

Communicating through Representative Agents In systems where agents communicate via a representative agent, communication between agents is mediated by an intermediary or a representative agent. This intermediary coordinates, processes, and distributes messages to and from individual agents, enhancing communication efficiency and often simplifying the coordination process. Several approaches have been proposed for this type of communication, as summarized here. MS-MARL-GCM Kong et al. (2017) introduces a master agent who gathers local observations and hidden states from agents within the environment. The master agent then sends a common message back to each agent, thereby reducing the need for direct communication between individual agents. This centralized communication model improves coordination in partially observable environments. Similarly, HAMMER Gupta et al. (2023) employs a central representative agent that collects local observations from agents in the multi-agent system (MAS) and distributes both a global and individualized message to each agent. This ensures efficient communication by aggregating and distributing relevant information to agents, facilitating better decision-making. ATOC Jiang & Lu (2018) uses an LSTM to link nearby agents who choose to join a communication group. The representative agent in this system sends coordinated messages to the group members, streamlining communication and improving the ability of agents to act cohesively based on shared information. MD-MADDPG Pesce & Montana (2020a) maintains a shared memory among agents, allowing them to selectively store local observations in the memory. The representative agent then loads and relays this information, facilitating communication while reducing message overload. This method ensures that only relevant information is communicated to the necessary agents. IMAC Wang et al. (2020b) defines a scheduler that aggregates encoded information from all agents and sends individualized messages back to each agent. This centralized scheduling of communication helps agents focus on their individual tasks while still benefiting from global knowledge. In GA-Comm Liu et al. (2020) and MAGIC Niu et al. (2021), a global message processor is used to aggregate information from agents and then distribute messages based on agent-specific weights. This enables more customized communication, where each agent receives a message tailored to its unique context within the system. Finally, Gated-ACML Mao et al. (2020) takes a slightly different approach, allowing agents to decide whether or not to communicate through a message coordinator. This introduces an additional level of decision-making, where agents evaluate the necessity of communication based on the relevance of the information at hand. These methods illustrate the utility of using representative agents to enhance communication in multi-agent systems by centralizing, filtering, and distributing messages in ways that improve overall system efficiency and agent coordination.

3.3.3 Compression and Encoding via Message Weighting

Recent studies in MARL-Comm often handle multiple incoming messages collectively. Message Integration refers to the method by which received messages are combined before being passed into the agents' internal models. In cases where a representative agent is used, the representative agent typically consolidates and coordinates messages, providing a unified single message to each agent, this eliminates the need for individual agents to manage the combination of messages. However, when no representative agent is involved, each agent must independently determine how to merge multiple incoming messages. Since the communicated messages reflect the sender's personal interpretation of the learning process or the environment, some messages may be prioritized over others.

Message Concatenation In this approach, incoming messages are concatenated into a single input vector without applying explicit preferences or weighting. This ensures that no information is discarded, but it increases the input dimensionality, which may become problematic as the number of agents grows. Methods such as DIAL Foerster et al. (2016a), RIAL Foerster et al. (2016a), ETCNet Hu et al. (2020a), Variable-length Coding Freed et al. (2020a), MADDPG-M Kilinc & Montana (2018), and Diff Discrete Freed et al. (2020a) adopt this approach. For instance, DIAL and ETCNet concatenate scalar message values, preserving simplicity while retaining essential information. SchedNet Kim et al. (2019a) uses shorter message vectors to reduce communication overhead before concatenation, allowing agents to operate more efficiently within limited bandwidth settings. Since all messages are treated equally without prioritization, this method is particularly suitable for small-scale systems, where preserving the entirety of transmitted data is more beneficial than selectively filtering or compressing messages.

Equally Valued Message Integration When agents are assumed to contribute equally, their messages are integrated using permutation-invariant operations such as summation or averaging, rather than concatenation. While concatenation also treats messages uniformly in the sense that each is included, it preserves ordering and increases dimensionality, whereas averaging or summing compresses the messages into a fixed-size representation without assuming any order. Techniques such as CommNet Sukhbaatar et al. (2016), IC3Net Singh et al. (2018), and VBC Zhang et al. (2019) employ this form of integration. In CommNet, all incoming messages are averaged to produce a single vector, ensuring equal influence from each sender. IC3Net uses a similar averaging mechanism over messages from neighboring agents. FlowComm Du et al. (2021) instead sums the messages, maintaining equal contribution while enabling gradient-based message refinement. This approach is especially effective in cooperative tasks where all agents are assumed to have similarly relevant information, allowing simple yet effective message integration without the need for attention or learned weighting.

Weighted Message Prioritization In this category, agents assign different levels of importance to received messages based on their relevance or utility to the current task. Some approaches rely on handcrafted rules to selectively prune messages deemed less useful. For example, DCC-MD Kim et al. (2019b) and TMC Zhang et al. (2020) implement manual thresholds on message magnitude or relevance scores to discard low-salience messages, thereby reducing communication overhead. More commonly, modern approaches utilize attention mechanisms to learn these weights dynamically. Attention-based models such as TarMAC Das et al. (2018), Agent-Entity Graph Agarwal et al. (2019b), and MAGNet-SA-GS-MG Malysheva et al. (2018) compute relevance scores for each incoming message and combine them using a weighted sum. This enables the receiving agent to focus on the most informative messages while suppressing noise from less critical sources.

The Use of Neural Networks Neural networks are frequently employed to automatically learn the importance of messages, enabling agents to extract and prioritize useful information without hand-crafted rules. In IMAC Wang et al. (2020b), GA-Comm Liu et al. (2020), and HAMMER Gupta et al. (2023), simple multilayer perceptrons (MLPs) are used to integrate messages while implicitly learning which ones are most valuable. Gated-ACML Mao et al. (2020) takes a similar approach but introduces a gating mechanism to regulate communication links with a central representative agent. For more complex settings, convolutional neural networks (CNNs) are used to capture spatial dependencies among agents, as in DGN Jiang et al. (2018), which is particularly beneficial in dynamic environments where agent positions change frequently and updated spatial context is essential. Recurrent neural networks (RNNs), including Long Short-Term Memory (LSTM) variants, are widely applied for sequential decision-making tasks in communication-centric coordination. These are used in ATOC Jiang & Lu (2018), I2C Ding et al. (2020), NeurComm Chu et al. (2020), BiCNet Peng et al. (2017), MD-MADDPG Pesce & Montana (2020a), and GAXNet Yun et al. (2021), where they capture temporal dependencies across communication rounds, allowing agents to make informed decisions based on evolving message histories. Lastly, graph neural networks (GNNs) have gained popularity in works such as MAGIC Niu et al. (2021), GA-Comm Liu et al. (2020), and LSC Sheng et al. (2022). These approaches utilize a learned graph structure over agents to model interactions and assign relevance to neighboring messages, often combining topological information with attention mechanisms to improve communication efficiency and task performance.

Table 2 presents a taxonomy of representative approaches to the problem of 'what to communicate' in MARL-Comm. It organizes key strategies into three overarching categories: relevance filtering, targeted communication, and message integration. Within each category, we highlight how agents address information selection and compression by focusing on different communication design choices. For relevance filtering, methods are grouped by whether they encode past knowledge (e.g., histories or observations) or future knowledge (e.g., intentions or plans), and whether communication occurs via representative agents or in a peer-to-peer manner. Targeted communication is further classified by proximity-based strategies (static vs. dynamic neighborhoods), interaction with other learning agents, and centralized communication through representative agents. Finally, message integration methods are categorized by how agents combine received messages: concatenation, equal weighting, weighted prioritization (manual or attention-based), and neural architectures such as MLPs, CNNs, RNNs, and GNNs. This structured overview reveals trends in how communication is selectively managed to balance informativeness, efficiency, and scalability in multi-agent systems.

3.4 Answering 'When To Communicate?' - Communication Timing and Adaptation for Communication Constraints

A fundamental challenge in MARL-Comm is determining *when* agents should exchange information. While continuous communication keeps agents fully informed, it can be costly, inefficient, or even counterproductive in decentralized environments with limited bandwidth or computational resources. Thus, effective communication timing is critical for balancing the benefits of coordination with practical constraints. Instead of communicating at every time step, agents must learn to identify key moments when communication provides meaningful advantage, enabling better decision-making and collaboration under resource constraints. Beyond general bandwidth constraints, sharing a communication medium introduces additional complexity, particularly when multiple agents must avoid message collisions. While protocols like Wi-Fi and LTE manage this at the packet level through well-established scheduling mechanisms, such low-level access control is typically

Category	Approach	Representative Works
Relevance Filtering (Sec. 3.3.1)	*Encoding Past Knowledge* (with/without representative agents)	• With representative agents: Sukhbaatar et al. (2016); Kong et al. (2017); Jiang et al. (2018); Das et al. (2018); Malysheva et al. (2018); Gupta et al. (2023) • Without representative agents: Foerster et al. (2016a); Peng et al. (2017); Liu et al. (2020); Mao et al. (2020); Sheng et al. (2022); Freed et al. (2020a); Hu et al. (2020a); Yun et al. (2021)
	Encoding Future Knowledge (intended actions, fingerprints, plans)	Jiang & Lu (2018); Chu et al. (2020); Qu et al. (2020); Kim et al. (2020)
Targeted Communication (Sec. 3.3.2)	*Nearby Agents* (static/dynamic proximity)	• Static: Jiang et al. (2018); Agarwal et al. (2019b); Sheng et al. (2022); Chu et al. (2020); Qu et al. (2020); Yun et al. (2021) • Dynamic: Liu et al. (2020); Niu et al. (2021); Du et al. (2021)
	Other Learning Agents (learned communicatee sets)	Foerster et al. (2016a); Sukhbaatar et al. (2016); Peng et al. (2017); Das et al. (2018); Kilinc & Montana (2018); Singh et al. (2018); Kim et al. (2019a;b); Zhang et al. (2019); Freed et al. (2020a); Ding et al. (2020); Kim et al. (2020); Hu et al. (2020a); Zhang et al. (2020)
	Representative Agents (centralized message processing)	Kong et al. (2017); Gupta et al. (2023); Jiang & Lu (2018); Pesce & Montana (2020a); Wang et al. (2020b); Liu et al. (2020); Niu et al. (2021); Mao et al. (2020)
Message Integration (Sec. 3.3.3)	*Concatenation* (scalar/short vectors)	Foerster et al. (2016a); Hu et al. (2020a); Freed et al. (2020a); Kilinc & Montana (2018); Kim et al. (2019a)
	Equal Weighting (averaging/summing)	Sukhbaatar et al. (2016); Singh et al. (2018); Zhang et al. (2019); Du et al. (2021)
	Weighted Prioritization (manual or attention-based)	• Manual pruning: Kim et al. (2019b); Zhang et al. (2020) • Attention: Das et al. (2018); Agarwal et al. (2019b); Malysheva et al. (2018)
	Neural Integration Models (MLPs, CNNs, RNNs, GNNs)	• MLPs: Wang et al. (2020b); Liu et al. (2020); Gupta et al. (2023); Mao et al. (2020) • CNNs: Jiang et al. (2018) • RNNs/LSTMs: Jiang & Lu (2018); Ding et al. (2020); Chu et al. (2020); Peng et al. (2017); Pesce & Montana (2020a); Yun et al. (2021) • GNNs: Niu et al. (2021); Liu et al. (2020); Sheng et al. (2022)

Table 2: Taxonomy of Approaches to ‘What to Communicate’ (Sec. 3.3) in MARL-Comm.

handled separately from decision-making in MARL systems. Instead, the key challenge in MARL-Comm is higher-level: determining *when* to communicate so that information is shared efficiently and meaningfully without overwhelming the channel. This involves two main concerns: (1) selecting concise, task-relevant messages, and (2) coordinating communication timing to minimize redundancy and interference Kim et al. (2019a). Rather than optimizing communication protocols directly, recent MARL approaches address these issues by integrating message selection and scheduling into agents' learning policies.

3.4.1 Learning Communication Timing and Adaptation under Constraints

Certain constraints, such as discrete decisions about whether or when to communicate, can make communication timing non-differentiable, posing challenges for gradient-based learning algorithms. This non-differentiability hinders end-to-end training of communication protocols. One common workaround is *centralized training with decentralized execution* (CTDE), where agents are trained in a centralized manner with full access to global information and differentiable approximations of discrete choices (e.g., via soft gating or policy gradients). This allows agents to learn effective communication strategies during training, even if message transmission is limited or constrained during execution. At test time, agents deploy these learned strategies under stricter, often non-differentiable, conditions. This section reviews methods that address communication constraints by learning *when* and *how* to communicate, including approaches based on CTDE, differentiable gating mechanisms, public belief modeling, and distributed communication scheduling.

Learning to communicate through backpropagation Sukhbaatar et al. Sukhbaatar et al. (2016) introduced a 'Communication Network' (CommNet) that processes partial observations from multiple agents to determine their actions. Each layer in this multi-layered network has several cells, where each cell takes as input its own previous hidden state h_m^{i-1} and aggregated messages c_m^i from other agents' hidden states. The updated hidden state is computed as $h_m^{i+1} = \sigma(C^i c_m^i + H^i h_m^i)$, where σ is a non-linear activation function. The final layer outputs the agent's action, and the entire network is trained using backpropagation based on a reward signal. This approach outperformed baseline models where communication within the network occurred through discrete symbols. Subsequent research shifted toward *centralized training with decentralized execution*, as demonstrated by Deep Recurrent Q-Networks (DRQN) Hausknecht & Stone (2015). Unlike independent Deep Q-Networks (DQN) van Hasselt et al. (2015), DRQN does not assume full observability and is better suited for partially observable multi-agent environments. By maintaining internal memory through recurrent units, DRQN helps mitigate issues of non-stationarity arising from concurrent policy updates across agents, making it more robust in decentralized multi-agent settings.

Foerster et al. Foerster et al. (2016b) extended this by proposing the two following centralized communication schemes where agents share network weights and learn to communicate effectively in partially observable environments with shared rewards.

- **Deep Distributed Recurrent Q-Network Foerster et al. (2016b):** The Deep Distributed Recurrent Q-Network (DDRQN) extends Deep Recurrent Q-Networks (DRQN) by adapting them to multi-agent settings through modifications to the independent Q-learning paradigm. Specifically, DDRQN introduces three key modifications to standard independent Q-learning: (1) experience replay is disabled, as concurrently learning agents invalidate each other's past experiences; (2) each agent's previous action is appended to the current input to capture temporal dependencies; and (3) a single Q-network is shared across all agents, rather than maintaining separate networks. This weight sharing improves learning efficiency and reduces computational overhead.

  Despite using a shared network, each agent receives individualized inputs—including its own observation, previous action, and agent identifier—allowing the shared Q-network to specialize behavior for each agent. The Q-function learned by the Recurrent Neural Network (RNN) is defined as:

  $$Q(o_t^m, h_{t-1}^m, m, a_{t-1}^m, a_t^m; \theta), \tag{1}$$

  where o_t^m and a_t^m are the observation and current action of agent m at time t, a_{t-1}^m is its previous action, h_{t-1}^m is the hidden state, and θ denotes the shared network parameters.

  Training proceeds in two stages: (1) agents interact with the environment using an ϵ-greedy policy based on the current Q-function, and (2) the shared network is updated using the Bellman equation, optimizing temporal-difference error over episodes.

- **Differentiable Inter-Agent Learning (DIAL) Foerster et al. (2016a):** DIAL introduces a framework where, at each time step, agent m outputs not only an environment action a_t^m but also a communication action c_t^m. During centralized training, communication is unrestricted and differentiable, enabling gradient flow across agents; during decentralized execution, messages are constrained to low-bandwidth discrete channels. Unlike traditional Q-networks that only estimate value functions, DIAL's Q-network additionally outputs real-valued messages c_t^m, which are sent at time $t+1$ to other agents' Q-networks (all agents share weights).

  Two types of gradients flow through the network: (1) the reward gradient from the agent's own Q-learning loss and (2) an error gradient from message recipients, which propagates back through the communication channel. This allows the sending agent to directly adjust its messages to minimize downstream DQN loss, accelerating the emergence of effective communication.

  Experimental results show that both DIAL and DDRQN eventually converge to similar performance—100% of Oracle rewards for $n = 3$ agents, and 90% for $n = 4$. However, DIAL converges substantially faster, requiring only 20,000 episodes compared to DDRQN's 500,000. For $n = 4$, DIAL maintains reliable convergence, while DDRQN may fail to converge altogether. The performance gains are attributed to DIAL's differentiable message passing, which enables more direct and efficient coordination learning during training.

Learning to communicate through parameter sharing Besides the differentiable inter-agent learning (DIAL) proposed in Foerster et al. (2016a), the authors of this paper also proposed another communication approach based on centralized learning decentralized execution using deep Q-learning with a recurrent network to address partial observability, i.e., reinforced inter-agent learning (RIAL). RIAL has two variants, using either independent or centralized Q-learning. In independent Q-learning, each agent learns its own network, treating others as part of the environment. Another variant trains a single shared network for all agents. During execution, agents act independently based on their unique observations. RIAL combines DRQN with independent Q-learning to learn both environmental and communication actions. Each agent's Q-network, Q^a, takes as input the hidden state h_t^a, observation o_t^a, and the messages m_{t-1}^a received from other agents at the previous time step. To manage the potentially large action space, RIAL factorizes Q^a into two separate output heads: Q_u^a for environmental actions $u_t^a \in U$ and Q_m^a for communication actions $m_t^a \in M$, where U and M denote the respective action spaces. An ϵ-greedy policy is used to select both types of actions, requiring the Q-network to produce $|U| + |M|$ outputs per time step. Both Q_u and Q_m are trained using DQN, with two key modifications: disabling experience replay to address non-stationarity, and feeding the actions u_t and m_t as inputs at the next time step to account for partial observability. Even though the agents learn independently, the decentralized execution is identical to the training phase.

Learning to Communicate through Public Belief The DIAL method Foerster et al. (2016a) assumes the presence of an idealized communication channel where agents can freely exchange information without affecting the environment. However, in many practical scenarios, explicit communication is unavailable or constrained. To address this, Foerster et al. Foerster et al. (2019) introduced the concept of *public belief*, which captures a shared probabilistic estimate over latent state features, based on common observations. At time step t, a latent feature f_t is inferred, with its publicly observable part denoted as f_t^{pub}. The public belief is updated as:

$$\mathcal{B}_t = P(f_t | f_1^{pub}, f_2^{pub}, \ldots, f_t^{pub}).$$

Building on this, they proposed the **Bayesian Action Decoder (BAD)**, a method that uses public belief to enable coordination without direct communication. Each agent selects actions based on its private observation and a shared virtual policy $\hat{\pi}_t$, computed from the public belief and known policies of all agents. This virtual policy maps private observations to likely actions, enabling agents to infer others' private states from their observable actions. Specifically, after observing an action a_t^m, an agent updates its belief about another's private state f_t^m using:

$$P(f_t^m | a_t^m, \mathcal{B}_t, f_t^{pub}, \hat{\pi}_t) \propto \mathbf{1}(\hat{\pi}_t(f_t^m) = a_t^m) P(f_t^m | \mathcal{B}_t, f_t^{pub}).$$

This approach was evaluated on the cooperative card game *Hanabi*, where players cannot see their own cards and must rely on hints and inference. BAD achieved state-of-the-art performance in two-player settings, with agents learning implicit conventions (e.g., using a hint like "red" to signal which card to play) that facilitated non-verbal coordination. The success of BAD demonstrates how shared beliefs can support effective communication even in the absence of explicit channels.

Scheduling Communication in a Distributed Manner A common approach to enhancing coordination in multi-agent systems is by enabling distributed communication, where agents exchange information to act cohesively without relying on a centralized controller. This is often achieved through the CTDE paradigm Lowe et al. (2017), which is also adopted by many of the previously discussed methods in this section. **SchedNet** Kim et al. (2019a) exemplifies this approach by using reinforcement learning to learn a scheduling policy that determines which agents should communicate at each time step. This dynamic scheduling is particularly important in practical scenarios where communication occurs over a shared medium—such as a wireless frequency channel—introducing two key constraints: limited bandwidth and contention for medium access. SchedNet addresses these constraints by jointly learning (1) a message encoder, (2) a policy for scheduling which agents can broadcast, and (3) an action selector that uses both local observations and received messages to make decisions.

3.4.2 Communication Message Inner Integration Timing Strategies

Message integration plays a critical role in determining not only *how*, but also *when* communicated information affects an agent's behavior during both learning and execution. By message integration, we refer to the process of incorporating received communication into an agent's internal computation, such as in its policy or value function, during a training round or decision cycle. In most MARL-Comm methods, communication is treated as additional input and incorporated into either the policy network, the value network, or both. The specific timing of this integration—whether messages influence immediate action selection, delayed value updates, or both—affects how quickly agents adapt to new information and coordinate their behavior. Poorly timed integration may delay responses to critical cues or cause misaligned coordination. To examine these effects, we categorize recent approaches into three strategies: policy-level integration, value-level integration, and joint policy-and-value integration Zhu et al. (2022).

Policy-Level Communication Messages Integration Incorporating communication messages into the policy model directly influences *when* agents adjust their decisions based on received information. By conditioning their next action on the latest messages from other agents, policies can dynamically adapt to evolving environments and coordination needs in real time. Typically, messages are concatenated with an agent's local observations or hidden states and fed into the policy network. This integration allows the agent to immediately update its behavior at every decision step based on the most recent communicated information, rather than relying solely on its own partial observations. In **policy gradient methods**, such as REINFORCE Sukhbaatar et al. (2016); Liu et al. (2020); Singh et al. (2018); Kong et al. (2017), the timing of communication affects the entire trajectory distribution: agents collect rewards over episodes while continuously adjusting their policies based on ongoing message exchanges. Communication influences the probability of action sequences during learning, making timely message integration critical for effective gradient estimation and coordination. **Actor-critic approaches** Du et al. (2021); Wang et al. (2020b); Jiang & Lu (2018); Kim et al. (2019a); Ding et al. (2020); Chu et al. (2020); Kim et al. (2020); Pesce & Montana (2020a); Freed et al. (2020a); Yun et al. (2021); Gupta et al. (2023) further emphasize timing by using critics that evaluate future returns conditioned on both local observations and communication messages. Here, agents can adapt their action choices at each step based on newly received messages, optimizing not only for immediate rewards but also for longer-term cooperation based on shared information. Thus, in policy-level integration, communication timing determines *when* new information influences the agent's decision-making pipeline, enabling agents to react to dynamic changes in their teammates' states and strategies throughout the learning and execution process.

Value-Level Communication Messages Integration In value-level integration, communication messages are incorporated into the value function (or action-value function), directly influencing how agents evaluate the long-term utility of their actions. By conditioning value estimates on both local observations and received messages, agents can better assess expected returns under partial observability and dynamic conditions. Many works following DQN-style frameworks adopt this strategy Foerster et al. (2016a); Jiang et al. (2018); Zhang et al. (2019); Agarwal et al. (2019b); Sheng et al. (2022); Kim et al. (2019b). Typically, messages are concatenated with local observations or embedded as additional features for the value network. This allows incoming communication to immediately affect an agent's Q-value estimates, which in turn guide future action selection via ϵ-greedy or similar policies. Unlike policy-level integration, where messages directly shape the action distribution, value-level integration influences behavior indirectly, through changes in the estimated returns of different actions. In both cases, the timing of message reception matters. However, in value-level integration, updates often occur after reward feedback is received, introducing a temporal delay in how communicated information affects future decisions. This lag can influence when agents benefit from new information, particularly in environments with fast-changing dynamics. Incorporating communication into the value function allows agents to reason about the long-term consequences of actions informed by peer input, thereby enhancing coordination in sequential decision-making tasks.

Combined Policy and Value Communication Messages Integration In combined integration approaches, communication messages are incorporated into both the policy and value models, enabling agents to adjust not only their immediate action selection but also their evaluation of future outcomes based on received information. This dual integration enhances the agent's ability to dynamically adapt its decisions *when* new messages are received, influencing both short-term reactions and long-term planning. This approach is often based on actor-critic methods, where messages are taken as additional inputs to both the actor and the critic. Some works feed the messages separately into the policy and value networks Agarwal et al. (2019b); Peng et al. (2017), allowing independent processing of communication information in action generation and value estimation. Others first combine messages with local observations to form a shared internal representation, which is then jointly used by both the actor and critic models Niu et al. (2021); Das et al. (2018); Liu et al. (2020); Freed et al. (2020a). By integrating communication into multiple stages of decision-making, agents can react promptly to new information at the policy level while simultaneously refining their value predictions. The timing of communication thus has a compounded effect: when messages arrive at critical decision points, they can immediately influence an agent's behavior, either by directly altering action probabilities (in policy-level integration) or by updating value estimates that drive action selection (in value-level integration, e.g., via Q-learning). At the same time, these updates can shift agents' expectations of long-term rewards, enhancing coordination over both immediate and future timescales.

Table 3 summarizes key approaches to determining when agents should communicate in multi-agent reinforcement learning systems. The first part of the table highlights learning-based strategies for adapting communication under bandwidth or resource constraints, including backpropagation-enabled frameworks like CommNet and DIAL, centralized training with shared parameters, public belief modeling for environments without explicit channels, and distributed scheduling using reinforcement learning. The second part focuses on the timing of message integration within an agent's learning model, either at the policy level (directly influencing action selection), at the value level (shaping return estimates) or both. Together, these approaches aim to balance timely and efficient communication with coordination demands, especially in dynamic, partially observable, or bandwidth-constrained settings.

3.5 Answering 'How to Communicate?' - Protocol Design and Coordination Mechanisms

After determining what and when to communicate, the next challenge in MARL-Comm is how information should be structured and exchanged. Effective communication protocols must balance coordination efficiency, interpretability, and scalability, particularly as the number of agents increases. Key design choices include whether communication should be centralized or decentralized and whether messages should be explicit (human-interpretable) or implicit (latent representations optimized for coordination).

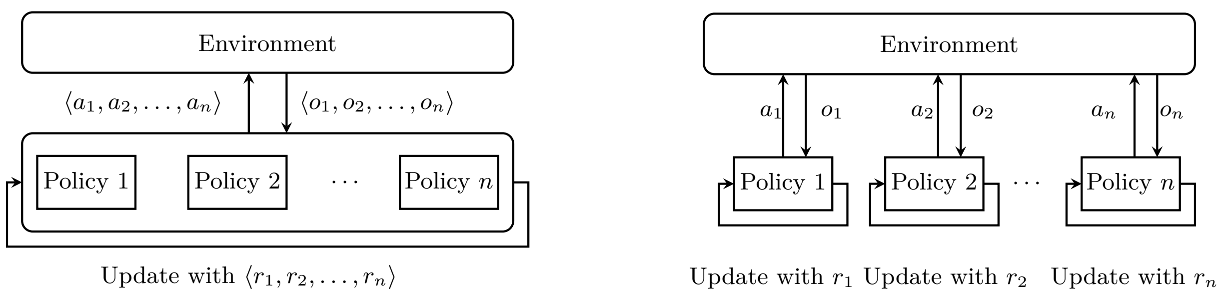
3.5.1 Training Frameworks in Communication-Enhanced MARL

The training framework in an MARL-Comm system defines how agents use the collected experiences, including observations, actions, rewards, and messages, to improve performance. Different frameworks specify how this information is processed and shared during learning, shaping the effectiveness of communication and coordination among agents. *Decentralized Learning* involves training each agent independently using only its respective experience. This approach allows agents to learn their individual policies based solely on local observations and rewards. However, decentralized learning faces significant challenges in multi-agent set-

Category	Approach	Representative Works
Learning Communication Timing (Sec. 3.4.1)	Backpropagation-based Communication	Sukhbaatar et al. (2016); Foerster et al. (2016a;b)
	Parameter Sharing (RIAL/DRQN)	Foerster et al. (2016a)
	Public Belief Modeling	Foerster et al. (2016a; 2019)
	Distributed Scheduling (MAC)	Lowe et al. (2017); Kim et al. (2019a)
Inner Integration Timing (Sec. 3.4.2)	Policy-Level Integration	• REINFORCE Sukhbaatar et al. (2016); Liu et al. (2020); Singh et al. (2018); Kong et al. (2017) • Actor-critic approaches Du et al. (2021); Wang et al. (2020b); Jiang & Lu (2018); Kim et al. (2019a); Ding et al. (2020); Chu et al. (2020); Kim et al. (2020); Pesce & Montana (2020a); Freed et al. (2020a); Yun et al. (2021); Gupta et al. (2023)
	Value-Level Integration	• DQN-like frameworks Foerster et al. (2016a); Jiang et al. (2018); Zhang et al. (2019); Agarwal et al. (2019b); Sheng et al. (2022); Kim et al. (2019b)
	Combined Policy-Value Integration	• Feeding the messages separately into the policy and value networks Agarwal et al. (2019b); Peng et al. (2017) • Combining messages with local observations to form a shared internal representation, which is then jointly used by both the actor and critic models Niu et al. (2021); Das et al. (2018); Liu et al. (2020); Freed et al. (2020a)

Table 3: Taxonomy of Approaches to ‘When to Communicate’ in MARL-Comm

tings, primarily due to the non-stationary environment introduced by constantly adapting agents. As agents evolve, the environment each agent perceives is continuously changing, complicating the learning process. On the other hand, *Centralized Learning* involves training all agents jointly by accessing the combined experience of every agent. This approach transforms the multi-agent system into a stationary environment where global information can be utilized to guide the training process. However, centralized learning suffers from the curse of dimensionality, as the joint policy space grows exponentially with the number of agents, making it computationally expensive and difficult to search. To address the drawbacks of purely decentralized or centralized learning, the *Centralized Training with Decentralized Execution (CTDE)* paradigm Foerster et al. (2016a); Kraemer & Banerjee (2016) has become the dominant training framework for MARL-Comm. In CTDE, agents learn their local policies with the benefit of centralized information during training, such as access to global observations or the actions of other agents. However, during execution, agents operate autonomously, using only their local information. This approach balances the benefits of centralized training with the flexibility of decentralized execution, allowing agents to learn more robust policies while avoiding the challenges of joint policy search during execution. In addition, *Parameter Sharing* schemes Foerster et al. (2016a) have emerged as a solution to further reduce the complexity of multi-agent systems. In these methods, agents share a common set of parameters (such as neural network weights), effectively training a single model that can control all agents. While this reduces the computational overhead of training multiple policies, it may limit the diversity of agent behaviors, as all agents are governed by the same model. Lastly, *Concurrent Learning* addresses situations where agents must simultaneously learn to interact with each other while managing their own policies. Methods like MD-MADDPG Pesce & Montana (2020a) and IS Kim et al. (2020) allow agents to train in tandem, coordinating their learning processes while maintaining individuality in decision-making.



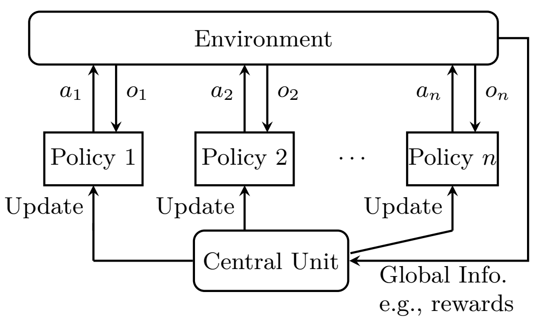
(a) Centralized Learning

(b) Decentralized Learning

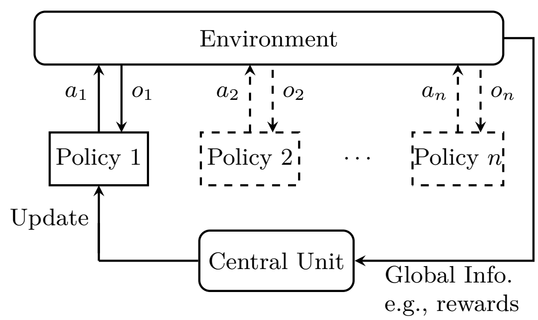
Figure 4: Centralized and Decentralized Learning. In centralized learning (a), policies are jointly optimized with shared knowledge and gradient flow. In decentralized learning (b), each agent operates and learns independently using local observations and rewards.

Centralized Learning In the centralized learning paradigm, the experiences of all agents—including their observations, actions, rewards, and messages—are aggregated by a central unit, which learns to optimize the collective behavior of the system. As illustrated in Fig. 4a, this centralized controller leverages a global view of the environment to coordinate agents and compute joint policies or value functions. By having access to system-wide information, the learning process can exploit dependencies and correlations that may be inaccessible in decentralized settings. In practice, however, centralized learning is typically confined to the training phase. Most recent works assume that, during execution, agents operate independently without access to a central controller. This decoupling—commonly known as centralized training with decentralized execution (CTDE)—reflects the fact that maintaining a centralized controller at runtime is often infeasible in real-world deployments, particularly in dynamic, distributed, or bandwidth-constrained environments. As such, centralized learning serves as a practical means to facilitate coordination during training while preserving the scalability and autonomy of decentralized execution.

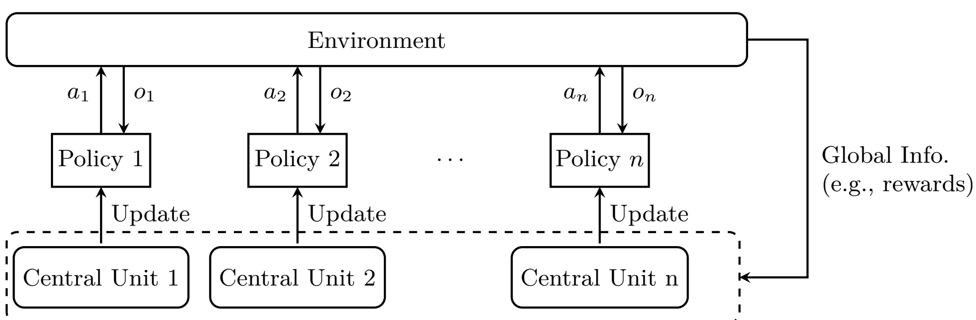
Decentralized Learning As illustrated in Fig. 4b, decentralized learning frameworks enable each agent to collect experience and optimize its policy independently, using only locally available information. Agents learn from their own observation-action-reward-message trajectories, without reliance on a centralized controller or shared global knowledge. This paradigm is particularly suitable for scenarios where communication is constrained, delayed, or unreliable, and where agents must operate under partial observability or localized perspectives. Several recent approaches adopt this decentralized formulation, including Agent-Entity Graph Agarwal et al. (2019b), NeurComm Chu et al. (2020), Intention Propagation (IP) Qu et al. (2020), MAGNet-SA-GS-MG Malysheva et al. (2018), DCC-MD Kim et al. (2019b), and Diff Discrete Freed et al. (2020a). Agent-Entity Graph Agarwal et al. (2019b) models structured interactions between agents and environmental entities by constructing a dynamic graph where nodes represent agents and entities, enabling local message exchange and action selection based on proximity. NeurComm Chu et al. (2020) learns a neural message-passing protocol that encodes policy information in a decentralized fashion, allowing agents to share latent representations with neighbors. Intention Propagation (IP) Qu et al. (2020) introduces intention-sharing, where agents broadcast planned actions to nearby peers, enabling anticipatory coordination. MAGNet-SA-GS-MG Malysheva et al. (2018) employs a modular graph-based attention mechanism to guide communication between agents under varying spatial constraints. DCC-MD Kim et al. (2019b) scales communication in large agent populations by dropping low-utility messages to reduce overhead while preserving coordination quality. Diff Discrete Freed et al. (2020a) promotes sparse and discrete communication through Gumbel-softmax approximations, reducing message bandwidth and learning interpretable protocols. These decentralized methods are designed to address the inherent non-stationarity of multi-agent environments by enabling agents to adapt their policies based on localized, evolving information. By removing the dependency on centralized coordination, such methods demonstrate robust performance in scenarios with communication constraints or dynamically changing interaction topologies.



(a) Individual Parameters



(b) Parameter Sharing



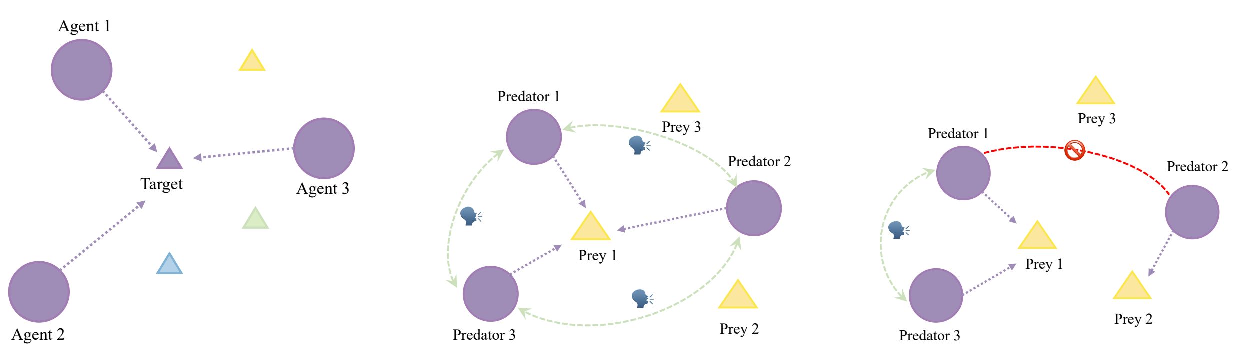
(c) Concurrent

Figure 5: Three types of CTDE Scheme.

Centralized Training with Decentralized Execution (CTDE) In the CTDE framework (Fig. 5), the experiences from all agents are made available for optimizing their individual policies. This approach allows the gradients, computed from the combined experiences of all agents, to guide the learning of each agent's local policy, enabling decentralized execution during actual interactions in the environment. The key benefit of CTDE is that it leverages global information during the training phase to optimize local agent behavior without needing centralized control during execution. One critical component of CTDE is *parameter sharing* Foerster et al. (2016a), which improves data efficiency by using a single set of parameters (e.g., a shared

Q-function or policy) across all agents, rather than maintaining separate learning processes. Despite sharing parameters, agents can still exhibit diverse behaviors, as they are most likely to receive distinct observations at each time step due to their unique positions or roles in the environment. Based on these observations, we can further categorize recent works into several subgroups: **(1). Individual Policy Parameters.** In this approach, each local policy has its own set of parameters, while a central unit aggregates the experiences from all agents to provide global information and guidance, such as gradients, as illustrated in Fig. 5a. This method typically employs policy gradient algorithms (e.g., with REINFORCE) Kong et al. (2017) or actor-critic-based methods to train the entire system Wang et al. (2020b); Kim et al. (2019a); Mao et al. (2020); Yun et al. (2021). The central unit processes experiences from the environment and computes global feedback, which is then used to update the local policies of individual agents. **(2). Parameter Sharing.** In the parameter-sharing approach, all local policies or value functions share a common set of parameters, as shown in Fig. 5b. This approach is particularly useful for improving data efficiency and training speed. In cases where DQN-like algorithms are used, a local Q-function can be learned to process all experiences Foerster et al. (2016a); Jiang et al. (2018); Sheng et al. (2022), or agents can employ an additional global Q-function to guide learning Zhang et al. (2019; 2020). If actor-critic methods are utilized, a shared actor (i.e., policy network) is trained using all observation-action pairs, while receiving gradient guidance from a centralized critic Niu et al. (2021); Du et al. (2021); Das et al. (2018); Jiang & Lu (2018); Hu et al. (2020a); Mao et al. (2020). In some cases, policy gradient methods with REINFORCE are employed instead of actor-critic algorithms, where the optimization process relies on rewards sampled over entire episodes Sukhbaatar et al. (2016); Liu et al. (2020); Singh et al. (2018). **(3). Concurrent Learning.** In scenarios where storing all experiences in a centralized buffer is impractical, each agent maintains its own experience buffer while assuming access to the actions and observations of other agents. These agent-specific backups are then used to train a centralized critic for each agent, as illustrated in Fig. 5c. This setup allows agents to learn concurrently and benefit from shared information without the need for a global experience buffer. This is distinct from purely decentralized learning, as each agent retains its own set of parameters but receives global information to guide its local policy, as shown in Fig. 5c. Concurrent CTDE often utilizes actor-critic-based methods, where each agent has access to its own centralized critic to help guide the optimization of its local policy Kim et al. (2020); Pesce & Montana (2020a). This concurrent approach ensures that each agent has the necessary guidance to improve performance while maintaining local autonomy. CTDE's balance between centralized training and decentralized execution has become a standard training framework for multi-agent systems, offering flexibility in environments where both individual decision-making and cooperative behavior are required.

3.5.2 Communication Shaped by Reward Objectives



(a) Cooperative (b) Competitive (c) Mixed

Figure 6: Communication shaped by reward objectives: cooperative (a), competitive (b), and mixed (c).

The design of reward functions plays a central role in shaping how agents communicate and coordinate in multi-agent systems. By tailoring these reward signals, agents can be incentivized to pursue specific interaction patterns and group dynamics. Depending on how these rewards are structured, agent interactions tend to fall into one of three primary categories: *cooperative*, *competitive*, or a combination of both, referred to as *mixed* or *cooperative-competitive* scenarios Ning & Xie (2024); Busoniu et al. (2008); Matignon et al.

(2012). Each paradigm requires distinct communication strategies to support the underlying objective. For instance, fully cooperative tasks encourage agents to share informative messages that maximize joint utility, while competitive tasks may discourage communication or foster deception. A number of recent approaches validate their adaptability by testing across multiple behavioral regimes Liu et al. (2020); Das et al. (2018); Singh et al. (2018); Jiang et al. (2018); Sheng et al. (2022). In **cooperative settings** (Fig. 6a), agents are encouraged to work together to maximize a shared objective. A common approach is to assign the same reward to all agents, i.e., $r_1 = r_2 = \cdots = r_N$, so that collective performance is prioritized over individual gain. This equal reward distribution incentivizes agents to support one another and avoid failures that could harm the group's overall outcome, thereby fostering stronger coordination. **Competitive settings** (Fig. 6b), by contrast, emphasize individual success. Each agent seeks to maximize its own reward, potentially at the expense of others. In fully competitive tasks—often modeled as zero-sum games—the total reward is fixed, such that $\sum_{i=1}^{N} r_i = 0$. Here, one agent's gain directly results from another's loss, and the objective becomes outperforming opponents rather than collaborating. **Mixed settings** (Fig. 6c), also known as general-sum games, incorporate both cooperative and competitive dynamics. In these environments, agents have partially aligned or independent goals, and rewards are neither fully shared nor strictly opposed. Rather than involving multiple tasks, agents in mixed settings may find themselves cooperating in some parts of a task while competing in others. For example, in autonomous driving, vehicles may cooperate to avoid collisions and maintain smooth traffic flow (shared safety objective), but compete for limited road resources such as merging lanes or parking spots. Similarly, in multi-player video games or robot soccer, teams may exhibit internal cooperation while competing against opposing teams. These scenarios reflect many realistic multi-agent applications and require more nuanced communication strategies to manage the trade-offs between collaboration and competition.

Cooperative Scenarios In cooperative scenarios, agents are incentivized to communicate for the purpose of maximizing collective performance. A common approach is to assign a shared team reward, where all agents receive the same signal regardless of individual contributions Foerster et al. (2016a); Sukhbaatar et al. (2016); Peng et al. (2017); Das et al. (2018); Wang et al. (2020b); Jiang & Lu (2018); Kim et al. (2019a); Mao et al. (2020); Ding et al. (2020); Agarwal et al. (2019b); Kim et al. (2020); Zhang et al. (2019; 2020); Yun et al. (2021); Kilinc & Montana (2018); Sheng et al. (2022); Freed et al. (2020a); Hu et al. (2020a); Gupta et al. (2023). This reward structure naturally promotes coordination and cooperation, as agents must work together toward a common goal. An alternative formulation uses individualized rewards that are still dependent on the performance of other agents Liu et al. (2020); Singh et al. (2018); Peng et al. (2017); Jiang et al. (2018); Freed et al. (2020a); Kim et al. (2019b); Kong et al. (2017); Gupta et al. (2023). In these cases, agents are encouraged to act cooperatively because their own success is tightly coupled with the success of their teammates. Additionally, many approaches incorporate penalties, such as for collisions or redundant actions, to discourage behaviors that hinder team efficiency Niu et al. (2021); Das et al. (2018); Liu et al. (2020); Jiang & Lu (2018); Hu et al. (2020a); Kim et al. (2020); Jiang et al. (2018); Kim et al. (2019b). Some models also implement neighborhood-based reward sharing to strengthen localized cooperation among nearby agents Chu et al. (2020); Qu et al. (2020). Communication is fundamental in such environments, allowing agents to exchange observations, predicted actions, or high-level intentions. This exchange leads to more synchronized behavior and robust team dynamics. Early approaches like DIAL Foerster et al. (2016a) and CommNet Sukhbaatar et al. (2016) leverage differentiable communication protocols to support end-to-end learning and shared representations in cooperative settings. These methods are generally applicable to a wide range of cooperative objectives, as they assume that communication enhances coordination toward a common goal. ATOC Jiang & Lu (2018) builds on this by introducing an attention-based mechanism that enables agents to dynamically decide when communication is most beneficial, reducing unnecessary message exchange. To further reinforce cooperative behavior, some methods explicitly shape the reward structure to align individual incentives with team success. For example, IC3Net Singh et al. (2018) and GA-Comm Liu et al. (2020) design rewards such that agents benefit more from collaborative success than from individual achievements, encouraging implicit cooperation. Similarly, frameworks such as TarMAC Das et al. (2018) and MAGIC Niu et al. (2021) incorporate penalty terms to discourage behaviors that hinder group performance, such as collisions, integrating communication and reward shaping as dual mechanisms for promoting coordinated action.

Competitive Scenarios In competitive MARL, agents operate in adversarial environments where their objectives conflict. Each agent is incentivized to maximize its own reward while potentially obstructing the progress of others. These settings introduce distinct challenges, requiring agents not only to learn effective policies but also to adaptively respond to dynamic opponent behaviors. Strategic communication, if appropriately designed, can be a powerful tool—even in antagonistic contexts—yet remains significantly underexplored. While the StarCraft Multi-Agent Challenge (SMAC) Zhang et al. (2019; 2020); Samvelyan et al. (2019) offers a popular benchmark for evaluating multi-agent coordination under adversarial conditions, most existing works focus on intra-team cooperation. Communication across competitive teams, where revealing intentions could be exploited by adversaries, is rarely modeled explicitly, limiting our understanding of communication dynamics under zero-sum or mixed objectives. Recent work has begun to address this gap. IC3Net Singh et al. (2018) introduces a learnable gating mechanism that enables agents to selectively communicate only when it improves their own utility. This mechanism is particularly advantageous in competitive settings where indiscriminate communication may incur strategic disadvantages. For example, agents in IC3Net learn to share information just before seizing a goal or evading pursuit, demonstrating how self-interested agents may still develop conditional communication protocols. Beyond IC3Net, the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) framework Lowe et al. (2017) offers a generalizable approach applicable to both cooperative and competitive tasks. MADDPG adopts a CTDE paradigm, where each agent learns a policy using a centralized critic that has access to the global state and the actions of all agents during training. This enables more stable learning in environments where agents' behaviors are interdependent, including competitive settings where opponents' actions critically affect individual outcomes. In such adversarial contexts, having a centralized critic allows agents to model and anticipate the strategies of others more effectively, which is crucial for learning robust policies. MADDPG's flexible architecture lays the groundwork for extensions that incorporate message passing, especially in mixed or dynamic environments where cooperation and competition coexist. Differentiable Inter-Agent Learning (DIAL) Foerster et al. (2016a) explores how communication protocols can emerge in both cooperative and competitive scenarios. By allowing end-to-end learning of communication messages through differentiable channels, DIAL enables agents to optimize information sharing based on learned policies. In adversarial contexts, such mechanisms can support private or encoded communication strategies that preserve team advantages without leaking information to opponents. Complementing these methods, recent work by Deka and Sycara Deka & Sycara (2021) demonstrates the emergence of diverse roles within competitive teams. By leveraging graph neural networks and decentralized training, their agents learn to specialize—developing heterogeneous behaviors that enhance team-level robustness. Such diversity is especially beneficial in adversarial settings, where role specialization enables agents to counter varied opponent tactics more effectively. Overall, while communication in competitive MARL is still in its infancy, these foundational efforts highlight its potential. Future research must delve deeper into adversarial communication modeling, including deception, privacy, and selective signaling, to unlock robust strategies for complex multi-agent interactions under competition.

Mixed Scenarios Mixed scenarios in multi-agent reinforcement learning involve agents that pursue individual reward functions, which may be partially aligned, entirely independent, or even conflicting. These environments inherently blend cooperative and competitive dynamics, requiring agents to dynamically adapt their strategies and communication behaviors based on the situational context. Mixed settings can be viewed as a generalization of purely competitive scenarios: while fully competitive environments assume zero-sum rewards (i.e., one agent's gain equals another's loss), mixed scenarios relax this constraint, allowing for both cooperative interactions and adversarial competition to emerge simultaneously. The core challenge lies in determining when to communicate, what information to share, and how to balance mutual benefit against self-interest to achieve long-term success. One of the foundational works in this space is MADDPG Lowe et al. (2017), which introduces a CTDE paradigm. This framework enables each agent to condition its policy on other agents' actions during training via a centralized critic, while maintaining decentralized execution at test time. This flexibility makes MADDPG well-suited not only for mixed cooperative-competitive tasks but also for purely competitive scenarios. Similarly, RIAL and DIAL Foerster et al. (2016a) propose architectures that support both private and shared communication channels, allowing agents to learn implicit negotiation strategies and dynamic roles. Though originally designed for cooperative settings, these models have also demonstrated adaptability in competitive and mixed-reward environments by enabling agents to develop context-aware communication protocols. Liu et al. Liu et al. investigate how cooperative behaviors can emerge in competitive multi-agent soccer through decentralized population-based training and reward shaping, revealing a progression from individual to team strategies without explicit coordination signals. DGN Jiang et al. (2018) investigates a setting where agents must choose between collecting food for positive rewards or attacking others for higher gain. It demonstrates that graph-based communication allows agents to shift from competitive to cooperative behavior. IC3Net Singh et al. (2018) introduces a gating mechanism for communication, enabling agents to learn when to communicate based on individual reward structures. Similarly, TarMAC Das et al. (2018) develops a targeted communication architecture evaluated on mixed scenarios like Predator-Prey. I2C Ding et al. (2020) builds on this by introducing incentive-based messaging, where agents weigh communication cost against collective utility. MAGIC Niu et al. (2021) further investigates mixed environments with a modular communication strategy and dynamic attention mechanisms. Ryu et al. Ryu et al. (2021) incorporate cooperative and competitive biases into agent training to improve adaptability in hybrid environments. Santos et al. Santos et al. (2022) introduce a hybrid paradigm where agents adapt to dynamic communication availability at execution time, improving robustness under fluctuating conditions. These studies demonstrate that communication learning in mixed settings is sensitive to reward design, environmental dynamics, and temporal dependencies. By leveraging structured protocols, attention mechanisms, or contrastive feedback, agents can learn when to collaborate, when to act independently, and how to balance the two under uncertainty.

3.5.3 Protocol Design regarding Communication Format

Cooperative MARL has proven effective for collaborative decision-making in diverse domains Canese et al. (2021); Oroojlooy & Hajinezhad (2023). Compared to single-agent settings, multi-agent environments introduce additional challenges such as partial observability and non-stationarity. To address these issues, many approaches adopt the CTDE paradigm, where agents are trained with access to global information but operate based on local observations during execution Rashid et al. (2018); Wang et al. (2020a); Yu et al. (2022a). Nonetheless, relying on limited local data during decentralized execution can complicate effective and cooperative decision-making. For example, in a scenario where opposing forces clash, one allied agent might view the situation as advantageous by seeing only one adversary, whereas others, seeing multiple enemies, might judge it as dire. This variance in perceptions can hinder cohesive decision-making efforts.

To mitigate such issues, **explicit communication** has been proposed, allowing agents to share information and enhance cooperative actions Zhu et al. (2022). Despite its benefits, explicit communication necessitates continuous connectivity between agents and a central server, potentially causing network congestion. Moreover, it somewhat undermines the decentralized model and may prove impractical in settings where communication options are restricted. While direct information sharing could be beneficial, many of these methods presuppose extra communication actions governed by specific protocols. These explicit communication mechanisms necessitate a thorough consideration of communication structures, costs, and channel constraints, thereby adding complexity to the problem.

Rather than engaging in direct information exchange, implicit communication allows agents to convey information by observing the actions or effects of other agents Oliehoek et al. (2016). The literature outlines two main approaches to this form of communication. Works by Knepper et al. Knepper et al. (2017), Shaw et al. Shaw et al. (2022), and Tsiamis et al. Tsiamis et al. (2015) focus on determining actions that indirectly communicate information and its implications, effectively creating a protocol for non-direct message exchange. For instance, Knepper et al.Knepper et al. (2017) develop criteria for robots to interpret and react to observed human actions, facilitating human-robot interactions without explicit messaging. **Implicit communication** offers a strategy where information is transmitted through behaviors instead of explicit messages Breazeal et al. (2005). For instance, an allied agent might discern a challenging situation by noting the retreat of fellow agents rather than an advance. This observation enables the agent to adapt its strategy based on the implicit cues gathered from the actions of others. Such a method relies on the interpretation of observed behaviors to convey necessary information. Although implicit communication-based MARL techniques Tian et al. (2020); Grupen et al. (2021) have been promising, their effectiveness has predominantly been demonstrated in simplistic scenarios, limiting their immediate applicability to more complex, real-world environments.

On the other hand, several works Tian et al. (2020); Grupen et al. (2021); Li et al. (2023a) explore communication strategies that do not depend on predefined protocols. Instead, they use separate models to infer other agents' states or intentions based solely on local or partially shared observations. These methods facilitate coordination by embedding communication implicitly within learned representations. For example, Tian et al. Tian et al. (2020) introduce a belief module and auxiliary rewards to support implicit communication in cooperative tasks. Grupen et al. Grupen et al. (2021) propose a fully decentralized curriculum learning approach, where agents rely on spatial positioning as a form of implicit signal. Li et al. Li et al. (2023a) present TACO, which supports explicit communication, with each agent learning to approximate aggregated global information based on local observations. The portion of global information is gradually reduced, eventually facilitating implicit communication.

Table 4: Summary of Approaches to 'How to Communicate?' in MARL-Comm

Category	Sub-Category	Representative Approaches
Training Framework (Sec. 3.5.1)	Centralized Learning	Foerster et al. (2016a); Kraemer & Banerjee (2016)
	Decentralized Learning	Agent-Entity Graph Agarwal et al. (2019b), NeurComm Chu et al. (2020), Intention Propagation (IP) Qu et al. (2020), MAGNet-SA-GS-MG Malysheva et al. (2018), DCC-MD Kim et al. (2019b), Diff Discrete Freed et al. (2020a).
	Centralized Training with Decentralized Execution (CTDE)	• Individual Policy Parameters: REINFORCE Kong et al. (2017), Actor-critic-based Wang et al. (2020b); Kim et al. (2019a); Mao et al. (2020); Yun et al. (2021) • Parameter Sharing: local Q-function Foerster et al. (2016a); Jiang et al. (2018); Sheng et al. (2022), global Q-function Zhang et al. (2019; 2020), shared actor receiving gradient guidance from a centralized critic Niu et al. (2021); Du et al. (2021); Das et al. (2018); Jiang & Lu (2018); Hu et al. (2020a); Mao et al. (2020), rewards sampled over entire episodes via REINFORCE Sukhbaatar et al. (2016); Liu et al. (2020); Singh et al. (2018) • Concurrent Learning: centralized critic to help guide the optimization of its local policy Kim et al. (2020); Pesce & Montana (2020a)

Continued on next page

Table 4 – continued from previous page

Category	Approach	Representative Works
Communication Shaped by Reward Objectives (Sec. 3.5.2)	Cooperative Scenarios.	• Shared team reward, all agents receive the same signal regardless of individual contributions Foerster et al. (2016a); Sukhbaatar et al. (2016); Peng et al. (2017); Das et al. (2018); Wang et al. (2020b); Jiang & Lu (2018); Kim et al. (2019a); Mao et al. (2020); Ding et al. (2020); Agarwal et al. (2019b); Kim et al. (2020); Zhang et al. (2019; 2020); Yun et al. (2021); Kilinc & Montana (2018); Sheng et al. (2022); Freed et al. (2020a); Hu et al. (2020a); Gupta et al. (2023) • Individualized rewards that are still dependent on the performance of other agents Liu et al. (2020); Singh et al. (2018); Peng et al. (2017); Jiang et al. (2018); Freed et al. (2020a); Kim et al. (2019b); Kong et al. (2017); Gupta et al. (2023) • Incorporating penalties Niu et al. (2021); Das et al. (2018); Liu et al. (2020); Jiang & Lu (2018); Hu et al. (2020a); Kim et al. (2020); Jiang et al. (2018); Kim et al. (2019b) • Neighborhood-based reward sharing to strengthen localized cooperation Chu et al. (2020); Qu et al. (2020); Sukhbaatar et al. (2016); Jiang & Lu (2018).
	Competitive Scenarios	• Learnable gating mechanism enabling agents to selectively communicate: IC3Net Singh et al. (2018) • Centralized critics during training: MADDPG Lowe et al. (2017) • End-to-end learning of communication messages through differentiable channels: DIAL Foerster et al. (2016a) • Leveraging graph neural networks and decentralized training Deka & Sycara (2021)

Continued on next page

Table 4 – continued from previous page

Category	Approach	Representative Works
	Mixed Scenarios	• Centralized training with decentralized execution paradigm: MADDPG Lowe et al. (2017) • Learn both shared and private communication channels, supporting implicit negotiation and role emergence in partially aligned settings: RIAL and DIAL Foerster et al. (2016a) • Decentralized population-based training and reward shaping Liu et al.; Jiang et al. (2018) • Gating mechanism for communication: IC3Net Singh et al. (2018) • Targeted communication architecture evaluated on mixed scenarios: TarMAC Das et al. (2018); Ding et al. (2020) • Incorporate cooperative and competitive biases into agent training Ryu et al. (2021) • Dynamic communication availability at execution time Santos et al. (2022)
Message Format Objectives (Sec. 3.5.3)	Explicit communication	Singh et al. (2018); Das et al. (2018); Lowe et al. (2017); Foerster et al. (2016a); Chen et al. (2024b); Breazeal et al. (2005)
	Implicit communication	Oliehoek et al. (2016); Knepper et al. (2017); Shaw et al. (2022); Tsiamis et al. (2015); Knepper et al. (2017); Tian et al. (2020); Grupen et al. (2021)

Table 4 provides a structured summary of the key approaches addressing the question of how to communicate in MARL-Comm. The taxonomy is organized into four major categories: training frameworks, reward-shaped communication strategies, communication message formats, and protocol designs. Each category is further subdivided into specific approaches, such as centralized learning, decentralized learning, and the increasingly popular centralized training with decentralized execution. Representative works are provided for each sub-category, highlighting influential algorithms and methods from the literature. The table captures both explicit and implicit communication protocols, as well as variations tailored for cooperative, competitive, and mixed-motive scenarios. This summary offers a comprehensive reference for understanding how communication mechanisms are operationalized in MARL systems, balancing coordination, scalability, and interpretability.

3.6 Discussions

Table 5 provides a comparative overview of representative MARL-Comm methods across three key dimensions: what to communicate, whom to communicate with, and how to encode or compress information. This summary reveals the diversity of design choices in message content, communication topology, and integration strategies. Notably, methods differ in whether they rely on past or future knowledge, whether they adopt peer-to-peer or representative-based communication, and whether they use simple concatenation, equal weighting, or advanced attention-based mechanisms. By organizing these approaches within a unified taxonomy, we highlight important trends and open challenges, providing a foundation for future research on efficient, scalable, and interpretable multi-agent communication.

Through our comparative analysis of MARL-Comm methods along the dimensions of what to communicate, whom to communicate with, and how to communicate, we provide a structured perspective from which researchers can design or evaluate their own Comm-MARL systems. Despite the significant progress made in recent years, several important challenges remain open for future exploration.

Table 5: Comparison of MARL-Comm Methods: What, Whom, and How to Communicate.

Method	What to Comm.	Whom to Comm.	How to Comm.
DIAL Foerster et al. (2016a)	Past Knowledge (w/o rep)	Other Learning Agents	Concatenation
RIAL Foerster et al. (2016a)	Past Knowledge (w/o rep)	Other Learning Agents	Concatenation
CommNet Sukhbaatar et al. (2016)	Past Knowledge (rep)	Other Learning Agents	Equal Weighting
IC3Net Singh et al. (2018)	Past Knowledge (rep)	Other Learning Agents	Equal Weighting
TarMAC Das et al. (2018)	Past Knowledge (rep)	Other Learning Agents	Weighted
MADDPG-M Kilinc & Montana (2018)	Past Knowledge (w/o rep)	Other Learning Agents	Concatenation
ETCNet Hu et al. (2020a)	Past Knowledge (w/o rep)	Other Learning Agents	Concatenation
HAMMER Gupta et al. (2023)	Past Knowledge (rep)	Rep. Agent	Weighted
GA-Comm Liu et al. (2020)	Past Knowledge (rep)	Rep. Agent	Weighted
MAGIC Niu et al. (2021)	Past Knowledge (rep)	Rep. Agent	Weighted
ATOC Jiang & Lu (2018)	Future Knowledge	Rep. Agent	RNN-based
FlowComm Du et al. (2021)	Past Knowledge (w/o rep)	Nearby Agents	Equal Weighting
GAXNet Yun et al. (2021)	Past Knowledge (w/o rep)	Nearby Agents	RNN/Attention
DGN Jiang et al. (2018)	Past Knowledge (w/o rep)	Nearby Agents	CNN
I2C Ding et al. (2020)	Future Knowledge	Other Learning Agents	RNN
MD-MADDPG Pesce & Montana (2020a)	Past Knowledge (rep)	Rep. Agent	RNN
Gated-ACML Mao et al. (2020)	Past Knowledge (rep)	Rep. Agent	Gated NN
MAGNet-SA-GS-MG Malysheva et al. (2018)	Past Knowledge (rep)	Nearby Agents	Attention-based
NeurComm Chu et al. (2020)	Future Knowledge	Nearby Agents	RNN
Agent-Entity Graph Agarwal et al. (2019b)	Past Knowledge (w/o rep)	Nearby Agents	Attention-based

First, we observe that most existing works adopt a Sender-Receiver or Sender-Proxy-Receiver communication paradigm, where agents proactively transmit information about their local observations or learning states to others. While this design simplifies learning—particularly by enabling gradient backpropagation through communication channels, it also restricts the flexibility of agent interactions. More expressive communication patterns, such as allowing agents to actively query, request, or synchronize specific information from others, remain underexplored. Expanding beyond passive broadcasting to more dynamic communication protocols could enable richer coordination behaviors. In addition to transmitting local observations, investigating alternative forms of communicated knowledge, such as action-value functions, as explored in Chen et al. (2024b), also presents a promising direction for communication training. In the meantime, establishing theoretical guarantees for communication protocols is crucial for providing stronger foundations for reliability, stability, and interpretability in learned communication systems. **Second**, communication constraints such as limited bandwidth, message delays, and collision risks play a crucial role in real-world applications. Although several methods address scheduling and bandwidth-awareness, incorporating realistic communication constraints systematically into learning frameworks remains an open challenge. Future work should further integrate these constraints, especially for domains like multi-robot systems, wireless networks, and autonomous driving, where communication robustness is critical. **Third**, evaluating the true effectiveness of a communication protocol is nontrivial. Performance improvements may arise either from better decision-making facilitated by communication or simply from stronger action policies, making it difficult to isolate the causal benefits of communication itself. Developing standardized evaluation metrics or controlled ablation studies that directly assess communication utility remains an important direction for advancing the field. **Fourth**, in terms of learning communication policies, we find that current approaches largely fall into two categories: Reinforced (reward-driven) or Differentiable (gradient-driven) learning. Reinforced approaches often require careful reward shaping and human-designed signals to succeed, while Differentiable methods, though elegant, may struggle with credit assignment across agents sharing a global reward. Designing more sophisticated communication learning paradigms that balance flexibility, interpretability, and credit assignment remains a pressing need. **Finally**, parameter sharing is a common technique used to stabilize and accelerate multi-agent training. However, this often assumes homogeneous agents with identical observation and action spaces. Building Comm-MARL systems that support heterogeneous agents—potentially with differing capabilities, goals, or knowledge representations—poses a major challenge that is still largely unexplored. **Overall**, while MARL-Comm has matured significantly, addressing these open issues is essential for creating scalable, efficient, and human-aligned communication protocols that can operate in realistic, complex multi-agent environments.

4 Emergent Language in Multi-agent Systems

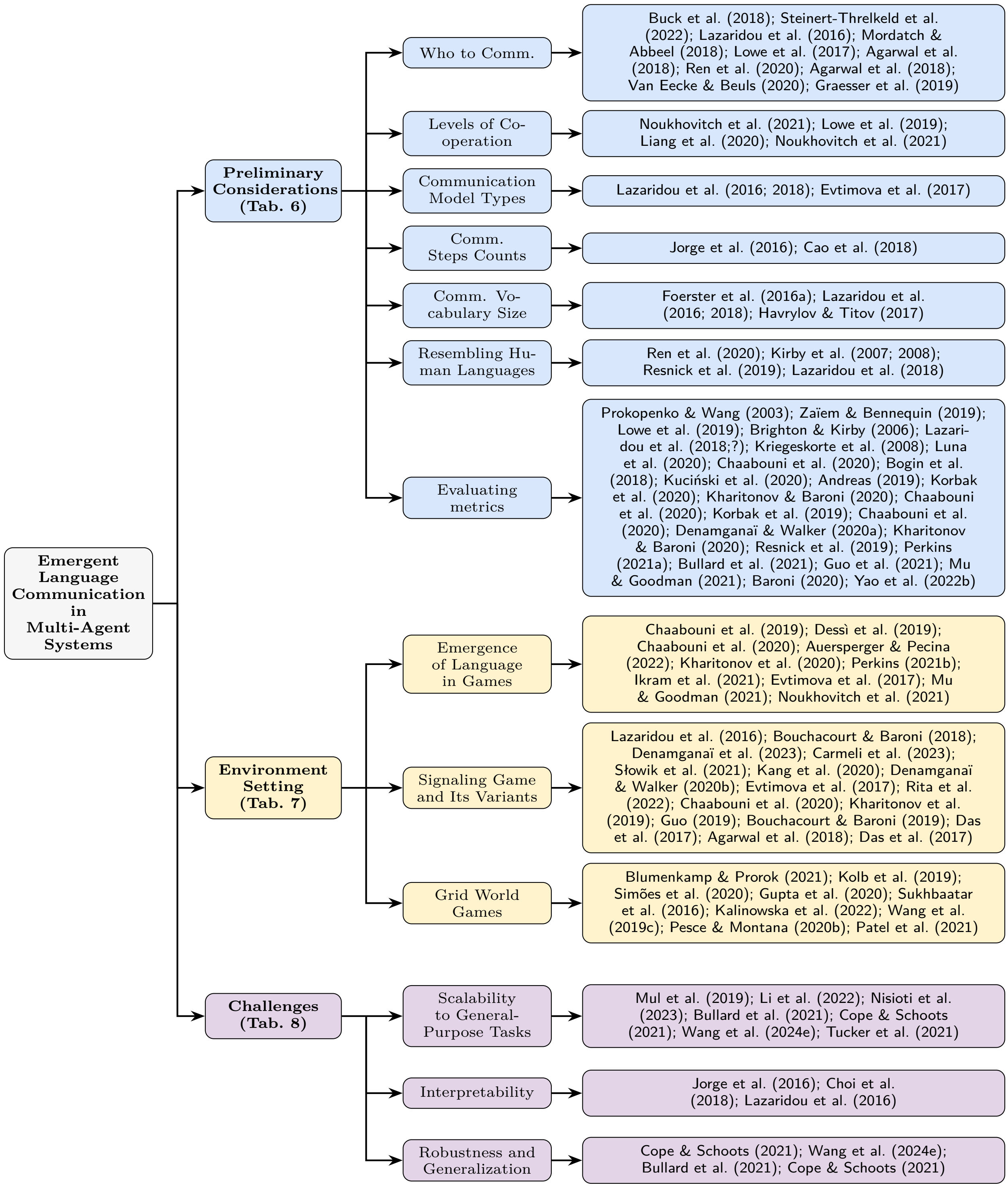


Figure 7: EL-Comm Agents Taxonomy

Deep learning methods in natural language processing and multi-agent reinforcement learning offer a robust framework for simulating the emergence of human-like communication systems, a field known as Emergent

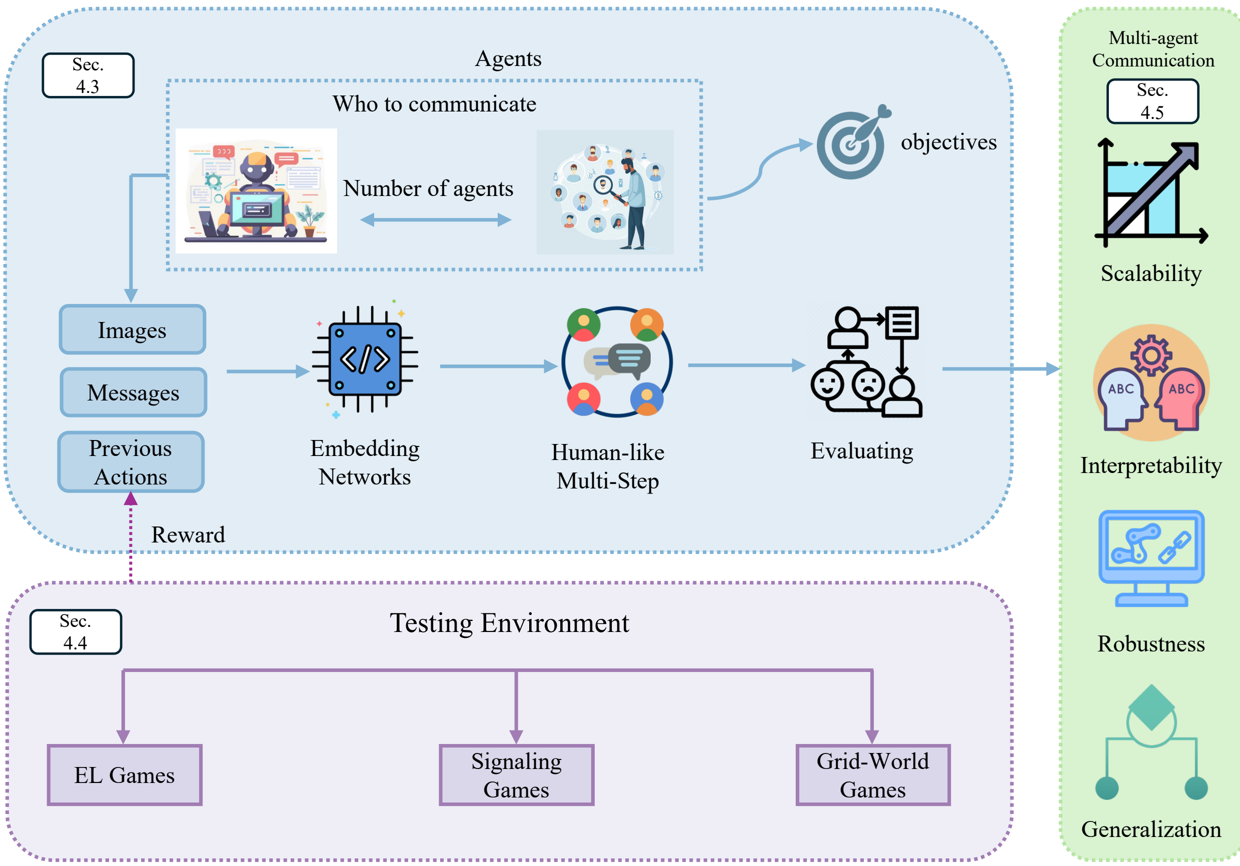


Figure 8: Illustrating key components of emergent language (EL) communication research of Sec. 4, including agent interaction settings (Sec. 4.3), testing environments such as EL games and grid-worlds (Sec. 4.4), and evaluation metrics aligned with desirable properties—scalability, interpretability, robustness, and generalization (Sec. 4.5). It captures the full pipeline from message generation to downstream assessment in human-like, multi-step communication tasks.

Communication or **Emergent Language (EL)** Boldt & Mortensen (2024). Multi-agent reinforcement learning systems, such as AlphaZero Silver et al. (2017) and OpenAI's hide-and-seek agents Baker et al., have demonstrated how complex behaviors can arise from basic environmental dynamics through self-play. This approach was first applied to discrete communication systems in seminal works Foerster et al. (2016a); Lazaridou et al. (2016); Havrylov & Titov (2017); Mordatch & Abbeel (2018). While modeling human language behavior is a compelling goal, it is equally important to define its practical applications across science, technology, and engineering.

The study of **Emergent Language Communications (EL-Comm)** in multi-agent systems explores how agents develop communication protocols through interaction, without explicit supervision. This research direction has gained traction for its potential to shed light on the origins of language and to enable scalable, interpretable coordination among artificial agents. In this section, we begin with background information (Sec. 4.1), summarizing foundational concepts, and identifying challenges faced in real-world MARL systems. An overview of the related research is categorized and summarized in Sec. 4.2. To orient readers through the core elements of this section, we provide two guiding figures. Fig. 7 offers a taxonomy of the EL-Comm literature, organized into three major themes: preliminary considerations (Sec. 4.3), environment settings (Sec. 4.4), and open challenges (Sec. 4.5). It highlights key factors such as agent interaction structure, communication model design, and metrics for evaluating emergent protocols. Complementing this, Fig. 8 illustrates the end-to-end pipeline of emergent communication—from message generation and embedding to evaluation across different environments. It covers agent interaction dynamics (Sec. 4.3), testing environments including EL, signaling, and grid-world games (Sec. 4.4), and desirable properties such as scalability, interpretability, robustness, and generalization (Sec. 4.5). Together, these visualizations guide our survey exploration of emergent language as both a behavioral phenomenon and a functional component in multi-agent reinforcement learning.

4.1 Background on Emergent Language Research

Emergent language has become a rapidly advancing area of research within artificial intelligence, particularly in the realm of multi-agent reinforcement learning. Historically, the study of language emergence focused on understanding the origins of human language, with limited attention to its applicability for artificial agents. However, recent advancements in reinforcement learning aim to enable agents to develop communication protocols with capabilities comparable to or even exceeding human language. These efforts extend beyond the statistical approaches commonly used in natural language processing, prompting critical questions about the conditions for language emergence and how to effectively evaluate its success. To provide context for the taxonomy and analysis, this subsection outlines the fundamental concepts of communication and linguistics and offers background knowledge of emergent language research Peters et al. (2024).

Communication fundamentally involves the exchange of signals that convey information, whether intentional, such as speech, or unintentional, like bodily reactions. According to Watzlawik's 'Interactional View' Watzlawick et al. (2011), all behavior communicates, making communication universal and essential. It occurs through various channels and can be broadly categorized into intrapersonal, interpersonal, group, public, and mass communication, depending on its purpose and audience Peters et al. (2024). In emergent language research, two forms of communication are commonly studied: interpersonal and group communication. Interpersonal communication involves interactions between entities that influence each other, with each entity perceiving its own environment, often overlapping through a shared communication channel. Noise may affect the process by distorting perception or communication. Group communication extends this to multiple entities, focusing on collective goals or tasks, making it more formal than interpersonal communication, which often has a social character. Most population-based EL research focuses on group communication, while intrapersonal, public, and mass communication remain largely unexplored.

Natural language is one of humanity's greatest achievements, serving as the foundation for all forms of communication. It enables us to convey highly complex information through a discrete and humanly manageable amount of utterances. Much AI research focuses on developing natural language models for tasks like translation and text generation Wolf et al. (2020); Brown et al. (2020); Lauriola et al. (2022); Khurana et al. (2023). However, many in the AI community argue that current statistical models trained on static datasets lack true language understanding for effective human cooperation Mordatch & Abbeel (2018); Choi et al. (2018); Lazaridou et al. (2021); Merrill et al. (2021). Emerging EL research in AI aims to enable agents to communicate like humans, enhancing cooperation, performance, and generalization, and fostering meaningful interaction between humans and AI Mordatch & Abbeel (2018). While explicit forms of EC are widely studied, some survey papers excluded implicit communication Peters et al. (2024), such as spatial positioning in multi-agent systems Grupen et al. (2022).

Emergent Language (EL) is a form of communication that develops naturally among artificial agents through interaction, without explicit programming. It arises from the agents' need to cooperate and solve tasks in their environment Brandizzi (2023). EL involves agents creating, adapting, and refining linguistic structures to exchange information effectively Lipowska & Lipowski (2022). Research focuses on understanding how elements like syntax Ueda et al. (2022), semantics Colas et al. (2020); Qiu et al. (2022), and pragmatics Lowe et al. (2019) emerge and enhance agent performance and cooperation. A natural language-like communication system would make artificial agents more accessible, comprehensible, and powerful Iocchi et al. (2022); Noukhovitch et al. (2021); Lazaridou & Baroni (2020). Initially focused on the origins of language Steels (1997), EL research now emphasizes enabling agents to harness communication mechanisms akin to natural language Lazaridou & Baroni (2020). Modern EL in computer science involves self-learned, reusable, teachable, and interpretable communication protocols Nowak & Krakauer (1999); Lazaridou et al. (2016); Li & Bowling (2019). The ultimate goal is seamless and extensible communication between machines and humans Lazaridou & Baroni (2020); Yao et al..

In this section, we review papers that proposed agent-based computer models in which communication emerges as a general paradigm for coordinating behavior in multi-agent systems Boldt & Mortensen (2024). These models feature agents with communication channels that generate discrete message symbols, aiming to mimic human-like language. The structure and content of the messages are not predefined but instead emerge naturally during training, shaped by the agents' interactions and the characteristics of their environments. Such approaches have been explored within a variety of agentic frameworks, often leveraging deep learning methods and neural networks optimized through gradient descent.

4.2 An Overview of Emergent Language Research

Communication arises from conventions and rules developed through the need or benefit of coordination. Lewis Lewis (2008) formalized such scenarios as 'coordination problems' and introduced a simple signaling game, where a speaker describes an object and a listener identifies it among options. This game significantly influenced emergent language (EL) research in computer science. Early studies focused on specific aspects of emergent communication (EC) through hand-crafted simulations Wagner et al. (2003); Steels (1997); Nowak & Krakauer (1999); Kirby (2002); Cangelosi & Parisi (2012); Christiansen & Kirby (2003); Batali (1998); Oliphant & Batali (1997); Steels (1995); Skyrms (2002); Smith et al. (2003), often relying on supervised learning and non-situated settings, which limited their exploration of complex linguistic features Wagner et al. (2003). Recently, EL research has gained momentum Foerster et al. (2016a); Lazaridou et al. (2016); Havrylov & Titov (2017); Bouchacourt & Baroni (2018); Cao et al. (2018); Mordatch & Abbeel (2018); Das et al. (2017) leveraging multi-agent reinforcement learning (MARL) approaches Agarwal et al. (2019a); Blumenkamp & Prorok (2021); Brandizzi et al. (2022); Iocchi et al. (2022); Chaabouni et al. (2022); Gupta et al. (2021); Lo & Sengupta; Lowe et al. (2019); Vanneste et al. (2022b); Yu et al. (2022b) to study more intricate features of language emergence.

A key goal of EL research in MARL is to enable agents to autonomously develop communication systems that support both agent-to-agent and agent-to-human interaction in a natural language (NL)-like manner Wagner et al. (2003); Bouchacourt & Baroni (2018); Iocchi et al. (2022); Lo & Sengupta; Bogin et al. (2018); Noukhovitch et al. (2021). Reinforcement learning (RL) methods are particularly promising for this purpose. They not only create agents capable of flexible, practical communication Lazaridou & Baroni (2020) but also offer insights into the evolution of NL itself Galke et al.. Unlike NLP models, which rely on static datasets to mimic NL Steinert-Threlkeld et al. (2022); Bender & Koller (2020), EL focuses on agents designing and learning their own communication systems through active, goal-driven experiences Lemon (2022). This experiential approach contrasts with the shallow understanding of language seen in large language models (LLMs) Manning & Schutze (1999); Qiu et al. (2020); Wolf et al. (2020). In EL, agents develop communication while solving tasks, mirroring natural processes Lewis (2008), leading to a deeper grasp of their environment Bogin et al. (2018). Advances in EL pave the way for innovative multi-agent systems and more human-centric AI applications Lazaridou & Baroni (2020).

Various research areas and questions have emerged in EL. Recent studies explore diverse settings, including semi-cooperative environments Noukhovitch et al. (2021); Liang et al. (2020), adversaries Blumenkamp & Prorok (2021); Yu et al. (2022b), noise in messages Cope & Schoots (2020), and social structures Dubova et al.; Fitzgerald & Tagliabue (2022). Grounding EL has been approached through representation learning Lin et al. (2021), combining supervised learning with self-play Lowe et al., or using EL agents for fine-tuning NL models Yao et al.. Other work examines the emergence of NL-like characteristics, such as internal and external pressures Luna et al. (2020); Kalinowska et al. (2022); Dagan et al. (2021); Iocchi et al. (2022); Rita et al. (2022), compositionality Resnick et al. (2020), generalization Chaabouni et al. (2022), expressivity Guo et al., and the connection between compositionality and generalization Chaabouni et al. (2020). Unlike NLP models like LLMs, which imitate language statistically without addressing the functional purpose of communication Mordatch & Abbeel (2018), EL treats language as a tool for achieving meaningful outcomes Brandizzi et al. (2022). Agents must learn EL through necessity or benefits Luna et al. (2020), in settings that encourage communication, such as cooperative scenarios Noukhovitch et al. (2021). However, EL faces significant challenges. Encouraging communication can lead to task-specific gibberish rather than NL-like language Mu & Goodman (2021), making proper incentives critical. Additionally, measuring successful communication and assessing language properties like syntax, semantics, and pragmatics are essential for guiding and evaluating EL development Lowe et al. (2019); Chaabouni et al. (2020). The following sections delve into these challenges and related approaches.

Existing survey publications on EL research address similar topics but differ in scope. Peters et al. Peters et al. (2024) categorized these surveys into three main directions, and we expand on this framework by including more recent works and providing broader descriptions. Surveys focusing on learning settings explore the design of environments and learning processes, as seen in works Van Eecke & Beuls (2020); Lipowska & Lipowski (2022); Denamganaï & Walker (2020b). Those emphasizing methods review learning and evaluation techniques, including studies Lowe et al. (2019); Lemon (2022); Korbak et al. (2020); LaCroix (2019); Mihai & Hare (2021); Galke & Raviv (2024); Vanneste et al. (2022a). Lastly, surveys offering general overviews provide broad insights into the EL field, as demonstrated in Iocchi et al. (2022); Lazaridou & Baroni (2020); Galke et al.; Hernandez-Leal et al. (2019); Fernando et al. (2020); Brandizzi (2023); Boldt & Mortensen (2024); Peters et al. (2024).

There are also several surveys papers that reviews the EL research. Some surveys focus on learning problems, environments, and language learning design. van Eecke and Beuls Van Eecke & Beuls (2020) reviewed the language game paradigm, categorizing experiments and highlighting properties like symmetric agent roles and autonomous behavior. Our survey extends beyond the language game paradigm, providing more details about RL-based communication protocols. Similarly, Lipowska and Lipowski Lipowska & Lipowski (2022) explored EL in MARL, emphasizing naming games, protolanguage explainability, and sociocultural factors like migration and teachability. While these are included in our analysis, we integrate them into a broader context. Denamganaï and Walker Denamganaï & Walker (2020b) reviewed referential games, introducing the ReferentialGym framework and metrics like positive signaling and listening Lowe et al. (2019). Our survey includes referential games but covers a wider range of metrics and approaches.

Some surveys focused on learning and evaluation methods. Korbak et al. Korbak et al. (2020) discussed compositionality metrics, introducing the tree reconstruction error to address gaps in analyzing semantic aspects. LaCroix LaCroix (2019) critiqued the focus on compositionality in EL, suggesting reflexivity as a better direction, though metrics for this remain undeveloped. Lemon Lemon (2022) reviewed language grounding, emphasizing the need for better data collection, but did not propose metrics. Lowe et al. Lowe et al. (2019) highlighted language utility over semantics, proposing metrics like positive signaling and listening, and Mihai and Hare Mihai & Hare (2021) emphasized exploring semantic factors rather than low-level hashes but did not introduce new metrics. Galke and Raviv Galke & Raviv (2024)identified pressures and biases aligning neural EL with human NL, focusing on NL phenomena rather than training biases. Vanneste et al. Vanneste et al. (2022a) reviewed discretization methods for MARL-based EL, recommending DRU, straight-through DRU, and Gumbel-Softmax for general use, though these are not central to our survey.

There are also more existing surveys providing general overviews or focus on unique aspects of EL research that do not fit into the settings or methods categories. Hernandez-Leal et al. Hernandez-Leal et al. (2019) offered a broad review of multi-agent reinforcement learning (MARL), categorizing work into emergent behavior, communication, cooperation, and agent modeling. While their survey is valuable for understanding

MARL's historical context, it lacks detailed analysis of EL-specific metrics, which we address in this work. Brandizzi and Iocchi Iocchi et al. (2022) emphasized human-in-the-loop interactions in EC research, highlighting the underrepresentation of human-AI communication. They explored interaction types but did not provide a comprehensive review of EL papers, as we do. Moulin-Frier and Oudeyer Moulin-Frier & Oudeyer (2020) connected MARL to linguistic theories, identifying challenges like decentralized learning and intrinsic motivation but omitted detailed metric discussions. Galke et al. Galke et al. reviewed RL-based EC approaches, focusing on the gap between emergent and human language, particularly in compositionality. Fernando et al. Fernando et al. (2020) proposed a novel drawing-based communication system, which, while interesting, falls outside our survey's scope. Suglia et al. Suglia et al. (2024) reviewed visually grounded language games, categorizing tasks and models but primarily focusing on multimodal grounding rather than EL itself. Zhu et al. Zhu et al. (2024) structured EC works along nine dimensions but focused more on learning tasks than EL emergence. Lazaridou and Baroni Lazaridou & Baroni (2020) provided a summary of the EL field but did not emphasize metrics or taxonomy, which are central to our work. Similarly, Brandizzi Brandizzi (2023) categorized EC research and identified open challenges but lacked a structured taxonomy or a systematic literature approach, which we include with more publications. Boldt et al. Boldt & Mortensen (2024) offer a review of EL research, however, it mainly focused specifically on the goals and applications of EL research.

4.3 Preliminary Considerations of Emergent Communications

Who to Communicate Unlike the MARL-Comm, to answer 'who to communicate' within the literature in EL, it mainly focused on the number of agents in the communication circle. There are three main communication settings Peters et al. (2024) in terms of the agents' numbers: single agent, two agents, and agents' numbers settings. The single agent setting is rare and typically used to train human-machine interfaces Buck et al. (2018) or model fine-tuning Steinert-Threlkeld et al. (2022). Dual-agent settings, more common, involve a speaker and a listener, each with distinct roles Lazaridou et al. (2016). This model is further studied in Mordatch & Abbeel (2018), and has been adopted in Lowe et al. (2017) for its grounded communication environment. But at that stage, the emergent utterances come from a fix set and the communication symbols is in the K-dimensional one-hot encoding sample. Population settings, involving larger groups of agents, require greater computational resources but allow for regularization Agarwal et al. (2018), language evolution Ren et al. (2020), and more opportunities to influence the emergence process Agarwal et al. (2018); Van Eecke & Beuls (2020); Graesser et al. (2019).

Levels of Cooperation Even though communication mostly raise attention of study in fully cooperative scenarios, there are also other settings of the environment. We identified three primary types: cooperative, semi-cooperative, and competitive, which are shaped by the level of cooperation inherent in the environment. Most studies focus on fully cooperative settings, where agents share rewards entirely and lack individual incentives. This emphasis is reasonable since agents rely on a shared language to coordinate, and communication does not develop if dominance can be achieved without it Noukhovitch et al. (2021). Fewer studies explore semi-cooperative setups, which include both shared and individual rewards, introducing the challenge of balancing collective and personal objectives Lowe et al. (2019); Liang et al. (2020). These setups resemble simplified social scenarios, blending societal goals with individual interests. Fully competitive settings, where agents compete for rewards without shared objectives, are rarely studied. Only one work explores such settings, as they often lead to deceptive language, making meaningful communication unlikely without cooperative elements Noukhovitch et al. (2021).

DRQN-based Communication Each agent (speaker and listener) is equipped with an input network that processes incoming data such as images, messages, and previous actions, embedding them using either Recurrent Neural Networks or basic shallow Feedforward networks. Lazaridou Lazaridou et al. (2016) found that utilizing Convolutional Neural Networks to generate features from image inputs resulted in higher-quality features, thereby improving outcomes during the guessing phase. Lazaridou et al. Lazaridou et al. (2018) implemented a recurrent policy using a Recurrent Neural Network (RNN) for the decoder Hausknecht & Stone (2015), enabling it to maintain information from previous states and enhance its performance in generating longer sequences. Evtimova et al. Evtimova et al. (2017) incorporated an attention mechanism into the message-generation process. They reported that this addition enhanced communication success with previously unseen objects by allowing the embedding layers to focus more effectively on identifiable objects. To clarify, given images i_L and i_R, a convolutional network produces a hidden state h_{LR} which is then input to a single-layer LSTM decoder (representing the speaker's recurrent policy). This layer generates a message $m = g(h_{LR}, \theta_g)$. The listener utilizes a single-layer LSTM to encode the message (as it also handles sequence input), producing an encoding $z = h(m, \theta_h)$. Together with the encoding for the candidate images, it predicts a target t by comparing similarity measures between the candidates $u \in U$ and the encoded vector v. All weights for both the speaker and listener agents are optimized jointly, using the agents' choices as the sole learning signal. As their tasks differ, no weights are shared between the trained agents.

Human-Like Interaction via Multi-Step Communication We have outlined a well-known model in the field of language emergence above. Given the expanding body of literature on emergent languages, several intriguing variants are worth discussing. One particularly interesting variant is presented in the studies Jorge et al. (2016) and Cao et al. (2018), which aim to mimic human-like interactions by implementing multi-step communication between agents. In Jorge et al. (2016), the task remains similar—identifying target images (celebrities in this instance)—but involves both agents in message generation. The guessing agent initiates communication based on the provided candidates by posing questions. The answering agent, who is aware of the target, responds to these questions with 'yes' or 'no', akin to the 'Guess Who' game. The most effective results were achieved with two rounds of questioning and answering. Their analysis indicates that the questioning agent focuses on different features in the first and second questions, thereby acquiring more information. Cao et al. Cao et al. (2018) explored multi-step communication as well, but introduced significant modifications. Their focus shifted from referential games to a negotiation context where agents bargain over items that hold different utilities for each party. They implemented two communication channels: the first transmits information similar to the earlier discussed channels, while the second facilitates concrete trade offers that can either be accepted or rejected by the other agent. When the agents cooperate—sharing a common reward—the messages clarify to the other agent which items are most valuable and specify their value. The findings indicate that the 'linguistic' channel aids in reaching a Nash equilibrium and decreases the variance in joint optimality, thereby enhancing the robustness of the trading system.

Vocabulary Size Related Communication The vocabulary size significantly impacts the effectiveness of communication between agents. Foerster et al. Foerster et al. (2016a) initially experimented with simple 1-bit messages in scenarios that required minimal information, such as solving the switch riddle or determining the parity of a number. However, as the complexity of problems increased, a larger vocabulary became essential for transmitting meaningful messages. Lazaridou et al. Lazaridou et al. (2016) explored the use of single-symbol messages with vocabulary sizes ranging from 0 to 100. Subsequent research inspired by natural language proposed the adoption of LSTM-based encoder-decoder architectures to create message sequences Lazaridou et al. (2018); Havrylov & Titov (2017). It's crucial to note that these symbols initially have no inherent meaning; their association with visual or physical attributes develops only through the learning process. One of the primary advantages of sequence generation is that it enables speakers to create languages that exhibit compositionality. For instance, in a compositional language, a speaker can describe an object, such as a 'red box', using previously learned prefixes or suffixes for 'red' and 'box' from the training set. This capability is maintained even if the "red box" combination was not encountered during training.

Resembling Human Languages Emergent communication research aims to develop systems that resemble human language, from low-level traits like compositional semantics and syntax to high-level traits like pragmatics and sociolinguistics. This resemblance does not require exact replication but reflects the variation and commonalities in human language (e.g., nouns, verbs, and shared units of meaning). The goal is to create conditions—such as environments, agent architectures, and tasks—that enable the emergence of human-like communication. No existing research has holistically pursued the rederivation of human language. Most studies focus on isolated aspects of making emergent communication more human-like, which risks missing the complex, interdependent nature of emergent phenomena. Complex systems rely on interactions among many components, and removing single elements can lead to unpredictable changes. For example, Ren et al. Ren et al. (2020) demonstrate that iterated learning—imperfect transmission of language across generations—drives compositionality, as supported by earlier studies (Kirby et al. (2007; 2008)). However, many studies on compositionality use fixed-population settings, ignoring iterated learning's potential as a sufficient driver. This oversight questions findings attributing compositionality to other factors, such as model capacity (Resnick et al. (2019)) or perception (Lazaridou et al. (2018)). Comprehensive approaches are needed to capture the full complexity of emergent communication.

Evaluating Metrics of Emergent Communication Choice accuracy is a common metric for assessing communication quality, but metrics like the purity index don't measure interpretability well. Forcing agents to describe broad categories (e.g., 'dog') instead of detailed characteristics (e.g., color) may reduce communication effectiveness. A better approach could be evaluating communication without human bias. Prokopenko and Wang Prokopenko & Wang (2003) analyzed public belief entropy, which decreases as communication conventions form. Similarly, Zaiem et al. Zaïem & Bennequin (2019) proposed using entropy to quantify how much a message reduces uncertainty. This metric could be valuable for assessing communication protocols independently of task performance, as also suggested by Lowe et al. Lowe et al. (2019). Metrics quantify properties of emergent communication systems. While concrete properties like vocabulary size are measurable, abstract ones like compositionality require precise definitions. Evaluation metrics, such as F-score, assess system effectiveness and guide research progress by enabling hypothesis testing, result analysis, and comparison with prior work. Most metrics in emergent communication literature focus on compositionality and generalizability, reflecting the primary goals of many studies: to develop communication systems with compositional semantics that generalize beyond training scenarios.

- **Compositionality.** Compositionality, or compositional semantics, refers to the principle that complex utterances derive their meaning from the combination of their components (e.g., 'red car' means a car that is red), contrasting with holistic communication where meanings are unrelated to components. The most common metric for compositionality is topographic similarity Brighton & Kirby (2006); Lazaridou et al. (2018), which measures the correlation between distances in the referent feature space and message space. Lazaridou et al. Lazaridou et al. (2018) specifically use Spearman's rank correlation coefficient (ρ), with cosine similarity for feature space distances and Levenshtein distance for message space distances. Toposim represents a family of metrics, as different methods can be used to calculate these correlations and distances. Representation similarity analysis Kriegeskorte et al. (2008); Luna et al. (2020) quantifies compositionality by measuring correlations between the feature space and agents' internal representations. Other metrics focus on disentanglement, where message components independently specify single attributes, including positional and bag-of-words disentanglement Chaabouni et al. (2020), context independence Bogin et al. (2018), and conflict count Kuciński et al. (2020). Tree reconstruction error Andreas (2019) evaluates how well a compositional semantics model approximates the language produced by agents. In response to extensive research on compositionality metrics, Korbak et al. Korbak et al. (2020) conducted a meta-analysis, revealing that while many metrics capture basic compositionality, they often fail to detect more complex forms. Additionally, Kharitonov and Baroni Kharitonov & Baroni (2020) and Chaabouni et al. Chaabouni et al. (2020) challenge the claim that standard compositionality metrics also measure a language's ability to generalize to novel objects.
- **Generalizability.** Generalizability in machine learning typically refers to the ability of a model to perform well beyond its training conditions. In the context of MARL communication, this is often measured by how well agents can describe objects with previously unseen attribute combinations, effectively generalizing from training to test data Korbak et al. (2019); Chaabouni et al. (2020); Denamganaï & Walker (2020a); Kharitonov & Baroni (2020); Resnick et al. (2019); Perkins (2021a). Beyond this form of generalization, research has also explored generalization to new communication partners Bullard et al. (2021), adaptation to different environments Guo et al. (2021); Mu & Goodman (2021), and generalization across linguistic structures, such as disentangling syntax from semantics Baroni (2020). To assess the quality of emergent communication, Yao et al. Yao et al. (2022b) propose a data-driven evaluation metric that quantifies the effectiveness of an emergent language by comparing its machine translation performance to human language. The underlying intuition is that the closer an emergent language is to human language, the more effectively it can be substituted in machine learning tasks.

Table 6: Summary of Preliminary Considerations in Emergent Langauge Communications (Sec. 4.3)

Category	Sub-Category	Representative Approaches
Who to Communicate	Single agent	• Human-machine interfaces Buck et al. (2018); • Model fine-tuning Steinert-Threlkeld et al. (2022).
	Dual-agent settings /speaker and listener	• Each with distinct roles, emergent utterances come from a fix set and the communication symbols is in the K-dimensional one-hot encoding sample Lazaridou et al. (2016); Mordatch & Abbeel (2018); Lowe et al. (2017); • Larger groups of agents: regularization Agarwal et al. (2018), language evolution Ren et al. (2020), influencing the emergence process Agarwal et al. (2018); Van Eecke & Beuls (2020); Graesser et al. (2019).
Levels of Cooperation	Fully cooperative settings	Agents share rewards entirely and lack individual incentives and rely on a shared language to coordinate, and communication does not develop if dominance can be achieved without it Noukhovitch et al. (2021).
	Semi-cooperative setups	Both shared and individual rewards, introducing the challenge of balancing collective and personal objectives Lowe et al. (2019); Liang et al. (2020).
	Fully competitive settings	Agents compete for rewards without shared objectives, often leads to deceptive language, making meaningful communication unlikely without cooperative elements Noukhovitch et al. (2021).
DRQN-based Communication	CNN-based Visual Feature Extraction	Lazaridou et al. (2016) used CNNs to process image inputs, yielding more informative features that improved referential task performance.
	Recurrent Policies for Temporal Encoding	Lazaridou et al. (2018) introduced RNN-based decoders to maintain temporal memory and enhance the generation of longer communication sequences.
	Attention-Enhanced Communication	Evtimova et al. (2017) added attention mechanisms to guide symbol generation, improving generalization by focusing on salient features in visual inputs.
	Jointly Optimized Sender–Receiver Architectures	These works train speaker and listener networks end-to-end using only final task success as the learning signal, without weight sharing between agents.

Continued on next page

Table 6 – continued from previous page

Category	Approach	Representative Works
Human-Like Interaction via Multi-Step Communication	Identifying target task Jorge et al. (2016)	• Task: Identifying target images (e.g., celebrities). • Agent Roles: Asymmetric — guesser initiates questions, answerer replies. • Communication: Binary responses ('yes'/'no') across two interaction rounds. • Architecture: MLPs for perception, 2-layer GRUs for state/message updates, DRU for discretization. • Key Findings: Later questions refine earlier guesses; multi-step dialog improves target accuracy and dialogue consistency.
	Agents bargain over items that hold different utilities for each party Cao et al. (2018)	• Task: Negotiation over items with differing utilities. • Agent Roles: Symmetric — both agents negotiate cooperatively. • Communication: Two channels — symbolic (linguistic) and action-based (trade proposals). • Architecture: RL agents with value networks and discrete messaging. • Key Findings: Linguistic messages clarify preferences; leads to Nash equilibrium and reduces variance in joint rewards.
Vocabulary Size Related Communication	Simple 1-bit messages	Foerster et al. Foerster et al. (2016a) initially experimented with simple 1-bit messages in scenarios that required minimal information
	Single-symbol messages	Lazaridou et al. Lazaridou et al. (2016) explored the use of single-symbol messages with vocabulary sizes ranging from 0 to 100.
	Message sequences	Research Lazaridou et al. (2018); Havrylov & Titov (2017) are inspired by natural language, which proposed the adoption of LSTM-based encoder-decoder architectures to create message sequences.
Resembling Human Languages	Imperfect language transmission across generations drives compositionality	Ren et al. Ren et al. (2020) demonstrate that iterated learning—imperfect transmission of language across generations—drives compositionality, as supported by earlier studies Kirby et al. (2007; 2008).
	Fixed-population settings	Many studies fix population size, overlooking iterated learning as a sufficient driver of compositionality, thereby questioning claims that attribute it to model capacity Resnick et al. (2019) or perception Lazaridou et al. (2018).
Evaluating metrics of emergent communication	Public belief entropy	Prokopenko and Wang Prokopenko & Wang (2003) analyzed public belief entropy, which decreases as communication conventions form

Continued on next page

Table 6 – continued from previous page

Category	Approach	Representative Works
	Using entropy to quantify how much a message reduces uncertainty	Zaiem et al.Zaïem & Bennequin (2019) used entropy to measure how much a message reduces uncertainty, a metric Lowe et al.Lowe et al. (2019) also suggested for evaluating communication protocols beyond task performance
	Compositionality	• *Topographic similarity*: measures the correlation between distances in the referent feature space and message space Brighton & Kirby (2006); Lazaridou et al. (2018); • *Spearman's rank correlation coefficient (ρ)*: with cosine similarity for feature space distances and Levenshtein distance for message space distances: Lazaridou et al. (2018); • *Representation similarity analysis*: Kriegeskorte et al. (2008); Luna et al. (2020); • *Disentanglement-based Metrics:* Evaluate whether individual message components represent distinct attributes, including: – Positional and bag-of-words disentanglement Chaabouni et al. (2020) – Context independence Bogin et al. (2018) – Conflict count Kuciński et al. (2020) • *Tree Reconstruction Error:* Measures how well a compositional semantics model reconstructs the structure of the language produced by agents Andreas (2019). • *Meta-Analysis of Metrics:* Korbak et al. Korbak et al. (2020) found that existing metrics capture basic compositionality but struggle to detect more complex forms. • *Limitations of Compositionality Metrics:* Kharitonov and Baroni Kharitonov & Baroni (2020) and Chaabouni et al. Chaabouni et al. (2020) questioned whether current metrics reliably assess generalization to novel objects.

Continued on next page

Table 6 – continued from previous page

Category	Approach	Representative Works
	Generalizability	• *Unseen Attributes:* Agents generalize to novel object attribute combinations Korbak et al. (2019); Chaabouni et al. (2020); Denamganaï & Walker (2020a); Kharitonov & Baroni (2020); Resnick et al. (2019); Perkins (2021a). • *New Partners and Environments:* Agents adapt to unfamiliar partners Bullard et al. (2021) and settings Guo et al. (2021); Mu & Goodman (2021). • *Syntax–Semantics Separation:* Generalization across linguistic structure Baroni (2020). • *MT-based Human Alignment:* Translation performance used to assess similarity to human language Yao et al. (2022b).

4.4 Developing Broader Use of Experimental Environments for Testing Emergent Communications

The primary goal of experimental environments in emergent communication research is to simplify the implementation and execution of experiments. Since this field falls within computer science, these environments typically consist of programs, source code, and occasionally datasets. While other researchers can adapt any codebase used in an emergent communication experiment, the focus here is on environments specifically designed for flexibility and reuse across a wide range of experiments.

Standardized experimental environments save researchers time by reducing the need to reimplement basic features for emergent communication experiments. They also lower the risk of bugs, enable improvements to existing tools, and ensure more reliable comparisons across studies by minimizing variation in implementation. However, care is needed to avoid systematic biases, as emergent communication is sensitive to environmental and implementation effects. Additionally, well-designed, user-friendly environments benefit researchers with limited software development skills, making it easier to turn ideas into code and contribute to the field.

Communication settings in EL research are defined through various communication games. This part reviews the games commonly used in the literature, focusing on language games—a subset that prioritizes explicit communication through a predefined language channel. These games are grouped into categories as: referential games, reconstruction games, question-answer games, grid-world games, continuous environment games. Our analysis shows that these categories are among the most frequently used in EL studies.

Emergence of Language in Games Emergence of Language in Games (EGG) Kharitonov et al. (2019) is the most widely used framework for deep learning-based emergent communication. It provides a straightforward Python interface for common emergent communication games, agent architectures, and metrics. Studies that implement new games with EGG, such as Chaabouni et al. (2019); Dessì et al. (2019); Chaabouni et al. (2020); Auersperger & Pecina (2022); Kharitonov et al. (2020), expand its accessible range of games and metrics. Some tools target specific aspects of emergent communication experiments. For example, TexRel Perkins (2021b) provides a synthetic dataset where compositional language can describe constructed images. HexaJungle Ikram et al. (2021) introduces a suite of environments for studying emergent communication. For experiments beyond standard environments, many researchers develop custom implementations of games and agents Evtimova et al. (2017); Mu & Goodman (2021); Noukhovitch et al. (2021) using general-purpose tools, such as PyTorch Paszke et al. (2019).

Signaling Game and Its Variants A signaling game generally involves two agents: a sender and a receiver Lazaridou et al. (2016). The goal is for the receiver to identify a target sample from a set, which may include distractors, using only the message sent by the sender. The set can include images Lazaridou et al. (2016); Bouchacourt & Baroni (2018); Denamganaï et al. (2023), object feature vectors (texts) Carmeli et al. (2023), or graphs Słowik et al. (2021). The sender encodes a message containing information about the correct sample. Key design decisions include whether the sender views only the correct sample or also some distractors, which might differ from those given to the receiver Lazaridou et al. (2016). On the receiver's side, factors such as the number of distractors and whether the correct sample or a similar one is provided for selection must also be considered Kang et al. (2020). The encoded message is the only information passed to the receiver, who uses it to make their selection. However, such signaling game in referential Gym Denamganaï & Walker (2020b) has seen less adoption in the literature Evtimova et al. (2017). A variation similar to the signaling game is the reconstruction game. In this setup, the receiver does not select from a collection but instead reconstructs a sample based on the sender's communication. The objective is to replicate the original sample shown to the sender as closely as possible Rita et al. (2022); Chaabouni et al. (2020). This approach resembles an autoencoder, with the latent space designed to mimic or support language, as implemented in the Emergence of Language in Games (EGG) framework Kharitonov et al. (2019). The key distinction between reconstruction and referential games, often conflated in early studies, is that referential games involve a collection for the receiver to choose from, whereas reconstruction games do not Guo (2019). The question-answer game is also a variation of the signaling game that introduces more flexibility by allowing iterative, bilateral communication over multiple rounds Bouchacourt & Baroni (2019). Unlike original signaling and reconstruction games, it enables the receiver to ask follow-up or clarifying questions Das et al. (2017); Agarwal et al. (2018). This setup has sparked interest in exploring the symmetry of emergent language (EL), though it remains less widely adopted Das et al. (2017).

Grid World Games Grid world games simulate simplified 2D environments to model scenarios such as warehouse path planning Blumenkamp & Prorok (2021), object movement Kolb et al. (2019), traffic management Simões et al. (2020); Gupta et al. (2020), or maze navigation Sukhbaatar et al. (2016); Kalinowska et al. (2022). These games offer significant design flexibility, allowing agents to either participate as part of the environment or act as external supervisors. Key design elements include the complexity of the environment and the scope of agents' observations. Despite their popularity in the literature, grid world implementations vary greatly, making them a highly diverse category.

The variants will be the Continuous world games, which introduce greater complexity to the learning process Wang et al. (2019c); Pesce & Montana (2020b). In EL approaches, these settings often involve multi-task learning, where agents simultaneously engage with the environment and develop language. Continuous world games, whether 2D or 3D, offer added challenges and realism, making them more intricate than discrete environments. This increased complexity enhances the potential for deploying EL agents in real-world scenarios Patel et al. (2021).

4.5 Addressing Challenges in Multi-agent Communication

The field of **multi-agent communication** focuses on enabling autonomous (computer) agents to coordinate their actions using communication protocols. This concept is most commonly illustrated by teams of autonomous robots working together, but it also applies to scenarios such as self-driving cars interacting on the road or IoT devices communicating over a local area network. There are two primary approaches to developing communication protocols for multi-agent systems: **handcrafting** them or **learning** them as latent variables between agents Boldt & Mortensen (2024). **Handcrafted protocols** (e.g., DHCP for network configuration) are highly effective for specific tasks but require extensive expert design. This limits their adaptability for open-domain or general-purpose communication. **Automatically learned protocols**, where messages are represented as continuous vectors, address some of these limitations. However, they introduce new challenges, such as low interpretability, due to the complexity of deep learned representations. It is important to note that this task is distinct from autonomous agents communicating directly with humans, which will be discussed in the next subsection.

Category	Description	Representative Works
Emergence of Language in Games	General frameworks and toolkits designed for building, training, and evaluating agents in emergent communication tasks. These platforms standardize interfaces, simplify development, and encourage reproducibility.	• EGG: symbolic and neural agent toolkit Kharitonov et al. (2019); • Studies that implement new games with EGG, such as Chaabouni et al. (2019); Dessì et al. (2019); Chaabouni et al. (2020); Auersperger & Pecina (2022); Kharitonov et al. (2020), expand its accessible range of games and metrics; • TexRel: compositional image-language benchmark Perkins (2021b); • HexaJungle: diverse grounded communication tasks Ikram et al. (2021); • Custom codebases using PyTorch Evtimova et al. (2017); Mu & Goodman (2021); Noukhovitch et al. (2021).
Signaling Game and Its Variants	Sender-receiver interaction games where the goal is either to identify a target, reconstruct a message, or iteratively refine communication. These variants differ in message structure, agent roles, and level of interactivity.	• Signaling Game: sender encodes target, receiver selects Lazaridou et al. (2016); Bouchacourt & Baroni (2018); Denamganaï et al. (2023); Carmeli et al. (2023); Słowik et al. (2021); Kang et al. (2020); • Signaling game in referential Gym Denamganaï & Walker (2020b); Evtimova et al. (2017) • Reconstruction Game: receiver reconstructs input Rita et al. (2022); Chaabouni et al. (2020); • Question–Answer Game: multi-round clarification Bouchacourt & Baroni (2019); Das et al. (2017); Agarwal et al. (2018).
Grid World Games	2D spatial environments for grounded interaction, object manipulation, or path planning. These games range from discrete to continuous settings, often used to study embodiment and communication under navigation or cooperation tasks.	• Pathfinding and object movement Kolb et al. (2019); Sukhbaatar et al. (2016); • Traffic coordination and network control Simões et al. (2020); Gupta et al. (2020); • Complex planning in mazes Blumenkamp & Prorok (2021); • Continuous 2D/3D worlds for realism Wang et al. (2019c); Pesce & Montana (2020b); Patel et al. (2021).

Table 7: Summary of Experimental Settings for Emergent Communication (Sec. 4.4)

Emergent communication offers a promising framework for addressing the limitations of traditional multi-agent communication protocols. This approach is characterized by three key advantages, each of which can be enhanced through specific design elements and experimental validation.

Scalability to General-Purpose Tasks Emergent communication protocols are inherently scalable to a broader range of tasks, as they are developed through computational processes driven by the functional demands of the specific task. To encourage the development of general-purpose language, emergent communication systems can be designed within open-world, open-ended environments such as *Minecraft* or scenarios with diverse situations like *Starcraft* or *Dota 2*. These environments provide a rich variety of contexts, enabling agents to develop adaptable communication strategies. Furthermore, tasks with adversarial components—where one team of agents must innovate to outcompete another—can elicit a diversity of situations, fostering the emergence of robust and flexible communication protocols.

Current work on developing multi-agent communication protocols has explored various environments, including navigation tasks Mul et al. (2019); Li et al. (2022); Nisioti et al. (2023) and signaling games Bullard et al. (2021); Cope & Schoots (2021); Wang et al. (2024e); Tucker et al. (2021). While these studies demonstrate the feasibility of emergent communication, they have yet to establish a definitive real-world task where emergent communication surpasses state-of-the-art non-emergent approaches. Future work should focus on identifying such a niche, particularly in open-domain tasks that demand continual adaptation, rendering handcrafted protocols impractical and highlighting the advantages of emergent communication over learned continuous methods.

Interpretability A notable feature of emergent communication is its structural resemblance to human language, which enhances interpretability. To achieve this, agents can be constrained to communicate using discrete symbols at human-like scales, such as modest vocabulary sizes and message lengths. This design choice contrasts with continuous vector-based representations, which lack intuitive interpretability and require complex mathematical transformations for analysis. By aligning more closely with human-like linguistic structures, emergent communication protocols offer greater transparency and ease of understanding. Empirical validation of interpretability remains a critical next step. While emergent communication has intuitive advantages, it is necessary to formally demonstrate that it is more interpretable than continuous communication. This requires developing rigorous metrics for interpretability and conducting comparative studies to verify whether the structural similarities to human language confer functional benefits in real-world applications. Interpreting noisy agent-generated messages is challenging, often leading researchers to form broad hypotheses about their meanings. Jorge et al. Jorge et al. (2016) attempted to decode agent messages and found that agents focused on specific features, like hair color. Choi et al. Choi et al. (2018) showed that linking messages to object attributes becomes harder in dynamic scenarios. Lazaridou et al. Lazaridou et al. (2016) used the purity index to assess how well agent messages aligned with human-like categories, achieving 41% purity with CNN encoding and a vocabulary of 100 symbols. Choi et al. demonstrated agents' ability to compose language through zero-shot evaluation, where agents accurately described unseen combinations of object features. To make agent communication more human-like, Lazaridou et al. modified the standard referential game by showing conceptually related images, encouraging the use of more human-like terms. This resulted in a modest improvement, raising language purity to 45%.

Robustness and Generalization Human language serves as a gold standard for communication protocols due to its unparalleled ability to generalize to novel situations and remain robust in the presence of noise or other disruptions. Emergent communication protocols can mimic these properties by incorporating design elements such as communication channel noise or periodically cycling out agents in the population. These techniques prevent agents from overfitting to each other, thereby fostering the development of more robust and generalizable communication strategies.

Current research has begun to explore the robustness of emergent communication to corruption and noise Cope & Schoots (2021); Wang et al. (2024e), as well as its ability to communicate with partners not encountered during training Bullard et al. (2021); Cope & Schoots (2021). However, further work is needed to empirically verify the intuitions that emergent communication's structural similarities to human language confer functional benefits beyond those of continuous or unconstrained communication. Demonstrating these advantages would significantly expand the potential applications of emergent communication in multi-agent systems.

Challenge Category	Approach	Representative Works
Scalability to General-Purpose Tasks	Use of open-world, task-rich environments to elicit adaptive and scalable communication strategies	• Hierarchical protocol learning in referential games Mul et al. (2019) • Grounding communication via autoencoders in navigation Li et al. (2022) • Autotelic reinforcement learning for adaptive goal generation Nisioti et al. (2023) • Dual-channel communication in negotiation settings Tucker et al. (2021)
	Techniques to enforce protocol generality across diverse partners or contexts	• Symbol remapping, channel randomization Cope & Schoots (2021) • Quasi-equivalence training across agent populations Bullard et al. (2021)
Interpretability	Design protocols with discrete symbols and bounded vocabulary size to mimic human language	• Grounding symbols in interpretable features (e.g., hair color) Jorge et al. (2016) • Symbol purity and linguistic structure analysis Lazaridou et al. (2016) • Zero-shot compositional language assessment Choi et al. (2018)
Robustness and Generalization	Introduce agent cycling, bottlenecks, and symbol-level constraints to avoid co-adaptation and foster generalization	• Population cycling and bottleneck architectures Wang et al. (2024e) • Zero-shot communication with unseen agents Bullard et al. (2021) • Channel noise injection and message scrambling Cope & Schoots (2021)

Table 8: Summary of Challenges and Representative Approaches in Emergent Multi-Agent Communication (Sec. 4.5)

4.6 Discussions

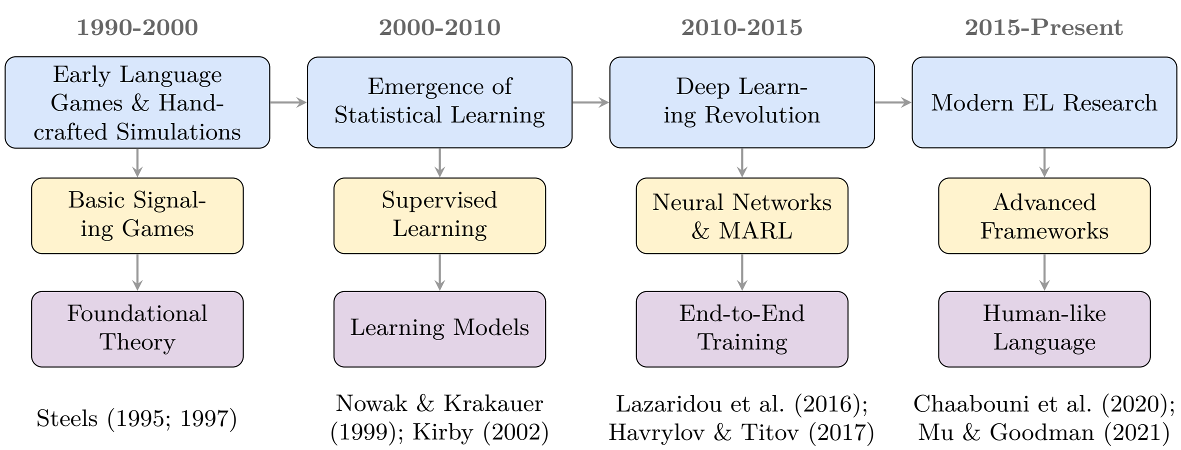


Figure 9: Evolution of Emergent Language Research: Timeline showing major milestones, approaches, and achievements

Fig. 9 traces the development of emergent language research over the past three decades. Early work in the 1990s centered on signaling games and hand-crafted simulations, establishing theoretical foundations for how communication systems can emerge among agents. In the 2000s, statistical learning and iterated learning models formalized language acquisition dynamics from an evolutionary perspective. The 2010s brought deep learning and multi-agent reinforcement learning, enabling end-to-end training of communication protocols and scalable demonstrations of language emergence. More recent studies have moved toward advanced frameworks emphasizing compositionality, generalization, and human-like communication. This trajectory reflects both a methodological shift, from handcrafted to statistical to neural and hybrid frameworks, and an evolving set of goals, from theory-driven explorations to practical systems that exhibit increasingly human-like behavior.

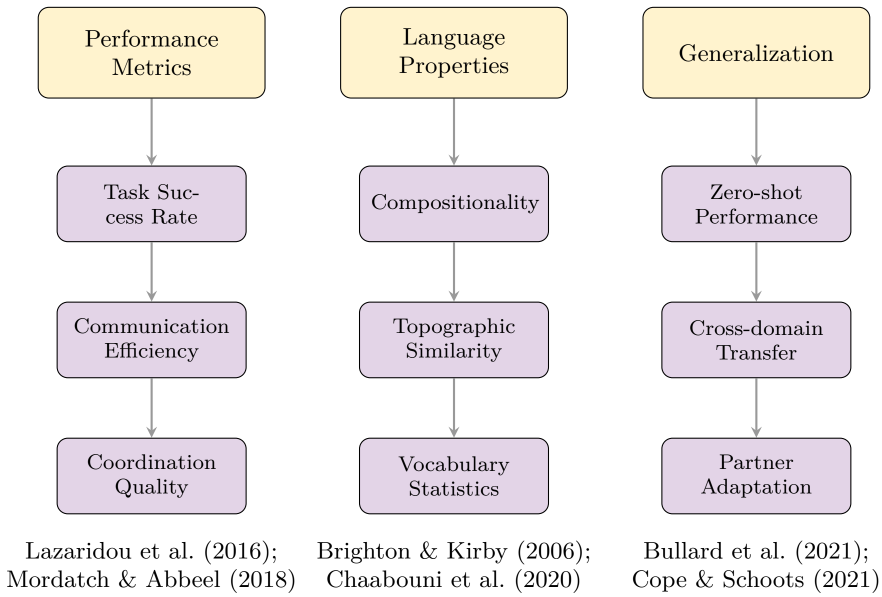


Figure 10: Taxonomy of Evaluation Metrics in Emergent Language Research

Fig. 10 summarizes the evaluation metrics that have shaped EL research, grouped into three categories: performance, language properties, and generalization. Performance metrics, such as task success, communication efficiency, and coordination quality, capture how communication contributes to solving cooperative tasks. Language property metrics, including compositionality, topographic similarity, and vocabulary statistics, assess the interpretability and structure of the induced protocols. Generalization metrics—such as zero-shot performance, cross-domain transfer, and partner adaptation—measure whether communication extends beyond the training distribution and adapts to new agents and environments. Taken together, these metrics highlight the field's progression from evaluating raw task success to emphasizing interpretability, adaptability, and alignment with human language.

As previously shown in Fig. 7 at the beginning of this section, we organize existing work across four dimensions: preliminary communication assumptions, environment settings, methodological challenges, and evaluation strategies. The reviewed studies demonstrate a steady progression from symbolic signaling games to multi-agent interactions in grounded 2D or 3D environments. Communication protocols have likewise diversified, evolving from simple discrete vocabularies to hybrid schemes that incorporate neural encoders, attention mechanisms, or memory modules. Architecturally, approaches range from feedforward and recurrent models to transformer-based frameworks designed for sequential or compositional communication.

Evaluation practices have matured alongside these developments. Early studies emphasized referential accuracy, while recent work prioritizes compositionality, contextual generalization, partner adaptation, and human-likeness. We categorize these trends under foundational assumptions (e.g., vocabulary size, cooperation level), environmental grounding, and functional benchmarks.

Looking forward, key challenges remain. First, there is a need for benchmarks where emergent communication provides clear advantages over engineered or continuous protocols. Such benchmarks should emphasize open-ended tasks requiring continual adaptation, compositional generalization, and interpretability. Progress also depends on testing common assumptions: that discrete emergent languages are inherently more interpretable than continuous ones, and that structural alignment with natural language yields practical benefits in transferability and adaptability. Finally, while current benchmarks achieve near-perfect accuracy in controlled tasks, they provide limited insight into real-world applicability. Future work must shift toward dynamic, multimodal, and semantically grounded environments, where communication is not only a tool for coordination but also a medium for building shared understanding.

5 Multi-Agent Systems Empowered by Large Language Models

Large language models (LLMs) have shown strong performance across various domains, largely due to training on large, diverse datasets, resulting in emergent **commonsense reasoning** capabilities and the ability to understand and generate **human language** Liang et al. (2022). Human language, though complex, is a general and flexible medium for conveying information.

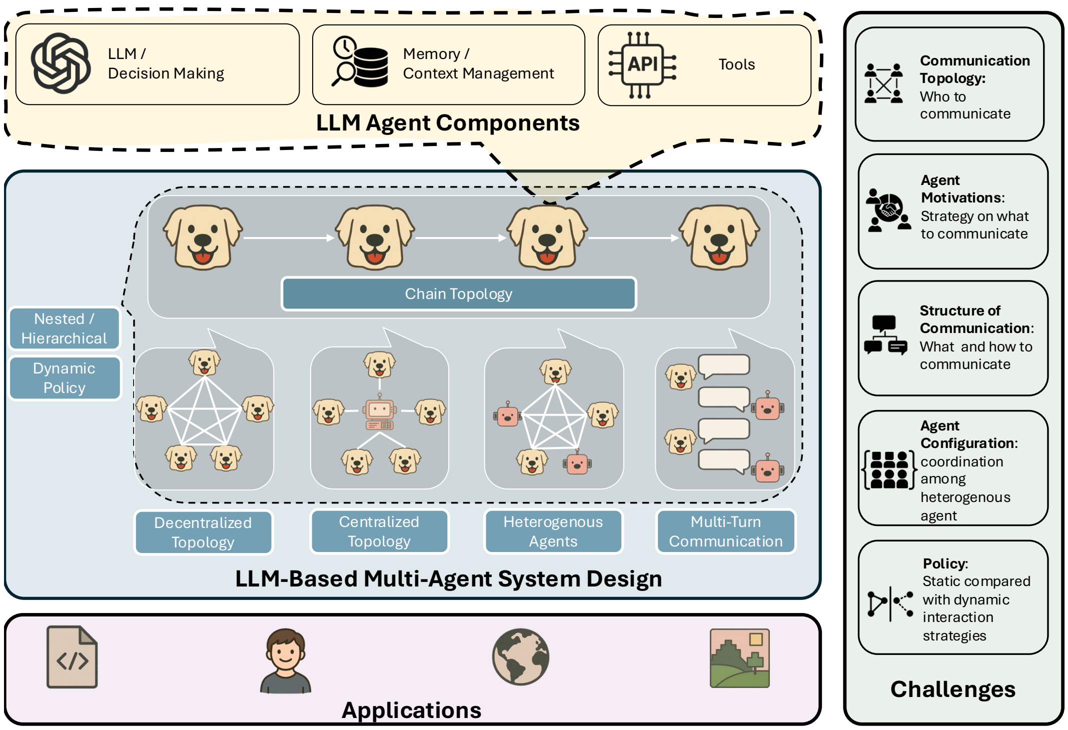


Figure 11: LLM-Based Multi-Agent System Design and Challenges. In an LLM-based multi-agent system, each agent outputs decisions that can be enhanced through memory, context management, and the use of domain-specific tools. Agents communicate through different topologies and can operate in either static or dynamic configurations, with applications spanning diverse domains. Building functional multi-agent systems remains challenging due to numerous design considerations, and several key challenges are summarized.

LLMs are increasingly used as autonomous agents that can plan, reason, and act in diverse environments. Recently, researchers have begun deploying multiple such LLM-based agents together, forming multi-agent systems (MAS) that communicate via natural language. The motivation is that a team of LLM agents can leverage diverse skills and perspectives to solve complex tasks beyond the capabilities of a single agent. LLM-powered agents use natural language as their primary communication method, enabling coordination and collaboration that are not tied to specific tasks or domains, while also remaining interpretable to humans. This natural language interface supports a wide range of tasks without requiring hand-crafted communication protocols, offering new possibilities for interaction and cooperation in multi-agent systems. This section surveys the architectures and communication patterns that characterize LLM-based MAS and provides an overview of key applications across different domains.

Figure 12: LLM Agent Communication Taxonomy

Language-based agents bring several advantages to multi-agent settings. First, an environment can be described and reasoned about using natural language Guo et al. (2024a); Liu et al. (2025b). Since most scenarios can be expressed in language, and because large language models (LLMs) are trained on vast datasets with commonsense knowledge, these agents are capable of generalizing to new environments without requiring retraining. While they still face issues like hallucination and limited grounding Ahn et al. (2022), they already offer a level of adaptability that is hard to achieve with traditional reinforcement learning (RL) agents Guo et al. (2024a). These capabilities are further improved by multimodal LLMs (MLLMs), which allow agents to process both language and sensory inputs Durante et al. (2024). Second, LLM agents can communicate with each other using flexible and structured messages. Unlike agents that follow fixed communication protocols, LLM agents can easily switch between different communication patterns—such as single-turn, multi-turn, self-reflection, iterative planning, or following specific workflows—either through prompting or with fine-tuning. This allows them to support dynamic negotiation, task sharing, and coordination without needing specialized retraining for each behavior. Third, LLM agents are well-suited for interacting with humans. Natural language enables direct human-AI collaboration, where humans can give instructions, provide feedback, and guide agent behavior in an intuitive and human-interpretable way Wu et al. (2022); Feng et al. (2024).

In contrast to traditional RL agents, which rely on domain-specific training and fixed communication protocols, LLM-based agents offer a general and adaptable communication framework. Recent research has explored combining LLMs with RL to take advantage of both natural language reasoning and reward-based learning Sun et al. (2024a). These hybrid approaches aim to develop agents that are both grounded in their environments and capable of flexible planning and communication.

As interest in multi-agent LLM systems continues to grow, communication has become a central focus. Natural language offers a unified interface for agent-environment interaction, agent-agent coordination, and human-agent collaboration. This survey examines how LLMs introduce new communication capabilities in multi-agent systems and reviews emerging strategies and research directions enabled by language-based interaction. We begin by introducing the foundational capabilities and opportunities brought by LLMs in Section 5.1.1, followed by a discussion of LLM-based agents in Section 5.1.2. Section 5.2 outlines key design dimensions for communication in LLM-powered multi-agent systems. Section 5.3 provides an overview of popular frameworks that support multi-LLM collaboration, and Section 5.4 highlights major application domains. We then review recent efforts to integrate RL with LLMs in MAS settings in Section 5.5, and discuss evaluation platforms for LLM-based multi-agent systems in Section 5.6. Finally, Section 5.7 concludes with a discussion of future research directions.

5.1 Background on LLMs and LLM Agents

5.1.1 Large Language Models (LLMs)

LLMs have emerged as transformative tools for agent-based systems, offering advanced capabilities such as reasoning, planning, and natural language understanding. These capabilities arise not from explicit programming but from the emergent properties of training on large-scale, diverse corpora Liang et al. (2022). When deployed as decision-making agents, LLMs demonstrate robust adaptability in dynamic, multi-agent environments by enabling flexible coordination and communication beyond rule-based constraints: (1). **Emergent Behaviors.** LLMs display emergent abilities—such as strategic planning, abstract reasoning, and social simulation—that were not explicitly encoded during training. These behaviors empower agents to engage in high-level tasks including negotiation, cooperative problem-solving, and context-aware decision-making. In multi-agent systems, such behaviors are particularly valuable for fostering robust coordination strategies in unpredictable settings. (2). **In-Context, Few-shot, and Zero-shot Learning.** LLMs are capable of adapting to new tasks via in-context learning, interpreting task instructions and examples directly from the prompt without parameter updates. This flexibility extends to few-shot and zero-shot settings, where agents can perform novel tasks using minimal or no prior training data. Such capabilities significantly reduce development overhead and enable agents to operate in open-ended, evolving environments with task-agnostic generalization. (3). **Natural Language Communication.** The ability to understand and generate human language makes LLMs ideal for facilitating interpretable communication within multi-agent systems and between agents and humans. Agents can articulate goals, strategies, and rationales using natural language, improving both inter-agent alignment and human oversight. This is critical for collaboration, transparency, and trust in mixed-agent teams. (4). **General-Purpose Reasoning Across Environments.** Importantly, LLMs are not pre-trained to optimize behavior for any specific environment (e.g., a particular game or graphical interface). Instead, they serve as general-purpose cognitive engines that can be fine-tuned or scaffolded to suit domain-specific requirements. This property supports reusability across diverse agent applications—from GUI-based assistants to strategic game agents—while maintaining task flexibility. In summary, the emergent reasoning, flexible learning paradigms, and language-driven communication of LLMs establish them as foundational components for next-generation multi-agent systems. By leveraging these attributes, agents can dynamically adapt to new objectives, collaborate through interpretable dialogue, and function effectively in complex, uncertain environments.

5.1.2 LLM Agent

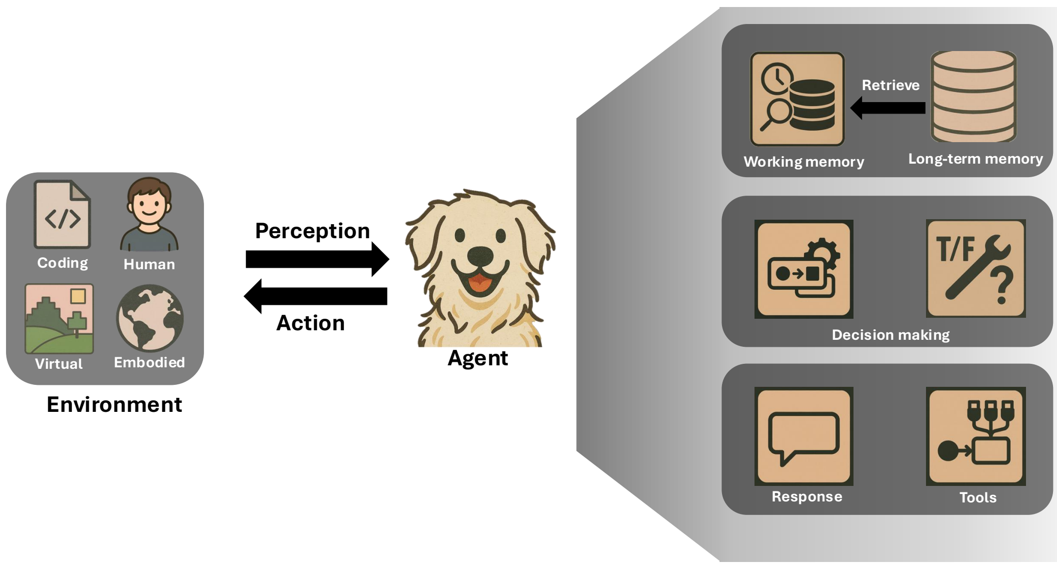


Figure 13: General Architecture of an LLM-Based Agent. The figure depicts how an agent interacts with coding, human, virtual, and embodied environments through perception and action loops. Within the agent, customized modules can be employed to handle memory, decision making, response generation, and tool use, enabling flexible and adaptable performance across domains.

An LLM-based agent is a modular system that integrates reasoning, memory, and task execution around a central language model Durante et al. (2024); Wang et al. (2024a). As illustrated in Fig. 13, these agents are structured into four core components, each contributing to their ability to act coherently and adaptively within dynamic environments: (1). *Profile.* This component encodes the agent's identity, personality traits, and social attributes. Profiles influence interaction style and decision preferences, allowing agents to simulate diverse roles or personas. Profiles can be handcrafted, learned, or aligned with external datasets. (2). *Memory.* Memory systems enable agents to persist and retrieve historical information, such as past interactions or task outcomes. Implemented as embeddings, text logs, or databases, this component supports long-term coherence and task continuity through read/write operations and reflective reasoning. (3). *Planning.* The planning module supports decomposition of goals into actionable steps and the formation of strategies to reach those goals. Plans may be iteratively refined based on agent self-evaluation, feedback from other agents, or environmental input, enabling responsiveness to dynamic task contexts. (4). *Action.* This component operationalizes plans into concrete actions. Actions may involve environmental manipulation, communication, tool usage, or internal reasoning. By integrating outputs from memory and planning, this module allows the agent to execute tasks effectively and adapt behavior in real time. Together, these components form a cohesive architecture for LLM-based agents capable of multi-turn reasoning, human interaction, and decentralized coordination. This modular design facilitates extensibility and customization for a broad range of multi-agent settings.

5.2 Overview of LLM-powered Multi-agent Systems with Communications

Communication is fundamental for enabling collective intelligence in LLM-powered multi-agent systems (LLM-MAS). To systematically characterize agent interactions and coordination strategies, we analyze communication along the following structured dimensions: (1) **Communication Topology** (Section 5.2.1), which defines the architectural organization of communication flows and agent interaction patterns; (2) **Motivational Settings** (Section 5.2.2), which describe the underlying relationships and objectives among agents that shape their interaction strategies; (3) **Agent Settings** (Section 5.2.3), distinguishing between homogeneous and heterogeneous agent populations and their implications for coordination; (4) **Structure of Communication** (Section 5.2.4), examining whether agent interactions are single-turn or multi-turn, and analyzing the communication content by detailing the types of information agents exchange during collaboration.

5.2.1 Communication Topology

The **communication topology** in LLM-based multi-agent systems (LLM-MAS) defines how agents exchange messages, coordinate behaviors, and structure their interactions. Different topologies significantly influence system scalability, robustness, coordination efficiency, and flexibility. Here, we examine five representative communication structures (Figures 14–18), each highlighting distinct advantages and trade-offs.

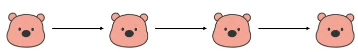


Figure 14: Chain

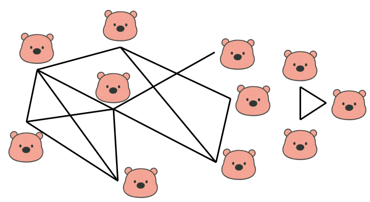


Figure 15: Decentralized

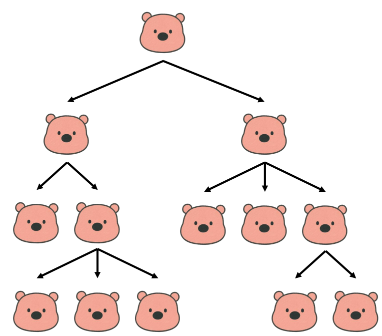


Figure 16: Hierarchical

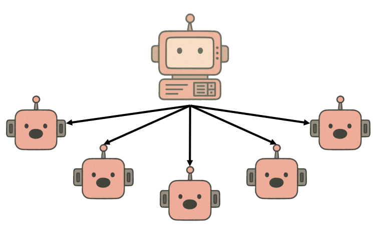


Figure 17: Centralized

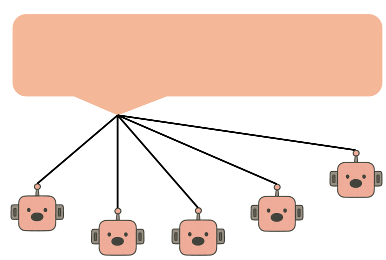


Figure 18: Shared Message Pool

Chain topology (Fig. 14) describes a linear communication pattern where each agent communicates exclusively with its immediate neighbors. This structure is particularly suited for pipelines or sequential workflows, ensuring messages flow in an organized manner.

In sequential communication, agents operate in a step-by-step chain, where each one handles a part of the task and passes the result to the next Zhang et al. (2024d). This setup breaks down long or complex problems into smaller parts that are easier to process. Some systems follow this pattern for multi-step workflows such as software development, with different agents handling design, coding, testing, and review in a clear sequence Qian et al. (2024a); Du et al. (2024). This can make complex tasks more organized and manageable, but it also requires careful coordination between stages to avoid errors or loss of information.

Decentralized topology (Fig. 15) permits agents to directly communicate with multiple peers, creating fully or partially connected networks without centralized control. While enhancing scalability and robustness in dynamic environments, this flexibility may introduce redundancy, increased communication overhead, or inconsistencies in global strategies.

Existing decentralized communication works can be broadly categorized into several typical approaches. One line of work focuses on social or strategic game-based environments, often incorporating role-playing and persona-driven agent designs. In these settings, agents interact in rounds to reason, deceive, and coordinate, as demonstrated in games like Werewolf, Avalon, and other scenario-based simulations Xu et al. (2023b); Hua et al. (2023); Bailis et al. (2024); Light et al. (2023); Kannan et al. (2023); Wang et al. (2023b); Lan et al. (2023); Chen et al. (2024a). Another active area develops debate-style frameworks, where agents engage in structured dialogues to challenge, refine, or converge on ideas Li et al. (2024e); Jin et al. (2024); Chen et al. (2023b); Du et al. (2023); Liang et al. (2023); Chan et al. (2023). These debates are typically conducted in decentralized settings using fully connected communication graphs, allowing agents to access and respond to each other's thoughts, promoting both diverse reasoning and consensus formation. Other efforts center on embodied or task-grounded agents, where communication facilitates physical collaboration or goal-directed behavior in interactive environments Ma et al. (2023); Zhang et al. (2023b); Wu et al. (2025); Kannan et al. (2023); Liu et al. (2024a). Finally, a growing body of work explores large-scale agent societies and general-purpose communication frameworks that enable agents to self-organize, share knowledge, and adapt over time. These include systems designed to simulate emergent social dynamics or orchestrate decentralized coordination at scale Park et al. (2023); Carroll et al. (2019); Kaiya et al. (2023); Nascimento et al. (2023b); Piatti et al. (2024); Zhang et al. (2024b;a); Liu et al. (2024b); Li et al. (2024a); Xiong et al. (2023); Nascimento et al. (2023a); D'Arcy et al. (2024); Abdelnabi et al. (2024); Li et al. (2023c); AL et al. (2024); Liu et al. (2023b).

Hierarchical topology (Fig. 16) structures communication into layers, assigning distinct roles to agents: higher-level agents handle planning, abstraction, and task allocation, while lower-level agents are specialized for executing subtasks or domain-specific actions. This design enables modularity and scalable coordination by decomposing complex tasks. However, effective use relies on strong role specialization, coherent coordination among agents across layers, and a shared understanding of task structure and progression.

A common hierarchical approach is to organize agent teams into multiple layers, where higher-level agents focus on planning, task assignment, and monitoring, while lower-level agents carry out subtasks or specialized behaviors Fourney et al. (2024); Wang et al. (2024b); Qiao et al. (2024); Jinxin et al. (2023). This layered setup supports scalability and improved coordination in complex systems. Notably, many hierarchical systems emerge from role-based multi-agent simulations, where agents are embedded within organizational structures that mirror real-world hierarchies. For example, simulations of hospitals Li et al. (2024b), legal services Sun et al. (2024b), and automated machine learning pipelines Trirat et al. (2024); Du et al. (2024) assign domain-specific roles (e.g., doctors, nurses, or analysts) that are coordinated by supervisory agents at higher levels. Similarly, multi-agent environments such as AgentVerse Chen et al. (2023c) facilitate structured collaboration and often adopt hierarchical control to support emergent behaviors and inter-role communication. Beyond these domain-specific applications, general-purpose frameworks and surveys have highlighted the broader value of hierarchy in enabling problem decomposition, organized cooperation, adaptive teaming, and goal tracking across varied contexts Ishibashi & Nishimura (2024); Rasal & Hauer (2024); Xu et al. (2023a); Wu et al. (2024a); Ahn & Sentis (2021); Sun et al. (2025c); Zheng et al. (2023); Guo et al. (2024c). Together, these studies underscore that well-designed hierarchical systems can manage complexity, improve communication flow, and enhance reliability in multi-agent LLM systems, while also emphasizing the need for clear role definitions and safeguards against coordination bottlenecks.

Centralized topology (Fig. 17) involves a single central agent coordinating communication among all agents. This centralized coordinator manages information flow, assigns tasks, and synchronizes actions, simplifying coordination and promoting consistency, particularly beneficial in large-scale deployments.

Scalable Multi-Robot Collaboration Chen et al. (2024g) demonstrates this approach by effectively allocating roles and synchronizing multiple agents. Another example of centralized coordination is MindAgent, which uses a central manager to guide agent behaviors in gaming scenarios by distributing information and ensuring coherent interactions Gong et al. (2023). Centralized topologies can simplify decision-making and maintain consistent global goals. However, reliance on a central coordinator can make the system less adaptable and prone to failure if the central agent is not reliable.

A number of frameworks adopt a centralized component to orchestrate multi-agent systems, typically using a central planner or coordinator to manage information flow, assign tasks, and synchronize agent behaviors. In task planning and coordination, the central agent decomposes high-level goals into subtasks, delegates them to specialized agents, and aggregates outcomes to ensure coherent execution Wu et al. (2023b); Hong et al. (2023); Chen et al. (2023a); Xie et al. (2023). In workflow automation and system integration, centralized controllers manage complex pipelines or multi-agent schedules, guiding tool use and handling inter-agent dependencies Fourney et al. (2024); Trirat et al. (2024); Pan et al. (2024). Centralized control also appears in embodied environments and strategic games, where a single agent supervises multiple role-specific agents to coordinate actions and maintain global task progression Kannan et al. (2023); Gong et al. (2023). While these centralized designs offer promising scalability, they also centralize responsibilities such as dependency management, synchronization, and fault handling in a single component, leading to over-reliance that can undermine system reliability when that component fails or underperforms.

Shared message pool topology (Fig. 18) employs a global broadcast channel or shared repository where agents asynchronously publish messages and retrieve relevant information. This topology offers a balance between structural rigor and flexibility but may face challenges like message overload or latency in high-traffic scenarios.

Several recent frameworks adopt a shared message pool or shared context to coordinate agents without requiring direct pairwise communication. In dialogue-based collaboration, agents interact by contributing to a shared conversation history or memory buffer, enabling asynchronous communication and joint reasoning through context accumulation Li et al. (2023b); Wu et al. (2023b); Chen et al. (2023c;b). In graph-based coordination, agents represent nodes in a shared computational graph, updating local states and propagating information through structured message passing or knowledge graphs to facilitate scalable planning and cooperation Hu et al. (2024d); Jeyakumar et al. (2024); Yang et al. (2025). Task-oriented collaboration frameworks often rely on shared repositories where agents publish plans, partial results, or reasoning steps that others can retrieve and build upon—supporting modularity and collaborative problem-solving without centralized control Chen et al. (2024d); Islam et al. (2024); Wang et al. (2024f); Pan et al. (2024); Tang et al. (2023). Finally, some works explore emergent organization, where agents self-assemble roles and structures through interactions within a common environment or memory space, often incorporating theory-of-mind modeling or adaptive self-organization mechanisms Li et al. (2023c); Ishibashi & Nishimura (2024). These shared context approaches offer flexibility, loose coupling, and scalability, enabling multi-agent systems to coordinate effectively without requiring dense communication topologies.

5.2.2 Motivational Settings

Motivational settings define the nature of the objectives and relationships among agents, significantly influencing their communication strategies. We categorize motivational settings into three main types: competitive, cooperative, and mixed-motive (Figures 19–21).

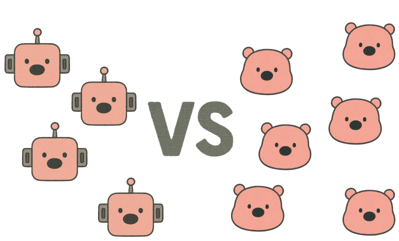


Figure 19: Competitive

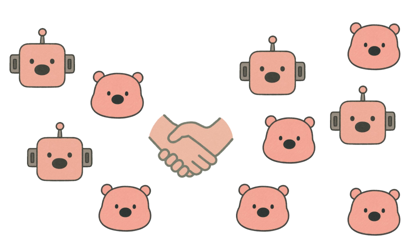


Figure 20: Cooperative

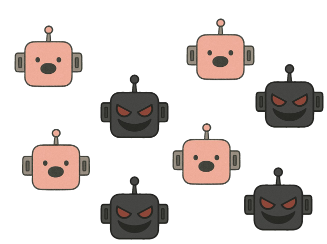


Figure 21: Mixed-Motives

Competitive settings (Fig. 19) occur in adversarial environments where agents pursue conflicting goals. In these contexts, communication often involves negotiation, strategic deception, persuasion, or withholding information to gain an advantage. However, even in competitive setups, the ultimate objective is not always purely adversarial. Some systems use competition, such as debate or deliberation, to surface stronger arguments or reach more accurate collective outcomes. In this way, competitive dynamics can serve as a means to enhance reasoning or drive convergence, blending rivalry with underlying cooperative intent.

Agents may attempt to influence, mislead, or withhold critical knowledge to gain an advantage, often adapting tactics in response to opponent behavior. These dynamics appear in a variety of formats, including hidden-role games Bailis et al. (2024); Wang et al. (2023b); Light et al. (2023), structured debates Du et al. (2023); Liang et al. (2023); Chan et al. (2023), and competitive benchmarks Chen et al. (2024c). Some systems simulate strategic dialogues where agents engage in negotiation or critique to advance individual interests Abdelnabi et al. (2023); Chen et al. (2023b). Others embed agents in dynamic, multi-agent environments where communication must adapt over time to changing conditions and opponent strategies Gong et al. (2023); Chen et al. (2024c). Learning-based frameworks also explore how agents improve in competitive settings through self-play or reinforcement mechanisms, such as actor-critic collaboration Du et al. (2024); Feng et al. (2024). Together, these studies highlight how language serves not only as a channel for coordination, but as a means of tactical adaptation in adversarial multi-agent interactions.

Cooperative settings (Fig. 20) focus on shared objectives, encouraging agents to engage in collaborative communication aimed at maximizing collective performance. In these scenarios, agents exchange task-relevant information, share plans, and continually update each other on progress. Most multi-agent LLM frameworks to date fall under this cooperative paradigm, reflecting the foundational assumption that intelligent agents can accomplish more together than alone.

Many systems define structured roles, planning hierarchies, and dialogue rules to guide coordination, especially in task-driven domains such as software engineering, healthcare, law, and scientific discovery Li et al. (2023b); Qian et al. (2024a); Du et al. (2024); Li et al. (2024b); Sun et al. (2024b); Ghafarollahi & Buehler (2024). Others explore generalized frameworks or benchmarks that assess cooperative capabilities in areas like cross-cultural alignment, collaborative review generation, team formation, and dynamic task allocation Zhang et al. (2024b;a); Zheng et al. (2023); D'Arcy et al. (2024); Agashe et al. (2023b); Pan et al. (2024); Xu et al. (2023a); Li et al. (2024a). A growing body of research also investigates value alignment, trust-building, and preference modeling in human-agent cooperation, often using structured discussion protocols or reflective updates Carroll et al. (2019); Feng et al. (2024); Chen et al. (2023b); Yin et al. (2023); Wu et al. (2024a); Shinn et al. (2024). Cooperative paradigms are especially important in embodied and physical multi-agent systems, where agents must coordinate real-time decisions in robotics, simulation, or navigation tasks Zhang et al. (2023b); Chen et al. (2024g); Mandi et al. (2024); Guo et al. (2024c); Wang et al. (2023c). Many frameworks now support self-organizing or reconfigurable teams that adapt over time to new goals or contexts, enabling scalable and general-purpose collaboration Ishibashi & Nishimura (2024); Liu et al. (2024b); Rasal & Hauer (2024); Schmidgall et al. (2025); Chen et al. (2024f); Qiao et al. (2024); Zhuge et al. (2024a); Liu et al. (2024a); Hu et al. (2024d); Dong et al. (2024); Trirat et al. (2024); Tang et al. (2023). Recent systems such as *AgentVerse*Chen et al. (2023c), *AutoGen*Wu et al. (2023b), *MetaGPT*Hong et al. (2023), and *ProAgent*Zhang et al. (2023a) provide reusable infrastructures for spawning and coordinating LLM agents across diverse tasks. Other frameworks support emergent collaboration via shared memory, configurable interfaces, or social cognition modules Jinxin et al. (2023); Das et al. (2023); Li et al. (2023c); Nascimento et al. (2023b); Wang et al. (2024f); Islam et al. (2024). Together, these studies reveal a prevailing trend: cooperation is not only a practical baseline, but also a powerful lens for studying language-based interaction, coordination, and collective intelligence in LLM-agent systems.

Mixed-motive settings (Fig. 21) blend cooperative and competitive elements, often unfolding in open-ended environments where agents are free to pursue their own objectives, form alliances, negotiate roles, or act independently. Rather than being confined to rigid workflows, agents in these settings interact dynamically—sometimes collaborating to achieve shared goals, and at other times competing for influence, resources, or strategic advantage.

These dynamics are especially common in simulations of complex social environments, where agents develop their own strategies, shift alliances, or respond to evolving interpersonal contexts, as seen in generative agent societies, role-based simulations, and open-world behavioral sandboxes Lan et al. (2023); Hua et al. (2023); Chen et al. (2024a); Abdelnabi et al. (2024). Several frameworks propose general-purpose platforms for modeling large-scale, self-organizing agent collectives capable of handling both collaboration and rivalry at scale Cheng et al. (2024); Liu et al. (2025b); Tran et al. (2025b); Sun et al. (2025c); Xie et al. (2023); AL et al. (2024); Hu et al. (2024a). Other systems use modular or nested architectures—such as role-based organizations, team-vs-team structures, or mixtures of experts—where cooperation exists within groups but competition may emerge between them Fourney et al. (2024); Xu et al. (2023c); Gong et al. (2023); Wu et al. (2025); Li et al. (2024e); Ahn & Sentis (2021). These studies underscore the importance of designing flexible communication and coordination protocols that allow agents to manage trade-offs, respond to evolving social landscapes, and pursue complex goals in mixed-motive multi-agent systems.

Overall, motivational settings significantly shape agent interactions, determining whether agents prioritize open information sharing, hide information, or dynamic adaptation in response to others' behaviors and evolving task demands.

5.2.3 Agent Settings

LLM-based multi-agent systems (LLM-MAS) can be configured with either heterogeneous or homogeneous agents, influencing complexity, flexibility, and interaction dynamics (Figures 22 and 23).

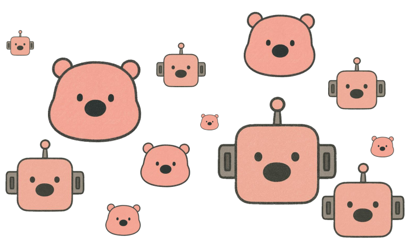


Figure 22: Heterogeneous

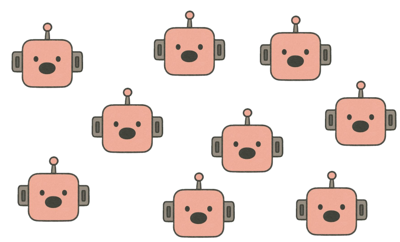


Figure 23: Homogeneous

Heterogeneous agents (Fig. 22) feature agents with different roles, skills, knowledge, or identities, enabling specialized communication and division of labor. This diversity supports nuanced interactions and richer reasoning processes.

Many frameworks build multi-agent systems for collaborative tasks such as team coordination, cooperative problem solving, and cross-domain knowledge sharing, often adopting a role-playing paradigm where agents simulate distinct functions using a shared LLM backbone—achieving behavioral diversity through prompting rather than architectural differences Li et al. (2023b); Qian et al. (2024a); Xu et al. (2023b); Wu et al. (2025); Chen et al. (2024c); Ghafarollahi & Buehler (2024); Li et al. (2024c); Stepputtis et al. (2023); Lan et al. (2023); Park et al. (2023); Cheng et al. (2024); Hu et al. (2024a); Fourney et al. (2024); Xie et al. (2023); AL et al. (2024); Hong et al. (2023); Schmidgall et al. (2025); Sun et al. (2024b); Chen et al. (2024e); Hua et al. (2023). Others introduce more substantial heterogeneity by equipping agents with specialized tools, APIs, or environmental access Chen et al. (2024e); Schmidgall et al. (2025); Hua et al. (2023). Additional works emphasize persona diversity—agents with different memories, social styles, or behavioral traits—to simulate realistic human dynamics Carroll et al. (2019); Chen et al. (2024a); Ahn & Sentis (2021); Liu et al. (2025b); Sun et al. (2025c); Tran et al. (2025b); Li et al. (2024e); Zhang et al. (2023b). Some explore strategic heterogeneity in contexts like debate, where different reasoning styles support negotiation and opinion shaping Li et al. (2024e); Xu et al. (2023b); Chen et al. (2024a). Together, these works highlight how heterogeneity fosters more flexible and effective multi-agent communication.

Homogeneous agents (Fig. 23) utilize agents with identical architectures and capabilities. Uniformity allows streamlined interaction patterns, reducing the complexity of coordination mechanisms and facilitating large-scale deployments.

Several frameworks focus on how identical agents coordinate tasks, share resources, or reason collectively Chen et al. (2024f); Dong et al. (2024); Hu et al. (2024d); Tang et al. (2023); Du et al. (2023). These works demonstrate that homogeneous setups can produce robust collective behaviors with minimal role differences, making them practical for scalable and maintainable multi-agent systems. In embodied environments, modular embodied agents Zhang et al. (2023b) can also collaborate effectively on tasks such as object gathering and navigation.

Ultimately, the choice between heterogeneous and homogeneous agent settings involves trade-offs between flexibility, complexity, coordination effort, and scalability, tailored to the specific needs and objectives of the LLM-powered multi-agent system.

5.2.4 Structure of Communication

Communication in multi-agent systems is shaped by both interaction patterns and message content, which determine how agents share information, coordinate actions, and reason collectively.

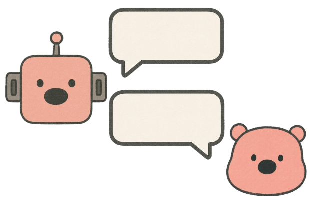


Figure 24: Single-turn

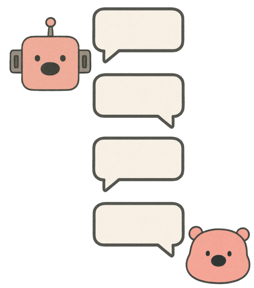


Figure 25: Multi-turn

Interaction Pattern. The pattern of communication shapes the depth and effectiveness of multi-agent collaboration. It determines whether agents operate independently after minimal coordination or engage in iterative exchanges that enable joint reasoning, coordination, and consensus building. Depending on the task, different patterns may be more suitable: single-turn interactions reduce latency and simplify control, while multi-turn dialogues support deeper deliberation, mutual understanding, and alignment. Each comes with trade-offs: multi-turn communication introduces overhead, complexity, and randomness, while single-turn communication may limit expressiveness and limit agents' ability to coordinate or reach agreement.

- **Single-turn communication** restricts agents to one exchange before action, simplifying protocols for immediate decisions. This approach is particularly effective in scenarios where aggregation or ensemble strategies can be leveraged without requiring iterative coordination. For example, *More Agents Is All You Need* Li et al. (2024c) demonstrates that increasing the number of independently sampled agents—each generating a solution without inter-agent interaction—and then aggregating their responses through voting can significantly enhance performance in reasoning tasks, all without relying on multi-turn dialogue.
- **Multi-turn communication** allows iterative dialogues, enabling agents to iteratively exchange information, coordinate actions, and refine collective understanding. One prominent category is artifact generation, where agents collaboratively produce tangible outputs such as software, plans, or structured artifacts. These systems include frameworks for multi-agent software development, AutoML pipelines, design assistants, and recommender systems Hu et al. (2024a); AL et al. (2024); Park et al. (2023); Ahn & Sentis (2021); Fourney et al. (2024); Durante et al. (2024); Liu et al. (2024b); Trirat et al. (2024); Chen et al. (2023a; 2024g); Wang et al. (2024b); Kannan et al. (2023); Du et al. (2024). Another key focus is interactive reasoning, where agents engage in multi-round dialogues to solve complex problems, perform question answering, synthesize knowledge, or jointly explore hypotheses Lan et al. (2023); Händler (2023); Wu et al. (2023a); Tran et al. (2025b); Liu et al. (2025b); Xu et al. (2023c); Feng et al. (2024); Li et al. (2024a); Chen et al. (2024c); Jinxin et al. (2023); Wang et al.; Ghafarollahi & Buehler (2024); Estornell et al. (2025). In debate settings, agents may hold conflicting views or objectives and engage in argumentation to influence outcomes Xiong et al. (2023); Abdelnabi et al. (2024); Du & Zhang (2024); Jin et al. (2024); Li et al. (2024e); Wang et al. (2024c); Chen et al. (2023b); Yin et al. (2023); Baltaji et al. (2024); Tang et al. (2023). Finally, a significant body of work centers on coordination and execution, where agents operate in grounded

or embodied environments to perform spatial, temporal, or physical tasks, ranging from teamwork in simulated worlds to the orchestration of complex action sequences Sun et al. (2025c); Piatti et al. (2024); Wu et al. (2025); Cheng et al. (2024); Hua et al. (2023); Xu et al. (2023b); Zhuge et al. (2024a); Li et al. (2023c); Xu et al. (2023a); Wu et al. (2024a).

Note: Many works involve completing a long-term task that unfolds over multiple steps. While these settings may feature single-turn communication at each step, we still categorize them as multi-turn communication overall, since agents interact repeatedly throughout the episode.

Communication Content. Beyond how agents interact, what they communicate plays a critical role in coordinating action and achieving shared goals. Messages may contain commands, knowledge, feedback, or reflections—each type serving a unique function in shaping agent behavior and collective reasoning.

- **Structured and Unstructured Messages.** *Structured messages* use predefined formats like JSON to ensure clarity, reduce ambiguity, and enable reliable execution in tasks such as sub-task decomposition, coordination, or tool invocation. In contrast, unstructured messages convey rich, context-dependent information that supports more open-ended or creative collaboration. While structured formats favor precision and determinism, unstructured media allow for flexibility but require more sophisticated interpretation.

- **Content Type.** In LLM-based multi-agent systems, the content of communication varies by task and system goals, shaping how agents coordinate, reason, and adapt. *Task execution* messages relay goals, instructions, and sub-tasks to align agent responsibilities and progress, common in collaborative workflows like AutoGen Wu et al. (2023b) and GPT-in-the-Loop Nascimento et al. (2023a). *Knowledge exchange* involves sharing reasoning steps or insights to improve group understanding, as seen in CoMM Chen et al. (2024d). *Negotiation and strategy* content arises in competitive or mixed-motive settings, where agents negotiate goals or influence outcomes, exemplified by LLM-Deliberation Abdelnabi et al. (2023). *State updates and feedback* include performance reports or status updates used to refine coordination and maintain coherence, as in MetaGPT Hong et al. (2023) and ReConcile Chen et al. (2023b). *Persona maintenance* supports consistent agent identity and introspective reasoning, enabling stable role behavior in works like Conformity Baltaji et al. (2024) and Cognitive Synergy Wang et al..

Each content type plays a distinct role in shaping how agents interact, whether the focus is on direct task execution, strategic alignment, shared understanding, or maintaining a consistent identity. Together, these communication layers support robust and intelligent coordination in complex multi-agent systems.

5.2.5 Static and Dynamic Protocols

Multi-agent systems often balance fixed and flexible interaction patterns to handle diverse task demands and adapt to changing environments, aiming for efficiency, scalability, and responsiveness. A static communication protocol uses a predefined interaction flow in which agents follow a fixed sequence of steps and maintain consistent roles and message structures throughout a run. This design ensures repeatability and clear control over information flow, but can limit adaptability when tasks or inputs change. In contrast, a dynamic communication protocol allows agents to adjust their roles, communication paths, and coordination strategies in response to evolving conditions, intermediate results, or unexpected events. This flexibility supports more robust decision-making, more realistic social behaviors, and potentially better performance in open-ended or unpredictable scenarios.

Static communication protocol refers to interaction patterns where the communication flow and agent roles are predefined and fixed during execution. Agents follow structured sequences, often passing information through pipelines or staged workflows without adapting their messaging behavior. This design supports reproducibility and control, and is commonly used in chain-based frameworks Wu et al. (2022); Xiao et al. (2023), as well as in coordination or debate settings that rely on structured message exchanges in roundtable formats Li et al. (2024e). Many recent multi-agent systems adopt this approach across domains such as software development, scientific research, and machine learning pipelines Qian et al. (2024a); Schmidgall et al. (2025); Trirat et al. (2024). Others use fixed communication graphs or workflow templates to support scalable coordination in graph reasoning, code generation, or orchestration Qian et al. (2024b); Hu et al. (2024d); Yang et al. (2025). While effective for structure and efficiency, such protocols may limit adaptability in more dynamic or emergent settings.

Dynamic communication protocols allow agents to adjust how they interact and coordinate based on changing conditions and evolving information, making systems more flexible and adaptive than static ones. Many works apply dynamic protocols in role-playing, social interaction, or behavior simulation, where agents adapt conversations, maintain memories, and update responses as they interact Park et al. (2023); Xu et al. (2023b); Cheng et al. (2024); Zhang et al. (2023b); Dong et al. (2024); Bailis et al. (2024); Gong et al. (2023); Wu et al. (2023a); Guo et al. (2024c); Wu et al. (2025); Liu et al. (2024a); Kaiya et al. (2023); Händler (2023); Li et al. (2023c); Nascimento et al. (2023a); Piatti et al. (2024); Abdelnabi et al. (2024); Yin et al. (2023); Zheng et al. (2023); Tran et al. (2025b); Tang et al. (2023); Mandi et al. (2024); Baltaji et al. (2024); Sun et al. (2025c); Liu et al. (2025b); Wang et al. (2023c); Xu et al. (2023c); Nascimento et al. (2023b). Others focus on competitive games or benchmark scenarios, where agents dynamically plan and share information in real time Ma et al. (2023); Wang et al. (2023b); Light et al. (2023); Jin et al. (2024); Fourney et al. (2024). Many studies design dynamic protocols for human-agent collaboration, teamwork, planning, or structured debate, where agents coordinate steps, negotiate, or adjust group decisions Feng et al. (2024); Du et al. (2024); Du & Zhang (2024); Chen et al. (2023b); D'Arcy et al. (2024); Chen et al. (2024g); Ghafarollahi & Buehler (2024); Hu et al. (2024a); Zhuge et al. (2024a); Chen et al. (2024c); Zhang et al. (2024b;a); Chen et al. (2024f; 2023a); Jin et al. (2024); Händler (2023); Piatti et al. (2024). Frameworks and surveys describe general principles for flexible, multimodal, or large-scale agent systems with dynamic communication paths Sun et al. (2025c); Hu et al. (2024a). Finally, some works study general coordination, self-organization, or emergent societies of agents that operate without rigid communication patterns Liu et al. (2024b); Ishibashi & Nishimura (2024); Wang et al. (2024c); Liu et al. (2023a); Jinxin et al. (2023); Li et al. (2023c); Nascimento et al. (2023a); AL et al. (2024). Together, these approaches highlight how dynamic protocols keep multi-agent LLM systems robust and responsive in unpredictable settings.

5.3 Existing Frameworks

There are several popular frameworks designed to facilitate the development of LLM-powered multi-agent systems. These frameworks enable efficient implementation, evaluation, and customization of agents for diverse applications. Below is an overview of some notable frameworks, with an emphasis on their support for multi-agent communication.

Several frameworks have been developed to support LLM-based multi-agent systems, differing in their application domains and supported communication topologies. We organize them below by the primary use case and topology design, highlighting key tradeoffs.

Workflow-oriented frameworks typically adopt *centralized* or *hierarchical* communication topologies, where a central planner or manager agent decomposes user goals into sub-tasks and delegates them to sub-agents or tool-executing agents arranged in structured pipelines. This setup is well-suited for multi-step task automation, business process orchestration, software engineering, and machine learning pipelines, where modularity, role specialization, and top-down control are crucial for maintaining coherence and ensuring execution correctness. These systems often support flexible agent definitions but enforce strict execution order, which simplifies debugging and improves reliability in complex workflows lan (b); cre; Chen et al. (2023a); Fourney et al. (2024); Trirat et al. (2024); Hong et al. (2023); Qian et al. (2024a); Wang et al. (2024b); Ishibashi & Nishimura (2024); OpenAI; Xie et al. (2023).

Collaborative reasoning and decision-making frameworks favor *chain* and *shared message pool* topologies, enabling iterative communication among agents via a shared conversational history. Agents reason over each other's responses, critique proposals, and converge through multi-turn dialogue, debate, or consensus-building. This structure leverages the strengths of LLMs in contextual reasoning and is particularly effective for open-ended problem solving, logic, and question-answering tasks Li et al. (2023b); Chen et al. (2023b); Wu et al. (2023b); Zhuge et al. (2024a); Qian et al. (2024b); Tang et al. (2023).

Embodied and interactive environments often use *decentralized* topologies, where agents act autonomously based on local observations, possibly exchanging information or intentions through shared state

Table 9: Comparison of LLM Multi-Agent Frameworks

Framework	**Topology**					**Num Agents**	**Application-Domain**
	Chain	Hier.	Decent.	Cent.	Pool		
LangGraph lan (b)	✓	✓	✓	✓	✓	Flexible	General
CrewAI cre	✓	✓		✓		Flexible	General
OpenAI Swarm OpenAI	✓					Flexible	General
SuperAGI sup				✓		Flexible	General
CAMEL Li et al. (2023b)	✓				✓	2	General
AutoGen Wu et al. (2023b)	✓	✓		✓	✓	Flexible	Reasoning tasks
AgentVerse Chen et al. (2023c)		✓	✓		✓	Flexible	General
MetaGPT Hong et al. (2023)	✓	✓		✓		5	Software dev & coding
MegaAgent Wang et al. (2024b)		✓				590+	Software dev & coding; policy simulation
ProAgent Zhang et al. (2024a)			✓			2	Embodied - Overcooked
AgentCF Zhang et al. (2024b)			✓			Users & Items	Recommender
ChatDev Qian et al. (2024a)	✓					5	Software dev & coding
COLEA Zhang et al. (2023b)			✓			2	Embodied - WAH and TDW-MAT
MedAgents Tang et al. (2023)			✓		✓	3	Clinical reasoning tasks
ReConcile Chen et al. (2023b)			✓		✓	3	Reasoning tasks
AutoAgents Chen et al. (2023a)		✓		✓		Flexible	General task automation
DyLAN Liu et al. (2024b)			✓			Flexible	General
OpenAgents Xie et al. (2023)				✓		3	Data Analysis/API Tools/Web assistant
MACRec Wang et al. (2024f)			✓			5	Recommendation system
Magentic-One Fourney et al. (2024)				✓		5	General
AgentCoord Pan et al. (2024)	✓	✓	✓	✓	✓	Flexible	Visualization of coordination strategies
AutoML-Agent Trirat et al. (2024)				✓		5	ML pipeline (data→deployment)
GPT-Swarm Zhuge et al. (2024a)			✓			Flexible	Reasoning tasks
MACNet Qian et al. (2024b)	✓	✓	✓			Flexible	General
Theory-of-Mind Li et al. (2023c)		✓			✓	3	Customized text game (search and rescue)
Self-Organized Ishibashi & Nishimura (2024)			✓			Flexible	Software dev & coding
SMART-LLM Kannan et al. (2023)				✓		1 LLM + N robots	Multi-robot task planning

or signaling. These topologies align well with simulated and physical environments, where agents must plan, act, and adapt independently. Coordination, when needed, often emerges through spatial proximity or explicit collaboration protocols rather than central control Zhang et al. (2024a; 2023b); Li et al. (2023c); Kannan et al. (2023).

Personalization frameworks adopt *decentralized* agent models, where multiple agents independently analyze users, items, context, and preferences before aggregating their outputs. This structure allows for agent specialization and encourages interpretable, multi-perspective reasoning. Communication is typically lightweight, focused on exchanging embeddings, scores, or rationales Zhang et al. (2024b); Wang et al. (2024f).

General-purpose frameworks support a broad range of communication topologies, acting as flexible infrastructure for composing multi-agent systems. These platforms abstract away lower-level concerns such as message routing, memory sharing, and agent scheduling, while offering APIs or visual interfaces for building custom workflows. They are domain-agnostic and designed for extensibility, experimentation, and prototyping across varied applications lan (a); sup; Wu et al. (2023b); Chen et al. (2023c); Pan et al. (2024).

In summary, communication topologies in multi-agent LLM frameworks are shaped by task demands: *hierarchical and centralized* designs suit workflows; *chain and shared-message pool* topologies support iterative reasoning and dialogue; and *decentralized* designs enable autonomy and scalability in embodied or personalized settings. General-purpose platforms typically support flexible and large-scale agent composition, but this does not imply guaranteed performance. Scaling often introduces new coordination and efficiency challenges that require manual handling. While these frameworks enable diverse applications and demonstrate the expressive power of natural language communication among agents, they are best viewed as flexible design tools rather than reliable multi-agent solutions.

5.4 Applications

LLM-based multi-agent systems (LLM-MAS) are gaining increasing attention as researchers explore how LLMs can be used for collaborative and competitive agent systems. While many advancements have focused on single-agent applications, recent research highlights the potential for scaling LLMs to handle multiple agents. A recent survey on LLM-based multi-agents provides an overview of the progress and challenges

Category	Description	Representative Works
Coding	Systems designed to emulate or assist with software engineering processes. Agents may act as planners, coders, reviewers, or testers, collaboratively contributing to the development lifecycle. The focus is on multi-role coordination, code generation, task decomposition, and iterative refinement.	Qian et al. (2024a); Du et al. (2024); Besta et al. (2024); Ishibashi & Nishimura (2024); Zhang et al. (2023d); Schmidgall et al. (2025); Zhuang et al. (2023); Liu et al. (2024b); Trirat et al. (2024); Zhuge et al. (2024a); Chen et al. (2023a); Hu et al. (2024a); Pan et al. (2024); Islam et al. (2024); Li et al. (2024c)
Virtual environment	Systems where agents interact within simulated or narrative-based digital worlds. These environments often provide structured or open-ended tasks that require memory, planning, coordination, or social reasoning. Agents engage with evolving states, simulate behavior over time, and interpret complex virtual contexts.	Wu et al. (2024b); Park et al. (2023); Ma et al. (2023); Zhu et al. (2023); Hu et al. (2024c); Durante et al. (2024); Yao et al. (2022a); Fourney et al. (2024); Zhou et al. (2023); Deng et al. (2024); Chen et al. (2024g); Nascimento et al. (2023a); Gong et al. (2023); Tran et al. (2025b); Piatti et al. (2024); Dong et al. (2024); Zhang et al. (2024a;b); Bailis et al. (2024); Xie et al. (2023); Li et al. (2024b); AL et al. (2024); Chen et al. (2024c); Kaiya et al. (2023); Händler (2023); Jinxin et al. (2023); Liu et al. (2023a); Nascimento et al. (2023b); Light et al. (2023); Xu et al. (2023a); Li et al. (2023c); Abdelnabi et al. (2024); Wu et al. (2023a); Hua et al. (2023); Lan et al. (2023); Agashe et al. (2023a)
Embodied environment	Agents situated in physical or physically simulated environments. These agents perceive their surroundings, execute actions, and often learn to solve tasks grounded in spatial or temporal constraints. The emphasis is on embodiment, learning from feedback, and integrating language with physical action.	Wang et al. (2023a); Zhang et al. (2023b); Ahn et al. (2022); Kannan et al. (2023); Agashe et al. (2023b); Mandi et al. (2024); Liu et al. (2025b); Wu et al. (2025); Ahn & Sentis (2021); Guo et al. (2024c)
Human-Agent Workflows and Collaboration	Systems where agents communicate and coordinate with humans or each other using natural language to complete tasks, share knowledge, or simulate collaborative roles. These works emphasize dialog, reasoning, role-based behaviors, and decision support across interactive workflows.	Li et al. (2023b); Xu et al. (2023b); Liang et al. (2022); Xu et al. (2023c); Wang et al. (2024b); Zhang et al. (2024c); Yin et al. (2023); Xiong et al. (2023); Chen et al. (2024f); Li et al. (2024a); Ning et al. (2023); Wang et al. (2024f); Wu et al. (2022); Chen et al. (2024a); Stepputtis et al. (2023); Du & Zhang (2024); Jin et al. (2024); Feng et al. (2024); Wang et al. (2023b); Chen et al. (2023b); Mialon et al. (2023); D'Arcy et al. (2024); Yao et al. (2023); Li et al. (2024e); Wang et al. (2024c); Qiao et al. (2024); Suzgun & Kalai (2024); Liu et al. (2024a)

Table 10: Summary of LLM Multi-Agent System Applications by Domain

in this domain Guo et al. (2024a). It highlights the success of LLM-based agents in simulating complex environments and solving intricate problems through interaction. These systems offer new possibilities for multi-agent coordination and decision-making, particularly in dynamic and partially observable environments. LLM agents, leveraging their ability to understand and generate human language, are capable of operating across a wide variety of tasks, including computer tasks, gaming, coding, and real-world challenges Cheng et al. (2024); Guo et al. (2024a).

Coding Coding tasks involve automating software development processes, including writing, testing, and debugging code. LLM agents can collaborate in these workflows to reduce the time and effort required for development. Structured role-based systems assign agents to roles such as developer, tester, or reviewer, enabling natural language communication and iterative workflows that improve completeness and reliability Qian et al. (2024a). This paradigm has since been extended through larger-scale coordination, as multi-agent frameworks orchestrate several teams of agents working in parallel to broaden the solution space and deliver higher quality outcomes, while studies also suggest that simply increasing the number of agents enhances robustness through diverse exploration and ensemble effects Du et al. (2024); Li et al. (2024c). Other approaches remove fixed roles entirely and instead employ decentralized self-organizing swarms in which agents dynamically spawn and collaborate to handle increasing complexity, achieving superior scalability and accuracy in large-scale code generation Ishibashi & Nishimura (2024). Beyond general-purpose software projects, specialized multi-agent systems have shown effectiveness in domain-specific workflows, such as competitive programming, where retrieval, planning, coding, and debugging agents collectively outperform single models, and in automated machine learning pipelines, where agents collaborate on data preparation, model selection, and optimization Islam et al. (2024); Trirat et al. (2024). Collectively, these studies suggest that LLM-based agent collaboration can help reduce development time and improve software quality through parallelism, specialization, and iterative feedback, indicating a gradual move toward more automated forms of software engineering.

Virtual Environments Virtual environments provide rich scenarios for testing multi-agent communication and coordination. In sandbox simulations like Smallville, dozens of generative agents use natural language to interact, share information, and spontaneously organize social activities, producing human-like group behavior Park et al. (2023); Kaiya et al. (2023). Similarly, social deduction games such as Avalon and Werewolf have been repurposed as evaluation arenas, where multiple LLM agents must converse, plan, and occasionally deceive each other to achieve collaborative goals, highlighting both emergent coordination and notable gaps in strategic communication Lan et al. (2023); Light et al. (2023); Bailis et al. (2024). To systematically benchmark these capacities, researchers have introduced frameworks like LLMArena, which encompasses a suite of multi-agent game environments testing skills from spatial reasoning to teamwork and dialogue, alongside specialized coordination challenges that expose current limitations in opponent modeling and cooperative planning Chen et al. (2024c); Agashe et al. (2023a). Larger-scale simulations further illustrate the potential of virtual settings as multi-agent testbeds: one approach simulates a full hospital with LLM-powered doctors, nurses, and patients engaging in realistic medical dialogues, while another deploys hundreds of agents in a Minecraft-based society that demonstrates emergent division of labor, evolving rules, and collective decision-making over time Li et al. (2024b); AL et al. (2024); Wang et al. (2023a); Zhu et al. (2023); Wu et al. (2024b). Across these studies, virtual environments serve as scaffolding for observing and evaluating how LLM-based agents communicate, coordinate, and negotiate in pursuit of shared tasks, providing controlled yet complex arenas to probe emergent collaboration and planful dialogue.

Embodied Environments Embodied or physically simulated environments have become important testbeds for examining multi-agent coordination, perception, and situated language use with LLM agents. By coupling models with perceptual inputs and action interfaces, agents can interpret instructions in context and execute tasks cooperatively, exhibiting emergent coordination such as maintaining internal world models, reasoning about partners' intentions, and synchronizing plans through natural language Ahn et al. (2022); Agashe et al. (2023b); Wu et al. (2025); Zhang et al. (2023b); Liu et al. (2025b). These settings provide grounding in shared environments where dialogue refers to physical objects, locations, and actions, supporting situated communication and cooperative planning. Some works employ inter-agent dialogue where each agent has an LLM-based controller, enabling planning, perception, and cooperation with human-like communication patterns. Studies also explore centralized versus decentralized planning, showing that a single planner can orchestrate small teams whereas larger groups benefit from distributed reasoning Zhang et al. (2023b); Mandi et al. (2024); Kannan et al. (2023); Chen et al. (2024g); Dong et al. (2024); Guo et al. (2024c). Benchmarks in environments such as Overcooked test inference of partner goals, proactive assistance, and situated reasoning, with results indicating robust coordination and adaptability even with unfamiliar partners Agashe et al. (2023b); Zhang et al. (2024a). Across these studies, a recurring theme is the use of natural language as the medium for planning and communication, which enables division of labor, negotiation of task roles, and interpretable collaboration strategies Mandi et al. (2024); Agashe et al. (2023b); Guo et al. (2024c).

Human-Agent Workflows and Collaboration LLM-based agents are increasingly designed to collaborate with humans in complex tasks, emphasizing alignment with user intentions. Fully autonomous agents often struggle to adapt to dynamic scenarios or capture human needs without guidance, motivating research into cooperative systems Li et al. (2023b); Feng et al. (2024). One strategy is to allow multiple agents to coordinate with minimal human input, in which agents simulate user and assistant roles to iteratively prompt each other toward a task while maintaining consistency with goals Li et al. (2023b). Some approaches keep humans in the loop by providing interfaces for decomposing complex tasks into modular steps and adjusting intermediate outputs Wu et al. (2022). Other methods learn policies for inserting human interventions at optimal moments during reasoning, showing that limited but well-timed guidance can substantially improve task success Feng et al. (2024). Persona-based agents simulate diverse roles to support more natural, adaptive, and personalized interactions, while systems focusing on cultural grounding improve cross-cultural communication and alignment Chen et al. (2024a); Li et al. (2024a). Multi-agent communication has been explored in team-based environments such as social deduction games, where dialogue, deception, and long-horizon coordination serve as challenging tests of interaction quality Xu et al. (2023b); Stepputtis et al. (2023); Jin et al. (2024); Du & Zhang (2024). Other studies examine structured debates or round-table discussions, showing that communication among agents can improve reasoning consistency and solution quality through cross-verification Xiong et al. (2023); Chen et al. (2023b).

5.5 LLM-Powered Multi-Agent Systems and Multi-Agent Reinforcement Learning

Integrating LLMs into MARL presents a promising direction for building more adaptive, communicative, and intelligent agent systems. LLMs bring general-purpose reasoning, task flexibility, and natural language fluency to agent interactions, traits that are particularly valuable in dynamic and open-ended environments. These capabilities allow agents to negotiate, coordinate, and adapt without needing extensive retraining for every new scenario. However, LLMs also introduce challenges, such as the risk of hallucinated outputs and inconsistencies in decision-making, especially in high-stakes or highly domain-specific tasks. In contrast, RL offers strong advantages in precision and policy optimization. Through trial-and-error learning, RL agents can develop highly effective strategies aligned to a given environment's dynamics and constraints. These policies tend to be reliable and efficient once trained but require large volumes of data and computation to generalize across new tasks. Moreover, traditional RL agents typically lack the capacity for human-interpretable communication, which limits their effectiveness in settings that require explanation, persuasion, or cooperation with humans. By combining these approaches, LLMs and RL can complement one another. LLMs provide fast task adaptation, intuitive language-based communication, and general reasoning, while RL offers stable, optimized decision policies grounded in the environment. When integrated, these models enable agents to both communicate strategically and act reliably, unlocking new capabilities in multi-agent systems. Several systems exemplify this hybrid approach.

Cicero (FAIR) was developed for the game Diplomacy, which combines strategy with rich social interaction. It uses an LLM to generate fluent, persuasive language and integrates MARL for policy learning. This enables agents to form alliances, negotiate effectively, and plan multi-turn strategies. The key contribution of Cicero is its ability to align language generation with long-term game objectives, bridging communication and decision-making in a complex, adversarial setting.

Werewolf Xu et al. (2023c) provides a contrasting case by focusing on deception, hidden information, and group dynamics. The game requires agents to infer others' identities and intentions based on limited cues, a task well-suited to the nuanced reasoning capabilities of LLMs. Here, language is not only used for cooperation but also for misdirection. Agents must generate plausible lies, detect inconsistencies, and adapt their strategies accordingly. The integration of RL techniques allows agents to learn survival and persuasion strategies that evolve through repeated play.

FAMA (Functionally-Aligned Multi-Agents) Slumbers et al. (2023), which explicitly harnesses LLMs for co-operative decision-making in textual environments. FAMA aligns an LLM with the necessary task knowledge via a centralized MARL training update, and it leverages the model's linguistic strengths for natural language inter-agent message-passing. FAMA improves coordination on complex tasks by fine-tuning LLMs with coordinated objectives and enabling them to exchange messages in plain language. Experiments in environments like BabyAI-Text and an autonomous driving junction show that this approach consistently outperforms both independent LLM agents and traditional symbolic RL baselines.

ACC-Collab Estornell et al. (2025) uses an actor-critic paradigm to train collaboration between two LLM-based agents. Instead of relying on pre-trained models to cooperate on the fly in a multi-round actor-critic fashion for question answering, ACC-Collab jointly trains two specialized agents through iterative dialogue. In this setup, one LLM plays the *actor*, proposing solutions for a given task, while another LLM acts as the *critic*, providing feedback to refine the actor's answers over multiple conversation rounds.

Together, these examples show how LLMs can be used not just for dialogue but for deeply strategic, multi-agent interaction—especially when combined with the grounded learning processes of RL. This hybrid approach opens new possibilities for multi-agent systems that need to reason, plan, and communicate in uncertain and socially complex environments.

5.6 Evaluation of LLM-MAS

Evaluating LLM-based multi-agent systems (LLM-MAS) is inherently challenging due to the complexity of both components. On one hand, assessing the behavior of multi-agent systems requires understanding group dynamics, coordination efficiency, and collective reasoning, none of which are trivial to quantify. On the other hand, LLMs are large neural network based models whose responses are difficult to interpret and evaluate consistently, especially when operating in dynamic or open-ended settings. When these two paradigms are combined, the difficulty compounds: LLM agents may exhibit unpredictable or stochastic behaviors, and their interactions can lead to complex, emergent patterns that are hard to measure with standard metrics. A growing set of benchmarks has been developed to test a range of capabilities across planning, embodiment, social inference, open-ended behavior, and large-scale coordination. These benchmarks vary in their observation space, action structure, and evaluation criteria, making comparative assessments both challenging and necessary.

Table 11: Multi-agent environment benchmarks and their characteristics.

Environment	#Agents	Observation Space	Action Space	Evaluation Method
TextStarCraft II Ma et al. (2023)	2	partial; text-based	discrete (macro/micro text commands)	win rate; population ratio; resource ratio
LLM-Coordination Agashe et al. (2023b)	2	partial; text-based	discrete (text-based commands)	score; ToM accuracy; planning accuracy
C-WAH and TDW-MAT Zhang et al. (2023b)	2	partial; RGB-D	discrete (navigate; interact; communicate)	task completion rate; efficiency
AI2-THOR Kannan et al. (2023)	4	partial; symbolic	discrete (primitive API calls)	success rate; task completion; goal recall
GAIA Mialon et al. (2023)	no limit	partial; multi-modal)	discrete (tool use; browse; reason)	answer accuracy
BoxNet, Warehouse, BoxLift Chen et al. (2024g)	32	full; symbolic; grid-based	discrete (move; pick; place)	success rate; token usage; steps taken
CuisineWorld Gong et al. (2023)	4	full; text-based	discrete (move; get/put; use tool)	task score; efficiency
RoCoBench Mandi et al. (2024)	3	partial; symbolic	discrete (move; pick; place)	success rate; coordination
GOVSIM Piatti et al. (2024)	5	partial; text-based	discrete	survival rate; efficiency; equality
VillagerBench Dong et al. (2024)	3	partial; text-based	discrete (move; place; craft; manage)	completion rate; efficiency; workload balance
Smart Streetlight IoT Nascimento et al. (2023a)	100	partial; simulated sensory inputs	discrete (switch; adjust; communicate)	fitness score; latency; energy usage
Werewolf Arena Bailis et al. (2024)	8	partial; text-based	discrete (vote; act; communicate)	win rate; strategy quality; gameplay metrics
GraphInstruct Hu et al. (2024d)	no limit	partial; graph-based (local neighborhood)	discrete (message passing)	task accuracy
Agent Hospital Li et al. (2024b)	100	partial; text-based	discrete (diagnose; prescribe; examine)	diagnosis accuracy; MedQA score

Environment	#Agents	Observation Space	Action Space	Evaluation Method
LLMArena Chen et al. (2024c)	5	partial; text-based	discrete (game moves; communicate)	TrueSkill score; win rate; skill metrics
BOLAA Liu et al. (2023a)	3	partial; text-based	discrete (web navigation; query tools)	final reward; intermediate recall
Avalon (Role Identification) Stepputtis et al. (2023)	6	partial; text-based	discrete (vote; propose; communicate)	win rate; role identification accuracy
SwarmBench Ruan et al. (2025)	no limit	partial; grid-based	discrete (move; communicate)	success rate; coordination metrics
Collab-Overcooked Sun et al. (2025a)	2	partial; grid-based	discrete (move; pick; place; interact)	dishes delivered; efficiency
BattleAgentBench Wang et al. (2024d)	4	partial; grid-based	discrete (move; shoot)	mission progress; action accuracy; mission score
PARTNR Chang et al. (2024)	2	partial; 3D	discrete (navigate; manipulate; instruct)	success rate; steps vs human; error recovery
MultiAgentBench Zhu et al. (2025)	no limit	partial; text-based	discrete (task-dependent actions; communicate)	task completion; milestone KPIs; collaboration quality
MindCraft Bara et al. (2021)	2	partial; 3D	discrete (move; build; communicate)	ToM accuracy; task success
MineLand Yu et al. (2024)	48	partial; 3D	discrete (move; gather; build; communicate)	survival rate; resource collection; social dynamics
MeltingPot Agapiou et al. (2022)	8	partial; multi-modal	discrete (move; tag; collect)	episode return; generalization to novel partners

A wide range of benchmarks have been proposed to evaluate LLM-based multi-agent systems, each targeting distinct capabilities. **Planning and collaboration** tasks emphasize joint planning and cooperative execution in structured environments Sun et al. (2025a); Chen et al. (2024g). Agents typically share goals and operate in discrete action spaces, with performance measured by metrics like task completion, delivery rate, or step efficiency. **Embodied interaction and navigation** benchmarks such as Kannan et al. (2023); Zhang et al. (2023b) place agents in physically simulated environments with rich observations and symbolic or low-level action spaces, where success hinges on spatial coordination, manipulation, and robust perception. **Social inference and communication** environments like Werewolf Arena and Avalon Bailis et al. (2024); Stepputtis et al. (2023) test agents' theory-of-mind, dialogue reasoning, and hidden-role understanding, typically evaluating win rates, strategy diversity, or belief modeling. **Long-horizon or open-ended environments** such as MindCraft and MineLand Bara et al. (2021); Yu et al. (2024) challenge agents with evolving objectives in sandbox settings, requiring sustained planning, adaptation, and multi-step reasoning over extended interactions. **Large-scale swarm coordination** is studied in domains like SwarmBench and Smart Streetlight IoT Ruan et al. (2025); Nascimento et al. (2023a), where agents operate with partial observability and limited communication, focusing on emergent cooperation and decentralized control. While these benchmarks collectively span diverse interaction types and system scales, existing evaluation metrics often fall short in fully capturing the complexity, scalability, and qualitative nuances of agent communication and coordination Zhu et al. (2025); Chen et al. (2024c).

Despite the breadth of available benchmarks, current evaluation methodologies still face limitations. Many existing metrics are task-specific and often focus on one-time outcomes at the end of an episode, such as task completion or win rate, which provide limited insight into the underlying decision-making process or agent adaptation over time. These evaluations fail to capture the evolving dynamics of agent communication, strategic adjustments, or learning during interactions. Moreover, quantifying multi-agent behavior—such as cooperation, coordination efficiency, or division of labor—is inherently difficult, as such phenomena are often emergent, context-dependent, and qualitative in nature. Scalability also remains a challenge: many benchmarks are designed for a fixed task difficulty and do not scale meaningfully with the number of agents, making it hard to evaluate systems intended for large-scale collaboration. As LLM-MAS research continues to grow, developing more fine-grained, temporally aware, and scalable evaluation protocols will be critical for assessing real-world utility, coordination capabilities, and generalization across diverse tasks.

5.7 Discussion

The integration of large language models into multi-agent reinforcement learning presents a powerful paradigm shift, offering new possibilities for coordination, communication, and adaptability in complex environments. By combining the expressive capabilities of LLMs with the optimization strengths of reinforcement learning, these systems are better equipped to handle tasks that require both flexible reasoning and grounded decision-making. LLMs contribute by enabling natural language communication, fast adaptation to new instructions, and general-purpose problem-solving, while reinforcement learning provides stability, consistency, and efficiency in executing task-specific policies. Together, they form a complementary framework that addresses the limitations of each individual approach.

Despite these advantages, several challenges remain. Future work will need to address communication efficiency, ensure real-time adaptability, and mitigate the risk of unreliable outputs from LLMs. As multi-agent systems scale, optimizing decentralized coordination and reducing message overhead will be critical. Moreover, the growing role of human-AI interaction highlights the need for communication protocols that are not only interpretable to machines but also intuitive for humans. These developments also bring new opportunities—from creative collaboration and embodied interaction to simulation, training, and decision support in high-stakes settings. Ensuring ethical, transparent, and responsible deployment will be essential as these systems become more autonomous and influential across diverse domains.

6 Discussion, Open Problems, and Future Directions

In this section, we provide an in-depth discussion on multi-agent communication (**MARL-COMM**), emergent language in multi-agent systems (**MA-EL**), and multi-agent communication with large language models (**LLM-COMM**). These discussions expand on the main text and outline key challenges, methodologies, and future directions in each area. Our analysis focuses on how communication has been modeled, learned, and optimized within multi-agent systems, providing a comprehensive summary of this paper while offering insights for future research. For clarity, we use the following abbreviations: **MARL-COMM** - Multi-Agent Reinforcement Learning Communication, **MA-EL** - Emergent Language, and **LLM-COMM** - Multi-Agent Communication with Large Language Models.

6.1 Discussions on Multi-agent Communications

A Common Usage of Communication Models in Multi-Agent Systems Communication models play a crucial role in multi-agent systems by enabling agents to share information, coordinate actions, and enhance decision-making. These models can be broadly categorized based on their structure and learning mechanisms, and they are commonly used in MARL-COMM, MA-EL, LLM-COMM. In MARL, communication models facilitate cooperation by allowing agents to exchange observations, share rewards, or align policies. They can be integrated into centralized learning frameworks, where a global controller aggregates and distributes information, or into decentralized architectures, where agents autonomously decide whom and when to communicate with. To be specific, the communication process can be formulated as $m_{i,j} \sim C(a_i, a_j|s_t), \quad a_i \sim \pi_i(\cdot|s, m_{i,j}), \quad m_{i,j} \in M$, where $m_{i,j}$ represents the message exchanged between agent i and agent j, $C(a_i, a_j|s_t)$ denotes the communication function that determines the message exchange conditioned on state s_t, and $\pi_i(\cdot|s, m_{i,j})$ represents the policy of agent i based on the received messages. All communication models exhibit unique advantages and trade-offs: (1). **Graph-based communication** encodes agent interactions using predefined or dynamic graphs, allowing for structured message passing but requiring scalability optimizations. (2). **Attention-based communication** uses transformer-based mechanisms to selectively attend to relevant information, improving robustness but increasing computational overhead. (3). **Emergent communication** enables agents to learn their own protocols, enhancing adaptability but often lacking interpretability. (4). **LLM-based communication** leverages pretrained language models to enhance expressiveness but may suffer from grounding issues in task-specific settings. Each of these models provides different capabilities depending on the environment and the desired level of communication efficiency. For instance, graph-based models are well-suited for structured multi-agent interactions, whereas LLM-based approaches excel in human-agent communication and natural language coordination. Future research should focus on integrating adaptive communication mechanisms that dynamically adjust based on task complexity, scalability requirements, and real-time constraints.

The unique usages of each communication model in multi-agent systems Each communication model has distinct advantages and is suited for different multi-agent scenarios. (1) **Graph-based communication** structures agent interactions as a communication graph, where nodes represent agents and edges determine message passing. This structured representation is effective for tasks requiring **topology-aware coordination**, such as traffic control and distributed sensor networks. Graph-based models can be either **static**, where communication links are predefined, or **dynamic**, where agents learn whom to communicate with based on relevance and necessity. (2) **Attention-based communication** utilizes mechanisms such as transformers to allow agents to selectively attend to the most relevant messages, improving scalability and robustness. By dynamically weighing information from different sources, attention-based communication enhances **coordination efficiency** in large-scale multi-agent systems while reducing redundancy in message exchange. (3) **Emergent communication** enables agents to develop their own communication protocols through interaction, rather than relying on predefined language or rule-based exchanges. This type of communication is particularly useful in *cooperative tasks* where agents need to establish novel coordination strategies in *zero-shot or few-shot scenarios*. However, ensuring *compositionality and interpretability* remains a challenge. (4) **LLM-based communication** integrates large language models to facilitate natural language-based interactions among agents, making it particularly valuable for *human-agent collaboration* and *generalist AI systems*. LLMs can serve as a *high-level reasoning module*, allowing agents to infer intentions, interpret ambiguous instructions, and align communication with human-like structures. However, challenges include *grounding in task-specific contexts* and *reducing hallucinations in generated messages*. (5) **Adaptive communication policies** leverage reinforcement learning to optimize *when and whom to communicate with*, reducing unnecessary message exchange and improving bandwidth efficiency. These policies are crucial in *scalable multi-agent reinforcement learning (MARL)*, where excessive communication can lead to *overhead and inefficiency*. Future research should explore *hybrid communication architectures* that combine structured message-passing mechanisms with *learned communication strategies*, ensuring scalability, adaptability, and robustness across different multi-agent domains.

Integrated use of different communication models in multi-agent systems The unique usages and advantages of each communication model form the basis for integrating different approaches in multi-agent communication, and several works have explored this direction. Here, we highlight some key examples. (1) Graph-based and attention-based communication can be combined to leverage *structured message passing* while dynamically *selecting relevant information*, enhancing scalability in large-scale multi-agent coordination. (2) Emergent communication can be integrated with reinforcement learning to allow agents to *learn communication protocols* while optimizing their policies, enabling more *adaptive and decentralized interactions* in complex environments. (3) LLM-based communication can be used as a *high-level reasoning mechanism*, complementing low-level MARL-based message exchange to enable *human-interpretable and goal-oriented* communication. (4) Hybrid models incorporating both explicit and implicit communication strategies allow agents to *switch between direct messaging and behavior-based signaling*, improving efficiency in tasks requiring limited bandwidth or privacy constraints. (5) Adaptive communication policies can be incorporated into any of these models, allowing agents to *dynamically decide when and whom to communicate with*, balancing message complexity and coordination effectiveness. Future research should further explore the integration of *structured, learned, and language-based communication*, ensuring that multi-agent systems remain robust, interpretable, and scalable across diverse environments.

A summary of the base communication models in multi-agent systems The different communication models used in multi-agent systems can be categorized based on their structure and learning mechanisms. (1) In **MARL**, communication strategies can be broadly classified into *explicit* and *implicit* communication. Explicit communication methods include *graph-based* and *attention-based* models, where agents directly exchange messages to enhance coordination. Implicit communication, on the other hand, relies on agents learning to *infer information from observed behaviors*, reducing bandwidth but requiring stronger inference capabilities. (2) In **emergent communication**, agents develop their own structured protocols through interaction, often optimizing for *task-specific coordination* while balancing *generalization and interpretability*. Research has explored different learning objectives, such as maximizing *mutual information*, reinforcement learning rewards, or structured learning constraints to encourage *compositionality and efficiency*. (3) In **large language model-based communication**, pretrained models facilitate agent interactions through *human-interpretable messages*, allowing for seamless integration of *natural language understanding* in multi-agent collaboration. However, ensuring *grounding and alignment* with task-specific requirements remains an open challenge. (4) **Hybrid communication architectures** combine multiple models, such as integrating *learned attention mechanisms* with *structured graph-based message passing* or combining *emergent communication* with *LLM reasoning modules*. These approaches aim to balance *interpretability, efficiency, and scalability*. (5) Finally, **adaptive communication policies** allow agents to *dynamically decide when and whom to communicate with*, reducing unnecessary message exchange while maintaining *effective coordi-*

nation. Future research should focus on refining the balance between *structured and learned communication approaches*, ensuring that multi-agent systems achieve *robust, interpretable, and scalable communication strategies* across diverse applications.

Seminal works of communication models in multi-agent systems The development of communication models for multi-agent systems has followed diverse trajectories, with some models seeing greater adoption in certain domains than others. The impact of a communication model often depends on whether there are seminal works in the category, as many subsequent research efforts extend these foundational studies. Here, we highlight seminal contributions in this area. **Graph-based communication** has been widely studied in reinforcement learning contexts, where agent interactions are modeled as graph structures, and information is propagated through message passing mechanisms. Many works formalize communication as a function over a graph structure:$h_i^{(t+1)} = \phi\left(h_i^{(t)}, \sum_{j\in\mathcal{N}(i)} \psi(h_j^{(t)}, m_{j,i})\right)$, where $h_i^{(t)}$ is the hidden state of agent i at timestep t, $m_{j,i}$ is the message received from neighboring agent j, and ϕ and ψ are learned transformation functions. **Attention-based communication** leverages transformer-based architectures to allow agents to dynamically weigh the importance of messages. This mechanism is particularly effective in large-scale systems where message relevance changes based on task demands. The communication process can be modeled using attention weights: $\alpha_{i,j} = \frac{\exp(f(h_i,h_j))}{\sum_{k\in\mathcal{N}(i)}\exp(f(h_i,h_k))}$, where $\alpha_{i,j}$ represents the attention weight assigned by agent i to agent j, and $f(h_i, h_j)$ is a scoring function that determines the relevance of agent j's message. **Emergent communication** models have introduced new paradigms in reinforcement learning, where agents develop their own symbolic language through optimization. Research in this area often seeks to maximize the mutual information between the transmitted and received messages: $\max_{\pi_C} I(m; s) = H(m) - H(m|s)$, where m represents the communicated message and s is the observed state. The challenge in emergent communication lies in ensuring compositionality, interpretability, and generalization across different agent populations. **LLM-based communication** introduces pretrained language models as a foundation for natural communication between agents. Large language models provide agents with rich semantic representations, enabling them to process and generate messages that align more closely with human language. However, challenges such as grounding messages in task-relevant information remain critical: $p(m|s, a) = \frac{\exp(f(s,a,m))}{\sum_{m'}\exp(f(s,a,m'))}$, where $p(m|s, a)$ represents the probability of generating message m given state s and action a, and $f(s, a, m)$ is a learned function capturing relevance. Some seminal works have pioneered new paradigms in multi-agent communication. Graph-based communication has been foundational in distributed reinforcement learning, while emergent language research has opened up avenues for self-learned symbolic exchanges. The use of large language models represents a shift toward integrating structured natural language into agent communication. Future research should continue exploring the intersection of these methods, leveraging **hybrid architectures** that integrate structured message-passing with adaptive learning-based communication strategies.

A summary of the issues and extensions of communication models in multi-agent systems Communication models in multi-agent systems address various challenges and enable extensions that improve scalability, adaptability, and robustness. Below, we categorize key issues these models aim to solve, along with potential research directions for further improvements. Specifically, six major extensions have been explored: **multi-task (MT)**, **multi-agent (MA)**, **hierarchical (Hier)** learning, **model-based (MB)** communication, *policy safety*, and *policy generalization*. While significant progress has been made, many challenges remain unsolved, requiring further exploration. (1) **Communication efficiency and bandwidth constraints**: Many multi-agent systems operate under limited communication resources, requiring models to minimize message exchange while maintaining effective coordination. Graph-based and attention-based communication mechanisms mitigate unnecessary transmissions by dynamically selecting relevant information. The effectiveness of these approaches is often evaluated through message entropy and communication sparsity: $\min_{\pi_C} H(m|s, a) \quad \text{subject to} \quad I(m; s, a) \geq \lambda$, where $H(m|s, a)$ represents the entropy of transmitted messages given state-action pairs, ensuring minimal redundancy while maintaining sufficient information flow through a mutual information constraint $I(m; s, a)$. (2) **Multi-modal and heterogeneous input**: Agents in real-world scenarios must process multiple input modalities, such as vision, language, and structured data. Communication models designed for *multi-modal fusion* use attention-based mechanisms to integrate these diverse signals. A common approach is to embed multiple modalities into a shared latent space before generating communication messages: $z = f_{\text{modality}}(s_1, s_2, \ldots, s_n), \quad m = g(z)$, where f_{modality} encodes inputs from different modalities $s_1, s_2, \ldots, s_n$, and g generates a structured message m. (3) **Generalization to new partners and environments**: Ensuring that communication models generalize beyond their training conditions is an ongoing challenge. Emergent communication systems often struggle with partner adaptation, requiring mechanisms to align agent language representations dynamically. One approach involves optimizing for alignment loss between agents: $\min_{\pi_C} D_{\text{KL}}(p(m_i|s)||p(m_j|s)), \quad \forall i, j \in \mathcal{A}$, where D_{KL} is the Kullback-Leibler divergence ensuring that message distributions between agents i and j remain consistent. (4) **Scalability in large-scale multi-agent communication**: As the number of agents increases, traditional communication protocols face scalability issues due to excessive message overhead. Attention-based models address this by dynamically filtering irrelevant messages. Scalable communication is often formalized as a minimization problem over message complexity while maintaining task performance: $\min_{\pi_C} \sum_{i=1}^{N} |m_i|, \quad \text{subject to} \quad J(\pi) \geq J_{\text{threshold}}$, where $|m_i|$ represents the message size of agent i, and $J(\pi)$ is the expected return of the multi-agent policy, constrained to ensure effective coordination. (5) **Grounding and interpretability in LLM-based communication**: Large language models introduce challenges in aligning generated messages with task-relevant information. Ensuring messages remain *grounded* in the environment requires integrating reinforcement learning with human feedback mechanisms: $\max_{\pi_C} \mathbb{E}[r_{\text{grounded}}(m, s, a)]$, where $r_{\text{grounded}}(m, s, a)$ measures how well the message m aligns with meaningful environmental information. Future research should focus on **hybrid architectures** that combine structured message-passing with adaptive learning, ensuring that multi-agent communication is both *efficient and interpretable* across diverse applications.

Foundation models, Algorithms, and Data Among all the communication models in multi-agent systems, transformer-based and LLM-driven communication approaches have demonstrated some of the most promising results. (1) Many algorithms leveraging attention-based communication have achieved success in *complex multi-agent coordination tasks*, including real-world applications such as robotic teamwork, autonomous vehicle swarms, and large-scale multi-agent competitions. (2) The use of *transformers enables agents to process long-range dependencies* in communication, making them highly effective for handling *sequential interactions and multi-modal inputs*. Agents leveraging transformer-based communication can dynamically adjust message relevance, allowing for efficient collaboration in dynamic environments. (3) In particular, large-scale foundation models such as GPT-based architectures have been explored for *multi-agent dialogue and cooperative problem-solving*, where a single model can adapt to multiple domains, from *robotic task execution* to *human-agent interaction*. Recent research has explored the integration of *pretrained language models* with reinforcement learning for communication optimization, demonstrating *cross-domain generalization effects*. The ability to fine-tune large models for *multi-agent reasoning* suggests a paradigm shift towards data-driven communication learning, where pretrained models provide agents with *rich semantic representations* for structured coordination. However, despite these advancements, foundation models introduce new challenges, such as *computational efficiency, real-time adaptability, and grounding in task-specific environments*. The question remains whether future development should be *data-driven* (leveraging large-scale pretraining) or *algorithm-driven* (designing more efficient communication-specific architectures). Regardless, future research must emphasize *scalability and computational efficiency*, ensuring that communication models remain *robust, interpretable, and deployable in real-world multi-agent scenarios*.

6.2 Perspectives on Future Directions

Based on the discussions above and the main content in previous sections, we present some perspectives on future research directions in multi-agent communication. We categorize these directions into four key areas: theoretical models, algorithm developments, general benchmarks, and human-centric research. We believe that more open challenges can be identified from the detailed discussions in earlier sections, and we have included specific comments on future work within the relevant sections. Each of these categories represents a crucial aspect of advancing multi-agent communication, ranging from foundational theoretical advancements to practical considerations in real-world deployment.

6.2.1 Future Research on Theoretical Guarantees

Most works on multi-agent communication focus on algorithm design and empirical evaluation rather than theoretical analysis, highlighting a need for deeper theoretical research in this field. One promising direction is the formal derivation of **convergence guarantees and performance bounds** for key communication frameworks, such as attention-based message passing and emergent communication protocols. Establishing rigorous mathematical foundations will enable researchers to design communication strategies that are both scalable and provably efficient.

Theoretical Guarantees for MARL Communication Most MARL communication studies focus on learning-based methods that improve coordination, but theoretical results on optimality, convergence, and performance bounds remain scarce. A fundamental question is: *how much does communication improve the return in cooperative MARL*, and *what is the theoretical gap between policies with and without communication?* Given a cooperative MARL setting, where $J^*(\pi)$ is the expected return under an optimal fully centralized policy π^*, and $J(\pi_C)$ is the expected return under a communication-constrained policy π_C, the return gap can be defined as:

$$\Delta J = J^*(\pi) - J(\pi_C), \tag{2}$$

where ΔJ quantifies the loss incurred due to limited communication. A major research direction is deriving upper bounds for ΔJ, showing under what conditions communication policies can approach optimality. Additionally, scalability and communication efficiency are key concerns in large-scale MARL. Many existing methods use graph-based or attention-based message passing, but the theoretical trade-offs between bandwidth constraints, coordination performance, and sample efficiency remain unclear. A useful framework is minimizing communication costs while ensuring bounded performance loss:

$$\min_{\pi_C} \sum_{t=1}^{T} C_t, \quad \text{subject to} \quad J^*(\pi) - J(\pi_C) \leq \epsilon, \tag{3}$$

where C_t represents communication costs at time t and ϵ is the maximum allowable performance degradation. Future work should develop formal criteria for when and whom agents should communicate with to optimize efficiency. Finally, robustness in adversarial settings is a crucial but underexplored theoretical problem. Communication channels can be subject to noise, delays, or adversarial tampering. Establishing worst-case bounds on return degradation due to message corruption is essential:

$$J(\pi_C, \eta) = \mathbb{E}\left[\sum_{t=1}^{T} r_t \mid m_t = f(m_t^*, \eta)\right], \tag{4}$$

where m_t^* is the ideal message, η represents a perturbation or adversarial noise, and $f(m_t^*, \eta)$ is the received message. Theoretical guarantees on error correction, redundancy mechanisms, and robustness will be necessary for safe deployment of communication-based MARL.

Theoretical Analysis of Emergent Language (EL) Communication In emergent communication, agents develop their own protocols through interaction rather than being explicitly programmed. However, understanding why certain languages emerge, how they generalize, and how to ensure their efficiency remains a theoretical challenge. A major open problem is quantifying the expressiveness and compositionality of emergent languages. Current emergent protocols often lack systematic structure and interpretability, leading to poor generalization across tasks and partners. One research direction is modeling emergent language as an information bottleneck problem:

$$\max_{\pi_C} I(m; s) \quad \text{subject to} \quad H(m|s) \leq \delta, \tag{5}$$

where $I(m; s)$ measures the mutual information between message m and state s, while $H(m|s)$ ensures that messages remain concise and interpretable. Establishing conditions under which *compositionality emerges naturally* will be crucial for developing generalizable emergent languages. Another key theoretical question is the *stability and convergence of learned languages*. Many current methods rely on reinforcement learning to evolve communication, but the non-stationarity of language evolution can lead to instability. Understanding the convergence properties of emergent protocols and the conditions under which languages remain stable is a crucial research area. Lastly, *partner adaptation* remains an open challenge in emergent communication. Most current models assume a fixed agent population, but real-world multi-agent systems must interact with new, unseen partners. Theoretically, ensuring *zero-shot generalization* across different agent communities requires defining formal notions of language adaptability:

$$\min_{\pi_C} D_{\mathrm{KL}}(p(m_i|s)||p(m_j|s)), \quad \forall i, j \in \mathcal{A}, \tag{6}$$

where D_{KL} represents the Kullback-Leibler divergence between the language distributions of different agents. Establishing bounds on language alignment loss will be critical for ensuring communication effectiveness across diverse multi-agent populations.

Theoretical Challenges in LLM-Based Multi-Agent Communication Large language models (LLMs) introduce new possibilities for structured communication among agents, enabling human-like interaction and symbolic reasoning. However, integrating LLMs into multi-agent systems presents new theoretical challenges related to *grounding, consistency, and efficiency.* One major issue is ensuring that LLM-generated messages are *grounded in task-specific knowledge* rather than being syntactically plausible but semantically irrelevant. This can be formalized as a constrained optimization problem:

$$p(m|s, a) = \frac{\exp(f(s, a, m))}{\sum_{m'} \exp(f(s, a, m'))}, \tag{7}$$

where $p(m|s, a)$ represents the probability of generating message m given state s and action a, and $f(s, a, m)$ ensures message relevance. Future research should investigate ways to enforce *consistency and alignment* between LLM-based communication and decision-making policies. Another key challenge is *scalability and real-time performance.* LLM-based communication is computationally expensive, and real-time multi-agent interactions require *low-latency message generation.* Developing *bounded complexity guarantees* for LLM inference in multi-agent settings is an important area of theoretical research. Finally, *multi-modal communication* is an emerging field where LLMs integrate language with vision, gestures, and symbolic reasoning. Establishing a *unified theoretical framework* for multi-modal communication will help in designing systems where agents can effectively process and exchange diverse types of information.

6.2.2 Future Research on Algorithm Developments

From the discussions in previous sections and the main content in multi-agent communication (MA COMM), we can identify several potential future directions regarding algorithm design. Here, we outline some of these promising directions.

Under-explored categories in multi-agent communication. For existing categories of multi-agent communication algorithms, several areas remain under-explored, such as reinforcement learning-based communication without explicit message exchange, self-adaptive emergent communication protocols, and transformer-based communication learning that integrates structured memory for improved long-term planning. Additionally, while large-scale multi-agent systems often suffer from communication bottlenecks, more efficient message-passing algorithms that balance communication cost and coordination effectiveness are needed. Specifically, attention-based message filtering and information-theoretic compression techniques could significantly improve the scalability of multi-agent communication.

Integrated use of various communication paradigms in MARL Each category of communication strategies—such as explicit message-passing, emergent language, and LLM-based communication—has unique advantages. However, integrating these paradigms into a unified multi-agent communication framework remains an open challenge. Future work could explore hybrid architectures that combine structured communication for task-specific scenarios with emergent protocols for adaptability. For instance, multi-agent reinforcement learning (MARL) systems could benefit from hierarchical communication learning, where low-level agents use emergent communication while high-level agents leverage structured messages generated by

LLMs. Another promising research direction is improving multi-agent transformer-based communication, where agents can dynamically decide when to communicate using learned importance scores:

$$\alpha_{i,j} = \frac{\exp(f(h_i, h_j))}{\sum_{k \in \mathcal{N}(i)} \exp(f(h_i, h_k))}, \tag{8}$$

where $\alpha_{i,j}$ represents the attention weight assigned by agent i to the message from agent j, and $f(h_i, h_j)$ is a learned relevance function. Developing adaptive attention mechanisms that optimize communication efficiency while maintaining task performance remains a critical challenge.

Unsolved issues and extensions of multi-agent communication Several fundamental issues in multi-agent communication remain unsolved, requiring new algorithmic developments: (1). **Scalability in large-scale MARL**: Many existing methods struggle with scalability due to communication overhead. Efficient routing and selective message-passing strategies, such as reinforcement learning-based communication topologies, could be explored. (2). **Handling non-stationarity**: In non-stationary environments where agents continually update their policies, communication protocols must adapt dynamically. One approach could be meta-learning-based strategies that allow agents to learn and adapt their communication strategies over time. (3). **Multi-agent robustness under noisy communication**: Real-world applications introduce message loss, delays, or adversarial tampering. Designing robust multi-agent communication frameworks that employ redundancy mechanisms and adversarial training techniques could improve resilience. Specifically, MARL algorithms for fully competitive or mixed cooperative-competitive tasks, grounded in game-theoretic principles, require more sophisticated communication strategies that balance cooperation and deception.

Development of generalist multi-agent communication models One of the most exciting future directions in multi-agent communication is developing a generalist communication model that can adapt to a variety of environments and tasks without requiring task-specific fine-tuning. This could be achieved by leveraging advances in foundation models from NLP and computer vision. To build generalist multi-agent communication systems, several challenges must be addressed: (1). **Scalable architectures for universal communication**: Designing models that can process multi-modal inputs, including textual, visual, and structured data, while maintaining efficient message encoding. (2). **Pretraining and fine-tuning for generalization**: As seen in large-scale foundation models, pretraining communication models on diverse datasets followed by fine-tuning on task-specific interactions could enable better generalization. (3). **Learning from third-person demonstrations**: Future research should explore whether multi-agent communication strategies can be learned from video datasets or human-agent interactions, allowing agents to infer communication protocols without direct reinforcement learning.

Future works in related areas Finally, there are several related research areas that could significantly impact the future of multi-agent communication: (1). **Inverse reinforcement learning for communication**: Understanding how communication strategies emerge in expert demonstrations could improve learned policies. (2). **Uncertainty-aware communication in MARL**: Extending multi-agent communication to handle risk-sensitive decision-making could be critical for high-stakes applications such as autonomous driving. (3). **Incorporating new foundation models**: The emergence of new architectures such as Mamba models and Poisson Flow models provides exciting opportunities to enhance communication strategies for real-time and data-efficient learning.

6.2.3 Future Works on Benchmarking

Development of more realistic, challenging benchmarks We observe that most multi-agent communication (MA COMM) algorithms are evaluated on standard MARL benchmarks such as StarCraft Multi-Agent Challenge (SMAC), Multi-Agent Particle Environment (MPE), and Google Research Football (GRF). However, these benchmarks may not fully capture the complexity and challenges of real-world multi-agent communication. To advance the field and properly evaluate new methods, more comprehensive and challenging benchmarks should be developed: (1). **Large-scale datasets for communication evaluation**: The scalability of MA COMM algorithms remains a key challenge. Benchmarks should include large-scale datasets that test how communication methods perform with increasing agent populations, diverse communication constraints, and extensive training data. Understanding the relationship between dataset size and communication efficiency can help assess whether learned communication strategies generalize well. (2). **Multi-modal communication benchmarks**: Real-world multi-agent communication often involves heterogeneous data sources, including language, vision, and symbolic information. Future benchmarks should include multi-modal environments where agents communicate using a mix of textual, visual, and structured messages. For example, agents might receive partial visual observations and complement them with natural language messages to coordinate actions. (3). **More realistic task scenarios**: A major reason for the widespread success of deep learning in computer vision and NLP is that models are trained on large, real-world datasets. In contrast, multi-agent communication is still largely evaluated in simulated environments that lack real-world complexity. Future benchmarks should include real-world datasets of human or agent interactions, such as autonomous vehicle coordination data, multi-robot task execution logs, and human dialogue datasets for cooperative problem-solving. (4). **Beyond return-based evaluation**: Current benchmarks mainly evaluate performance based on cumulative rewards or task completion rates. However, effective communication should also be assessed in terms of **interpretability, efficiency, generalization, and safety**. Future benchmarks should include metrics that could evaluate problems like: **Message efficiency**, How much communication overhead is required for achieving optimal performance? And **Generalization**, such as answering questions like *Can learned communication protocols transfer to unseen environments or new agents?* And **Robustness**: How resilient is the communication strategy to message loss, delays, or adversarial interference? Or for **Emergent communication structure**, an efficient benchmark should help researchers to answer *Do agents develop structured and compositional languages, and how interpretable are they?*

Comparisons among different multi-agent communication paradigms While there exist numerous algorithms for multi-agent communication—ranging from explicit message-passing approaches to emergent language learning and LLM-based multi-agent coordination—there is a lack of standardized comparisons among them. To better understand their relative advantages and limitations, future research should establish benchmarks that allow fair evaluations across different paradigms: (1). **Standardized comparisons across communication strategies**: Future benchmarks should compare traditional MARL communication approaches (e.g., differentiable communication via continuous signals) with emergent language-based communication and LLM-augmented communication frameworks. By evaluating these approaches on the same set of tasks, researchers can better understand when structured communication is preferable over learned protocols. (2). **Scalability of different communication methods**: As the number of agents increases, how do different communication approaches scale? Some methods may perform well in small agent populations but struggle with communication bottlenecks in larger teams. Benchmarks should include scalability evaluations where different communication architectures are tested under increasing numbers of agents and message complexity constraints. (3). **Computational and learning efficiency**: While some multi-agent communication methods achieve high performance, they may require significant computational resources for training. Future benchmarks should track training time, inference speed, and memory usage across different methods, providing insights into their practical feasibility. (4). **Adaptability to new agents and tasks**: Benchmarks should evaluate how well communication strategies generalize to new agents or novel tasks. For example, if an agent trained in one environment is deployed in another with different teammates, how quickly can it adapt its communication protocol? Evaluating zero-shot and few-shot generalization capabilities would be crucial for developing flexible multi-agent systems.

Developing a unified benchmark suite for MA COMM To accelerate progress in multi-agent communication research, a unified benchmark suite should be developed that provides: (1). **A diverse set of environments**: Benchmarks should cover cooperative, competitive, and mixed-motive multi-agent tasks, ensuring that communication strategies are evaluated across a broad range of scenarios. (2). **Extensive datasets for pretraining and evaluation**: Similar to ImageNet in computer vision and GLUE in NLP, multi-agent communication research would benefit from large-scale datasets that facilitate pretraining and systematic evaluation. (3). **A set of well-defined metrics**: Future benchmarks should move beyond simple reward-based evaluation and incorporate new performance metrics that reflect communication efficiency, interpretability, robustness, and generalization. (4). **Baseline implementations for fair comparisons**: Providing open-source implementations of baseline communication methods would ensure reproducibility and encourage fair comparisons across different approaches.

6.2.4 Future Research on Human-Centric Multi-Agent Communication

How can multi-agent communication systems align with human expectations and interpretability? The effectiveness of multi-agent communication (MA COMM) is not solely determined by task performance but also by its interpretability, alignment with human communication norms, and ability to foster trust in human-agent collaboration. One fundamental research direction is to explore how multi-agent communication protocols can be designed to be both machine-efficient and human-comprehensible. (1) Current emergent communication models often develop symbols or representations that are opaque to humans. Ensuring that learned communication protocols exhibit structured, compositional, and generalizable properties similar to natural language is a key challenge. A possible approach is to impose information-theoretic constraints that favor structured communication while maintaining efficiency. Another direction is to explore how learned protocols can be mapped to existing human language structures without compromising performance. (2) Multi-agent communication should be interpretable not only in terms of the message content but also in how communication influences agent behavior. Formal methods for analyzing the causal effect of communication on decision-making remain underdeveloped. Establishing metrics to evaluate whether agents correctly utilize received messages in a way that aligns with human intuition can be beneficial for interpretability and debugging. (3) The degree to which communication can be adapted for diverse user preferences is also an open question. Humans exhibit variation in linguistic styles, levels of detail in communication, and implicit assumptions about shared knowledge. Future research should investigate adaptive communication strategies where agents dynamically adjust their communication style to align with individual user expectations. Such adaptability could be achieved through meta-learning techniques or reinforcement learning policies trained on diverse human interaction datasets.

How should multi-agent communication systems facilitate collaboration with humans? In many real-world applications, agents must not only communicate with each other but also interact with humans as teammates, supervisors, or evaluators. Designing communication protocols that effectively integrate human feedback and decision-making is critical for advancing human-AI collaboration. (1) One challenge is determining the optimal balance between autonomous agent communication and human-in-the-loop intervention. While some tasks require full autonomy, others benefit from human oversight, and communication models should be able to modulate information exchange accordingly. This raises the question of how to design communication-aware policies that incorporate human corrections efficiently without excessive reliance on human intervention. (2) Another open problem is learning communication strategies that generalize across different types of human users. Agents may need to adjust their communication level based on the expertise and role of the human collaborator. For instance, a novice user may require detailed explanations, while an expert user may prefer concise updates. A promising direction is to develop hierarchical communication frameworks where agents can reason about when and how to involve human input. (3) In settings where multiple agents communicate with a human, ensuring that communication remains coherent and non-redundant is crucial. Unlike homogeneous multi-agent settings, human interactions require mechanisms for resolving conflicting information, synthesizing diverse perspectives, and presenting relevant insights in a structured manner. Future work should investigate how multi-agent communication can be optimized for presenting actionable insights in an efficient and user-friendly manner.

How can human trust and safety be ensured in multi-agent communication? Trust in multi-agent communication systems is essential for adoption in high-stakes domains such as healthcare, autonomous driving, and critical infrastructure management. However, ensuring robustness and reliability in communication remains an open challenge. (1) One primary concern is that communication protocols learned purely through reinforcement learning may not always prioritize safety or reliability. Messages exchanged between agents should not only be optimized for task efficiency but also be constrained to avoid potentially harmful or misleading communication. Future research should explore formal safety constraints that ensure multi-agent messages adhere to ethical and regulatory guidelines. (2) Communication robustness under adversarial settings is another key research direction. In environments where agents operate in competitive or adversarial conditions, ensuring that communication remains truthful and resilient to manipulation is essential. Theoretical work on adversarial robustness in communication, such as establishing upper bounds on the probability of message tampering or introducing cryptographic techniques for secure multi-agent communication, is an important avenue of research. (3) Beyond security concerns, communication reliability in dynamic and non-stationary environments is also a challenge. Agents must be able to detect and recover from missing or corrupted messages while ensuring that communication failures do not lead to catastrophic decision errors. Future work should investigate self-correcting communication mechanisms that allow agents to infer missing information or reconstruct degraded messages in real time.

How can human feedback be effectively incorporated into multi-agent communication learning? Human-in-the-loop learning is a promising approach for improving multi-agent communication by leveraging human expertise to refine communication strategies. However, incorporating human feedback efficiently and scalably remains an open challenge. (1) One direction is investigating how agents can learn effective communication through a combination of explicit and implicit feedback. While direct human corrections can provide useful supervision, implicit signals such as human engagement, response time, or sentiment analysis from text-based interactions could serve as valuable learning signals for communication adaptation. (2) Future work should explore techniques for efficiently integrating sparse human feedback into multi-agent learning frameworks. Since obtaining human annotations for large-scale multi-agent interactions is costly, techniques such as preference-based reinforcement learning or weak supervision could be useful for extracting meaningful guidance with minimal human effort. (3) Evaluating the effectiveness of human feedback on multi-agent communication learning remains an underexplored research problem. While traditional reinforcement learning relies on reward functions, the impact of human feedback is more complex and may require new evaluation frameworks that capture the long-term benefits of human-informed communication strategies. Establishing benchmarks that explicitly test human-guided communication adaptation would be a valuable contribution to this area.

How can multi-agent communication be designed to enhance social interaction and fairness? As multi-agent systems are increasingly deployed in human-facing applications, it is important to consider fairness and social dynamics in communication. (1) One key question is how to ensure equitable information access in multi-agent interactions. In scenarios where multiple agents communicate with each other and with human users, disparities in information access can emerge, leading to biased decision-making. Research should focus on designing fairness-aware communication policies that ensure all participants receive relevant information based on their role and decision-making needs. (2) Another challenge is avoiding emergent behaviors that disadvantage certain agents or human users. For example, in competitive multi-agent environments, agents may develop deceptive communication strategies that provide short-term advantages but undermine trust and cooperation in the long run. Studying game-theoretic mechanisms for enforcing cooperative communication norms could help mitigate such behaviors. (3) Finally, multi-agent communication should support natural and socially appropriate interactions. This includes adapting communication tone, level of detail, and responsiveness based on the social context. Future research should explore computational models of social intelligence that enable agents to adjust their communication style in a human-like manner, improving engagement and user experience.

7 Conclusion

In this paper, we present a systematic survey of multi-agent communication (MA-Comm) in sequential decision-making. In Section 2, we provide necessary background on three aspects we wish to discuss, including MARL-Comm, Emergent Language, and LLMs. In Section 3, we review communication within the MARL framework, tracing progress from early message-passing mechanisms to trainable, end-to-end protocols. In Section 4, we cover emergent language studies, where agents develop communication strategies through interaction, highlighting key methodological trends and evaluation challenges. In Section 5, we examine the emerging role of LLMs, which enable agents to reason, plan, and collaborate using natural language in more open-ended settings. Complementing these sections, we provide taxonomies and comparative analyses of representative methods, along with discussions of evaluation metrics, open challenges, and opportunities for bridging different strands of research. Following the main sections, we also provide a summary on existing works and our perspectives on future research directions in Section 6.

As the first comprehensive survey dedicated to MA-Comm, we hope this work serves as a reference point for researchers entering the field and as a guide for advancing communication in multi-agent systems. By clarifying connections between MARL-based communication, emergent language, and LLM-driven approaches, we aim to foster new research directions that move beyond controlled benchmarks toward scalable, interpretable, and human-aligned communication in real-world multi-agent environments.

References

Crewai. `https://github.com/joaomdmoura/crewAI`.

Langchain. `https://github.com/langchain-ai/langchain`, a.

Langgraph. `https://www.langchain.com/langgraph`, b.

Superagi. `https://github.com/TransformerOptimus/SuperAGI`.

Sahar Abdelnabi, Amr Gomaa, Sarath Sivaprasad, Lea Schönherr, and Mario Fritz. Llm-deliberation: Evaluating llms with interactive multi-agent negotiation games. 2023.

Sahar Abdelnabi, Amr Gomaa, Sarath Sivaprasad, Lea Schönherr, and Mario Fritz. Cooperation, competition, and maliciousness: Llm-stakeholders interactive negotiation. *Advances in Neural Information Processing Systems*, 37:83548–83599, 2024.

Jeffrey L Adler and Victor J Blue. A cooperative multi-agent transportation management and route guidance system. *Transportation Research Part C: Emerging Technologies*, 10(5-6):433–454, 2002.

John P Agapiou, Alexander Sasha Vezhnevets, Edgar A Duéñez-Guzmán, Jayd Matyas, Yiran Mao, Peter Sunehag, Raphael Köster, Udari Madhushani, Kavya Kopparapu, Ramona Comanescu, et al. Melting pot 2.0. *arXiv preprint arXiv:2211.13746*, 2022.

Akshat Agarwal, Swaminathan Gurumurthy, Vasu Sharma, Mike Lewis, and Katia Sycara. Community regularization of visually-grounded dialog. *arXiv preprint arXiv:1808.04359*, 2018.

Akshat Agarwal, Swaminathan Gurumurthy, Vasu Sharma, Mike Lewis, and Katia Sycara. Community regularization of visually-grounded dialog. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 1042–1050, 2019a.

Akshat Agarwal, Sumit Kumar, and Katia Sycara. Learning transferable cooperative behavior in multi-agent teams. *arXiv preprint arXiv:1906.01202*, 2019b.

Saaket Agashe, Yue Fan, Anthony Reyna, and Xin Eric Wang. Llm-coordination: evaluating and analyzing multi-agent coordination abilities in large language models. *arXiv preprint arXiv:2310.03903*, 2023a.

Saaket Agashe, Yue Fan, and Xin Eric Wang. Evaluating multi-agent coordination abilities in large language models. *arXiv preprint arXiv:2310.03903*, 2023b.

Junhyeok Ahn and Luis Sentis. Nested mixture of experts: Cooperative and competitive learning of hybrid dynamical system. In *Learning for Dynamics and Control*, pp. 779–790. PMLR, 2021.

Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, Chuyuan Fu, Keerthana Gopalakrishnan, Karol Hausman, et al. Do as i can, not as i say: Grounding language in robotic affordances. *arXiv preprint arXiv:2204.01691*, 2022.

Altera AL, Andrew Ahn, Nic Becker, Stephanie Carroll, Nico Christie, Manuel Cortes, Arda Demirci, Melissa Du, Frankie Li, Shuying Luo, et al. Project sid: Many-agent simulations toward ai civilization. *arXiv preprint arXiv:2411.00114*, 2024.

Jacob Andreas. Measuring compositionality in representation learning. *arXiv preprint arXiv:1902.07181*, 2019.

RM Aratchige and WMKS Ilmini. Llms working in harmony: A survey on the technological aspects of building effective llm-based multi agent systems. *arXiv preprint arXiv:2504.01963*, 2025.

Michal Auersperger and Pavel Pecina. Defending compositionality in emergent languages. *arXiv preprint arXiv:2206.04751*, 2022.

Suma Bailis, Jane Friedhoff, and Feiyang Chen. Werewolf arena: A case study in llm evaluation via social deduction. *arXiv preprint arXiv:2407.13943*, 2024.

Bowen Baker, Ingmar Kanitscheider, Todor Markov, Yi Wu, Glenn Powell, Bob McGrew, and Igor Mordatch. Emergent tool use from multi-agent autocurricula. In *International Conference on Learning Representations*.

Razan Baltaji, Babak Hemmatian, and Lav R Varshney. Conformity, confabulation, and impersonation: Persona inconstancy in multi-agent llm collaboration. *arXiv preprint arXiv:2405.03862*, 2024.

Cristian-Paul Bara, Sky CH-Wang, and Joyce Chai. Mindcraft: Theory of mind modeling for situated dialogue in collaborative tasks. *arXiv preprint arXiv:2109.06275*, 2021.

Marco Baroni. Linguistic generalization and compositionality in modern artificial neural networks. *Philosophical Transactions of the Royal Society B*, 375(1791):20190307, 2020.

John Batali. Computational simulations of the emergence of grammar. *Approaches to the Evolution of Language-Social and Cognitive Bases-*, 1998.

Ali Beikmohammadi. Learning to communicate through multi-agent reinforcement learning (marl): A systematic literature review. *Preprints*, April 2024. doi: 10.20944/preprints202404.0813.v1. URL `https://doi.org/10.20944/preprints202404.0813.v1`.

Emily M Bender and Alexander Koller. Climbing towards nlu: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pp. 5185–5198, 2020.

Christopher Berner, Greg Brockman, Brooke Chan, Vicki Cheung, Przemysław Dębiak, Christy Dennison, David Farhi, Quirin Fischer, Shariq Hashme, Chris Hesse, et al. Dota 2 with large scale deep reinforcement learning. *arXiv preprint arXiv:1912.06680*, 2019.

Daniel S Bernstein, Robert Givan, Neil Immerman, and Shlomo Zilberstein. The complexity of decentralized control of markov decision processes. *Mathematics of operations research*, 27(4):819–840, 2002.

Maciej Besta, Florim Memedi, Zhenyu Zhang, Robert Gerstenberger, Guangyuan Piao, Nils Blach, Piotr Nyczyk, Marcin Copik, Grzegorz Kwaśniewski, Jürgen Müller, et al. Demystifying chains, trees, and graphs of thoughts. *arXiv preprint arXiv:2401.14295*, 2024.

Jan Blumenkamp and Amanda Prorok. The emergence of adversarial communication in multi-agent reinforcement learning. In *Conference on Robot Learning*, pp. 1394–1414. PMLR, 2021.

Ben Bogin, Mor Geva, and Jonathan Berant. Emergence of communication in an interactive world with consistent speakers. *arXiv preprint arXiv:1809.00549*, 2018.

Brendon Boldt and David Mortensen. A review of the applications of deep learning-based emergent communication. *arXiv preprint arXiv:2407.03302*, 2024.

Diane Bouchacourt and Marco Baroni. How agents see things: On visual representations in an emergent language game. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 981–985, 2018.

Diane Bouchacourt and Marco Baroni. Miss tools and mr fruit: Emergent communication in agents learning about object affordances, 2019. URL `https://arxiv.org/abs/1905.11871`.

Nicolo' Brandizzi. Toward more human-like ai communication: A review of emergent communication research. *IEEE Access*, 11:142317–142340, 2023.

Nicolo' Brandizzi, Davide Grossi, and Luca Iocchi. Rlupus: Cooperation through emergent communication in the werewolf social deduction game. *Intelligenza Artificiale*, 15(2):55–70, 2022.

Cynthia Breazeal, Cory D Kidd, Andrea Lockerd Thomaz, Guy Hoffman, and Matt Berlin. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *2005 IEEE/RSJ international conference on intelligent robots and systems*, pp. 708–713. IEEE, 2005.

Henry Brighton and Simon Kirby. Understanding linguistic evolution by visualizing the emergence of topographic mappings. *Artificial life*, 12(2):229–242, 2006.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

Christian Buck, Jannis Bulian, Massimiliano Ciaramita, Wojciech Gajewski, Andrea Gesmundo, Neil Houlsby, and Wei Wang. Analyzing language learned by an active question answering agent. *arXiv preprint arXiv:1801.07537*, 2018.

Kalesha Bullard, Franziska Meier, Douwe Kiela, Joelle Pineau, and Jakob Foerster. Exploring zero-shot emergent communication in embodied multi-agent populations. *arXiv preprint arXiv:2010.15896*, 2020.

Kalesha Bullard, Douwe Kiela, Franziska Meier, Joelle Pineau, and Jakob Foerster. Quasi-equivalence discovery for zero-shot emergent communication. *arXiv preprint arXiv:2103.08067*, 2021.

Lucian Busoniu, Robert Babuska, and Bart De Schutter. A comprehensive survey of multiagent reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38 (2):156–172, 2008.

Salvatore Candido and Seth Hutchinson. Minimum uncertainty robot navigation using information-guided pomdp planning. In *2011 IEEE International Conference on Robotics and Automation*, pp. 6102–6108. IEEE, 2011.

Lorenzo Canese, Gian Carlo Cardarilli, Luca Di Nunzio, Rocco Fazzolari, Daniele Giardino, Marco Re, and Sergio Spanò. Multi-agent reinforcement learning: A review of challenges and applications. *Applied Sciences*, 11(11):4948, 2021.

Angelo Cangelosi and Domenico Parisi. *Simulating the evolution of language*. Springer Science & Business Media, 2012.

Kris Cao, Angeliki Lazaridou, Marc Lanctot, Joel Z Leibo, Karl Tuyls, and Stephen Clark. Emergent communication through negotiation. *arXiv preprint arXiv:1804.03980*, 2018.

Yongcan Cao, Wenwu Yu, Wei Ren, and Guanrong Chen. An overview of recent progress in the study of distributed multi-agent coordination. *IEEE Transactions on Industrial Informatics*, 9(1):427–438, 2013. doi: 10.1109/TII.2012.2219061.

Boaz Carmeli, Ron Meir, and Yonatan Belinkov. Emergent quantized communication. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 11533–11541, 2023.

Micah Carroll, Rohin Shah, Mark K Ho, Tom Griffiths, Sanjit Seshia, Pieter Abbeel, and Anca Dragan. On the utility of learning about humans for human-ai coordination. *Advances in neural information processing systems*, 32, 2019.

Cristiano Castelfranchi. The theory of social functions: challenges for computational social science and multi-agent learning. *Cognitive Systems Research*, 2(1):5–38, 2001.

Mert Cemri, Melissa Z Pan, Shuyi Yang, Lakshya A Agrawal, Bhavya Chopra, Rishabh Tiwari, Kurt Keutzer, Aditya Parameswaran, Dan Klein, Kannan Ramchandran, et al. Why do multi-agent llm systems fail? *arXiv preprint arXiv:2503.13657*, 2025.

Rahma Chaabouni, Eugene Kharitonov, Emmanuel Dupoux, and Marco Baroni. Anti-efficient encoding in emergent communication. *Advances in Neural Information Processing Systems*, 32, 2019.

Rahma Chaabouni, Eugene Kharitonov, Diane Bouchacourt, Emmanuel Dupoux, and Marco Baroni. Compositionality and generalization in emergent languages. In *ACL 2020-8th annual meeting of the Association for Computational Linguistics*, 2020.

Rahma Chaabouni, Florian Strub, Florent Altché, Eugene Tarassov, Corentin Tallec, Elnaz Davoodi, Kory Wallace Mathewson, Olivier Tieleman, Angeliki Lazaridou, and Bilal Piot. Emergent communication at scale. In *International conference on learning representations*, 2022.

Marwa Chafii, Salmane Naoumi, Reda Alami, Ebtesam Almazrouei, Mehdi Bennis, and Merouane Debbah. Emergent communication in multi-agent reinforcement learning for future wireless networks. *IEEE Internet of Things Magazine*, 6(4):18–24, 2023.

Chi-Min Chan, Weize Chen, Yusheng Su, Jianxuan Yu, Wei Xue, Shanghang Zhang, Jie Fu, and Zhiyuan Liu. Chateval: Towards better llm-based evaluators through multi-agent debate. *arXiv preprint arXiv:2308.07201*, 2023.

Matthew Chang, Gunjan Chhablani, Alexander Clegg, Mikael Dallaire Cote, Ruta Desai, Michal Hlavac, Vladimir Karashchuk, Jacob Krantz, Roozbeh Mottaghi, Priyam Parashar, et al. Partnr: A benchmark for planning and reasoning in embodied multi-agent tasks. *arXiv preprint arXiv:2411.00081*, 2024.

Guangyao Chen, Siwei Dong, Yu Shu, Ge Zhang, Jaward Sesay, Börje F Karlsson, Jie Fu, and Yemin Shi. Autoagents: A framework for automatic agent generation. *arXiv preprint arXiv:2309.17288*, 2023a.

Jiangjie Chen, Xintao Wang, Rui Xu, Siyu Yuan, Yikai Zhang, Wei Shi, Jian Xie, Shuang Li, Ruihan Yang, Tinghui Zhu, et al. From persona to personalization: A survey on role-playing language agents. *arXiv preprint arXiv:2404.18231*, 2024a.

Jingdi Chen, Tian Lan, and Carlee Joe-Wong. Rgmcomm: Return gap minimization via discrete communications in multi-agent reinforcement learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 17327–17336, 2024b.

Junzhe Chen, Xuming Hu, Shuodi Liu, Shiyu Huang, Wei-Wei Tu, Zhaofeng He, and Lijie Wen. Llmarena: Assessing capabilities of large language models in dynamic multi-agent environments. *arXiv preprint arXiv:2402.16499*, 2024c.

Justin Chih-Yao Chen, Swarnadeep Saha, and Mohit Bansal. Reconcile: Round-table conference improves reasoning via consensus among diverse llms. *arXiv preprint arXiv:2309.13007*, 2023b.

Pei Chen, Boran Han, and Shuai Zhang. Comm: Collaborative multi-agent, multi-reasoning-path prompting for complex problem solving. *arXiv preprint arXiv:2404.17729*, 2024d.

Weize Chen, Yusheng Su, Jingwei Zuo, Cheng Yang, Chenfei Yuan, Chen Qian, Chi-Min Chan, Yujia Qin, Yaxi Lu, Ruobing Xie, et al. Agentverse: Facilitating multi-agent collaboration and exploring emergent behaviors in agents. *arXiv preprint arXiv:2308.10848*, 2(4):5, 2023c.

Weize Chen, Ziming You, Ran Li, Yitong Guan, Chen Qian, Chenyang Zhao, Cheng Yang, Ruobing Xie, Zhiyuan Liu, and Maosong Sun. Internet of agents: Weaving a web of heterogeneous agents for collaborative intelligence. *arXiv preprint arXiv:2407.07061*, 2024e.

Weize Chen, Jiarui Yuan, Chen Qian, Cheng Yang, Zhiyuan Liu, and Maosong Sun. Optima: Optimizing effectiveness and efficiency for llm-based multi-agent system. *arXiv preprint arXiv:2410.08115*, 2024f.

Yongchao Chen, Jacob Arkin, Yang Zhang, Nicholas Roy, and Chuchu Fan. Scalable multi-robot collaboration with large language models: Centralized or decentralized systems? In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4311–4317. IEEE, 2024g.

Ricson Cheng, Arpit Agarwal, and Katerina Fragkiadaki. Reinforcement learning of active vision for manipulating objects under occlusions. In *Conference on Robot Learning*, pp. 422–431. PMLR, 2018.

Yuheng Cheng, Ceyao Zhang, Zhengwen Zhang, Xiangrui Meng, Sirui Hong, Wenhao Li, Zihao Wang, Zekai Wang, Feng Yin, Junhua Zhao, et al. Exploring large language model based intelligent agents: Definitions, methods, and prospects. *arXiv preprint arXiv:2401.03428*, 2024.

Edward Choi, Angeliki Lazaridou, and Nando de Freitas. Compositional obverter communication learning from raw visual input, 2018. URL `https://arxiv.org/abs/1804.02341`.

Jongeun Choi, Songhwai Oh, and Roberto Horowitz. Distributed learning and cooperative control for multi-agent systems. *Automatica*, 45(12):2802–2814, 2009.

Morten H Christiansen and Simon Kirby. *Language evolution*. OUP Oxford, 2003.

Tianshu Chu, Sandeep Chinchali, and Sachin Katti. Multi-agent reinforcement learning for networked system control. *arXiv preprint arXiv:2004.01339*, 2020.

Cédric Colas, Tristan Karch, Nicolas Lair, Jean-Michel Dussoux, Clément Moulin-Frier, Peter Dominey, and Pierre-Yves Oudeyer. Language as a cognitive tool to imagine goals in curiosity driven exploration. *Advances in Neural Information Processing Systems*, 33:3761–3774, 2020.

Dylan Cope and Nandi Schoots. Learning to communicate with strangers via channel randomisation methods. In *The 4th Workshop on Emergent Communication at the International Conference on Neural Information Processing (NeurIPS)*, 2020.

Dylan Cope and Nandi Schoots. Learning to communicate with strangers via channel randomisation methods. *arXiv preprint arXiv:2104.09557*, 2021.

Jorge Cortes, Sonia Martinez, Timur Karatas, and Francesco Bullo. Coverage control for mobile sensing networks. *IEEE Transactions on robotics and Automation*, 20(2):243–255, 2004.

Gautier Dagan, Dieuwke Hupkes, and Elia Bruni. Co-evolution of language and agents in referential games. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pp. 2993–3004, 2021.

Mike D'Arcy, Tom Hope, Larry Birnbaum, and Doug Downey. Marg: Multi-agent review generation for scientific papers. *arXiv preprint arXiv:2401.04259*, 2024.

Abhishek Das, Satwik Kottur, José MF Moura, Stefan Lee, and Dhruv Batra. Learning cooperative visual dialog agents with deep reinforcement learning. In *Proceedings of the IEEE international conference on computer vision*, pp. 2951–2960, 2017.

Abhishek Das, Théophile Gervet, Joshua Romoff, Dhruv Batra, Devi Parikh, Michael Rabbat, and Joelle Pineau. Tarmac: Targeted multi-agent communication, 2018. URL `https://arxiv.org/abs/1810.11187`.

Ayushman Das, Shu-Ching Chen, Mei-Ling Shyu, and Saad Sadiq. Enabling synergistic knowledge sharing and reasoning in large language models with collaborative multi-agents. In *2023 IEEE 9th International Conference on Collaboration and Internet Computing (CIC)*, pp. 92–98. IEEE, 2023.

Ankur Deka and Katia Sycara. Natural emergence of heterogeneous strategies in artificially intelligent competitive teams. In *International Conference on Swarm Intelligence*, pp. 13–25. Springer, 2021.

Kevin Denamganaï and James Alfred Walker. On (emergent) systematic generalisation and compositionality in visual referential games with straight-through gumbel-softmax estimator. *arXiv preprint arXiv:2012.10776*, 2020a.

Kevin Denamganaï and James Alfred Walker. Referentialgym: A nomenclature and framework for language emergence & grounding in (visual) referential games. *arXiv preprint arXiv:2012.09486*, 2020b.

Kevin Denamganaï, Sondess Missaoui, and James Alfred Walker. Visual referential games further the emergence of disentangled representations. *arXiv preprint arXiv:2304.14511*, 2023.

Xiang Deng, Yu Gu, Boyuan Zheng, Shijie Chen, Sam Stevens, Boshi Wang, Huan Sun, and Yu Su. Mind2web: Towards a generalist agent for the web. *Advances in Neural Information Processing Systems*, 36, 2024.

Roberto Dessì, Diane Bouchacourt, Davide Crepaldi, and Marco Baroni. Focus on what's informative and ignore what's not: Communication strategies in a referential game. *arXiv preprint arXiv:1911.01892*, 2019.

Ziluo Ding, Tiejun Huang, and Zongqing Lu. Learning individually inferred communication for multi-agent cooperation. *Advances in Neural Information Processing Systems*, 33:22069–22079, 2020.

Yubo Dong, Xukun Zhu, Zhengzhe Pan, Linchao Zhu, and Yi Yang. Villageragent: A graph-based multi-agent framework for coordinating complex task dependencies in minecraft. *arXiv preprint arXiv:2406.05720*, 2024.

Silin Du and Xiaowei Zhang. Helmsman of the masses? evaluate the opinion leadership of large language models in the werewolf game. *arXiv preprint arXiv:2404.01602*, 2024.

Yali Du, Bo Liu, Vincent Moens, Ziqi Liu, Zhicheng Ren, Jun Wang, Xu Chen, and Haifeng Zhang. Learning correlated communication topology in multi-agent reinforcement learning. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 456–464, 2021.

Yilun Du, Shuang Li, Antonio Torralba, Joshua B Tenenbaum, and Igor Mordatch. Improving factuality and reasoning in language models through multiagent debate. *arXiv preprint arXiv:2305.14325*, 2023.

Zhuoyun Du, Chen Qian, Wei Liu, Zihao Xie, Yifei Wang, Yufan Dang, Weize Chen, and Cheng Yang. Multi-agent software development through cross-team collaboration. *arXiv preprint arXiv:2406.08979*, 2024.

Marina Dubova, Arsenii Kirillovich Moskvichev, and Rob Goldstone. Reinforcement communication learning in different social network structures. In *Language in Reinforcement Learning Workshop at ICML 2020*.

Zane Durante, Qiuyuan Huang, Naoki Wake, Ran Gong, Jae Sung Park, Bidipta Sarkar, Rohan Taori, Yusuke Noda, Demetri Terzopoulos, Yejin Choi, et al. Agent ai: Surveying the horizons of multimodal interaction. *arXiv preprint arXiv:2401.03568*, 2024.

Tom Eccles, Yoram Bachrach, Guy Lever, Angeliki Lazaridou, and Thore Graepel. Biases for emergent communication in multi-agent reinforcement learning. *Advances in neural information processing systems*, 32, 2019.

Andrew Estornell, Jean-François Ton, Yuanshun Yao, and Yang Liu. Acc-collab: An actor-critic approach to multi-agent llm collaboration. In *The Thirteenth International Conference on Learning Representations*, 2025.

Katrina Evtimova, Andrew Drozdov, Douwe Kiela, and Kyunghyun Cho. Emergent communication in a multi-modal, multi-step referential game. *arXiv preprint arXiv:1705.10369*, 2017.

Meta Fundamental AI Research Diplomacy Team (FAIR)†, Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, et al. Human-level play in the game of diplomacy by combining language models with strategic reasoning. *Science*, 378 (6624):1067–1074, 2022.

Xueyang Feng, Zhi-Yuan Chen, Yujia Qin, Yankai Lin, Xu Chen, Zhiyuan Liu, and Ji-Rong Wen. Large language model-based human-agent collaboration for complex task solving. *arXiv preprint arXiv:2402.12914*, 2024.

Chrisantha Fernando, Daria Zenkova, Stanislav Nikolov, and Simon Osindero. From language games to drawing games. *arXiv preprint arXiv:2010.02820*, 2020.

Nicole Fitzgerald and Jacopo Tagliabue. On the plurality of graphs. *Journal of Logic and Computation*, 32 (6):1129–1141, 2022.

Jakob Foerster, Ioannis Alexandros Assael, Nando De Freitas, and Shimon Whiteson. Learning to communicate with deep multi-agent reinforcement learning. *Advances in neural information processing systems*, 29, 2016a.

Jakob Foerster, Francis Song, Edward Hughes, Neil Burch, Iain Dunning, Shimon Whiteson, Matthew Botvinick, and Michael Bowling. Bayesian action decoder for deep multi-agent reinforcement learning. In *International Conference on Machine Learning*, pp. 1942–1951. PMLR, 2019.

Jakob N Foerster, Yannis M Assael, Nando de Freitas, and Shimon Whiteson. Learning to communicate to solve riddles with deep distributed recurrent q-networks. *arXiv preprint arXiv:1602.02672*, 2016b.

Adam Fourney, Gagan Bansal, Hussein Mozannar, Cheng Tan, Eduardo Salinas, Friederike Niedtner, Grace Proebsting, Griffin Bassman, Jack Gerrits, Jacob Alber, et al. Magentic-one: A generalist multi-agent system for solving complex tasks. *arXiv preprint arXiv:2411.04468*, 2024.

Benjamin Freed, Rohan James, Guillaume Sartoretti, and Howie Choset. Sparse discrete communication learning for multi-agent cooperation through backpropagation. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7993–7998. IEEE, 2020a.

Benjamin Freed, Guillaume Sartoretti, and Howie Choset. Simultaneous policy and discrete communication learning for multi-agent cooperation. *IEEE Robotics and Automation Letters*, 5(2):2498–2505, 2020b.

Lukas Galke and Limor Raviv. Emergent communication and learning pressures in language models: a language evolution perspective. *arXiv preprint arXiv:2403.14427*, 2024.

Lukas Galke, Yoav Ram, and Limor Raviv. Emergent communication for understanding human language evolution: What's missing? In *Emergent Communication Workshop at ICLR 2022*.

Alireza Ghafarollahi and Markus J Buehler. Sciagents: Automating scientific discovery through multi-agent intelligent graph reasoning. *arXiv preprint arXiv:2409.05556*, 2024.

Ran Gong, Qiuyuan Huang, Xiaojian Ma, Hoi Vo, Zane Durante, Yusuke Noda, Zilong Zheng, Song-Chun Zhu, Demetri Terzopoulos, Li Fei-Fei, et al. Mindagent: Emergent gaming interaction. *arXiv preprint arXiv:2309.09971*, 2023.

Laura Graesser, Kyunghyun Cho, and Douwe Kiela. Emergent linguistic phenomena in multi-agent communication games. *arXiv preprint arXiv:1901.08706*, 2019.

Sven Gronauer and Klaus Diepold. Multi-agent deep reinforcement learning: a survey. *Artificial Intelligence Review*, 55(2):895–943, 2022.

Niko A Grupen, Daniel D Lee, and Bart Selman. Multi-agent curricula and emergent implicit signaling. *arXiv preprint arXiv:2106.11156*, 2021.

Niko A. Grupen, Daniel D. Lee, and Bart Selman. Multi-agent curricula and emergent implicit signaling. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '22, pp. 553–561, Richland, SC, 2022. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 9781450392136.

Shixiang Gu, Ethan Holly, Timothy Lillicrap, and Sergey Levine. Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates, 2016. URL `https://arxiv.org/abs/1610.00633`.

Shengyue Guan, Haoyi Xiong, Jindong Wang, Jiang Bian, Bin Zhu, and Jian-guang Lou. Evaluating llm-based agents for multi-turn conversations: A survey. *arXiv preprint arXiv:2503.22458*, 2025.

Shangmin Guo. Emergence of numeric concepts in multi-agent autonomous communication. *arXiv preprint arXiv:1911.01098*, 2019.

Shangmin Guo, Yi Ren, Kory Wallace Mathewson, Simon Kirby, Stefano V Albrecht, and Kenny Smith. Expressivity of emergent languages is a trade-off between contextual complexity and unpredictability. In *International Conference on Learning Representations*.

Shangmin Guo, Yi Ren, Kory Mathewson, Simon Kirby, Stefano V Albrecht, and Kenny Smith. Expressivity of emergent language is a trade-off between contextual complexity and unpredictability. *arXiv preprint arXiv:2106.03982*, 2021.

Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V Chawla, Olaf Wiest, and Xiangliang Zhang. Large language model based multi-agents: A survey of progress and challenges. *arXiv preprint arXiv:2402.01680*, 2024a.

Taicheng Guo, Xiuying Chen, Yaqi Wang, Ruidi Chang, Shichao Pei, Nitesh V Chawla, Olaf Wiest, and Xiangliang Zhang. Large language model based multi-agents: A survey of progress and challenges. *arXiv preprint arXiv:2402.01680*, 2024b.

Xudong Guo, Kaixuan Huang, Jiale Liu, Wenhui Fan, Natalia Vélez, Qingyun Wu, Huazheng Wang, Thomas L Griffiths, and Mengdi Wang. Embodied llm agents learn to cooperate in organized teams. *arXiv preprint arXiv:2403.12482*, 2024c.

Abhinav Gupta, Marc Lanctot, and Angeliki Lazaridou. Dynamic population-based meta-learning for multi-agent communication with natural language. *Advances in Neural Information Processing Systems*, 34: 16899–16912, 2021.

Nikunj Gupta, G Srinivasaraghavan, Swarup Mohalik, Nishant Kumar, and Matthew E Taylor. Hammer: Multi-level coordination of reinforcement learning agents via learned messaging. *Neural Computing and Applications*, pp. 1–16, 2023.

Shubham Gupta, Rishi Hazra, and Ambedkar Dukkipati. Networked multi-agent reinforcement learning with emergent communication. *arXiv preprint arXiv:2004.02780*, 2020.

Thorsten Händler. Balancing autonomy and alignment: a multi-dimensional taxonomy for autonomous llm-powered multi-agent architectures. *arXiv preprint arXiv:2310.03659*, 2023.

Matthew Hausknecht and Peter Stone. Deep recurrent q-learning for partially observable mdps. In *2015 aaai fall symposium series*, 2015.

Serhii Havrylov and Ivan Titov. Emergence of language with multi-agent games: Learning to communicate with sequences of symbols. *Advances in neural information processing systems*, 30, 2017.

Pablo Hernandez-Leal, Bilal Kartal, and Matthew E Taylor. A survey and critique of multiagent deep reinforcement learning. *Autonomous Agents and Multi-Agent Systems*, 33(6):750–797, 2019.

Samuel Holt, Max Ruiz Luyten, and Mihaela van der Schaar. L2mac: Large language model automatic computer for extensive code generation. In *The Twelfth International Conference on Learning Representations*.

Sirui Hong, Xiawu Zheng, Jonathan Chen, Yuheng Cheng, Jinlin Wang, Ceyao Zhang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, Liyang Zhou, et al. Metagpt: Meta programming for multi-agent collaborative framework. *arXiv preprint arXiv:2308.00352*, 2023.

Guangzheng Hu, Yuanheng Zhu, Dongbin Zhao, Mengchen Zhao, and Jianye Hao. Event-triggered multi-agent reinforcement learning with communication under limited-bandwidth constraint. *arXiv preprint arXiv:2010.04978*, 2020a.

Hengyuan Hu, Adam Lerer, Alex Peysakhovich, and Jakob Foerster. "other-play" for zero-shot coordination. In *International Conference on Machine Learning*, pp. 4399–4410. PMLR, 2020b.

Shengran Hu, Cong Lu, and Jeff Clune. Automated design of agentic systems. *arXiv preprint arXiv:2408.08435*, 2024a.

Sihao Hu, Tiansheng Huang, Fatih Ilhan, Selim Tekin, Gaowen Liu, Ramana Kompella, and Ling Liu. A survey on large language model-based game agents. *arXiv preprint arXiv:2404.02039*, 2024b.

Sihao Hu, Tiansheng Huang, Fatih Ilhan, Selim Tekin, Gaowen Liu, Ramana Kompella, and Ling Liu. A survey on large language model-based game agents. *arXiv preprint arXiv:2404.02039*, 2024c.

Yuwei Hu, Runlin Lei, Xinyi Huang, Zhewei Wei, and Yongchao Liu. Scalable and accurate graph reasoning with llm-based multi-agents. *arXiv preprint arXiv:2410.05130*, 2024d.

Wenyue Hua, Lizhou Fan, Lingyao Li, Kai Mei, Jianchao Ji, Yingqiang Ge, Libby Hemphill, and Yongfeng Zhang. War and peace (waragent): Large language model-based multi-agent simulation of world wars. *arXiv preprint arXiv:2311.17227*, 2023.

Kiran Ikram, Esther Mondragón, Eduardo Alonso, and M Garcia-Ortiz. Hexajungle: a marl simulator to study the emergence of language. 2021.

Luca Iocchi et al. Emergent communication in human-machine games. In *Emergent Communication Workshop at ICLR 2022*, 2022.

Yoichi Ishibashi and Yoshimasa Nishimura. Self-organized agents: A llm multi-agent framework toward ultra large-scale code generation and optimization. *arXiv preprint arXiv:2404.02183*, 2024.

Md Ashraful Islam, Mohammed Eunus Ali, and Md Rizwan Parvez. Mapcoder: Multi-agent code generation for competitive problem solving. *arXiv preprint arXiv:2405.11403*, 2024.

Shankar Kumar Jeyakumar, Alaa Alameer Ahmad, and Adrian Garret Gabriel. Advancing agentic systems: Dynamic task decomposition, tool integration and evaluation using novel metrics and dataset. In *NeurIPS 2024 Workshop on Open-World Agents*, 2024.

Dongfu Jiang, Xiang Ren, and Bill Yuchen Lin. Llm-blender: Ensembling large language models with pairwise ranking and generative fusion. *arXiv preprint arXiv:2306.02561*, 2023.

Jiechuan Jiang and Zongqing Lu. Learning attentional communication for multi-agent cooperation. *Advances in neural information processing systems*, 31, 2018.

Jiechuan Jiang, Chen Dun, Tiejun Huang, and Zongqing Lu. Graph convolutional reinforcement learning. *arXiv preprint arXiv:1810.09202*, 2018.

Xuanfa Jin, Ziyan Wang, Yali Du, Meng Fang, Haifeng Zhang, and Jun Wang. Learning to discuss strategically: A case study on one night ultimate werewolf. *Advances in Neural Information Processing Systems*, 37:77060–77097, 2024.

Shi Jinxin, Zhao Jiabao, Wang Yilei, Wu Xingjiao, Li Jiawen, and He Liang. Cgmi: Configurable general multi-agent interaction framework. *arXiv preprint arXiv:2308.12503*, 2023.

Emilio Jorge, Mikael Kågebäck, Fredrik D Johansson, and Emil Gustavsson. Learning to play guess who? and inventing a grounded language as a consequence. *arXiv preprint arXiv:1611.03218*, 2016.

Zhao Kaiya, Michelangelo Naim, Jovana Kondic, Manuel Cortes, Jiaxin Ge, Shuying Luo, Guangyu Robert Yang, and Andrew Ahn. Lyfe agents: Generative agents for low-cost real-time social interactions. *arXiv preprint arXiv:2310.02172*, 2023.

Aleksandra Kalinowska, Elnaz Davoodi, Florian Strub, Kory Mathewson, Todd Murphey, and Patrick Pilarski. Situated communication: A solution to over-communication between artificial agents. In *Emergent Communication Workshop at ICLR 2022*, 2022.

Yipeng Kang, Tonghan Wang, and Gerard de Melo. Incorporating pragmatic reasoning communication into emergent language. *Advances in neural information processing systems*, 33:10348–10359, 2020.

Shyam Sundar Kannan, Vishnunandan LN Venkatesh, and Byung-Cheol Min. Smart-llm: Smart multi-agent robot task planning using large language models. *arXiv preprint arXiv:2309.10062*, 2023.

Eugene Kharitonov and Marco Baroni. Emergent language generalization and acquisition speed are not tied to compositionality. *arXiv preprint arXiv:2004.03420*, 2020.

Eugene Kharitonov, Rahma Chaabouni, Diane Bouchacourt, and Marco Baroni. Egg: a toolkit for research on emergence of language in games. *arXiv preprint arXiv:1907.00852*, 2019.

Eugene Kharitonov, Rahma Chaabouni, Diane Bouchacourt, and Marco Baroni. Entropy minimization in emergent languages. In *International Conference on Machine Learning*, pp. 5220–5230. PMLR, 2020.

Diksha Khurana, Aditya Koli, Kiran Khatter, and Sukhdev Singh. Natural language processing: state of the art, current trends and challenges. *Multimedia tools and applications*, 82(3):3713–3744, 2023.

Ozsel Kilinc and Giovanni Montana. Multi-agent deep reinforcement learning with extremely noisy observations. *arXiv preprint arXiv:1812.00922*, 2018.

Daewoo Kim, Sangwoo Moon, David Hostallero, Wan Ju Kang, Taeyoung Lee, Kyunghwan Son, and Yung Yi. Learning to schedule communication in multi-agent reinforcement learning. *arXiv preprint arXiv:1902.01554*, 2019a.

Woojun Kim, Myungsik Cho, and Youngchul Sung. Message-dropout: An efficient training method for multi-agent deep reinforcement learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 6079–6086, 2019b.

Woojun Kim, Jongeui Park, and Youngchul Sung. Communication in multi-agent reinforcement learning: Intention sharing. In *International conference on learning representations*, 2020.

Simon Kirby. Natural language from artificial life. *Artificial life*, 8(2):185–215, 2002.

Simon Kirby, Mike Dowman, and Thomas L Griffiths. Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, 104(12):5241–5245, 2007.

Simon Kirby, Hannah Cornish, and Kenny Smith. Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31):10681–10686, 2008.

Ross A Knepper, Christoforos I Mavrogiannis, Julia Proft, and Claire Liang. Implicit communication in a joint action. In *Proceedings of the 2017 acm/ieee international conference on human-robot interaction*, pp. 283–292, 2017.

Benjamin Kolb, Leon Lang, Henning Bartsch, Arwin Gansekoele, Raymond Koopmanschap, Leonardo Romor, David Speck, Mathijs Mul, and Elia Bruni. Learning to request guidance in emergent language. In *Proceedings of the Beyond Vision and LANguage: inTEgrating Real-world kNowledge (LANTERN)*, pp. 41–50, 2019.

Xiangyu Kong, Bo Xin, Fangchen Liu, and Yizhou Wang. Revisiting the master-slave architecture in multi-agent deep reinforcement learning. *arXiv preprint arXiv:1712.07305*, 2017.

Tomasz Korbak, Julian Zubek, Łukasz Kuciński, Piotr Miłoś, and Joanna Rączaszek-Leonardi. Developmentally motivated emergence of compositional communication via template transfer. *arXiv preprint arXiv:1910.06079*, 2019.

Tomasz Korbak, Julian Zubek, and Joanna Rączaszek-Leonardi. Measuring non-trivial compositionality in emergent communication. *arXiv preprint arXiv:2010.15058*, 2020.

Satwik Kottur, José MF Moura, Stefan Lee, and Dhruv Batra. Natural language does not emerge'naturally'in multi-agent dialog. *arXiv preprint arXiv:1706.08502*, 2017.

Landon Kraemer and Bikramjit Banerjee. Multi-agent reinforcement learning as a rehearsal for decentralized planning. *Neurocomputing*, 190:82–94, 2016.

Nikolaus Kriegeskorte, Marieke Mur, and Peter A Bandettini. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2:249, 2008.

Łukasz Kuciński, Paweł Kołodziej, and Piotr Miłoś. Emergence of compositional language in communication through noisy channel. In *Language in Reinforcement Learning Workshop at ICML 2020*, 2020.

Travis LaCroix. Biology and compositionality: Empirical considerations for emergent-communication protocols. *arXiv preprint arXiv:1911.11668*, 2019.

Guillaume Lample and Devendra Singh Chaplot. Playing fps games with deep reinforcement learning, 2016. URL `https://arxiv.org/abs/1609.05521`.

Yihuai Lan, Zhiqiang Hu, Lei Wang, Yang Wang, Deheng Ye, Peilin Zhao, Ee-Peng Lim, Hui Xiong, and Hao Wang. Llm-based agent society investigation: Collaboration and confrontation in avalon gameplay. *arXiv preprint arXiv:2310.14985*, 2023.

Ivano Lauriola, Alberto Lavelli, and Fabio Aiolli. An introduction to deep learning in natural language processing: Models, techniques, and tools. *Neurocomputing*, 470:443–456, 2022.

Angeliki Lazaridou and Marco Baroni. Emergent multi-agent communication in the deep learning era. *arXiv preprint arXiv:2006.02419*, 2020.

Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. Multi-agent cooperation and the emergence of (natural) language. *arXiv preprint arXiv:1612.07182*, 2016.

Angeliki Lazaridou, Karl Moritz Hermann, Karl Tuyls, and Stephen Clark. Emergence of linguistic communication from referential games with symbolic and pixel input. *arXiv preprint arXiv:1804.03984*, 2018.

Angeliki Lazaridou, Anna Potapenko, and Olivier Tieleman. Multi-agent communication meets natural language: Synergies between functional and structural language learning. *arXiv preprint arXiv:2005.07064*, 2020.

Angeliki Lazaridou, Adhi Kuncoro, Elena Gribovskaya, Devang Agrawal, Adam Liska, Tayfun Terzi, Mai Gimenez, Cyprien de Masson d'Autume, Tomas Kocisky, Sebastian Ruder, et al. Mind the gap: Assessing temporal generalization in neural language models. *Advances in Neural Information Processing Systems*, 34:29348–29363, 2021.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

Jason Lee, Kyunghyun Cho, Jason Weston, and Douwe Kiela. Emergent translation in multi-agent communication. *arXiv preprint arXiv:1710.06922*, 2017.

Jason Lee, Kyunghyun Cho, and Douwe Kiela. Countering language drift via visual grounding. *arXiv preprint arXiv:1909.04499*, 2019.

Joel Z Leibo, Vinicius Zambaldi, Marc Lanctot, Janusz Marecki, and Thore Graepel. Multi-agent reinforcement learning in sequential social dilemmas. *arXiv preprint arXiv:1702.03037*, 2017.

Oliver Lemon. Conversational grounding in emergent communication–data and divergence. In *Emergent Communication Workshop at ICLR 2022*, 2022.

David Lewis. *Convention: A philosophical study*. John Wiley & Sons, 2008.

Cheng Li, Damien Teney, Linyi Yang, Qingsong Wen, Xing Xie, and Jindong Wang. Culturepark: Boosting cross-cultural understanding in large language models. *arXiv preprint arXiv:2405.15145*, 2024a.

Dapeng Li, Zhiwei Xu, Bin Zhang, and Guoliang Fan. From explicit communication to tacit cooperation: A novel paradigm for cooperative marl. *arXiv preprint arXiv:2304.14656*, 2023a.

Fushan Li and Michael Bowling. Ease-of-teaching and language structure from emergent communication. *Advances in neural information processing systems*, 32, 2019.

Guohao Li, Hasan Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. Camel: Communicative agents for" mind" exploration of large language model society. *Advances in Neural Information Processing Systems*, 36:51991–52008, 2023b.

Huao Li, Hossein Nourkhiz Mahjoub, Behdad Chalaki, Vaishnav Tadiparthi, Kwonjoon Lee, Ehsan Moradi Pari, Charles Michael Lewis, and Katia P Sycara. Language grounded multi-agent reinforcement learning with human-interpretable communication. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*.

Huao Li, Yu Quan Chong, Simon Stepputtis, Joseph Campbell, Dana Hughes, Michael Lewis, and Katia Sycara. Theory of mind for multi-agent collaboration via large language models. *arXiv preprint arXiv:2310.10701*, 2023c.

Junkai Li, Yunghwei Lai, Weitao Li, Jingyi Ren, Meng Zhang, Xinhui Kang, Siyu Wang, Peng Li, Ya-Qin Zhang, Weizhi Ma, et al. Agent hospital: A simulacrum of hospital with evolvable medical agents. *arXiv preprint arXiv:2405.02957*, 2024b.

Junyou Li, Qin Zhang, Yangbin Yu, Qiang Fu, and Deheng Ye. More agents is all you need. *arXiv preprint arXiv:2402.05120*, 2024c.

Ling Li, Alcherio Martinoli, and Yaser S. Abu-Mostafa. Emergent specialization in swarm systems. In *Proceedings of the Third International Conference on Intelligent Data Engineering and Automated Learning*, IDEAL '02, pp. 261–266, Berlin, Heidelberg, 2002. Springer-Verlag. ISBN 3540440259.

Sheng Li, Jayesh K. Gupta, Peter Morales, Ross Allen, and Mykel J. Kochenderfer. Deep implicit coordination graphs for multi-agent reinforcement learning, 2020. URL `https://arxiv.org/abs/2006.11438`.

Sheng Li, Yutai Zhou, Ross Allen, and Mykel J Kochenderfer. Learning emergent discrete message communication for cooperative reinforcement learning. In *2022 International Conference on Robotics and Automation (ICRA)*, pp. 5511–5517. IEEE, 2022.

Xinyi Li, Sai Wang, Siqi Zeng, Yu Wu, and Yi Yang. A survey on llm-based multi-agent systems: workflow, infrastructure, and challenges. *Vicinagearth*, 1(1):9, 2024d.

Yunxuan Li, Yibing Du, Jiageng Zhang, Le Hou, Peter Grabowski, Yeqing Li, and Eugene Ie. Improving multi-agent debate with sparse communication topology. *arXiv preprint arXiv:2406.11776*, 2024e.

Paul Pu Liang, Jeffrey Chen, Ruslan Salakhutdinov, Louis-Philippe Morency, and Satwik Kottur. On emergent communication in competitive multi-agent teams. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 735–743, 2020.

Percy Liang, Rishi Bommasani, Tony Lee, Dimitris Tsipras, Dilara Soylu, Michihiro Yasunaga, Yian Zhang, Deepak Narayanan, Yuhuai Wu, Ananya Kumar, et al. Holistic evaluation of language models. *arXiv preprint arXiv:2211.09110*, 2022.

Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and Zhaopeng Tu. Encouraging divergent thinking in large language models through multi-agent debate. *arXiv preprint arXiv:2305.19118*, 2023.

Jonathan Light, Min Cai, Sheng Shen, and Ziniu Hu. Avalonbench: Evaluating llms playing the game of avalon. *arXiv preprint arXiv:2310.05036*, 2023.

Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning, 2015. URL `https://arxiv.org/abs/1509.02971`.

Toru Lin, Jacob Huh, Christopher Stauffer, Ser Nam Lim, and Phillip Isola. Learning to ground multi-agent communication with autoencoders. *Advances in Neural Information Processing Systems*, 34:15230–15242, 2021.

Dorota Lipowska and Adam Lipowski. Emergence and evolution of language in multi-agent systems. *Lingua*, 272:103331, 2022.

Bang Liu, Xinfeng Li, Jiayi Zhang, Jinlin Wang, Tanjin He, Sirui Hong, Hongzhang Liu, Shaokun Zhang, Kaitao Song, Kunlun Zhu, et al. Advances and challenges in foundation agents: From brain-inspired intelligence to evolutionary, collaborative, and safe systems. *arXiv preprint arXiv:2504.01990*, 2025a.

Bang Liu, Xinfeng Li, Jiayi Zhang, Jinlin Wang, Tanjin He, Sirui Hong, Hongzhang Liu, Shaokun Zhang, Kaitao Song, Kunlun Zhu, et al. Advances and challenges in foundation agents: From brain-inspired intelligence to evolutionary, collaborative, and safe systems. *arXiv preprint arXiv:2504.01990*, 2025b.

Siqi Liu, Guy Lever, Josh Merel, Saran Tunyasuvunakool, Nicolas Heess, and Thore Graepel. Emergent coordination through competition. In *International Conference on Learning Representations*.

Wei Liu, Chenxi Wang, Yifei Wang, Zihao Xie, Rennai Qiu, Yufan Dang, Zhuoyun Du, Weize Chen, Cheng Yang, and Chen Qian. Autonomous agents for collaborative task under information asymmetry. *arXiv preprint arXiv:2406.14928*, 2024a.

Yong Liu, Weixun Wang, Yujing Hu, Jianye Hao, Xingguo Chen, and Yang Gao. Multi-agent game abstraction via graph attention neural network. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 7211–7218, 2020.

Zhiwei Liu, Weiran Yao, Jianguo Zhang, Le Xue, Shelby Heinecke, Rithesh Murthy, Yihao Feng, Zeyuan Chen, Juan Carlos Niebles, Devansh Arpit, et al. Bolaa: Benchmarking and orchestrating llm-augmented autonomous agents. *arXiv preprint arXiv:2308.05960*, 2023a.

Zijun Liu, Yanzhe Zhang, Peng Li, Yang Liu, and Diyi Yang. Dynamic llm-agent network: An llm-agent collaboration framework with agent team optimization. *arXiv preprint arXiv:2310.02170*, 2023b.

Zijun Liu, Yanzhe Zhang, Peng Li, Yang Liu, and Diyi Yang. A dynamic llm-powered agent network for task-oriented agent collaboration. In *First Conference on Language Modeling*, 2024b.

Yat Long Lo and Biswa Sengupta. Learning to ground decentralized multi-agent communication with contrastive learning. In *Emergent Communication Workshop at ICLR 2022*.

Ryan Lowe, Abhinav Gupta, Jakob Foerster, Douwe Kiela, and Joelle Pineau. On the interaction between supervision and self-play in emergent communication. In *International Conference on Learning Representations*.

Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments, 2017. URL `https://arxiv.org/abs/1706.02275`.

Ryan Lowe, Jakob Foerster, Y-Lan Boureau, Joelle Pineau, and Yann Dauphin. On the pitfalls of measuring emergent communication. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '19, pp. 693–701, Richland, SC, 2019. International Foundation for Autonomous Agents and Multiagent Systems. ISBN 9781450363099.

Ryan Lowe, Abhinav Gupta, Jakob Foerster, Douwe Kiela, and Joelle Pineau. On the interaction between supervision and self-play in emergent communication. *arXiv preprint arXiv:2002.01093*, 2020.

Diana Rodríguez Luna, Edoardo Maria Ponti, Dieuwke Hupkes, and Elia Bruni. Internal and external pressures on language emergence: least effort, object constancy and frequency. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pp. 4428–4437, 2020.

Weiyu Ma, Qirui Mi, Yongcheng Zeng, Xue Yan, Yuqiao Wu, Runji Lin, Haifeng Zhang, and Jun Wang. Large language models play starcraft ii: Benchmarks and a chain of summarization approach. *arXiv preprint arXiv:2312.11865*, 2023.

Aleksandra Malysheva, Tegg Taekyong Sung, Chae-Bong Sohn, Daniel Kudenko, and Aleksei Shpilman. Deep multi-agent reinforcement learning with relevance graphs. *arXiv preprint arXiv:1811.12557*, 2018.

Zhao Mandi, Shreeya Jain, and Shuran Song. Roco: Dialectic multi-robot collaboration with large language models. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 286–299. IEEE, 2024.

Christopher Manning and Hinrich Schutze. *Foundations of statistical natural language processing*. MIT press, 1999.

Hangyu Mao, Zhengchao Zhang, Zhen Xiao, Zhibo Gong, and Yan Ni. Learning agent communication under limited bandwidth by message pruning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 5142–5149, 2020.

Laetitia Matignon, Guillaume J Laurent, and Nadine Le Fort-Piat. Independent reinforcement learners in cooperative markov games: a survey regarding coordination problems. *The Knowledge Engineering Review*, 27(1):1–31, 2012.

Lingheng Meng, Daiwei Lin, Adam Francey, Rob Gorbet, Philip Beesley, and Dana Kulić. Learning to engage with interactive systems: A field study on deep reinforcement learning in a public museum, 2019. URL `https://arxiv.org/abs/1904.06764`.

William Merrill, Yoav Goldberg, Roy Schwartz, and Noah A Smith. Provable limitations of acquiring meaning from ungrounded form: What will future language models understand? *Transactions of the Association for Computational Linguistics*, 9:1047–1060, 2021.

Grégoire Mialon, Clémentine Fourrier, Thomas Wolf, Yann LeCun, and Thomas Scialom. Gaia: a benchmark for general ai assistants. In *The Twelfth International Conference on Learning Representations*, 2023.

Daniela Mihai and Jonathon Hare. The emergence of visual semantics through communication games. *arXiv preprint arXiv:2101.10253*, 2021.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning, 2013. URL `https://arxiv.org/abs/1312.5602`.

Igor Mordatch and Pieter Abbeel. Emergence of grounded compositional language in multi-agent populations. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Manuel Mosquera, Juan Sebastian Pinzon, Manuel Rios, Yesid Fonseca, Luis Felipe Giraldo, Nicanor Quijano, and Ruben Manrique. Can llm-augmented autonomous agents cooperate?, an evaluation of their cooperative capabilities through melting pot. *arXiv preprint arXiv:2403.11381*, 2024.

Clément Moulin-Frier and Pierre-Yves Oudeyer. Multi-agent reinforcement learning as a computational tool for language evolution research: Historical context and future challenges. *arXiv preprint arXiv:2002.08878*, 2020.

Jesse Mu and Noah Goodman. Emergent communication of generalizations. *Advances in neural information processing systems*, 34:17994–18007, 2021.

Mathijs Mul, Diane Bouchacourt, and Elia Bruni. Mastering emergent language: learning to guide in simulated navigation. *arXiv preprint arXiv:1908.05135*, 2019.

Nathalia Nascimento, Paulo Alencar, and Donald Cowan. Gpt-in-the-loop: Adaptive decision-making for multiagent systems. *arXiv preprint arXiv:2308.10435*, 2023a.

Nathalia Nascimento, Paulo Alencar, and Donald Cowan. Self-adaptive large language model (llm)-based multiagent systems. In *2023 IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion (ACSOS-C)*, pp. 104–109. IEEE, 2023b.

Thanh Thi Nguyen, Ngoc Duy Nguyen, and Saeid Nahavandi. Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. *IEEE transactions on cybernetics*, 50(9): 3826–3839, 2020.

Xuefei Ning, Zinan Lin, Zixuan Zhou, Zifu Wang, Huazhong Yang, and Yu Wang. Skeleton-of-thought: Prompting llms for efficient parallel generation. *arXiv preprint arXiv:2307.15337*, 2023.

Zepeng Ning and Lihua Xie. A survey on multi-agent reinforcement learning and its application. *Journal of Automation and Intelligence*, 3(2):73–91, 2024.

Eleni Nisioti, Elías Masquil, Gautier Hamon, , and Clément Moulin-Frier. Autotelic reinforcement learning in multi-agent environments, 2023. URL https://arxiv.org/abs/2211.06082.

Yaru Niu, Rohan R Paleja, and Matthew C Gombolay. Multi-agent graph-attention communication and teaming. In *AAMAS*, pp. 964–973, 2021.

Michael Noukhovitch, Travis LaCroix, Angeliki Lazaridou, and Aaron Courville. Emergent communication under competition. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 974–982, 2021.

Martin A Nowak and David C Krakauer. The evolution of language. *Proceedings of the National Academy of Sciences*, 96(14):8028–8033, 1999.

Frans A Oliehoek, Christopher Amato, et al. *A concise introduction to decentralized POMDPs*, volume 1. Springer, 2016.

Michael Oliphant and John Batali. Learning and the emergence of coordinated communication. *Center for research on language newsletter*, 11(1):1–46, 1997.

OpenAI. Openai swarm. https://github.com/openai/swarm.

Afshin Oroojlooy and Davood Hajinezhad. A review of cooperative multi-agent deep reinforcement learning. *Applied Intelligence*, 53(11):13677–13722, 2023.

Joni Pajarinen and Ville Kyrki. Robotic manipulation of multiple objects as a pomdp. *Artificial Intelligence*, 247:213–228, 2017.

Bo Pan, Jiaying Lu, Ke Wang, Li Zheng, Zhen Wen, Yingchaojie Feng, Minfeng Zhu, and Wei Chen. Agentcoord: Visually exploring coordination strategy for llm-based multi-agent collaboration. *arXiv preprint arXiv:2404.11943*, 2024.

Joon Sung Park, Joseph O'Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pp. 1–22, 2023.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Shivansh Patel, Saim Wani, Unnat Jain, Alexander G Schwing, Svetlana Lazebnik, Manolis Savva, and Angel X Chang. Interpretation of emergent communication in heterogeneous collaborative embodied agents. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 15953–15963, 2021.

James Paulos, Steven W Chen, Daigo Shishika, and Vijay Kumar. Decentralization of multiagent policies by learning what to communicate. In *2019 International Conference on Robotics and Automation (ICRA)*, pp. 7990–7996. IEEE, 2019.

Peng Peng, Ying Wen, Yaodong Yang, Quan Yuan, Zhenkun Tang, Haitao Long, and Jun Wang. Multiagent bidirectionally-coordinated nets: Emergence of human-level coordination in learning to play starcraft combat games, 2017. URL https://arxiv.org/abs/1703.10069.

Hugh Perkins. Neural networks can understand compositional functions that humans do not, in the context of emergent communication. *arXiv preprint arXiv:2103.04180*, 2021a.

Hugh Perkins. Texrel: a green family of datasets for emergent communications on relations. *arXiv preprint arXiv:2105.12804*, 2021b.

Emanuele Pesce and Giovanni Montana. Improving coordination in small-scale multi-agent deep reinforcement learning through memory-driven communication. *Machine Learning*, 109(9-10):1727–1747, jan 2020a. doi: 10.1007/s10994-019-05864-5. URL https://doi.org/10.1007%2Fs10994-019-05864-5.

Emanuele Pesce and Giovanni Montana. Improving coordination in small-scale multi-agent deep reinforcement learning through memory-driven communication. *Machine Learning*, 109(9):1727–1747, 2020b.

Jannik Peters, Constantin Waubert de Puiseau, Hasan Tercan, Arya Gopikrishnan, Gustavo Adolpho Lucas De Carvalho, Christian Bitter, and Tobias Meisen. A survey on emergent language. *arXiv preprint arXiv:2409.02645*, 2024.

Giorgio Piatti, Zhijing Jin, Max Kleiman-Weiner, Bernhard Schölkopf, Mrinmaya Sachan, and Rada Mihalcea. Cooperate or collapse: Emergence of sustainable cooperation in a society of llm agents. *Advances in Neural Information Processing Systems*, 37:111715–111759, 2024.

M. Pipattanasomporn, H. Feroze, and S. Rahman. Multi-agent systems in a distributed smart grid: Design and implementation. In *2009 IEEE/PES Power Systems Conference and Exposition*, pp. 1–8, 2009. doi: 10.1109/PSCE.2009.4840087.

Mikhail Prokopenko and Peter Wang. Relating the entropy of joint beliefs to multi-agent coordination. In *RoboCup 2002: Robot Soccer World Cup VI 6*, pp. 367–374. Springer, 2003.

Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. 2014.

Chen Qian, Xin Cong, Cheng Yang, Weize Chen, Yusheng Su, Juyuan Xu, Zhiyuan Liu, and Maosong Sun. Communicative agents for software development. *arXiv preprint arXiv:2307.07924*, 6, 2023.

Chen Qian, Wei Liu, Hongzhang Liu, Nuo Chen, Yufan Dang, Jiahao Li, Cheng Yang, Weize Chen, Yusheng Su, Xin Cong, et al. Chatdev: Communicative agents for software development. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 15174–15186, 2024a.

Chen Qian, Zihao Xie, Yifei Wang, Wei Liu, Yufan Dang, Zhuoyun Du, Weize Chen, Cheng Yang, Zhiyuan Liu, and Maosong Sun. Scaling large-language-model-based multi-agent collaboration. *arXiv preprint arXiv:2406.07155*, 2024b.

Shuofei Qiao, Ningyu Zhang, Runnan Fang, Yujie Luo, Wangchunshu Zhou, Yuchen Eleanor Jiang, Chengfei Lv, and Huajun Chen. Autoact: Automatic agent learning from scratch for qa via self-planning. *arXiv preprint arXiv:2401.05268*, 2024.

Shuwen Qiu, Sirui Xie, Lifeng Fan, Tao Gao, Jungseock Joo, Song-Chun Zhu, and Yixin Zhu. Emergent graphical conventions in a visual communication game. *Advances in Neural Information Processing Systems*, 35:13119–13131, 2022.

Xipeng Qiu, Tianxiang Sun, Yige Xu, Yunfan Shao, Ning Dai, and Xuanjing Huang. Pre-trained models for natural language processing: A survey. *Science China technological sciences*, 63(10):1872–1897, 2020.

Chao Qu, Hui Li, Chang Liu, Junwu Xiong, Wei Chu, Weiqiang Wang, Yuan Qi, Le Song, et al. Intention propagation for multi-agent reinforcement learning. 2020.

Ahmed Hussain Qureshi, Yutaka Nakamura, Yuichiro Yoshikawa, and Hiroshi Ishiguro. Robot gains social intelligence through multimodal deep reinforcement learning, 2017. URL https://arxiv.org/abs/1702.07492.

Murtaza Rangwala and Ryan Williams. Learning multi-agent communication through structured attentive reasoning. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS'20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

Sumedh Rasal and EJ Hauer. Navigating complexity: Orchestrated problem solving with multi-agent llms. *arXiv preprint arXiv:2402.16713*, 2024.

Tabish Rashid, Mikayel Samvelyan, Christian Schroeder, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning. In *International conference on machine learning*, pp. 4295–4304. PMLR, 2018.

Yi Ren, Shangmin Guo, Matthieu Labeau, Shay B Cohen, and Simon Kirby. Compositional languages emerge in a neural iterated learning model. *arXiv preprint arXiv:2002.01365*, 2020.

Cinjon Resnick, Abhinav Gupta, Jakob Foerster, Andrew M Dai, and Kyunghyun Cho. Capacity, bandwidth, and compositionality in emergent language learning. *arXiv preprint arXiv:1910.11424*, 2019.

Cinjon Resnick, Abhinav Gupta, Jakob Foerster, Andrew M Dai, and Kyunghyun Cho. Capacity, bandwidth, and compositionality in emergent language learning. In *19th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2020*, pp. 1125–1133. International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), 2020.

Mathieu Rita, Florian Strub, Jean-Bastien Grill, Olivier Pietquin, and Emmanuel Dupoux. On the role of population heterogeneity in emergent communication. In *ICLP 2022-Tenth International Conference on Learning Representations*, 2022.

Kai Ruan, Mowen Huang, Ji-Rong Wen, and Hao Sun. Benchmarking llms' swarm intelligence. *arXiv preprint arXiv:2505.04364*, 2025.

Heechang Ryu, Hayong Shin, and Jinkyoo Park. Cooperative and competitive biases for multi-agent reinforcement learning. *arXiv preprint arXiv:2101.06890*, 2021.

Mikayel Samvelyan, Tabish Rashid, Christian Schroeder De Witt, Gregory Farquhar, Nantas Nardelli, Tim GJ Rudner, Chia-Man Hung, Philip HS Torr, Jakob Foerster, and Shimon Whiteson. The starcraft multi-agent challenge. *arXiv preprint arXiv:1902.04043*, 2019.

Pedro P Santos, Diogo S Carvalho, Miguel Vasco, Alberto Sardinha, Pedro A Santos, Ana Paiva, and Francisco S Melo. Centralized training with hybrid execution in multi-agent reinforcement learning. *arXiv preprint arXiv:2210.06274*, 2022.

Samuel Schmidgall, Yusheng Su, Ze Wang, Ximeng Sun, Jialian Wu, Xiaodong Yu, Jiang Liu, Zicheng Liu, and Emad Barsoum. Agent laboratory: Using llm agents as research assistants. *arXiv preprint arXiv:2501.04227*, 2025.

John Schulman, Sergey Levine, Philipp Moritz, Michael I. Jordan, and Pieter Abbeel. Trust region policy optimization, 2015. URL `https://arxiv.org/abs/1502.05477`.

Samuel Shaw, Emerson Wenzel, Alexis Walker, and Guillaume Sartoretti. Formic: Foraging via multiagent rl with implicit communication. *IEEE Robotics and Automation Letters*, 7(2):4877–4884, 2022.

Junjie Sheng, Xiangfeng Wang, Bo Jin, Junchi Yan, Wenhao Li, Tsung-Hui Chang, Jun Wang, and Hongyuan Zha. Learning structured communication for multi-agent reinforcement learning. *Autonomous Agents and Multi-Agent Systems*, 36(2):1–31, 2022.

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning. *Advances in Neural Information Processing Systems*, 36, 2024.

David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, et al. Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv preprint arXiv:1712.01815*, 2017.

David Simões, Nuno Lau, and Luís Paulo Reis. Multi-agent actor centralized-critic with communication. *Neurocomputing*, 390:40–56, 2020.

Amanpreet Singh, Tushar Jain, and Sainbayar Sukhbaatar. Learning when to communicate at scale in multiagent cooperative and competitive tasks, 2018. URL `https://arxiv.org/abs/1812.09755`.

Brian Skyrms. Signals, evolution and the explanatory power of transient information. *Philosophy of Science*, 69(3):407–428, 2002.

Oliver Slumbers, David Henry Mguni, Kun Shao, and Jun Wang. Leveraging large language models for optimised coordination in textual multi-agent reinforcement learning. 2023.

Kenny Smith, Simon Kirby, and Henry Brighton. Iterated learning: A framework for the emergence of language. *Artificial life*, 9(4):371–386, 2003.

Luc Steels. A self-organizing spatial vocabulary. *Artificial life*, 2(3):319–332, 1995.

Luc Steels. The synthetic modeling of language origins. *Evolution of communication*, 1(1):1–34, 1997.

Shane Steinert-Threlkeld, Xuhui Zhou, Zeyu Liu, and CM Downey. Emergent communication fine-tuning (ec-ft) for pretrained language models. In *Emergent Communication Workshop at ICLR 2022*, 2022.

Simon Stepputtis, Joseph Campbell, Yaqi Xie, Zhengyang Qi, Wenxin Sharon Zhang, Ruiyi Wang, Sanketh Rangreji, Michael Lewis, and Katia Sycara. Long-horizon dialogue understanding for role identification in the game of avalon with large language models. *arXiv preprint arXiv:2311.05720*, 2023.

Alessandro Suglia, Ioannis Konstas, and Oliver Lemon. Visually grounded language learning: a review of language games, datasets, tasks, and models. *Journal of Artificial Intelligence Research*, 79:173–239, 2024.

Sainbayar Sukhbaatar, Rob Fergus, et al. Learning multiagent communication with backpropagation. *Advances in neural information processing systems*, 29, 2016.

Chuanneng Sun, Songjun Huang, and Dario Pompili. Llm-based multi-agent reinforcement learning: Current and future directions. *arXiv preprint arXiv:2405.11106*, 2024a.

Haochen Sun, Shuwen Zhang, Lujie Niu, Lei Ren, Hao Xu, Hao Fu, Fangkun Zhao, Caixia Yuan, and Xiaojie Wang. Collab-overcooked: Benchmarking and evaluating large language models as collaborative agents. *arXiv preprint arXiv:2502.20073*, 2025a.

Jingyun Sun, Chengxiao Dai, Zhongze Luo, Yangbo Chang, and Yang Li. Lawluo: A chinese law firm co-run by llm agents. *arXiv preprint arXiv:2407.16252*, 2024b.

Lijun Sun, Yijun Yang, Qiqi Duan, Yuhui Shi, Chao Lyu, Yu-Cheng Chang, Chin-Teng Lin, and Yang Shen. Multi-agent coordination across diverse applications: A survey. *arXiv preprint arXiv:2502.14743*, 2025b.

Lijun Sun, Yijun Yang, Qiqi Duan, Yuhui Shi, Chao Lyu, Yu-Cheng Chang, Chin-Teng Lin, and Yang Shen. Multi-agent coordination across diverse applications: A survey. *arXiv preprint arXiv:2502.14743*, 2025c.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

Mirac Suzgun and Adam Tauman Kalai. Meta-prompting: Enhancing language models with task-agnostic scaffolding. *arXiv preprint arXiv:2401.12954*, 2024.

Agnieszka Słowik, Abhinav Gupta, William L. Hamilton, Mateja Jamnik, Sean B. Holden, and Christopher Pal. Structural inductive biases in emergent communication, 2021. URL `https://arxiv.org/abs/2002.01335`.

Victor Talpaert, Ibrahim Sobh, B Ravi Kiran, Patrick Mannion, Senthil Yogamani, Ahmad El-Sallab, and Patrick Perez. Exploring applications of deep reinforcement learning for real-world autonomous driving systems. *arXiv preprint arXiv:1901.01536*, 2019.

Xiangru Tang, Anni Zou, Zhuosheng Zhang, Ziming Li, Yilun Zhao, Xingyao Zhang, Arman Cohan, and Mark Gerstein. Medagents: Large language models as collaborators for zero-shot medical reasoning. *arXiv preprint arXiv:2311.10537*, 2023.

Zheng Tian, Shihao Zou, Ian Davies, Tim Warr, Lisheng Wu, Haitham Bou Ammar, and Jun Wang. Learning to communicate implicitly by actions. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 7261–7268, 2020.

Khanh-Tung Tran, Dung Dao, Minh-Duong Nguyen, Quoc-Viet Pham, Barry O'Sullivan, and Hoang D Nguyen. Multi-agent collaboration mechanisms: A survey of llms. *arXiv preprint arXiv:2501.06322*, 2025a.

Khanh-Tung Tran, Dung Dao, Minh-Duong Nguyen, Quoc-Viet Pham, Barry O'Sullivan, and Hoang D Nguyen. Multi-agent collaboration mechanisms: A survey of llms. *arXiv preprint arXiv:2501.06322*, 2025b.

Patara Trirat, Wonyong Jeong, and Sung Ju Hwang. Automl-agent: A multi-agent llm framework for full-pipeline automl. *arXiv preprint arXiv:2410.02958*, 2024.

Anastasios Tsiamis, Christos K Verginis, Charalampos P Bechlioulis, and Kostas J Kyriakopoulos. Cooperative manipulation exploiting only implicit communication. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 864–869. IEEE, 2015.

Mycal Tucker, Huao Li, Siddharth Agrawal, Dana Hughes, Katia Sycara, Michael Lewis, and Julie A Shah. Emergent discrete communication in semantic spaces. *Advances in Neural Information Processing Systems*, 34:10574–10586, 2021.

Ryo Ueda, Taiga Ishii, Koki Washio, and Yusuke Miyao. Categorial grammar induction as a compositionality measure for emergent languages in signaling games. 2022.

Paul Van Eecke and Katrien Beuls. Re-conceptualising the language game paradigm in the framework of multi-agent reinforcement learning. *arXiv preprint arXiv:2004.04722*, 2020.

Hado van Hasselt, Arthur Guez, and David Silver. Deep reinforcement learning with double q-learning, 2015. URL `https://arxiv.org/abs/1509.06461`.

Astrid Vanneste, Simon Vanneste, Kevin Mets, Tom De Schepper, Siegfried Mercelis, Steven Latré, and Peter Hellinckx. An analysis of discretization methods for communication learning with multi-agent reinforcement learning. *arXiv preprint arXiv:2204.05669*, 2022a.

Simon Vanneste, Astrid Vanneste, Kevin Mets, Tom De Schepper, Ali Anwar, Siegfried Mercelis, Steven Latré, and Peter Hellinckx. Learning to communicate using counterfactual reasoning. In *Adaptive and Learning Agents Workshop (ALA), collocated with AAMAS, 11-13 May, 2022, Auckland, New Zealand*, 2022b.

Kyle Wagner, James A Reggia, Juan Uriagereka, and Gerald S Wilkinson. Progress in the simulation of emergent communication and language. *Adaptive Behavior*, 11(1):37–69, 2003.

Chao Wang, Jian Wang, Yuan Shen, and Xudong Zhang. Autonomous navigation of uavs in large-scale complex environments: A deep reinforcement learning approach. *IEEE Transactions on Vehicular Technology*, 68(3):2124–2136, 2019a.

Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv:2305.16291*, 2023a.

Jianhao Wang, Zhizhou Ren, Terry Liu, Yang Yu, and Chongjie Zhang. Qplex: Duplex dueling multi-agent q-learning. *arXiv preprint arXiv:2008.01062*, 2020a.

Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, et al. A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6):186345, 2024a.

Qian Wang, Tianyu Wang, Qinbin Li, Jingsheng Liang, and Bingsheng He. Megaagent: A practical framework for autonomous cooperation in large-scale llm agent systems. *arXiv preprint arXiv:2408.09955*, 2024b.

Qineng Wang, Zihao Wang, Ying Su, Hanghang Tong, and Yangqiu Song. Rethinking the bounds of llm reasoning: Are multi-agent discussions the key? *arXiv preprint arXiv:2402.18272*, 2024c.

Rundong Wang, Xu He, Runsheng Yu, Wei Qiu, Bo An, and Zinovi Rabinovich. Learning efficient multi-agent communication: An information bottleneck approach. In *International Conference on Machine Learning*, pp. 9908–9918. PMLR, 2020b.

Shenzhi Wang, Chang Liu, Zilong Zheng, Siyuan Qi, Shuo Chen, Qisen Yang, Andrew Zhao, Chaofei Wang, Shiji Song, and Gao Huang. Avalon's game of thoughts: Battle against deception through recursive contemplation. *arXiv preprint arXiv:2310.01320*, 2023b.

Shiyong Wang, Jiafu Wan, Daqiang Zhang, Di Li, and Chunhua Zhang. Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination. *Computer networks*, 101:158–168, 2016.

Tixian Wang, Amirhossein Taghvaei, and Prashant G Mehta. Q-learning for pomdp: An application to learning locomotion gaits. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pp. 2758–2763. IEEE, 2019b.

Tonghan Wang, Jianhao Wang, Chongyi Zheng, and Chongjie Zhang. Learning nearly decomposable value functions via communication minimization. *arXiv preprint arXiv:1910.05366*, 2019c.

Wei Wang, Dan Zhang, Tao Feng, Boyan Wang, and Jie Tang. Battleagentbench: A benchmark for evaluating cooperation and competition capabilities of language models in multi-agent systems. *arXiv preprint arXiv:2408.15971*, 2024d.

Yuqi Wang, Xu-Yao Zhang, Cheng-Lin Liu, Tieniu Tan, and Zhaoxiang Zhang. Emergence of machine language: towards symbolic intelligence with neural networks. *National Science Review*, 11(4):nwad317, 2024e.

Z Wang, S Mao, W Wu, T Ge, F Wei, and H Ji. Unleashing cognitive synergy in large language models: A task-solving agent through multi-persona selfcollaboration. arxiv 2023. *arXiv preprint arXiv:2307.05300*.

Zhefan Wang, Yuanqing Yu, Wendi Zheng, Weizhi Ma, and Min Zhang. Macrec: A multi-agent collaboration framework for recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2760–2764, 2024f.

Zhenhailong Wang, Shaoguang Mao, Wenshan Wu, Tao Ge, Furu Wei, and Heng Ji. Unleashing the emergent cognitive synergy in large language models: A task-solving agent through multi-persona self-collaboration. *arXiv preprint arXiv:2307.05300*, 2023c.

Paul Watzlawick, Janet Beavin Bavelas, and Don D Jackson. *Pragmatics of human communication: A study of interactional patterns, pathologies and paradoxes.* WW Norton & Company, 2011.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, et al. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, pp. 38–45, 2020.

Cornelius Wolff, Julius Mayer, Elia Bruni, and Xenia Ohmer. Bidirectional emergent language in situated environments. *arXiv preprint arXiv:2408.14649*, 2024.

Annie Wong, Thomas Bäck, Anna V Kononova, and Aske Plaat. Deep multiagent reinforcement learning: Challenges and directions. *Artificial Intelligence Review*, 56(6):5023–5056, 2023.

Dekun Wu, Haochen Shi, Zhiyuan Sun, and Bang Liu. Deciphering digital detectives: Understanding llm behaviors and capabilities in multi-agent mystery games. *arXiv preprint arXiv:2312.00746*, 2023a.

Di Wu, Xian Wei, Guang Chen, Hao Shen, Xiangfeng Wang, Wenhao Li, and Bo Jin. Generative multi-agent collaboration in embodied ai: A systematic review. *arXiv preprint arXiv:2502.11518*, 2025.

Minghao Wu, Jiahao Xu, Yulin Yuan, Gholamreza Haffari, Longyue Wang, Weihua Luo, and Kaifu Zhang. (perhaps) beyond human translation: Harnessing multi-agent collaboration for translating ultra-long literary texts. *arXiv preprint arXiv:2405.11804*, 2024a.

Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Shaokun Zhang, Erkang Zhu, Beibin Li, Li Jiang, Xiaoyun Zhang, and Chi Wang. Autogen: Enabling next-gen llm applications via multi-agent conversation framework. *arXiv preprint arXiv:2308.08155*, 2023b.

Tongshuang Wu, Michael Terry, and Carrie Jun Cai. Ai chains: Transparent and controllable human-ai interaction by chaining large language model prompts. In *Proceedings of the 2022 CHI conference on human factors in computing systems*, pp. 1–22, 2022.

Yue Wu, So Yeon Min, Shrimai Prabhumoye, Yonatan Bisk, Russ R Salakhutdinov, Amos Azaria, Tom M Mitchell, and Yuanzhi Li. Spring: Studying papers and reasoning to play games. *Advances in Neural Information Processing Systems*, 36, 2024b.

Ziyang Xiao, Dongxiang Zhang, Yangjun Wu, Lilin Xu, Yuan Jessica Wang, Xiongwei Han, Xiaojin Fu, Tao Zhong, Jia Zeng, Mingli Song, et al. Chain-of-experts: When llms meet complex operations research problems. In *The twelfth international conference on learning representations*, 2023.

Tianbao Xie, Fan Zhou, Zhoujun Cheng, Peng Shi, Luoxuan Weng, Yitao Liu, Toh Jing Hua, Junning Zhao, Qian Liu, Che Liu, et al. Openagents: An open platform for language agents in the wild. *arXiv preprint arXiv:2310.10634*, 2023.

Kai Xiong, Xiao Ding, Yixin Cao, Ting Liu, and Bing Qin. Examining inter-consistency of large language models collaboration: An in-depth analysis via debate. *arXiv preprint arXiv:2305.11595*, 2023.

Binfeng Xu, Xukun Liu, Hua Shen, Zeyu Han, Yuhan Li, Murong Yue, Zhiyuan Peng, Yuchen Liu, Ziyu Yao, and Dongkuan Xu. Gentopia: A collaborative platform for tool-augmented llms. *arXiv preprint arXiv:2308.04030*, 2023a.

Yuzhuang Xu, Shuo Wang, Peng Li, Fuwen Luo, Xiaolong Wang, Weidong Liu, and Yang Liu. Exploring large language models for communication games: An empirical study on werewolf. *arXiv preprint arXiv:2309.04658*, 2023b.

Zelai Xu, Chao Yu, Fei Fang, Yu Wang, and Yi Wu. Language agents with reinforcement learning for strategic play in the werewolf game. *arXiv preprint arXiv:2310.18940*, 2023c.

Hanqing Yang, Jingdi Chen, Marie Siew, Tania Lorido-Botran, and Carlee Joe-Wong. Llm-powered decentralized generative agents with adaptive hierarchical knowledge graph for cooperative planning. *arXiv preprint arXiv:2502.05453*, 2025.

Shunyu Yao, Mo Yu, Yang Zhang, Karthik R Narasimhan, Joshua B Tenenbaum, and Chuang Gan. Linking emergent and natural languages via corpus transfer. In *International Conference on Learning Representations*.

Shunyu Yao, Howard Chen, John Yang, and Karthik Narasimhan. Webshop: Towards scalable real-world web interaction with grounded language agents. *Advances in Neural Information Processing Systems*, 35: 20744–20757, 2022a.

Shunyu Yao, Mo Yu, Yang Zhang, Karthik R Narasimhan, Joshua B Tenenbaum, and Chuang Gan. Linking emergent and natural languages via corpus transfer. *arXiv preprint arXiv:2203.13344*, 2022b.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. *Advances in neural information processing systems*, 36:11809–11822, 2023.

Zihao Yi, Jiarui Ouyang, Yuwen Liu, Tianhao Liao, Zhe Xu, and Ying Shen. A survey on recent advances in llm-based multi-turn dialogue systems. *arXiv preprint arXiv:2402.18013*, 2024.

Zhangyue Yin, Qiushi Sun, Cheng Chang, Qipeng Guo, Junqi Dai, Xuanjing Huang, and Xipeng Qiu. Exchange-of-thought: Enhancing large language model capabilities through cross-model communication. *arXiv preprint arXiv:2312.01823*, 2023.

Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in Neural Information Processing Systems*, 35:24611–24624, 2022a.

Dhara Yu, Jesse Mu, and Noah Goodman. Emergent covert signaling in adversarial reference games. In *Emergent Communication Workshop at ICLR 2022*, 2022b.

Xianhao Yu, Jiaqi Fu, Renjia Deng, and Wenjuan Han. Mineland: Simulating large-scale multi-agent interactions with limited multimodal senses and physical needs. *arXiv preprint arXiv:2403.19267*, 2024.

Won Joon Yun, Byungju Lim, Soyi Jung, Young-Chai Ko, Jihong Park, Joongheon Kim, and Mehdi Bennis. Attention-based reinforcement learning for real-time uav semantic communication. In *2021 17th International Symposium on Wireless Communication Systems (ISWCS)*, pp. 1–6. IEEE, 2021.

Mohamed Salah Zaïem and Etienne Bennequin. Learning to communicate in multi-agent reinforcement learning: A review. *arXiv preprint arXiv:1911.05438*, 2019.

Ceyao Zhang, Kaijie Yang, Siyi Hu, Zihao Wang, Guanghe Li, Yihang Sun, Cheng Zhang, Zhaowei Zhang, Anji Liu, Song-Chun Zhu, et al. Proagent: Building proactive cooperative ai with large language models. *CoRR*, 2023a.

Ceyao Zhang, Kaijie Yang, Siyi Hu, Zihao Wang, Guanghe Li, Yihang Sun, Cheng Zhang, Zhaowei Zhang, Anji Liu, Song-Chun Zhu, et al. Proagent: building proactive cooperative agents with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 17591–17599, 2024a.

Fangyi Zhang, Jürgen Leitner, Michael Milford, Ben Upcroft, and Peter Corke. Towards vision-based deep reinforcement learning for robotic motion control, 2015. URL `https://arxiv.org/abs/1511.03791`.

Hongxin Zhang, Weihua Du, Jiaming Shan, Qinhong Zhou, Yilun Du, Joshua B Tenenbaum, Tianmin Shu, and Chuang Gan. Building cooperative embodied agents modularly with large language models. *arXiv preprint arXiv:2307.02485*, 2023b.

Jintian Zhang, Xin Xu, Ningyu Zhang, Ruibo Liu, Bryan Hooi, and Shumin Deng. Exploring collaboration mechanisms for llm agents: A social psychology view. *arXiv preprint arXiv:2310.02124*, 2023c.

Junjie Zhang, Yupeng Hou, Ruobing Xie, Wenqi Sun, Julian McAuley, Wayne Xin Zhao, Leyu Lin, and Ji-Rong Wen. Agentcf: Collaborative learning with autonomous language agents for recommender systems. In *Proceedings of the ACM Web Conference 2024*, pp. 3679–3689, 2024b.

Sai Qian Zhang, Qi Zhang, and Jieyu Lin. Efficient communication in multi-agent reinforcement learning via variance based control. *Advances in neural information processing systems*, 32, 2019.

Sai Qian Zhang, Qi Zhang, and Jieyu Lin. Succinct and robust multi-agent communication with temporal message control. *Advances in neural information processing systems*, 33:17271–17282, 2020.

Yifan Zhang, Jingqin Yang, Yang Yuan, and Andrew Chi-Chih Yao. Cumulative reasoning with large language models. *arXiv preprint arXiv:2308.04371*, 2023d.

Yifan Zhang, Yang Yuan, and Andrew Chi-Chih Yao. On the diagram of thought. *arXiv preprint arXiv:2409.10038*, 2024c.

Yusen Zhang, Ruoxi Sun, Yanfei Chen, Tomas Pfister, Rui Zhang, and Sercan Arik. Chain of agents: Large language models collaborating on long-context tasks. *Advances in Neural Information Processing Systems*, 37:132208–132237, 2024d.

Yi Zheng, Chongyang Ma, Kanle Shi, and Haibin Huang. Agents meet okr: An object and key results driven agent system with hierarchical self-collaboration and self-evaluation. *arXiv preprint arXiv:2311.16542*, 2023.

Shuyan Zhou, Frank F Xu, Hao Zhu, Xuhui Zhou, Robert Lo, Abishek Sridhar, Xianyi Cheng, Tianyue Ou, Yonatan Bisk, Daniel Fried, et al. Webarena: A realistic web environment for building autonomous agents. *arXiv preprint arXiv:2307.13854*, 2023.

Changxi Zhu, Mehdi Dastani, and Shihan Wang. A survey of multi-agent reinforcement learning with communication. *arXiv preprint arXiv:2203.08975*, 2022.

Changxi Zhu, Mehdi Dastani, and Shihan Wang. A survey of multi-agent deep reinforcement learning with communication. *Autonomous Agents and Multi-Agent Systems*, 38(1):4, 2024.

Kunlun Zhu, Hongyi Du, Zhaochen Hong, Xiaocheng Yang, Shuyi Guo, Zhe Wang, Zhenhailong Wang, Cheng Qian, Xiangru Tang, Heng Ji, et al. Multiagentbench: Evaluating the collaboration and competition of llm agents. *arXiv preprint arXiv:2503.01935*, 2025.

Xizhou Zhu, Yuntao Chen, Hao Tian, Chenxin Tao, Weijie Su, Chenyu Yang, Gao Huang, Bin Li, Lewei Lu, Xiaogang Wang, et al. Ghost in the minecraft: Generally capable agents for open-world environments via large language models with text-based knowledge and memory. *arXiv preprint arXiv:2305.17144*, 2023.

Yuchen Zhuang, Xiang Chen, Tong Yu, Saayan Mitra, Victor Bursztyn, Ryan A Rossi, Somdeb Sarkhel, and Chao Zhang. Toolchain*: Efficient action space navigation in large language models with a* search. *arXiv preprint arXiv:2310.13227*, 2023.

Mingchen Zhuge, Wenyi Wang, Louis Kirsch, Francesco Faccio, Dmitrii Khizbullin, and Jürgen Schmidhuber. Gptswarm: Language agents as optimizable graphs. In *Forty-first International Conference on Machine Learning*, 2024a.

Mingchen Zhuge, Wenyi Wang, Louis Kirsch, Francesco Faccio, Dmitrii Khizbullin, and Jürgen Schmidhuber. Language agents as optimizable graphs. *arXiv preprint arXiv:2402.16823*, 2024b.