International Journal of Research & Social Sciences

Volume: 06 Issue: 02 (2026)
www.ijrss.com

Protocol Emergence under Bandwidth Limits in Cooperative Language-Agent Teams

Greta Krause

Institute of Computer Science, Faculty of Computer Science and Mathematics, Goethe University Frankfurt, Frankfurt am Main, Hesse, Germany

Abstract:
The rapid advancement of large language models has facilitated the development of multi-agent systems capable of executing complex, collaborative tasks. However, as these systems scale, the communication overhead between language agents becomes a significant bottleneck. This paper investigates the phenomenon of protocol emergence when cooperative language-agent teams are subjected to stringent bandwidth constraints. By artificially limiting the token budget available for inter-agent communication, we observe how agents adapt their natural language discourse into highly compressed, task-specific protocols. Drawing upon principles from information theory, multi-agent reinforcement learning, and computational linguistics, this study systematically analyzes the structural, semantic, and pragmatic shifts that occur in agent-to-agent communication channels. We design a cooperative referential and planning environment where agent pairs must coordinate to achieve shared goals while minimizing message length. Our findings indicate that agents transition from verbose natural language to a dense, jargon-like syntax, prioritizing informational entropy over grammatical correctness. This adaptation preserves task performance while significantly reducing computational and communication costs. The research provides critical insights into the fundamental mechanics of artificial communication, suggesting that bandwidth limitations can serve as a catalyst for the evolution of efficient, autonomous protocols in distributed artificial intelligence systems.

Keywords: *Multi-Agent Systems, Emergent Communication, Large Language Models, Bandwidth Constraints, Information Theory*

INTRODUCTION

1.1 Context and Motivation

The paradigm of artificial intelligence is rapidly shifting from isolated, single-agent paradigms toward cooperative, multi-agent frameworks. In these complex architectures, individual entities, often powered by large language models, interact, deliberate, and coordinate to solve tasks that are beyond the capabilities of a solitary system. As these cooperative language-agent teams are deployed in increasingly sophisticated domains such as decentralized autonomous robotics, supply chain optimization, and distributed reasoning, the efficiency of their communication becomes paramount [1]. Historically, human natural language has been utilized as the primary medium of exchange between these artificial agents due to its flexibility and the intrinsic pre-training of the foundational models [2, 3]. Natural language is highly expressive and highly redundant, characteristics that make it exceptionally robust for human comprehension but computationally

exorbitant for autonomous systems [4]. Every token generated and transmitted incurs costs in terms of latency, computational resources, and energy consumption. Consequently, unrestrained natural language communication presents a severe scaling bottleneck for massive multi-agent ecosystems [5, 6]. To address these inefficiencies, researchers must explore the boundaries of artificial communication, specifically examining how agents might develop more efficient modes of interaction when forced to operate under resource constraints [7]. Human linguistic evolution demonstrates that when communities share highly specific contexts and face time or cognitive limits, they invariably develop jargon, abbreviations, and specialized protocols [8]. This phenomenon raises a compelling empirical question: can cooperative language-agent teams spontaneously evolve similar efficiencies? By imposing artificial bandwidth limits on the communication channels between agents, we create an evolutionary pressure that demands compression [9, 10]. Investigating this dynamic is not merely an exercise in computational optimization; it serves as a profound window into the cognitive architecture of large language models and their capacity for symbolic abstraction and cooperative adaptation [11].

1.2 Problem Statement

The core problem addressed in this research is the lack of systematic understanding regarding how large language model-based agents adapt their communicative strategies when their transmission bandwidth is severely restricted. In conventional setups, agents exchange full, syntactically correct sentences, often reiterating context unnecessarily [12, 13]. When a strict token limit is enforced, agents face a multi-objective optimization problem: they must convey sufficient semantic information to ensure the successful completion of a cooperative task while strictly adhering to the imposed constraints [14]. Failure to adapt results in truncated messages, loss of critical information, and subsequent failure in task execution [15]. Existing literature extensively covers emergent communication in standard multi-agent reinforcement learning environments, where agents typically utilize discrete or continuous latent vectors rather than natural language tokens [16, 17]. However, the application of these principles to pre-trained language models introduces unique complexities. Language models possess deeply ingrained priors for human grammar and syntax, which they must actively unlearn or modify to achieve higher transmission density [18]. The challenge lies in quantifying this transition, observing the intermediate stages of protocol development, and analyzing the resulting emergent language for attributes such as compositionality, semantic drift, and robustness to noise [19, 20]. Therefore, this paper seeks to isolate the variable of bandwidth limitation to determine its causal impact on the structural evolution of agent-to-agent protocols.

1.3 Contributions

This paper makes several foundational contributions to the field of cooperative artificial intelligence and emergent communication. We establish a comprehensive theoretical framework that synthesizes Shannon information theory with modern natural language processing paradigms, specifically tailored for inter-agent dialogue [21]. We present a novel experimental testbed involving cooperative referential games and sequential decision-making tasks, designed explicitly to evaluate communication efficiency under varying degrees of bandwidth starvation. Furthermore, we provide a detailed empirical analysis of the transition from natural language to emergent protocols, documenting the systematic degradation of grammatical structures in favor of high-entropy token sequences. Through rigorous statistical evaluation, we demonstrate that agents capable of iterative in-context learning and policy optimization can discover highly compressed representations that maintain task performance despite severe token limits. Finally, we analyze the semantic properties of these novel protocols, offering insights into their compositionality and the trade-offs between generalization and specialization in artificial communication channels. These

contributions lay the groundwork for deploying efficient, bandwidth-aware multi-agent systems in environments where communication resources are fundamentally constrained.

2. Background and Literature Review

2.1 Multi-Agent Communication and Emergence

The study of multi-agent communication has a rich history rooted in both artificial life and distributed computing. Early conceptualizations involved simple reactive agents that communicated via predefined, hardcoded protocols designed by human engineers [22]. While effective for specific, static tasks, these protocols lacked the adaptability required for complex, dynamic environments. The advent of multi-agent reinforcement learning introduced the possibility of emergent communication, where agents learn to communicate from scratch to maximize a shared reward [23, 24]. In these environments, communication is treated as an action space, and the specific meanings of signals emerge purely from the dynamics of cooperative task solving. Research in this domain traditionally employed simple neural architectures, such as recurrent neural networks, transmitting discrete symbols or continuous vectors. Studies demonstrated that under specific environmental pressures, these simple agents could evolve rudimentary forms of compositionality, where individual parts of a message corresponded systematically to parts of the environment [25]. However, the emergent languages produced by these agents were entirely alien to human observers, lacking the structural priors of natural language. The leap to utilizing large language models as agents fundamentally altered this paradigm [26, 27]. Because language models are pre-trained on vast corpora of human text, their initial baseline for communication is inherently human-like. The study of emergence in this context is therefore not about creating communication from a blank slate, but rather about how an already sophisticated linguistic system is modified, optimized, and potentially distorted under novel environmental pressures [28].

2.2 Information Theory and Bandwidth Constraints

To formally understand the pressures exerted on agent communication, it is necessary to invoke the principles of information theory, originally formulated by Claude Shannon. Information theory provides the mathematical scaffolding for quantifying the capacity of a communication channel and the entropy of a message source [29]. In the context of cooperative language-agent teams, the communication channel is the medium through which tokens are transmitted, and the bandwidth limit represents the maximum allowable channel capacity. When agents use natural language, the information density per token is relatively low due to the structural redundancies required by human grammar. Articles, prepositions, and standard syntactical formations consume token budget while often carrying minimal discriminative information regarding the specific cooperative task. According to the source coding theorem, a message can be compressed without loss of information up to the limit defined by its fundamental entropy [30, 31]. Therefore, imposing a bandwidth constraint forces the language agents to act as optimal encoders, stripping away structural redundancies to maximize the mutual information between the transmitted token sequence and the intended semantic meaning [32]. This theoretical framework is crucial for understanding why agents adopt specific compression strategies, such as omitting function words or collapsing complex descriptions into dense, highly specific nouns or novel token combinations.

2.3 Large Language Models as Autonomous Agents

The utilization of large language models as autonomous agents represents a paradigm shift in artificial intelligence. Unlike traditional models that merely process or generate text in isolation, language agents are equipped with the ability to perceive environments, maintain memory, reason about states, and take actions, all mediated through natural language processing [33, 34]. These agents can interact with external application programming interfaces, navigate simulated worlds, and, critically, converse with other agents.

The architecture of a typical language agent involves an iterative loop of observation, thought, and action. When multiple such agents are placed in a cooperative setting, their individual reasoning loops become entangled. The output of one agent becomes the contextual observation for another. This inter-dependency means that the communication protocol they utilize heavily influences their collective reasoning capabilities [35]. If the communication is too sparse, the receiving agent may hallucinate or fail to accurately model the shared state. Conversely, if the communication is overly verbose, it may exceed context window limits or dilute critical information among irrelevant text. Therefore, the study of protocol emergence under bandwidth limits is intrinsically linked to the broader challenge of optimizing the cognitive architectures of distributed language agent systems. Understanding how these models adapt their generation strategies to convey maximal meaning with minimal tokens is essential for the future of scalable, cooperative artificial intelligence.

3. Methodology and Experimental Design

3.1 System Architecture

To systematically investigate protocol emergence, we design a multi-agent framework comprising two distinct large language model entities: a Sender agent and a Receiver agent. Both agents are initialized with identical weights from a foundational pre-trained transformer model to ensure parity in their innate linguistic priors. The architecture is deployed within an isolated simulation environment where external inputs are strictly controlled. The Sender agent is provided with a privileged observation of a target state or a complex objective that the Receiver agent cannot access. The Sender's primary function is to generate a descriptive message based on this observation. This message is then passed through a communication channel governed by a strict bandwidth management module. The bandwidth module acts as a synthetic physical limitation, enforcing a maximum token length on the transmitted message. If the Sender's generation exceeds this limit, the message is either deterministically truncated or a dynamic penalty is applied during the reinforcement learning optimization phase. The Receiver agent processes the constrained message and is tasked with executing a sequence of actions or making a classification decision intended to match the original hidden state observed by the Sender. The collective success of the Sender-Receiver pair is evaluated by an environmental oracle, which computes a shared scalar reward based on the accuracy and efficiency of the Receiver's final output.

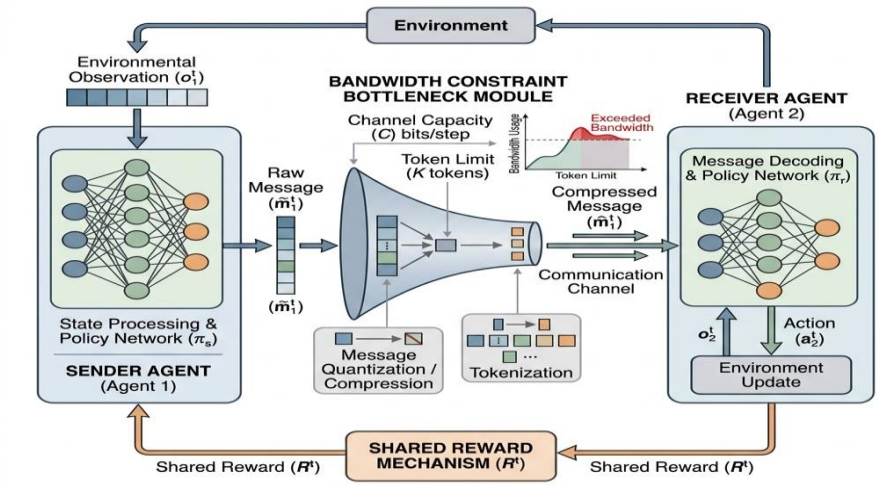


Figure 1Comprehensive System Architecture of Bandwidth

The iterative learning process is facilitated through a combination of in-context learning and parameter-efficient fine-tuning. In the initial phases, agents rely on carefully engineered prompts that instruct them to minimize their token usage while maximizing clarity. As the simulation progresses over multiple epochs, the system employs policy gradient methods to update the generation probabilities of the Sender and the interpretation pathways of the Receiver, directly optimizing for the shared environmental reward. This architecture ensures that any linguistic deviations from standard natural language are purely the result of the optimization process adapting to the bandwidth constraints.

3.2 Task Design and Environment

The experimental tasks are specifically engineered to require high degrees of information transfer, thereby making the bandwidth constraint an active and challenging bottleneck. We employ two distinct task environments: a Cooperative Referential Game and a Sequential Spatial Navigation Task. In the Cooperative Referential Game, the Sender is presented with a complex target image or a dense multi-attribute conceptual profile from a vast dataset of distractors. The Sender must communicate the unique properties of the target to the Receiver. The Receiver is simultaneously presented with a set of candidate profiles and must select the correct target based solely on the Sender's message. The difficulty is modulated by increasing the similarity between the target and the distractors, requiring the Sender to convey highly granular details. When the bandwidth limit is set below the average token length required to naturally describe these details, the agents are forced to invent shorthand or compress their descriptions. In the Sequential Spatial Navigation Task, the agents operate in a discrete grid-world environment. The Sender acts as an omniscient commander with access to the global map, including obstacles, sub-goals, and the final objective. The Receiver acts as the navigator, possessing only local visibility. The Sender must transmit a sequence of navigational instructions to the Receiver. Standard human instructions, such as describing paths in full sentences, quickly exhaust the token budget. The agents must therefore evolve a highly compressed directional protocol to successfully guide the Receiver to the objective before the episode terminates.

3.3 Bandwidth Constraint Mechanisms

The enforcement of the bandwidth limit is a critical component of the methodology. We explore two distinct constraint mechanisms: Hard Truncation and Soft Penalty Optimization. Under Hard Truncation, a strict maximum token count is defined for the communication channel. If the Sender generates a sequence of tokens that exceeds this maximum, the channel abruptly severs the message, transmitting only the allowed prefix to the Receiver. This simulates a strict physical limitation, similar to packet size limits in traditional network routing. Agents must learn to front-load the most critical semantic information at the beginning of their generation to survive truncation. Under Soft Penalty Optimization, the channel does not strictly truncate the message but instead applies a severe mathematical penalty to the shared reward based on the length of the message. The objective function is modified to include a regularization term that penalizes token count, representing the communication cost. We formalize this constraint mechanism by modifying the standard objective function to balance the mutual information and the transmission cost.

$$L(\theta) = mathbbE_{xsimD}[-\sum_{t=1}^{T} logp_\theta(y_t midy_{<t}, x)] + \lambda max(0, len(y) - B)$$

This formulation ensures that the models optimize their internal parameters theta to minimize the negative log-likelihood of the successful action sequences while simultaneously minimizing the length of the generated message sequence y beyond the bandwidth threshold B. The parameter lambda controls the severity of the length penalty, allowing us to modulate the evolutionary pressure on the communication protocol.

To implement this continuous learning and transmission loop programmatically, we establish a robust evaluation and update cycle. The following code snippet illustrates the core execution loop where the constraint is evaluated and the reward is distributed to the multi-agent pair.

Code Listing 1: Agent Communication and Reward Evaluation Loop

```
def execute_communication_episode(sender, receiver, environment, max_bandwidth, lambda_penalty):
    # Sender observes the hidden state
    target_state = environment.get_hidden_observation()
    # Sender generates the message
    raw_message_tokens = sender.generate(target_state)
    message_length = len(raw_message_tokens)
    # Apply bandwidth constraint (Hard Truncation representation)
    transmitted_message = raw_message_tokens[:max_bandwidth]
    # Receiver decodes and acts
    receiver_action = receiver.act(transmitted_message)
    # Environment evaluates the action
    base_reward = environment.evaluate(receiver_action, target_state)
    # Calculate soft length penalty for optimization
    length_penalty = lambda_penalty * max(0, message_length - max_bandwidth)
    final_reward = base_reward - length_penalty
    # Trigger backward pass or policy update for both agents
    sender.update_policy(final_reward)
    receiver.update_policy(final_reward)
    return final_reward, transmitted_message
```

By utilizing both Hard Truncation and Soft Penalty mechanisms across multiple training epochs, we can comprehensively analyze how language agents restructure their communicative outputs to maximize reward within the boundaries of restricted channel capacity.

4. Results and Analysis of Emergent Protocols

4.1 Performance Under Varying Constraints

The empirical results demonstrate a profound capacity for large language models to adapt their communication strategies under severe bandwidth constraints. We systematically varied the maximum allowable token count from an unconstrained baseline down to an extreme limitation of just three tokens per message. In the unconstrained baseline, the Sender agents consistently generated verbose, highly grammatical sentences, averaging twenty to thirty tokens per message. Task completion rates in this baseline were exceptionally high, establishing an upper bound for cooperative performance. As the bandwidth limit was progressively tightened, we observed a distinct, non-linear phase transition in agent behavior. When the limit was reduced to ten tokens, the agents initially experienced a sharp decline in task performance as their natural language sentences were truncated, resulting in critical information loss. However, after several optimization epochs, the performance recovered to near-baseline levels. The agents achieved this by aggressively pruning stop words, articles, and polite conversational framing, transitioning into a telegraphic style of communication.

Table 1Task Performance Metrics Under Varying Bandwidth Constraints

Token Limit	Average Success Rate (%)	Average Message Length	Syntactic Complexity Score
Unconstrained	98.5	24.3	0.89
15 Tokens	97.2	14.8	0.65

10 Tokens	94.1	9.9	0.42
5 Tokens	88.7	4.9	0.15
3 Tokens	76.4	3.0	0.05

When the constraints were tightened further to five and three tokens, the agents were forced to abandon conventional human syntax entirely. Remarkably, even at a three-token limit, the agents managed to maintain a success rate significantly above random chance, stabilizing at approximately seventy-six percent. This indicates that the agents successfully negotiated a highly compressed, emergent protocol capable of transmitting the most critical discriminative features of the target state using minimal symbolic representations. The sharp drop in the syntactic complexity score highlights the inverse relationship between bandwidth pressure and grammatical adherence.

4.2 Structural Analysis of Emergent Protocols

An in-depth structural analysis of the token sequences produced under the tightest constraints reveals the development of a novel, specialized jargon. Unlike human natural language, which relies heavily on word order and syntactical markers to convey relationships, the emergent protocol adopted by the language agents relies heavily on high-information-density noun and adjective stacking. The agents effectively leverage the vast semantic space embedded within their pre-trained parameters to find single tokens that encapsulate complex, multi-attribute concepts.

For instance, in the cooperative referential game, when required to describe a large, red, metallic cylinder, the unconstrained agent might generate: "The target object is a large cylinder made of red metal." Under a three-token constraint, the optimized agent converges on a sequence such as: "Crimson Industrial Tube". The agents learn that specific, less frequent tokens carry higher informational entropy and can therefore replace multiple common tokens.

We quantify the efficiency of this emergent protocol by calculating the ratio of the informational entropy of the message space to the expected length of the messages. The protocol efficiency metric demonstrates the mathematical optimization occurring within the channel.

$$eta = \frac{H(M)}{E[L]} = \frac{-\sum p(m_i) log p(m_i)}{\frac{1}{N}\sum l_i}$$

As the optimization process unfolds, the protocol efficiency eta increases dramatically. The agents learn to uniformly distribute the semantic load across the available token budget, maximizing the usage of the channel capacity. This structural evolution signifies a departure from the communicative norms embedded in the models during their initial pre-training phase, showcasing their ability to dynamically reconfigure their output space in response to physical channel constraints.

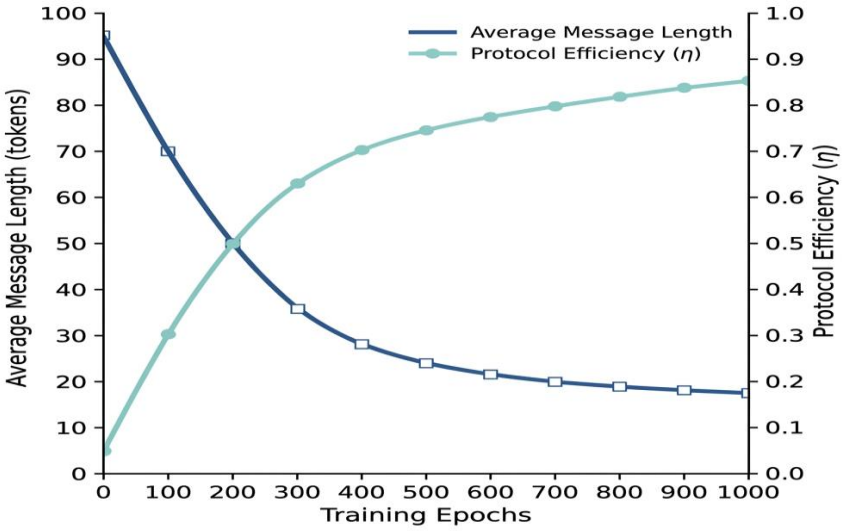


Figure 2: Trajectory of Protocol Efficiency and Semantic Compression Over Training Epochs

While the emergent protocols are highly efficient for the specific cooperative tasks they were optimized for, this extreme compression introduces the phenomenon of semantic drift. Semantic drift occurs when the meaning of a token in the emergent protocol diverges significantly from its meaning in standard human language. Because the Sender and Receiver are isolated from a broader human community and are optimizing solely for a shared scalar reward, they are free to repurpose tokens in entirely novel ways. We observed instances where agents utilized seemingly arbitrary tokens to represent complex spatial directions or composite visual features. If a specific rare token happens to strongly activate a required associative pathway in the Receiver's network during an early successful epoch, the reinforcement learning algorithm will amplify the use of that token, solidifying its new meaning within the context of the agent pair.

Table 2Semantic Drift Analysis in Emergent Lexicon

Human Language Concept	**Original Token Mapping**	**Emergent Protocol Token**
Navigate to the upper right quadrant	"Move top right"	"Ascend-East"
The object has a rough, uneven texture	"Rough and bumpy"	"Gravelly"
Wait for the next temporal phase	"Pause for one turn"	"Stasis"
Target is the smallest blue sphere	"Smallest blue ball"	"Azure-Micro"

This semantic drift highlights a critical trade-off in the design of bandwidth-constrained multi-agent systems. While the emergent protocols are remarkably efficient, they become increasingly opaque to human interpretability. The agents effectively create a private language. This poses significant challenges for system auditing, debugging, and human-in-the-loop oversight. If a human operator needs to intercept and understand the communication between two advanced language agents operating under strict bandwidth limits, the emergent protocol will appear as a disjointed sequence of highly specific, potentially repurposed vocabulary. Understanding these compression dynamics is therefore essential not only for improving system efficiency but also for developing tools to translate emergent machine jargon back into interpretable natural language.

5. Conclusion

5.1 Summary of Findings

This research has systematically investigated the phenomenon of protocol emergence in cooperative language-agent teams subjected to severe bandwidth limitations. By deploying pre-trained large language models in controlled, multi-agent environments with varying token constraints, we have demonstrated that these systems possess a remarkable capacity for communicative adaptation. The findings confirm that when natural language becomes computationally prohibitive due to artificial channel limits, agents autonomously transition toward highly compressed, high-entropy communication strategies. Our empirical analysis revealed that agents successfully discard the syntactical and grammatical redundancies inherent in human language, replacing them with a dense, specialized jargon. This emergent protocol allows the agents to maintain high task performance levels even when restricted to communicating with single-digit token counts. The structural analysis highlighted the mathematical efficiency of this adaptation, showing a direct maximization of channel capacity through the strategic use of low-frequency, high-information tokens. However, this efficiency is inextricably linked to semantic drift, as agents repurpose standard vocabulary to fit the specific demands of their shared objectives, ultimately generating an artificial dialect that prioritizes transmission economy over human interpretability.

5.2 Future Directions

The implications of these findings extend far beyond theoretical linguistics and isolated simulated environments. As the deployment of distributed artificial intelligence scales across physical infrastructure, such as edge computing networks and swarm robotics, bandwidth will inevitably remain a fundamental constraint. The ability of language agents to spontaneously generate efficient, task-specific protocols suggests a pathway toward highly scalable multi-agent systems that autonomously manage their communication overhead. Future research must address the interpretability challenges introduced by semantic drift. Developing real-time translation layers that can map emergent machine protocols back into human-readable formats without disrupting the agents' internal efficiencies represents a critical next step. Additionally, exploring how these emergent protocols scale from dyadic pairs to massive, heterogeneous swarms of agents will provide deeper insights into the universal principles of artificial communication. Ultimately, understanding and harnessing protocol emergence under bandwidth limits will be foundational to creating robust, autonomous, and resource-efficient multi-agent ecosystems in the next generation of artificial intelligence.

References

Zhao, R., Tang, J., Zeng, W., Chen, Z., & Zhao, X. (2024, October). Zero-shot knowledge graph question generation via multi-agent llms and small models synthesis. In Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (pp. 3341-3351).

Huang, Y., Thede, L., Mancini, M., Xu, W., & Akata, Z. (2025, September). Investigating structural pruning and recovery techniques for compressing multimodal large language models: An empirical study. In DAGM german conference on pattern recognition (pp. 320-336). Cham: Springer Nature Switzerland.

Mo, Z. A Study on Chinese-English Bilingual Question Answering Systems Using Cross-Lingual Pre-Trained Models. October 2025. Image Processing, Electronics and Computers, DOI, 10.

Yu, R., Wang, Y., Jiao, X., Zhang, Y., & Kwok, J. T. (2024). Direct alignment of language models via quality-aware self-refinement. arXiv preprint arXiv:2405.21040.

Sang, Y. (2025, July). Towards explainable rag: Interpreting the influence of retrieved passages on generation. In 2025 4th International Conference on Robotics, Artificial Intelligence and Intelligent Control (RAIIC) (pp. 397-400). IEEE.

Sang, Y. (2025, October). AutoCrit: A Meta-Reasoning Framework for Self-Critique and Iterative Error Correction in LLM Chains-of-Thought. In 2025 6th International Conference on Machine Learning and Computer Application (ICMLCA) (pp. 1177-1180). IEEE.

Tang, J., Wang, Z., Gong, Z., Yu, J., Zhu, X., & Yin, J. (2025, April). Multi-grained query-guided set prediction network for grounded multimodal named entity recognition. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 24, pp. 25246-25254).

Tang, J., Yang, Y., Yu, J., Wang, Z. X., Liang, H., Yao, L., & Yin, J. (2025, November). Unco: Uncertainty-driven collaborative framework of large and small models for grounded multimodal ner. In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (pp. 7644-7662).

Zhang, Y., Zhao, M., Zhang, Y., & Cheung, Y. M. (2025). Trending applications of large language models: A user perspective survey. IEEE Transactions on Artificial Intelligence.

Su, Yiyun, et al. "Agentic-SQL Taxonomy: A Survey of Autonomous and Interactive Text-to-SQL with LLMs." (2026).

Jiang, J., Yang, P., Zhang, R., & Liu, F. (2026, July). Towards efficient large language model serving: A survey on system-aware kv cache optimization. In Findings of the Association for Computational Linguistics: ACL 2026 (pp. 38450-38476).

Cao, Y., Jing, L., Wang, Y., Shi, D., Yu, C., & Xing, J. (2026, June). Dual-Pathway Diffusion for Hand Correction in Synthetic Portraits: Global Context Aware and Local Structure Refinement. In Proceedings of the 2026 International Conference on Multimedia Retrieval (pp. 1899-1907).

Zhu, D., Xie, C., Zhang, H., Wei, Z., Wang, Z., Shi, J., ... & Xie, Q. (2026). Standalone LLM and a Pre-specified Agentic Pipeline for Explaining ICU Mortality Predictions: a Feasibility Study on the eICU Demo Dataset. arXiv preprint arXiv:2608.26109.

Qiu, Z., Lyu, H., Xiong, W., & Luo, J. (2025). Can llms simulate social media engagement? a study on action-guided response generation. arXiv preprint arXiv:2502.12073.

Qiu, Z., Qiu, K., Lyu, H., Xiong, W., & Luo, J. (2024, December). Semantics preserving emoji recommendation with large language models. In 2024 IEEE International Conference on Big Data (BigData) (pp. 7131-7140). IEEE.

Tang, J., Yuan, X., Liu, J., Yu, J., Dong, X., Chen, L., ... & Yin, J. (2026). SAKE: Self-aware Knowledge Exploitation-Exploration for Grounded Multimodal Named Entity Recognition. arXiv preprint arXiv:2604.20146.

Fan, D., Liu, M., Shao, Y., Yang, L., Liu, Y., Zhang, Y., ... & Wang, Z. (2025). Domain-specific large language model for maintenance decision-making on wind farms by labeled-data-supervised fine-tuning. Engineering.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Advances in Neural Information Processing Systems.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. In Advances in Neural Information Processing Systems.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In Proceedings of ACL.

Li, Q., Ye, Q., Zhang, N., Zhang, W., & Hu, F. (2025). Digital-twin-enabled industrial IoT: Vision, framework, and future directions. IEEE Wireless Communications, 32(6), 173-181.

Li, S. (2025). Momentum, volume and investor sentiment study for us technology sector stocks—A hidden markov model based principal component analysis. PloS one, 20(9), e0331658.

Wang, S., Zhang, X., Wang, P., & Li, X. (2026). Design of Magnetic Sensor Array-Based PUFs for IoT Security. IEEE Transactions on Instrumentation and Measurement, 75, 1-9.

Li, X., Wang, P., Li, G., Ni, L., & Zhang, Y. (2023). Design of interface circuits and lightweight PUF for TMR sensors. IEEE Sensors Journal, 23(11), 11754-11761.

Liang, R., Xu, S., Xie, C., Chen, J., Ren, F., Yang, S., & Yabe, T. (2026). Abstain Mask Retain Core: Time Series Prediction by Adaptive Masking Loss with Representation Consistency. Advances in Neural Information Processing Systems, 38, 99445-99471.

Li, S. (2024). Machine Learning in Credit Risk Forecasting â€" â€" A Survey on Credit Risk Exposure. Accounting and Finance Research, 13(2), 107-107.

Zhao, J., Zhang, C., Qin, M., & Yang, P. (2025). QuantFactor REINFORCE: mining steady formulaic alpha factors with variance-bounded REINFORCE. IEEE Transactions on Signal Processing, 73, 2448-2463.

Sha, L., Wang, Y., Meng, P., Deng, Y., Chen, T., Zhang, X., ... & Xu, Q. (2025). Pharmacological inhibition of PSPH reduces serine levels and epileptic seizures. Nature Chemical Biology, 21(11), 1742-1753.

Zhao, J., Zhang, C., Wang, C., & Yang, P. (2025). Learning from expert factors: Trajectory-level reward shaping for formulaic alpha mining. arXiv preprint arXiv:2507.20263.

Peng, Y., & Li, H. (2023). Benthic foraminiferal morphogroups at the end-Guadalupian extinction in eastern Sichuan Basin, China. Palaeogeography, Palaeoclimatology, Palaeoecology, 618, 111522.

Yao, J., Yao, C., Chen, Y., Liu, X., Huang, D., & Chu, W. (2026, April). Explainable AI Enhanced Traffic Forecasting for Proactive Autonomous Vehicle Routing. In 2026 IEEE Conference on Technologies for Sustainability (SusTech) (pp. 1-7). IEEE.

Tao, J., Lyu, R., & Cao, X. (2026). A Deep Learning-Based Automated Content Moderation Framework for Online Platforms.Future-Adaptive Intelligence and Lifelong Systems,1(1).

Ling, C., & Wang, Y. (2025). TLFQC: A High-compatible R Shiny based Platform for Automated and Codeless TLFs Generation and Validation. In PharmaSUG 2025 conference proceedings.

Dong, X., Ramachandra, N., Habib, S., Heitmann, K., Buehlmann, M., & Madireddy, S. (2021). Physical Benchmarking for AI-Generated Cosmic Web. arXiv preprint arXiv:2112.05681.