

A Hybrid Multi-Agent Prompting Approach for Simplifying Complex Sentences

Pratibha Zunjare and Michael S. Hsiao
Electrical and Computer Engineering, Virginia Tech
Virginia Tech, Blacksburg VA, USA

Abstract—This paper addresses the challenge of transforming complex sentences into sequences of logical, simplified sentences while preserving semantic and logical integrity with the help of Large Language Models. We propose a hybrid approach that combines advanced prompting with multi-agent architectures to enhance the sentence simplification process. Experimental results show that our approach was able to successfully simplify 70% of the complex sentences written for video game design application. In comparison, a single-agent approach attained a 48% success rate on the same task.

Index Terms—LLMs, Multi-Agent Systems, Advanced Prompting, Sentence Simplification, Natural Language Processing.

I. INTRODUCTION

Sentence simplification is a challenging task in computational linguistics, aiming to transform complex sentences into simpler structures while preserving the original meaning. Effective sentence simplification has significant applications across numerous domains like education, content accessibility for individuals with cognitive disabilities, automated content creation, robotics, coding, legal documents, etc.

Traditional approaches to sentence simplification have relied on rule-based systems, statistical methods, and more recently neural network architectures [1]. However, these methods often struggle with maintaining semantic equivalence while achieving appropriate simplification levels.

Complex sentences present significant challenges in action-oriented contexts, particularly when attempting to derive executable/actionable functionalities such as robotics, legal documents, and video games. These sentences frequently combine conditional logic (cause-effect constructs), nested dependencies, and multifaceted instructions within intricate syntactic structures. The crafting of unambiguous logic involves converting complex instructions written in natural language into precise, step-by-step actions. However, when sentences are complicated and interdependent on one another, advanced simplification and rewriting techniques are needed, which involve strategically breaking down complex sentences into simple, manageable steps while maintaining the underlying cause-and-effect relationships and procedural flow. For example, consider the following sentence:

When a yellow alien touches a wall, it has 5 seconds before it explodes.

While this sentence is comprehensible to human readers, it contains a temporal constraint of 5 seconds that demands the involvement of one or more intermediate states. This intermediate state helps in arbitrating when the alien touches the wall and when it explodes. Using the running example, one possible sequence of simplified sentences is as follows.

When a yellow alien touches a wall, the alien becomes sad for 5 seconds, and the alien becomes ready_to_explode for 6 seconds. When the alien is ready_to_explode and the alien is not sad, it explodes.

The implicit delay (5 seconds) is transformed into explicit state changes with durations (*sad for 5 seconds*, *ready_to_explode for 6 seconds*). The resulting action (*explode*) is tied to explicit state conditions instead of a timer.

While one might try to ask a Large Language Model (LLM) to perform the re-writing task, the vast variation of simplification strategies renders LLMs ineffective. Likewise, fine-tuning can also result in overfitting due to the limited training set. As a result, something new is needed.

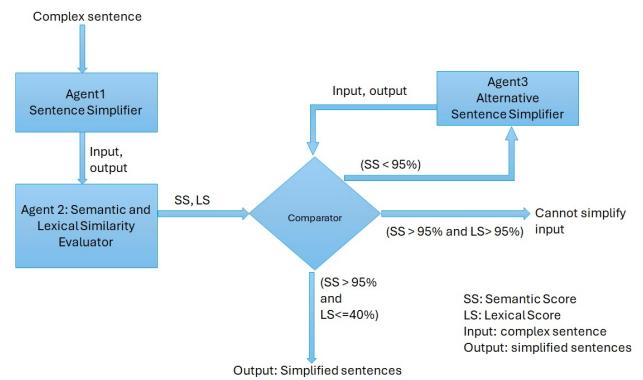


Fig. 1. Multi-Agent prompting workflow for sentence simplification

This paper introduces a novel approach, whose flow-chart is shown in Figure 1, that uses a multi-agent system with advanced prompt engineering to produce simplified sentences. We demonstrate how breaking down complex sentences into clear, actionable sentences enhances both understanding and execution. Our key contributions are: (1) a framework for optimizing prompts using multi-agent collaboration, (2) methods

for translating complex instructions into executable steps, and (3) empirical evaluation of these techniques across a range of scenarios for video game design applications.

While LLMs may be effective in generating simple games, they often struggle with complex games with complex logic. GameChangineer [2] is an educational platform designed to help students develop fundamental logic and critical thinking skills. The proposed simplification process can serve as a learning tool to help students decompose complex scenarios into logical steps, a necessary precursor to learning to code. This platform is also used to evaluate the simplified sentences against the corresponding generated games. The results show that the proposed simplification method was able to successfully decompose and simplify 70% of the complex sentences for gaming applications. In contrast, a single-agent approach achieved a 48% success rate on the same task.

The rest of the paper is organized as follows. Section II reviews the relevant background. Section III describes the proposed hybrid multi-agent prompting framework. Section IV analyzes the experimental results, comparing the performance of LLM-generated games using both simplified and complex sentences on the selected dataset. Section V concludes the paper and discusses directions for future work.

II. BACKGROUND

Sentence Simplification

Sentence simplification methods have evolved from rule-based approaches—using hand-crafted linguistic rules for tasks like sentence splitting and lexical substitution—to data-driven techniques that leverage parallel corpora for learning simplification patterns [3]. More recently, neural models, particularly sequence-to-sequence and transformer-based architectures, have become prevalent, enabling complex operations such as paraphrasing, sentence splitting, and controlled simplification through syntactic and lexical constraints. Modern approaches to sentence simplification are predominantly data-driven, utilizing parallel corpora of aligned complex and simplified sentences to learn effective simplification transformations, such as word substitution, sentence splitting, and reordering [4]. Despite progress, today's LLMs still struggle with semantic and lexical challenges in text simplification [5].

LLMs for Sentence Simplification

LLMs have achieved remarkable capabilities in recent years on natural language processing tasks like text generation and summarization, sentiment analysis, information retrieval and question answering, etc., presenting researchers with diverse methodologies to fully leverage their potential. LLMs like GPT-4o, GPT-4.1 [6], Gemini [7], Perplexity [8], etc. can engage in fluent, contextual, and some reasonable conversations. As LLMs continue to evolve, researchers have developed numerous approaches to fully harness their potential, including prompt engineering, fine-tuning, and multi-agent systems [9]. Despite these advances, certain linguistic challenges remain

particularly difficult, especially those requiring nuanced understanding and transformation of language, including transforming complex sentences found in technical or operational domains—into actionable, logically coherent sentences [10]. Furthermore, large language models are prone to hallucination and often struggle to accurately break down sentences for precise game logic implementation [11].

Advanced Prompting Techniques

Prompt engineering refers to the systematic design of instructions to optimize an LLM's output for specific tasks. Effective prompts typically exhibit key characteristics including specificity, clarity, appropriate structure, and task decomposition [12]. Recent studies demonstrate that LLMs achieve superior performance when utilizing well-structured advanced prompting techniques, leveraging their inherent contextual reasoning capacities to produce more nuanced and reliable outputs compared to static instruction-based approaches [13]. Notable strategies include:

- **Chain-of-Thought (CoT) Prompting:** CoT encourages models to generate intermediate reasoning steps, improving their performance on tasks requiring multi-step inference and decomposition of complex instructions [14].
- **In-Context Learning with Demonstrations:** By providing carefully curated examples within the prompt, LLMs can generalize to new domains and adapt to specific simplification tasks [15] [16].
- **Persona-Based Prompting:** An advanced technique in prompt engineering that involves assigning a detailed role or persona to a LLM to guide its responses in terms of tone, style, reasoning and domain expertise [17].
- **Meta-Prompting and Dynamic Prompt Adaptation:** Recent work explores prompts that evolve dynamically based on model feedback or input complexity, further enhancing robustness in real-world scenarios [18].

As per the result from previous research, prompt design strongly influences the result of large language generative model [17]. These prompting methods can be used based on the type and complexity of the tasks. There are various settings involved in the behavior of LLM for text generation like temperature and top-p. Temperature parameter controls the randomness to the model's output whereas top-p controls the nucleus sampling which adds randomness to the model's output [19]. The design of the prompt and model's setting plays a crucial role in the effectiveness of generated text.

Multi-Agent Architectures

While single-agent LLM offers remarkable fluency, they often struggle with both maintaining operational coherence and handling the hierarchical structure inherent in decomposition and simplification tasks [20]. Multi-agent architectures address these limitations by distributing the simplification process across specialized agents, each responsible for distinct subtasks such as parsing, content extraction, logical validation, and refinement. Multi-agent architectures are categorized based on three key dimensions: the roles of LLMs as actors,

the interaction strategies they employ, and the structural types of the systems [21].

- **Role-Based:** Agents are assigned focused roles (e.g., decomposition, evaluations, validation), mirroring human collaborative workflows and reducing cognitive overload on any single model component [21].
- **Hierarchical and Decentralized Coordination:** Architectures range from centralized controllers orchestrating agent interactions to decentralized, self-organizing agent networks that adapt to task complexity [22].
- **Iterative Cross-Agent Validation:** Agents iteratively review and refine each other’s outputs, leveraging contrastive prompting and feedback mechanisms to enhance output quality [23].

III. METHODOLOGY

We propose a hybrid approach that uses multiple LLMs to break down complex sentences written in natural language into simple, logical, and coherent sentences. We specifically target complex sentences written for video game design application and ensure that the simplified sentences are acceptable by the GameChangineer engine [2]. However, the proposed flow is applicable to other application domains as well.

Algorithm 1 Hybrid Multi-Agent Sentence Simplification Workflow

```

1: Agent 1: Sentence Simplifier
2:  $S \leftarrow \text{LLM\_Breakdown}(Q)$  {LLM generates break-down sentences}
3: Agent 2: Semantic and Lexical score Evaluator
4:  $\text{semantic\_score} \leftarrow \text{Semantic\_Evaluator}(S)$ 
5:  $\text{lexical\_score} \leftarrow \text{Lexical\_Evaluator}(S)$ 
   {Returns the semantic and lexical scores between input and output sentences}
   //Comparator block starts
6: if  $\text{semantic\_score} > 95$  and  $\text{lexical\_score} \leq 40$  then
7:   return  $S$ 
8: end if
9: if  $\text{semantic\_score} > 95$  and  $\text{lexical\_score} > 95$  then
10:  return cannot convert
11: end if
12: if  $\text{semantic\_score} < 95$  then
13:   Agent 3: Alternative sentence Simplifier
14:    $S' \leftarrow \text{LLM\_RevisedBreakdown}(Q)$ 
15:   Call Agent 2
16:   Go to step 6
17: end if //Comparator block ends

```

Our algorithm is based on advanced multi-agent prompting techniques. This framework coordinates with three specialized agents for the following tasks: (1) the decomposition of sentences, (2) the semantic and lexical evaluation, and (3) the iterative revision, to ensure both contextual equivalence and linguistic quality of the resultant sentences with input sentence. The workflow is illustrated in Algorithm 1.

Agent Descriptions

Agent 1: Sentence Simplifier

The sentence simplification agent receives the user query, which typically consists of complex sentences that encompass multiple steps, interdependent causes, and corresponding actions. The input query is incorporated into advanced, carefully designed prompts for GPT-4o, which then decomposes the complex sentence into a sequence of simpler, logically ordered instructions. The meta-instruction-based prompt [13] handles generalized form of complex sentences to decompose into simple, logical cause-action sentences. The output sentences are evaluated based on semantic and syntactic similarity with the input sentence. For example, consider the example shown below in *Example: Reasoning in Prompt1*.

Example 1: Reasoning in Prompt1

Query: *When the rabbit is yellow, the fox that is touched by the rabbit will die. Let’s think step by step:*

- *Rabbit is yellow.* (cause)
- *Rabbit touches the fox.* (cause)
- *Fox dies.* (action)

Answer: *When the rabbit is yellow and the rabbit touches the fox, the fox dies.*

In this example, we illustrate how a step-by-step reasoning process facilitates the logical and sequential decomposition of a complex sentence based on our Agent 1. The prompt is crafted such that the input query is examined to answer the separate causes and actions. The specifics of the prompt are provided in the provided page¹. Such an approach is crucial for translating natural language game specifications into unambiguous, sequential rules that can be reliably interpreted and executed by a game engine. Consider another example,

Example 2: Temporal / Referential Sentence Simplification

Original Query:

When a fox sees the rabbit touch a carrot, it chases it until the rabbit moves.

Transformation Steps:

- Resolve pronouns (*it* → fox or rabbit).
- Separate embedded actions: “sees the rabbit touch” → two actions:
 - *The fox sees the rabbit*
 - *The rabbit touches a carrot*
- Make temporal dependency explicit by splitting into distinct conditions.

Simplified Output (Sentence Simplifier Agent):

- *When a fox sees the rabbit and the rabbit touches a carrot, the fox chases the rabbit.*
- *When the rabbit moves, the fox stops chasing.*

¹https://github.com/pratibhazunj/CS_Simplification

The example 2 shows that pronouns are resolved and temporal dependencies are made explicit by dividing the sentence into two: one describing the chase and the other describing when it stops.

To promote clarity and logical coherence, we systematically examine each sentence to identify primary causal factors, actions, and dependencies. Complex statements are subsequently decomposed into single-action steps articulated in straightforward language, deliberately minimizing the use of technical terminology and modifiers. Conditional relationships are made explicit, and terms like “increase” or “decrease” are standardized. The output follows a strict format, presenting both the original query and the simplified answer. As per our dataset, this agent handles most of various conditional sentences based on cause-effect sentence structure.

Agent 2: Semantic and Lexical Similarity Evaluator

Semantic similarity evaluator checks how closely the meanings of input complex sentence and the generated simplified sentences align, regardless of their structure. It gives a score of 100 if the meaning of the input (complex) and output (simplified) sentences are equivalent; lower score shows divergence in the context.

Lexical similarity score is calculated based on the structure of sentences by ignoring the similar words. Lower lexical scores are preferred because they represent the different structures between the simplified sentences and the input sentence. We provide 3 examples below.

Agent 2 Example 1

Query: *When a fox sees the rabbit touch a carrot, it chases it until the rabbit moves.*

Answer:

- *When the fox sees the rabbit and rabbit touches a carrot, the fox turns silver for 0.1 second.*
- *When the fox is silver and a rabbit touches a carrot, the fox chases the rabbit.*
- *When rabbit moves, the fox stops.*

Semantic Similarity: 100

Lexical Similarity: 40

Agent 2 Example 2

Query: *When the rabbit is red, the fox cannot eat the rabbit.*

Answer: *When the rabbit is red, the fox cannot eat the rabbit.*

Semantic Similarity: 100

Lexical Similarity: 100

This pair is considered a *CannotConvert sentence* as both sentences are identical.

These three examples above illustrate that while semantic similarity remains high (the meaning is preserved), lexical

similarity can vary significantly depending on the structural changes introduced during simplification.

Agent 2 Example 3

Query: *When a rabbit is touched, score adds 1.*

Answer: *When a rabbit is touched, the score increases by 1.*

Semantic Similarity: 90

Lexical Similarity: 70

The Semantic (95%) and Lexical (95% and 40%) score thresholds were empirically determined based on experiments to achieve optimal accuracy and responsiveness for a downstream gaming application(GameChangineer). These values balance precision and flexibility to ensure reliable intent matching while accommodating varied user input.

Agent 3: Alternative Sentence Simplifier

The Alternative Sentence Simplifier is a context-aware language model agent designed to address specialized sentence simplification scenarios involving temporal dynamics and mathematical conditional logic. This approach integrates hybrid prompting by combining directional stimulus prompting, meta-prompting [19], and chain-of-thought (CoT) [14] techniques. Unlike generic simplification approaches, rule-guided transformations are used to resolve ambiguities in sentences where outcomes depend on time-sensitive parameters or arithmetic constraints.

For instance, in interactive gaming narratives, the system accurately simplifies statements like “The character’s movement speed increases if 3 minutes pass without taking damage” by applying domain-specific rules that map temporal triggers (e.g., elapsed time thresholds) and mathematical conditions (e.g., damage counters) to unambiguous causal relationships. This targeted approach ensures faithful preservation of operational logic while reducing syntactic complexity, which is critical for applications requiring precision in dynamic environments.

The Agent 3 is invoked only if the Agent 2 reports a semantic similarity less than 95%, as shown in Algorithm 1.

The specific prompts for Agents 1 and 3 were designed through an iterative process, during which each prompt was refined and assessed to ensure the generation of intended simplified sentences are suitable for downstream code generation in game development contexts.

This agent specializes in analyzing complex sentences that involve dynamic mathematical relationships, such as time sensitive speed adjustments (e.g., “increase the speed by 20% every 5 seconds”) or logic combinations requiring precise numerical reasoning (e.g., “reduce the size by 15% if the rabbit collects three carrot within 10 seconds”). These structures demand explicit handling of temporal dependencies, rate calculations, and conditional arithmetic operations inherent in game mechanics.

Multi-Agent vs. Single-Agent Approach

While one might think that Agents 1 and 3 could be combined into a single agent, this might not work well because single agent models are constrained by the monolithic nature of single-prompt instruction handling. Thus, they often struggle to reconcile multiple objectives within a unified representation. As the complexity of simplification increases, these models tend to exhibit degraded performance and diminished interpretability [23]. The limitations of the single-agent formulation underscore the efficacy and practical benefits of adopting a multi-agent approach for complex sentence simplification.

In contrast, multi-agent approaches offer methodological advantages over single-agent systems, particularly in terms of modularity, adaptability, and scalability. By partitioning responsibilities across multiple specialized agents, the architecture supports fine-grained control over distinct subtasks, enabling more interpretable, maintainable, and extensible systems [22]. This structural decomposition not only improves the performance but also enhances the system's capacity to generalize across a broader spectrum of sentence types.

IV. EXPERIMENTAL RESULTS

Dataset

The dataset for sentence simplification in video gaming was meticulously curated by human annotators. A set of 100 complex sentences containing a range of linguistic and semantic features relevant to game logic and player interaction is used. Each sentence was evaluated for degree of complexity, presence of cause-effect relationships, descriptions of speed or size changes, and the occurrence of multiple actions within a single sentence. The dataset comprises a diverse distribution of sentence types, with conditional sentences forming the majority (50%), followed by sequential (25%), miscellaneous (15%), and a notable subset labeled as '*Cannot Convert*' (10%). Each category was defined: (1) conditional sentences specify actions contingent on in-game conditions (e.g., "When the rabbit touches a gem, it gets more speed."); (2) sequential sentences mandate ordered execution of actions; (3) miscellaneous sentences encompass uncategorized patterns not fitting other classes. The '*Cannot Convert*' subset represents sentences where simplification risks critical information loss or logical incoherence, reflecting inherent limitations in preserving game logic during rephrasing. The curation process prioritized the identification and annotation of cause-and-action structures, as these are pivotal for downstream code generation tasks in gaming environments. For all of the experiments, GPT-4o was used.

Results on dataset

The proposed multi-agent system achieved a successful simplification rate of 70%, with these sentences exhibiting a semantic similarity greater than 95% and a lexical similarity less than or equal to 40%, indicating that the simplified sentences retained the original meaning while employing substantially different wording. Additionally, 10% of the dataset was classified as

'Cannot Convert', reflecting cases where simplification would compromise essential information or logical structure. The 70% of sentences successfully simplified by the system underwent further human evaluation, confirming their acceptability for integration into the GameChangineer platform and ensuring alignment with the requirements of executable game logic.

Model	Acceptance Rate
Baseline LLM	10%
Single Agent Prompt Engineering	48%
Multi-Agent Workflow	70%

Table1: GameChangineer Acceptance Rates

The multi-agent prompting system fails to simplify 30% of the sentences, particularly when handling abstract actions or ambiguous roles. For example, with "*The cobras try to make sure the fox does not touch the bunny*" the system struggles to interpret "try" and translate it into a concrete game action. Similarly, sentences like "*It is not possible to win without eating all the carrots*" lack a clearly defined main player, leading to further simplification failures.

We further evaluated system performance by disabling Agent 3 and relying solely on Agent 1 for simplification tasks. A single-agent approach achieved only a 48% success rate on the same task. Agent 1 struggled with sentences requiring analytical computations, such as "*when the rabbit touches a diamond, its speed increases by 20% every 10 seconds*." and was similarly unable to effectively simplify sentences involving positional and temporal reasoning, such as, "*The rabbit starts at the bottom and moves up four times and drops a carrot every time it moves*." In contrast, Agent 3 effectively managed sentences with mathematical, spatial, and temporal complexities, underscoring the benefits of a multi-agent approach for simplifying intricate instructions.

Game A: With Complex Sentences

Game A (Complex Sentences)

- A fox and a rabbit wander in a field where 10 carrots and a diamond are scattered.
- Arrow keys control the rabbit.
- When a fox sees a rabbit but not a diamond, it chases the rabbit.
- When a fox is near, the speed of the rabbit increases by 0.1.
- When a fox sees the rabbit move, it chases it until the rabbit touches a diamond.
- When the rabbit touches a diamond, it becomes yellow.
- When the rabbit is yellow, it can eat the fox it touches and you win.
- When a fox touches a diamond, it has 5 seconds before it explodes.
- When the fox explodes, the game is over.

The game below consists of complex sentences, such as those explained earlier in the paper. For generating code for the

given game description, logical clarity is essential to derive the correct behavior of all the characters.

Game B: Simplified Version of Game A

The following version is derived using our multi-agent sentence simplification system. Sentences are made lexically simpler and structured for unambiguous logic.

Game B (Simplified Sentences)

- There are 10 carrots, a diamond, a fox, and a rabbit.
- Carrots are scattered randomly.
- Fox and rabbit wander.
- When the left arrow is pressed, the rabbit moves left.
- When the right arrow is pressed, the rabbit moves right.
- When the up arrow is pressed, the rabbit moves up.
- When the down arrow is pressed, the rabbit moves down.
- When a fox sees a rabbit and does not see a diamond, the fox chases the rabbit.
- When a fox is near, the rabbit's speed increases by 0.1.
- When a fox sees the rabbit and the rabbit moves, the fox chases the rabbit.
- When the rabbit touches a diamond, the fox stops chasing.
- When the rabbit is yellow and touches the fox, the rabbit eats the fox.
- When the rabbit eats the fox, you win.
- When a fox touches a diamond, it becomes sad for 5 seconds and `ready_to_explode` for 6 seconds.
- When the fox is `ready_to_explode` and not sad, it explodes.
- When the fox explodes, the game is over.

Game B statements are simplified for code generation with explicit state transitions and unambiguous lexical structure. The revised rules introduce phased behavioral states (e.g., *sad*, *ready_to_explode*) with temporally bounded conditions, creating a finite state machine structure that directly maps to programmable logic.

Unlike Game A's ambiguous temporal chain (e.g., "5 seconds before explosion"), Game B uses clear state labels and logical triggers (e.g., *ready_to_explode AND not sad*) to avoid race conditions. Explicit key mappings (left/right/up/down) also resolve vagueness, enabling direct keycode-to-method binding in implementation.

Human Evaluation on Code generated for Game Description

Finally, we manually crafted more than ten game descriptions, each containing a subset of complex sentences from the data set alongside their simplified counterparts, to assess the impact of sentence complexity on the generation of code using LLMs. When generating game code from complex sentences, LLMs frequently missed constraints and misinterpreted instructions,

resulting in games that ran longer but were inaccurate. Simplified sentences more effectively conveyed constraints, reducing ambiguity and improving code accuracy, yet LLMs can still struggle to consistently fulfill user intentions in both scenarios. Our evaluation indicates that while complex sentences increases ambiguity and reduces accuracy, the simplified sentences increases the number of constraints. Both complex sentences and increased number of constraints pose challenges to general-purpose LLMs.

More specifically, we find that LLMs struggle to generate precise code from games described in natural language that contain complex sentences as the LLMs need to rely on implicit assumptions in complex sentences and may not fully capture the intended logic. Conversely, when a game is rewritten with simplified sentences that offer greater number of constraints that capture both precision and explicit state conditions, the LLMs can overlook these critical conditions if the description contains too many constraints.

Although general-purpose LLMs struggle generating these complex games, domain-specific engines such as the GameChangineer [24] platform not only generated more accurate games from simplified sentences, but also provided detailed debug messages and guidelines when it failed to understand a sentence. Such a feature can enable users to revise and correct errors in the game description. In contrast, general LLMs did not offer any error messages or feedback when encountering problematic sentences, making it difficult for users to identify and resolve issues in the generated code.

V. CONCLUSIONS AND FUTURE WORK

Our approach leverages a hybrid multi-agent framework, wherein distinct LLM-based agents are assigned specialized roles such as sentence simplifier, lexical and semantic similarity evaluator, and alternative sentence simplifier. Two different sentence simplifying agents enable the system to address a wide variety of sentence types and linguistic phenomena, which are often challenging for single-agent models. By allowing agents to operate in parallel to enhance scalability and efficiency, as well as the overall quality of the simplification output. Empirical results show that our hybrid multi-agent model achieves consistently stronger performance than single-agent baselines in both automatic and human evaluations for sentence simplification.

A key challenge of sentence simplification is in simplifying sentences with abstract actions or ambiguous roles, which can often lead to failures in accurately conveying the game logic. Additionally, our current prompt design is closely tied to the dataset, which can lead to overfitting. Future work could address these by incorporating multiple agents to better handle more diverse sentence structures.

REFERENCES

[1] H. Saggion and G. Hirst, *Automatic Text Simplification*, ser. Synthesis Lectures on Human Language Technologies. San Rafael, CA, USA: Morgan & Claypool, 2017, vol. 32.

- [2] M. S. Hsiao, “Automated program synthesis from object-oriented natural language for computer games,” in *Proc. Int. Workshop on Controlled Natural Language*, Aug. 2018.
- [3] F. Alva-Manchego, C. Scarton, and L. Specia, “Data-driven sentence simplification: Survey and benchmark,” *Computational Linguistics*, vol. 46, no. 1, pp. 135–187, 2020.
- [4] —, “Data-driven sentence simplification: Survey and benchmark,” *Computational Linguistics*, vol. 46, no. 1, pp. 135–187, 2020.
- [5] Y. Feng, J. Qiang, Y. Li, Y. Yuan, and Y. Zhu, “Sentence simplification via large language models,” *arXiv preprint arXiv:2302.11957*, 2023.
- [6] OpenAI, “Gpt-4o: Multimodal language model,” OpenAI, Tech. Rep., May 2024, technical Report.
- [7] G. Team *et al.*, “Gemini: A family of highly capable multimodal models,” *arXiv preprint arXiv:2312.11805*, 2023.
- [8] P. AI, “Introducing perplexity deep research,” Perplexity.ai, Feb. 2025, <https://www.perplexity.ai/hub/blog/introducing-perplexity-deep-research>.
- [9] T. Guo, X. Chen, Y. Wang, R. Chang, S. Pei, N. V. Chawla, O. Wiest, and X. Zhang, “Large language model based multi-agents: A survey of progress and challenges,” *arXiv preprint arXiv:2402.01680v2*, Apr. 2024.
- [10] S. Gooding, “On the ethical considerations of text simplification,” in *Proc. Ninth Workshop on Speech and Language Processing for Assistive Technologies (SLPAT-2022)*, 2022, pp. 50–57.
- [11] G. Agrawal, T. Kumarage, Z. Alghamdi, and H. Liu, “Can knowledge graphs reduce hallucinations in llms? a survey,” in *Proc. NAACL*, Mexico City, Mexico, 2024, pp. 3947–3960.
- [12] S. Schulhoff, M. Ilie, N. Balepur, K. Kahadze, A. Liu, C. Si, Y. Li, A. Gupta, H. Han, S. Schulhoff *et al.*, “The prompt report: A systematic survey of prompt engineering techniques,” *arXiv preprint arXiv:2406.06608*, 2024.
- [13] S. M. Bsharat, A. Myrzakhan, and Z. Shen, “Principled instructions are all you need for questioning llama-1/2, gpt-3.5/4,” *arXiv preprint arXiv:2312.16171*, 2024.
- [14] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou *et al.*, “Chain-of-thought prompting elicits reasoning in large language models,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 35, 2022, pp. 24 824–24 837.
- [15] Q. Dong, L. Li, D. Dai, C. Zheng, J. Ma, R. Li, H. Xia, J. Xu, Z. Wu, T. Liu *et al.*, “A survey on in-context learning,” *arXiv preprint arXiv:2301.00234*, 2022.
- [16] Z. Wu, Y. Wang, J. Ye, and L. Kong, “Self-adaptive in-context learning: An information compression perspective for in-context example selection and ordering,” *arXiv preprint arXiv:2212.10375*, 2022.
- [17] M. Cheng, E. O. Durmus, and D. Jurafsky, “Marked personas: Using natural language prompts to measure stereotypes in language models,” *arXiv preprint arXiv:2305.18189*, 2023.
- [18] J. Coda-Forno, M. Binz, Z. Akata, M. Botvinick, J. X. Wang, and E. Schulz, “Meta-in-context learning in large language models,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, 2023.
- [19] M. Suzgun and A. T. Kalai, “Meta-prompting: Enhancing language models with task-agnostic scaffolding,” *arXiv preprint arXiv:2401.12954*, 2024.
- [20] P. Agarwal, H. Dave, J. Bandlamudi, R. Sindhwatta, and K. Mukherjee, “Multi-stage prompting for next best agent recommendations in adaptive workflows,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 21, Mar. 2024, pp. 22 843–22 849.
- [21] K.-T. Tran, D. Dao, M.-D. Nguyen, Q.-V. Pham, B. O’Sullivan, and H. D. Nguyen, “Multi-agent collaboration mechanisms: A survey of llms,” *arXiv preprint arXiv:2501.06322*, Jan. 2025.
- [22] Z. Wang, S. Moriyama, W. Y. Wang, B. Gangopadhyay, and S. Takamatsu, “Talk structurally, act hierarchically: A collaborative framework for llm multi-agent systems,” *arXiv preprint arXiv:2502.11098*, 2025.
- [23] S. Han, Q. Zhang, Y. Yao, W. Jin, Z. Xu, and C. He, “Llm multi-agent systems: Challenges and open problems,” *arXiv preprint arXiv:2402.03578*, 2024.
- [24] M. S. Hsiao, “Multi-phase context vectors for generating feedback for natural-language based programming,” in *Proc. Int. Workshop on Controlled Natural Language*, Sept. 2021.