SID: Multi-LLM Debate Driven by Self Signals

Jianan Zhang | Yanlin Wang | Dongmei Zhang | Zhenan Fan | Jian Shi | Zibin Zheng
SE for AI, Sun Yat-sen University, China | SE for AI, Sun Yat-sen University, China | SE for AI, Sun Yat-sen University, China | Microsoft Research Asia, China | Alibaba Group, China | SE for AI, Sun Yat-sen University, China

ID: 2025-1-Research-Track-Paper-Regular-Paper-Full-Research-Paper
26 Nov 2025 (modified: 26 Nov 2025) | 2025 ACM Conference on Software Testing and Analysis (ISSTA) | Keywords: Software Engineering, Multi-Agent Systems, Large Language Models, Code Generation, Program Repair | 22 pt. | 22 pt.

Abstract. Multi-agent debate (MAD) has emerged as a promising multi-model inference capability across diverse application domains. Recent work has explored that LLM-based MAD offers a way to enhance performance by leveraging the reasoning and critique capabilities of multiple models. However, most existing studies still rely on fixed, human-designed debate protocols. This rule-driving approach limits adaptivity to task complexity and model capabilities, which may hinder both reasoning quality and computational efficiency. We propose Self-Information Debate (SID), a single-model multi-agent debate framework that induces debate behaviors from intrinsic signals rather than predefined protocols. SID uses token-level log-probabilities to allocate compute adaptively, prioritizing uncertain reasoning steps while pruning low-confidence branches. We also introduce Self-Information Gain (SIG), which leverages log-probs of self-signals to model debate quality and introduce a Self-Signals Driven Debate Protocol (SSDD), which leverages log-probs of self-signals to model debate quality and introduce a Self-Signals Driven Debate Protocol (SSDD), which leverages these signals to guide agent allocation, early stopping, and pruning. We evaluate SID on four reasoning benchmarks across three LLM backbones. Results show that SID achieves competitive accuracy while reducing token consumption by up to 19.8% on average. We also find that uniform compute allocation hurts performance, highlighting the effectiveness of utilizing self-signals for enhancing both the performance and efficiency of multi-agent reasoning.

Paper type: Innovative or Creative Models, including LLMs
Artifacts: We plan to make our artifacts available at the time of the event have read and commit to adhering to the ECA Code of Ethics.
Estimated Submitted: 26 Jun 2025 (Pre-AE Artifacts Submitted) | **Artifacts Submitted:** 26 Jun 2025 (Artifacts Submitted)
Artifacts Evaluated: 26 Jun 2025 (Artifacts Evaluated)
Reproducibility: We do have access to and expertise in the artifacts.
Accessibility: We do
Hardware: Any
Software: Any
Justification: We believe the artifact is in LFS, (e.g., github, google) that could be used to find authentic identity.
Artifacts Evaluated: We, the Artifact Evaluation Committee, have evaluated the artifacts submitted for the artifacts track.
Large Language Models: We, the Artifact Evaluation Committee, have evaluated the artifacts submitted for the artifacts track.
Artifacts Evaluated: Yes, the artifacts have been evaluated and reviewed.

Filter by title, date, ... | Sort by relevance | Search keywords ... | Apply Selected Filter(s) | 17 / 17 replies shown

Selected: 0 | Add Official Comment | Add Official Comment

Official Review of Submission 101 by Reviewer 650L
Official Review | Rev. Reviewer 650L | 15 Nov 2025 at 09:06 (modified: 26 Nov 2025 at 12:01) | Keywords: | Revisions:

Summary:
Existing Multi-Agent Debate (MAD) methods rely on external structures and prompts, which limits adaptivity and leads to redundant discussion and extra token consumption. The authors leverage intrinsic token-level log-probs and confidence-based stop-loss signals to dynamically allocate compute, prune low-confidence branches, and stop debate when confidence is high.
Strengths: 1. Clarity
Weaknesses: 1. Novelty
Questions:
1. Do the authors discuss the algorithm reduces token consumption (up to ~19%)
2. Is per-round, per-agent temperature sampling used?
3. The paper is generally well written.
4. Is the Self-Signals Driven Debate Protocol (SSDD) a novel contribution?
5. Do the authors discuss the limitations of using self-signals (log-probs) as a proxy for reasoning quality?
References:
[1] Chen, Z., Wang, W., Tian, Z., Jiang, Y., et al. Countersuffix Attack Language Modeling. NeurIPS, 2025.
[2] Zhang, J., Shi, J., Yang, D., Fan, G., Hao, H., Wang, Z. Identifying and mitigating vulnerabilities in LLM-integrated IoT systems. ICSE-NIER, 2025.
[3] Yang, S., Zu, Q., Shao, Y. and Yang, Y. Multi-Agent Reasoning Benchmark. arXiv:2509.15293, 2025.
[4] Liu, Z., Piwowarski, P. ROCK: Prompt-guided hard negative mining for dense retrieval pre-training. Transactions of the Association for Computational Linguistics (TACL), 2024.
[5] Q. Zhu, Hui, Binhuan, Y., et al. Qwen2.5-Coder technical report. arXiv preprint arXiv:2409.12186, 2024.
Comments:
1. Discussion: Discuss the limitations of using self-signals (log-probs) as a proxy for reasoning quality.
2. The algorithm appears to consume tokens within homogeneous rounds, and calibrating confidence gating across different models (Claude/Gemini/Qwen) may require additional tuning. Please clarify whether the reported token savings account for this overhead.
3. The evaluation focuses on accuracy and token consumption. Adding latency measurements would strengthen the claim that the method reduces wall-clock time in addition to fewer tokens.
4. Minor: Fig. 1 caption: "Self-Info Gain" → "Self-Info Gain".
Rating: 3 - Marginally above the acceptance threshold. Not enough novel impact or significance.
Confidence: 4 - I am fairly confident in my assessment. It is possible that I may have missed some details, but I do not believe that any of my errors are critical enough to change my overall assessment.
Code of Conduct: Yes

Add Official Comment | Add Official Comment

Response Part 1/2
Official Comment | Rev. Reviewer 650L | 26 Nov 2025 at 12:00 | Keywords: | Revisions:

Comment:
Dear Reviewer 650L, Thank you for your patience. Here's our response. We've also revised our manuscript to address your suggestions.

W1. Code availability
Thank you very much for pointing out. We now provide a stripped-down experiment repository containing our full implementation: https://anonymous.4open.science/r/SID-4F40/. This includes the SSDD protocol, baselines, and evaluation scripts for reproducing the main results on GSM8K and MATH with GPT-4o-mini.

W2. Application to heterogeneous debate
We acknowledge that we only used homogeneous backbones for clean ablation. We've added a preliminary experiment in the revised version, summarized below. We plan to extend this in future work.

Setup: We evaluate SID on a heterogeneous ensemble of three agents: GPT-4o-mini, Claude-3.5-Haiku, and Gemini-1.5-Flash, on GSM8K and MATH (500 samples each, 3 debate rounds).

Method	GSM8K	MATH	Avg Tokens
Homogeneous (GPT-4o-mini)	91.2	68.4	1,247
Heterogeneous	92.8	71.6	1,389

Preliminary observation: Heterogeneous ensemble improves accuracy by +1.6 on GSM8K and +3.2 on MATH, at ~11% token overhead. This suggests that model diversity provides complementary reasoning signals, consistent with prior work on multi-agent diversity [1,2].

Thank you for the suggestion. We will include this ablation in the revised version.

Q1. Why does Self-Consistency (SC) appear relatively weak?
SC performs strong on GSM8K (91.2) but drops on MATH (68.4), while SID shows more consistent gains across both datasets. This is expected: SC relies on majority voting over independent samples, which works well when the model's mode is correct but struggles when reasoning errors are systematic. SID's advantage on MATH suggests that iterative refinement with confidence-weighted aggregation helps correct errors that SC cannot.

Q2. Is per-round, per-agent temperature sampling used?
Yes. Temperature is sampled independently for each agent at each round, with a base value of 0.7 and a random offset in [−0.1, +0.1]. This prevents all agents from collapsing to the same reasoning trajectory.

Q3. Is the Self-Signals Driven Debate Protocol (SSDD) a novel contribution?
We believe the following combination is novel: using token-level log-probs from the same model as both a confidence signal for early stopping and a pruning signal for branch allocation, without any external reward model or verifier. Prior work uses confidence for voting [3] or routing [4], but not for joint stopping and pruning in multi-agent debate.

Q4. Do the authors discuss the limitations of using self-signals (log-probs) as a proxy for reasoning quality?
Yes, we discuss this in Section 6.2. Briefly: self-signals can be misleading when the model is confidently wrong (hallucination), and our method inherits this failure mode. We plan to explore ensemble-based signals in future work.

Thank you for the suggestion. We will include latency measurements in the revised version.

Minor: Fig. 1 caption "Self-Info Gain" → "Self-Info Gain". Thanks, fixed.

Rating: 3 - Marginally above the acceptance threshold. Not enough novel impact or significance.
Confidence: 4 - I am fairly confident in my assessment. It is possible that I may have missed some details, but I do not believe that any of my errors are critical enough to change my overall assessment.

Add Official Comment | Add Official Comment

Response Part 2/2
Official Comment | Rev. Reviewer 650L | 26 Nov 2025 at 12:01 | Keywords: | Revisions:

Comment:
Dear Author,
Thank you very much for your thoughtful and constructive rebuttal. We appreciate your time and helpful suggestions, and we have revised the manuscript accordingly. Responses to specific points follow.

W1. Relation to EoT
Thank you for indicating the paper. We will cite the paper and discuss it in the related work section. The discussion is as follows.
Thank you for the suggestion. We will include this ablation in the revised version. The proposed mechanism is intuitive: using token-level log-probs from the same model as both a confidence signal for early stopping and a pruning signal for branch allocation, without any external reward model or verifier. Such signals have been used individually in prior work (e.g., confidence for routing [4], group of models) consistently converge on an important pattern: the resulting "brittleness" will still be high, even though the underlying reasoning is flawed. This risk is amplified when the same model family is used for both generation and evaluation. Your observation that this failure mode is inherited by self-signal based methods is well taken. Ensemble-based signals are a promising direction for future work.
Thank you for the suggestion. We will include latency measurements in the revised version.
Minor: Fig. 1 caption "Self-Info Gain" → "Self-Info Gain". Thanks, fixed.
Rating: 3 - Marginally above the acceptance threshold. Not enough novel impact or significance.
Confidence: 4 - I am fairly confident in my assessment. It is possible that I may have missed some details, but I do not believe that any of my errors are critical enough to change my overall assessment.

Add Official Comment | Add Official Comment

Official Review of Submission 101 by Reviewer Unpaid
Official Review | Rev. Reviewer Unpaid | 15 Nov 2025 at 17:00 | Keywords: | Revisions:

Summary:
The paper proposes SID, a self-signal-driven framework for multi-LLM debate that replaces external controllers (e.g., LLM-as-judge) with two internal cues from the model: Self-Info Gain (SIG) and Self-Confidence. These signals guide three components: a Self-Signals Driven Debate Protocol (SSDD) that allocates compute to uncertain reasoning steps, early stopping when confidence is high, and pruning low-confidence branches. Across four reasoning benchmarks and three LLM backbones, SID achieves competitive accuracy while reducing token consumption by up to 19.8% on average. The authors also introduce SID-1, a variant trained on GSM8K that generalizes well to out-of-distribution tasks, consistently improving accuracy over MADSDM baselines while using token-level signals for up to ~40% of tokens.
Strengths: 1. Clarity
Weaknesses: 1. Novelty
Questions: 1. Novelty
Comments:
1. Well-motivated by improving MAD from several critical perspectives.
2. The evaluation is comprehensive, covering a wide range of LLM/MAS/LoRA ablation/ablation experiments.
3. The presentation is clear and well-organized, with a well-structured Discussion and Conclusion.
Minor/Typos:
1. Missing baselines. There are several previous works with related contributions/motivations, e.g., [1-7] (also incorporated a model-level confidence/reward/preference/ensemble/routing/cascade/...).
2. Typo: "Self-Info Gain" vs "Self-Info Gain".
Questions:
1. Is the conditional Mutual Information Function? Helps LLMs focus on the points of disagreement within a debate. Can we have the model do this by itself, rather than relying on external attention/...?
Rating: 3 - Marginally above the acceptance threshold.
Confidence: 4 - I am fairly confident in my assessment. It is possible that I may have missed some details, but I do not believe that any of my errors are critical enough to change my overall assessment.
Code of Conduct: Yes

Add Official Comment | Add Official Comment

Response Part 1/2
Official Comment | Rev. Reviewer Unpaid | 26 Nov 2025 at 12:00 | Keywords: | Revisions:

Comment:
Dear Reviewer Unpaid,
Thank you very much for your thoughtful and constructive review. We appreciate your time and helpful suggestions, and we have revised the manuscript accordingly. Responses to specific points follow.

W1. Relation to EoT
Thank you for indicating the paper. We will cite the paper and discuss it in the related work section. The discussion is as follows.

W2. Can the model perform disagreement-focused extraction by itself?
Thank you for the suggestion. We will include this ablation in the revised version. The proposed mechanism is intuitive: using token-level log-probs from the same model as both a confidence signal for early stopping and a pruning signal for branch allocation, without any external reward model or verifier. Such signals have been used individually in prior work (e.g., confidence for routing [4], group of models) consistently converge on an important pattern: the resulting "brittleness" will still be high, even though the underlying reasoning is flawed. This risk is amplified when the same model family is used for both generation and evaluation. Your observation that this failure mode is inherited by self-signal based methods is well taken. Ensemble-based signals are a promising direction for future work.

Q1. Can the model perform disagreement-focused extraction by itself?
Thank you for the suggestion. We will include this ablation in the revised version. The proposed mechanism is intuitive: using token-level log-probs from the same model as both a confidence signal for early stopping and a pruning signal for branch allocation, without any external reward model or verifier. Such signals have been used individually in prior work (e.g., confidence for routing [4], group of models) consistently converge on an important pattern: the resulting "brittleness" will still be high, even though the underlying reasoning is flawed. This risk is amplified when the same model family is used for both generation and evaluation. Your observation that this failure mode is inherited by self-signal based methods is well taken. Ensemble-based signals are a promising direction for future work.

Minor/Typos: Thanks, fixed.

Rating: 3 - Marginally above the acceptance threshold.
Confidence: 4 - I am fairly confident in my assessment. It is possible that I may have missed some details, but I do not believe that any of my errors are critical enough to change my overall assessment.

Add Official Comment | Add Official Comment

Response Part 2/2
Official Comment | Rev. Reviewer Unpaid | 26 Nov 2025 at 12:01 | Keywords: | Revisions:

Comment:
Dear Reviewer Unpaid,
Thank you very much for your thoughtful and constructive review. We appreciate your time and helpful suggestions, and we have revised the manuscript accordingly. Responses to specific points follow.

W1. Relation to EoT
Thank you for indicating the paper. We will cite the paper and discuss it in the related work section. The discussion is as follows.

W2. Can the model perform disagreement-focused extraction by itself?
Thank you for the suggestion. We will include this ablation in the revised version. The proposed mechanism is intuitive: using token-level log-probs from the same model as both a confidence signal for early stopping and a pruning signal for branch allocation, without any external reward model or verifier. Such signals have been used individually in prior work (e.g., confidence for routing [4], group of models) consistently converge on an important pattern: the resulting "brittleness" will still be high, even though the underlying reasoning is flawed. This risk is amplified when the same model family is used for both generation and evaluation. Your observation that this failure mode is inherited by self-signal based methods is well taken. Ensemble-based signals are a promising direction for future work.

Q1. Can the model perform disagreement-focused extraction by itself?
Thank you for the suggestion. We will include this ablation in the revised version. The proposed mechanism is intuitive: using token-level log-probs from the same model as both a confidence signal for early stopping and a pruning signal for branch allocation, without any external reward model or verifier. Such signals have been used individually in prior work (e.g., confidence for routing [4], group of models) consistently converge on an important pattern: the resulting "brittleness" will still be high, even though the underlying reasoning is flawed. This risk is amplified when the same model family is used for both generation and evaluation. Your observation that this failure mode is inherited by self-signal based methods is well taken. Ensemble-based signals are a promising direction for future work.

Minor/Typos: Thanks, fixed.

Rating: 3 - Marginally above the acceptance threshold.
Confidence: 4 - I am fairly confident in my assessment. It is possible that I may have missed some details, but I do not believe that any of my errors are critical enough to change my overall assessment.

Add Official Comment | Add Official Comment

Official Review of Submission 101 by Reviewer Unpaid
Official Review | Rev. Reviewer Unpaid | 15 Nov 2025 at 19:00 | Keywords: | Revisions:

Summary:
The paper proposes SID, a multi-agent debate framework for multi-LLM inference that uses token-level log-probs to allocate compute adaptively, prune low-confidence branches, and stop debate when confidence is high. The authors also introduce Self-Information Gain (SIG) and a Self-Signals Driven Debate Protocol (SSDD), which leverage these signals to guide agent allocation, early stopping, and pruning. Across four reasoning benchmarks and three LLM backbones, SID achieves competitive accuracy while reducing token consumption by up to 19.8% on average. The authors also find that uniform compute allocation hurts performance, highlighting the effectiveness of utilizing self-signals for enhancing both the performance and efficiency of multi-agent reasoning.
Strengths: 1. Clarity
Weaknesses: 1. Novelty
Questions:
1. How much of the efficiency is coming from the early stop and reduction?
2. Is the Self-Signals Driven Debate Protocol (SSDD) a novel contribution? Individual components (token-level log-probs, early stopping, pruning) have been explored in prior work. Is the novelty in the combination of these signals? Or is there something more specific about how the signals are used to coordinate the debate?
3. Does the hyperparameter sweep (Fig. 4) suggest the method is sensitive to the threshold? Are the chosen values robust across tasks and models?
Comments:
1. Clarity: The paper is well written and easy to follow. The motivation is clear and the ablations are thoroughly presented.
2. Minor: Fig. 1 caption: "Self-Info Gain" → "Self-Info Gain".
Rating: 3 - Marginally above the acceptance threshold. Not enough novel impact or significance.
Confidence: 4 - I am fairly confident in my assessment. It is possible that I may have missed some details, but I do not believe that any of my errors are critical enough to change my overall assessment.
Code of Conduct: Yes

Add Official Comment | Add Official Comment

Response Part 1/2
Official Comment | Rev. Reviewer Unpaid | 26 Nov 2025 at 12:00 | Keywords: | Revisions:

Comment:
Dear Reviewer Unpaid,
Thank you very much for your thoughtful and constructive review. We appreciate your time and helpful suggestions, and we have revised the manuscript accordingly. Responses to specific points follow.

W1. Relation to EoT
Thank you for indicating the paper. We will cite the paper and discuss it in the related work section. The discussion is as follows.

W2. Can the model perform disagreement-focused extraction by itself?
Thank you for the suggestion. We will include this ablation in the revised version. The proposed mechanism is intuitive: using token-level log-probs from the same model as both a confidence signal for early stopping and a pruning signal for branch allocation, without any external reward model or verifier. Such signals have been used individually in prior work (e.g., confidence for routing [4], group of models) consistently converge on an important pattern: the resulting "brittleness" will still be high, even though the underlying reasoning is flawed. This risk is amplified when the same model family is used for both generation and evaluation. Your observation that this failure mode is inherited by self-signal based methods is well taken. Ensemble-based signals are a promising direction for future work.

Q1. Can the model perform disagreement-focused extraction by itself?
Thank you for the suggestion. We will include this ablation in the revised version. The proposed mechanism is intuitive: using token-level log-probs from the same model as both a confidence signal for early stopping and a pruning signal for branch allocation, without any external reward model or verifier. Such signals have been used individually in prior work (e.g., confidence for routing [4], group of models) consistently converge on an important pattern: the resulting "brittleness" will still be high, even though the underlying reasoning is flawed. This risk is amplified when the same model family is used for both generation and evaluation. Your observation that this failure mode is inherited by self-signal based methods is well taken. Ensemble-based signals are a promising direction for future work.

Minor/Typos: Thanks, fixed.

Rating: 3 - Marginally above the acceptance threshold.
Confidence: 4 - I am fairly confident in my assessment. It is possible that I may have missed some details, but I do not believe that any of my errors are critical enough to change my overall assessment.

Add Official Comment | Add Official Comment

Response Part 2/2
Official Comment | Rev. Reviewer Unpaid | 26 Nov 2025 at 12:01 | Keywords: | Revisions:

Comment:
Dear Reviewer Unpaid,
Thank you very much for your thoughtful and constructive review. We appreciate your time and helpful suggestions, and we have revised the manuscript accordingly. Responses to specific points follow.

W1. Relation to EoT
Thank you for indicating the paper. We will cite the paper and discuss it in the related work section. The discussion is as follows.

W2. Can the model perform disagreement-focused extraction by itself?
Thank you for the suggestion. We will include this ablation in the revised version. The proposed mechanism is intuitive: using token-level log-probs from the same model as both a confidence signal for early stopping and a pruning signal for branch allocation, without any external reward model or verifier. Such signals have been used individually in prior work (e.g., confidence for routing [4], group of models) consistently converge on an important pattern: the resulting "brittleness" will still be high, even though the underlying reasoning is flawed. This risk is amplified when the same model family is used for both generation and evaluation. Your observation that this failure mode is inherited by self-signal based methods is well taken. Ensemble-based signals are a promising direction for future work.

Q1. Can the model perform disagreement-focused extraction by itself?
Thank you for the suggestion. We will include this ablation in the revised version. The proposed mechanism is intuitive: using token-level log-probs from the same model as both a confidence signal for early stopping and a pruning signal for branch allocation, without any external reward model or verifier. Such signals have been used individually in prior work (e.g., confidence for routing [4], group of models) consistently converge on an important pattern: the resulting "brittleness" will still be high, even though the underlying reasoning is flawed. This risk is amplified when the same model family is used for both generation and evaluation. Your observation that this failure mode is inherited by self-signal based methods is well taken. Ensemble-based signals are a promising direction for future work.

Minor/Typos: Thanks, fixed.

Rating: 3 - Marginally above the acceptance threshold.
Confidence: 4 - I am fairly confident in my assessment. It is possible that I may have missed some details, but I do not believe that any of my errors are critical enough to change my overall assessment.

Add Official Comment | Add Official Comment

Official Review of Submission 101 by Reviewer mH9G
Official Review | Rev. Reviewer mH9G | 15 Nov 2025 at 20:00 | Keywords: | Revisions:

Summary:
The paper proposes Self-Information Debate (SID), a single-model multi-agent debate framework that induces debate behaviors from intrinsic signals rather than predefined protocols. SID uses token-level log-probs to allocate compute adaptively, prunes low-confidence branches, and stops debate when confidence is high. Across four reasoning benchmarks and three LLM backbones, SID achieves competitive accuracy while reducing token consumption by up to 19.8% on average. The authors also find that uniform compute allocation hurts performance, highlighting the effectiveness of utilizing self-signals for enhancing both the performance and efficiency of multi-agent reasoning.
Strengths: 1. Clarity
Weaknesses: 1. Novelty
Questions:
1. How much of the efficiency is coming from the early stop and reduction?
2. Is the Self-Signals Driven Debate Protocol (SSDD) a novel contribution? Individual components (token-level log-probs, early stopping, pruning) have been explored in prior work. Is the novelty in the combination of these signals? Or is there something more specific about how the signals are used to coordinate the debate?
3. Does the hyperparameter sweep (Fig. 4) suggest the method is sensitive to the threshold? Are the chosen values robust across tasks and models?
Comments:
1. Clarity: The paper is well written and easy to follow. The motivation is clear and the ablations are thoroughly presented.
2. Minor: Fig. 1 caption: "Self-Info Gain" → "Self-Info Gain".
Rating: 3 - Marginally above the acceptance threshold. Not enough novel impact or significance.
Confidence: 4 - I am fairly confident in my assessment. It is possible that I may have missed some details, but I do not believe that any of my errors are critical enough to change my overall assessment.
Code of Conduct: Yes

Add Official Comment | Add Official Comment

Response Part 1/2
Official Comment | Rev. Reviewer mH9G | 26 Nov 2025 at 12:00 | Keywords: | Revisions:

Comment:
Dear Reviewer mH9G,
Thank you very much for your thoughtful and constructive review. We appreciate your time and helpful suggestions, and we have revised the manuscript accordingly. Responses to specific points follow.

W1. Relation to EoT
Thank you for indicating the paper. We will cite the paper and discuss it in the related work section. The discussion is as follows.

W2. Can the model perform disagreement-focused extraction by itself?
Thank you for the suggestion. We will include this ablation in the revised version. The proposed mechanism is intuitive: using token-level log-probs from the same model as both a confidence signal for early stopping and a pruning signal for branch allocation, without any external reward model or verifier. Such signals have been used individually in prior work (e.g., confidence for routing [4], group of models) consistently converge on an important pattern: the resulting "brittleness" will still be high, even though the underlying reasoning is flawed. This risk is amplified when the same model family is used for both generation and evaluation. Your observation that this failure mode is inherited by self-signal based methods is well taken. Ensemble-based signals are a promising direction for future work.

Q1. Can the model perform disagreement-focused extraction by itself?
Thank you for the suggestion. We will include this ablation in the revised version. The proposed mechanism is intuitive: using token-level log-probs from the same model as both a confidence signal for early stopping and a pruning signal for branch allocation, without any external reward model or verifier. Such signals have been used individually in prior work (e.g., confidence for routing [4], group of models) consistently converge on an important pattern: the resulting "brittleness" will still be high, even though the underlying reasoning is flawed. This risk is amplified when the same model family is used for both generation and evaluation. Your observation that this failure mode is inherited by self-signal based methods is well taken. Ensemble-based signals are a promising direction for future work.

Minor/Typos: Thanks, fixed.

Rating: 3 - Marginally above the acceptance threshold.
Confidence: 4 - I am fairly confident in my assessment. It is possible that I may have missed some details, but I do not believe that any of my errors are critical enough to change my overall assessment.

Add Official Comment | Add Official Comment

Response Part 2/2
Official Comment | Rev. Reviewer mH9G | 26 Nov 2025 at 12:01 | Keywords: | Revisions:

Comment:
Dear Reviewer mH9G,
Thank you very much for your thoughtful and constructive review. We appreciate your time and helpful suggestions, and we have revised the manuscript accordingly. Responses to specific points follow.

W1. Relation to EoT
Thank you for indicating the paper. We will cite the paper and discuss it in the related work section. The discussion is as follows.

W2. Can the model perform disagreement-focused extraction by itself?
Thank you for the suggestion. We will include this ablation in the revised version. The proposed mechanism is intuitive: using token-level log-probs from the same model as both a confidence signal for early stopping and a pruning signal for branch allocation, without any external reward model or verifier. Such signals have been used individually in prior work (e.g., confidence for routing [4], group of models) consistently converge on an important pattern: the resulting "brittleness" will still be high, even though the underlying reasoning is flawed. This risk is amplified when the same model family is used for both generation and evaluation. Your observation that this failure mode is inherited by self-signal based methods is well taken. Ensemble-based signals are a promising direction for future work.

Q1. Can the model perform disagreement-focused extraction by itself?
Thank you for the suggestion. We will include this ablation in the revised version. The proposed mechanism is intuitive: using token-level log-probs from the same model as both a confidence signal for early stopping and a pruning signal for branch allocation, without any external reward model or verifier. Such signals have been used individually in prior work (e.g., confidence for routing [4], group of models) consistently converge on an important pattern: the resulting "brittleness" will still be high, even though the underlying reasoning is flawed. This risk is amplified when the same model family is used for both generation and evaluation. Your observation that this failure mode is inherited by self-signal based methods is well taken. Ensemble-based signals are a promising direction for future work.

Minor/Typos: Thanks, fixed.

Rating: 3 - Marginally above the acceptance threshold.
Confidence: 4 - I am fairly confident in my assessment. It is possible that I may have missed some details, but I do not believe that any of my errors are critical enough to change my overall assessment.

Add Official Comment | Add Official Comment