

Self-reflecting Large Language Models: A Hegelian Dialectical Approach

Sara Abdali¹, Michael Solodko¹, Can Goksen¹, Saeed Amizadeh¹,

Julie E. Maybee², Kazuhito Koishida¹, Pashmina Cameron¹

¹Applied Sciences Group (ASG), Microsoft, Redmond, WA, USA

²Department of Philosophy, Lehman College, City University of New York, Bronx, NY, USA

saraabdali@microsoft.com

Abstract

In this paper, we introduce a self-reflection framework for Large Language Models (LLMs) grounded in the *Hegelian Dialectic*, a philosophical method in which an initial proposition is challenged by a generated opposition, and both are reconciled into a unified, more comprehensive idea. We formalize this process as an iterative operator over the space of consistent theories and apply it to two complementary tasks: (i) generating novel scientific ideas across domains such as mathematics, physics, economics, and philosophy, and (ii) improving reasoning by enabling LLMs to identify and correct their own errors through structured self-critique. We study generation temperature through two configurations (a dynamic annealing schedule that shifts from creative exploration to refinement, and a constant temperature), to examine the effect of fixed versus dynamic temperature rather than advocate either. To evaluate ideas without domain experts, we introduce Multi-Agent Majority Voting (MAMV), in which multiple LLMs independently assess the validity and novelty of each synthesis. Our experiments show significant gains over baselines on mathematical (GSM-8k, GSM-hard), symbolic (GSM-Symbolic), and knowledge-intensive (MMLU Pro) reasoning, with promising qualitative results in open-ended scientific ideation.

1 Introduction

LLMs have driven rapid progress in reasoning, scientific discovery, and creative problem solving (Zhang et al., 2024c; Wu et al., 2023; Zhang et al., 2024a; Zheng et al., 2023), with In-Context Learning (ICL) (Dong et al., 2024) enabling zero- and few-shot reasoning from instructions or demonstrations without additional training. Yet ensuring factual accuracy and logical consistency in multi-step reasoning remains challenging (Abdali et al., 2024a,c): LLMs neglect key conditions, misinterpret context, and hallucinate (Shayegani et al.,

2023; Millièrè and Buckner, 2024; Abdali et al., 2024b), which is especially costly in domains like mathematics and science where one logical error can invalidate an entire chain.

Many strategies aim to mitigate this. *Training-based* approaches include fine-tuning on curated data (Lewkowycz et al., 2022; Rajani et al., 2019; Zelikman et al., 2022), RLHF (Ziegler et al., 2019; Christiano et al., 2017a), and truthfulness-oriented data pruning (Christiano et al., 2017b). *Inference-time* methods include Chain-of-Thought (CoT) (Wei et al., 2022), Reversing CoT (Xue et al., 2023), self-consistency (Wang et al., 2023), scratchpads (Cobbe et al., 2021; Nye et al., 2022), and retrieval (Gua et al., 2020). More recently, *multi-agent debate* (MAD) frameworks, Society of Minds (Du et al., 2024), Multi-Persona (Liang et al., 2023), and ChatEval (Chan et al., 2024), show that structured argumentation between agents improves factual accuracy in QA (Smit et al., 2024).

A complementary direction is *iterative self-reflection* (Shinn et al., 2023; Madaan et al., 2023), where a model critiques and improves its own outputs. However, naive self-reflection suffers from *degeneracy-of-thought*: once converged on a high-confidence answer, later iterations add no meaningful revision (Liang et al., 2023; Wang et al., 2024b), motivating a more structured form of self-critique.

We view self-reflection as a form of *self-dialectic*, a discourse that advances by considering and resolving opposing views (Cambridge University Press, n.), with meaning varying across traditions (Bobzien and Duncombe, 2023; Maybee, 2020). In particular, the *Hegelian Dialectic* (Hegel, 1807, 1951; Maybee, 2020) has an initial proposition generate its own opposition by exposing its defects, then reconcile the two into a higher-order unified idea preserving the strengths of both. This structure (understanding, sublation, speculation) overcomes degeneracy-of-thought, as it *requires* a substantive opposition before any synthesis.

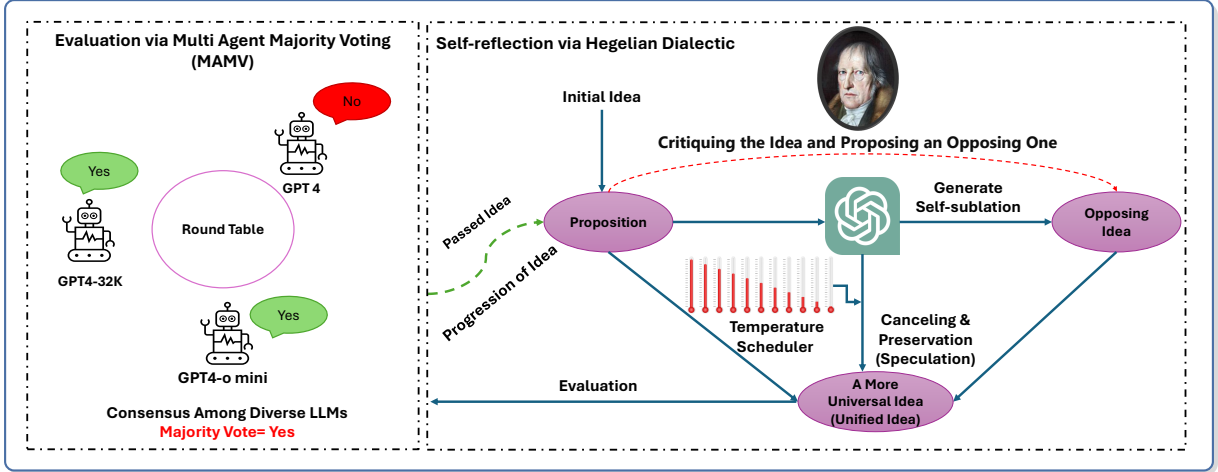


Figure 1: **Overview of Hegelian Dialectical Self-Reflection.** A proposition is challenged via sublation and reconciled via speculation under two temperature configurations. MAMV evaluates novelty and validity of ideation.

Philosophy-inspired NLP has attracted growing interest, connecting computational methods with classical methodologies (Millière and Buckner, 2024; Milliere and Buckner, 2024). We propose a Hegelian dialectical approach to structured self-reflection that enables LLMs to generate novel ideas, identify errors, and correct them during problem solving. We formalize the dialectic as an iterative operator over consistent theories (Section 3), instantiate it with LLM-based sublation and speculation operators (Section 1), and explore the effect of generation temperature via two configurations: a *dynamic annealing* schedule formalizing the shift from creative exploration to focused refinement (Delahaye et al., 2019), and a *constant temperature* setting, treating both as configurations to study rather than proposing either as our method. To evaluate generated ideas without domain experts, we introduce *Multi-Agent Majority Voting (MAMV)* (Minsky, 1988; Zhuge et al., 2023; Amirizani et al., 2024), in which multiple LLMs independently vote on validity and novelty. Figure 1 overviews our method. In summary:

- We introduce a self-reflection framework grounded in Hegel’s dialectic, formalized as an iterative operator with provable consistency, preservation, and progress properties.
- We study the effect of fixed vs. dynamic temperature through extensive analysis.
- We propose MAMV, a scalable framework to assess validity/novelty without experts.
- We achieve significant gains across multiple models and benchmarks, plus promising scientific ideation across topics.

2 Related Work

Self-Reflection in LLMs In the context of LLMs, self-reflection involves evaluating and refining a model’s outputs through iterative cycles of feedback and adjustment (Chen et al., 2024; Zhang et al., 2024b). Li et al. (Zhang et al., 2024b) explore diverse perspectives, summarize discrepancies, and organize them into a checklist for refinement. Madaan et al. (Madaan et al., 2023) improve LLM outputs through iterative feedback, while Reflexion (Shinn et al., 2023) enables agents to learn from trial-and-error. Self-reflection has also been applied to Retrieval Augmented Generation (Asai et al., 2024), and translation (Wang et al., 2024a). However, naive self-reflection, where the model is simply instructed to reconsider its answer, suffers from *degeneracy-of-thought*: after converging on a high-confidence response, the model fails to diverge in subsequent iterations (Liang et al., 2023; Wang et al., 2024b). Our work addresses this limitation by replacing unconstrained self-critique with *structured dialectical opposition*, ensuring that each iteration introduces a substantive counter-argument grounded in identifiable defects.

Philosophy-Inspired NLP Researchers have recently examined NLP through a philosophical lens, connecting computational methods with classical debates about cognition, semantics, and intelligence (Millière and Buckner, 2024; Milliere and Buckner, 2024). Philosophical prompting techniques have emerged as a means to enhance LLM reasoning: the *Socratic method* (Chang, 2023) uses structured questioning to stimulate critical think-

ing, while the Kantian-inspired *UPAR* framework (Understand, Plan, Act, Reflect) (Geng et al., 2023) structures prompting to emulate human cognitive processes. Our work extends this line of research by drawing on Hegel’s dialectical method, a richer philosophical framework that, unlike Socratic questioning or Kantian reflection, provides a formal mechanism for *opposition generation* and *speculative unification*, naturally suited to iterative refinement in generative models. To our knowledge, ours is the first to formalize Hegel’s dialectic as a computational operator and instantiate it within an LLM-based reasoning framework.

3 Background

Hegelian Dialectical Method The term *Dialectic* denotes a logical discourse that advances through the consideration and resolution of opposing views (Cambridge University Press, n.), a concept central to Hegel’s dialectic (1807) (Hegel, 1807, 2010). Unlike Plato’s *elenchus*, where a contradiction is terminal and forces the premises to be rejected, Hegel treats contradictions as *productive tensions* that drive ideas toward higher unity and synthesis (Popper, 1940; Maybee, 2020). And unlike Fichte’s rigid thesis–antithesis–synthesis triad, Hegel described an organic process of *self-sublation* (*Aufhebung*) in which each stage simultaneously negates, preserves, and elevates the preceding one (Maybee, 2020).

From Philosophy to Computation This method is naturally suited to computational instantiation: it is *iterative* (each cycle’s output seeds the next, mirroring auto-regressive generation), *constructive* (generates new content via speculative unification rather than merely eliminating, as in *elenchus*), and *self-correcting* (the sublation step explicitly surfaces defects, addressing the degeneracy-of-thought problem of naive self-reflection (Liang et al., 2023)). These motivate our formalization.

Formal Characterization We ground the dialectical process in first-order logic (FOL), following the formalization of Inoue (Inoue, 2014). Let $\mathcal{L}$ denote a first-order language and let $\text{Th}(\mathcal{L})$ be the set of consistent theories expressible in $\mathcal{L}$. A *dialectical trajectory* is a sequence $\langle T_0, (A_0, S_0), T_1, (A_1, S_1), \dots, T_N \rangle$ where each T_i is a proposition, A_i is its opposition, and $S_i = T_{i+1}$ is the speculative unification.

Definition 3.1 (Dialectical Opposition (Inoue, 2014)). Let L_1 and L_2 be FOL languages, and let $T_1 \in \text{Th}(L_1)$, $T_2 \in \text{Th}(L_2)$ be consistent theories. For a sentence $w \in L_1 \cap L_2$:

- (i) w *dialectically opposes* T_2 relative to T_1 iff $T_1 \vdash w$ and $T_2 \not\vdash w$.
 - (ii) w *dialectically contradicts* T_2 relative to T_1 iff $T_1 \vdash w$ and $T_2 \vdash \neg w$.
- We write $T_1 \stackrel{w}{\asymp} T_2$ for opposition and $T_1 \stackrel{w}{\bowtie} T_2$ for contradiction.

Definition 3.2 (Dialectical Operators). Let $\mathcal{L}$ be a first-order language and $\text{Th}(\mathcal{L})$ the set of consistent theories in $\mathcal{L}$. We define:

- (i) *Sublation*: $\mathcal{A} : \text{Th}(\mathcal{L}) \rightarrow \text{Th}(\mathcal{L})$ maps T_i to $A_i = \mathcal{A}(T_i)$ s.t. $\exists w : T_i \vdash w \wedge A_i \vdash \neg w$.
- (ii) *Speculation*: $\mathcal{S} : \text{Th}(\mathcal{L})^2 \rightarrow \text{Th}(\mathcal{L})$ maps (T_i, A_i) to a unified theory $T_{i+1} = \mathcal{S}(T_i, A_i)$.
- (iii) *Dialectical*: $\mathcal{D} = \mathcal{S} \circ (\text{id} \times \mathcal{A})$, i.e.:

$$T_{i+1} = \mathcal{D}(T_i) = \mathcal{S}(T_i, \mathcal{A}(T_i)) \quad (1)$$

Proposition 3.3 (Speculation Constraints). Let $S_i = \mathcal{S}(T_i, A_i)$. Then S satisfies $\forall T_i, A_i \in \text{Th}(\mathcal{L})$:

- (a) **Consistency**: $S_i \not\vdash \perp$ (non-trivial)
- (b) **Preservation**: $S_i \supseteq T_i \cap A_i$ (common ground retained)
- (c) **Progress**: $S_i \not\subseteq T_i \wedge S_i \not\subseteq A_i$ (advances beyond either)

The iterative application of $\mathcal{D}$ generates a dialectical trajectory $\{T_0, T_1, T_2, \dots\}$ with monotonically non-decreasing common ground between successive proposition–opposition pairs:

$$|T_0 \cap A_0| \leq |T_1 \cap A_1| \leq \dots \quad (2)$$

This monotonicity reflects the accumulation of reconciled premises: as the dialectic progresses, a unified theory T_{i+1} inherits an increasingly large set of validated claims from prior iterations, progressively narrowing the space of admissible oppositions while broadening the scope of the unification.

Remark (Non-determinism and Temperature). The speculation operator $\mathcal{S}$ is not uniquely determined. Multiple valid unifications may exist for a given pair (T_i, A_i) , forming a feasible set:

$$\mathcal{F}_i = \{T \in \text{Th}(\mathcal{L}) \mid T \supseteq T_i \cap A_i, T \not\vdash \perp\}$$

In our LLM instantiation, we sample from $\mathcal{F}_i$ via the generation temperature τ , which governs the entropy of the token-level distribution:

- **Low** τ (exploitation): concentrates probability mass on the most likely unification.
- **High** τ (exploration): encourages diverse speculative syntheses.

This connects the philosophical notion of *speculative freedom* to the information-theoretic concept of *distributional entropy*.

Hegel’s Three Moments Hegel frames the process as three “moments” that map directly onto our operators (Maybee, 2020; Hegel, 1807): **Understanding** ($\mathfrak{U}$), the stable current proposition T_i ; the **Dialectical moment / Sublation** ($\mathfrak{A}$), in which the flaws of T_i *itself* drive an opposition $A_i = \mathcal{A}(T_i)$ (hence *self-sublation*, which we enforce by conditioning $\mathcal{A}$ on T_i alone); and the **Speculative moment** ($\mathfrak{S}$), a unified theory $T_{i+1} = \mathcal{S}(T_i, A_i)$ that integrates and transcends both, deriving its character from their specific opposition. The cycle is:

$$\underbrace{T_i}_{\mathfrak{U}} \xrightarrow{\mathcal{A}} \underbrace{A_i}_{\mathfrak{A}} \xrightarrow{\mathcal{S}} \underbrace{T_{i+1}}_{\mathfrak{S}} \quad (3)$$

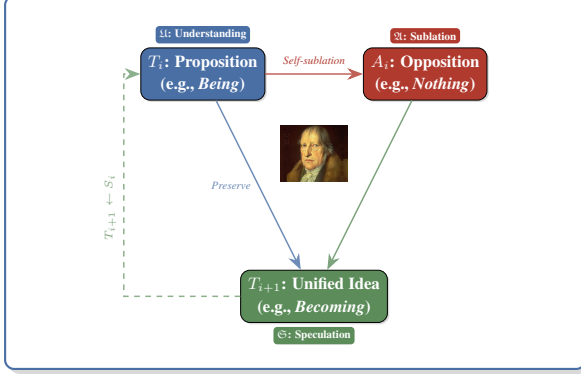


Figure 2: **The Hegelian Dialectic.** A proposition T_i ($\mathcal{U}$) generates its opposition A_i via self-sublation ($\mathcal{A}$); their interaction produces a unified idea T_{i+1} through speculation ($\mathcal{S}$), which becomes the next proposition.

Crucially, $\mathcal{S}$ sublates *both* prior moments, negating them with a new concept yet still depending on them for its definition (Maybee, 2020).

Consistency and Paraconsistency A distinctive feature of the speculative moment is that oppositions need not be fully resolved, admitting ambiguity among competing perspectives (Hegel, 1951, 1807). Classically, $T_i \cup A_i$ may be inconsistent ($\vdash \perp$), which by the *principle of explosion* would trivialize the theory ($\vdash \psi$ for any ψ); we avoid this via the constraints of Proposition 3.3, resolving oppositions through qualification while accumulating their generating premises. This unresolved-opposition stance drew criticism from Popper (Popper, 1940) and Russell (Russell, 1959) for violating non-contradiction (Gottlieb, 2023), but *paraconsistent logics* (Inoue, 2014; Maybee, 2020) admit both φ and $\neg\varphi$ without explosion ($\varphi, \neg\varphi \not\vdash \psi$), providing a natural foundation for Hegel’s unifying step and framing our consistency constraint (Prop. 3.3a) as a design choice rather than a necessity.

Applications of the Dialectic The dialectic has been applied across politics (Boukhatem, 2022), economics (Neschen, 2013), physical sciences (e.g., geocentric vs. heliocentric astronomy, classical vs. relativistic mechanics) (Inoue, 2014), and ecological dynamics (Angeler and Maybee, 2025), in each case synthesizing conflicting hypotheses into more comprehensive models.

4 Proposed Hegelian Self-reflection

Following the Hegelian Dialectic, an LLM $\mathcal{M}$ iteratively critiques its own *Proposition* via a generated *Opposition* and fuses both into a *Unified response*.

Step 1: Understanding The process begins with an initial proposition $T_0 \in \text{Th}(\mathcal{L})$.

Step 2: Dialectical Self-sublation At iteration i , $\mathcal{M}$ generates an opposition A_i at temperature τ_A :

Sublation. $A_i = \mathcal{M}(T_i, \tau_A, p_A)$ s.t. $\exists w: T_i \vdash w \wedge A_i \vdash \neg w$. The constant τ_A ensures iteration-invariant opposition.

Step 3: Speculation with Annealing $\mathcal{M}$ produces a unified response S_i :

Speculation. $S_i = \mathcal{M}(T_i, A_i, \tau(i), p_S)$
s.t. $S_i \supseteq (T_i \cap A_i) \wedge S_i \not\vdash \perp$, where:

$$\tau(i) = \tau_0 \cdot e^{-\theta i}, \quad \theta \geq 0, i \in \mathbb{N} \quad (4)$$

Setting $\theta = 0$ recovers constant temperature $\tau(i) = \tau_0 \forall i$.

The compositional dependency is:

$$\begin{aligned} A_i &= \mathcal{A}_{\tau_A}(T_i), S_i = \mathcal{S}_{\tau(i)}(T_i, A_i) \\ \implies S_i &= \mathcal{S}_{\tau(i)}(T_i, \mathcal{A}_{\tau_A}(T_i)) \end{aligned} \quad (5)$$

τ governs exploration-exploitation: high τ promotes diverse syntheses, low τ concentrates on refinement. The proposition is updated: $T_{i+1} \leftarrow S_i$.

Probabilistic Formalization We formalize our complete framework as a joint probabilistic model. At each iteration t , the joint distribution over the dialectical triple (T_t, A_t, S_t) factorizes according to the three Hegelian moments:

$$P_{\tau(t), \tau_A}(S_t, A_t, T_t) = \underbrace{P(T_t)}_{\mathcal{U}} \cdot \underbrace{P_{\tau_A}(A_t | T_t)}_{\mathcal{A}} \cdot \underbrace{P_{\tau(t)}(S_t | T_t, A_t)}_{\mathcal{S}} \quad (6)$$

where $\tau(t) \in \{\tau_0, \tau_0 e^{-\theta}, \dots, \tau_0 e^{-\theta n}\}$ and $\tau_A \in \mathbb{R}^+$. The Markov property $T_{t+1} = S_t$ induces the chain:
 $T_0 \rightarrow (A_0, S_0) \rightarrow T_1 \rightarrow (A_1, S_1) \rightarrow \dots \rightarrow T_N$

4.1 Measuring Validity and Novelty

Inspired by Minsky’s “*Society of Mind*” (Minsky, 1988; Zhuge et al., 2023), we employ K LLMs to independently assess validity and novelty via majority voting. We adopt a greedy stopping heuristic: if any step fails MAMV, the process terminates. Algorithm 1 formalizes our self-reflection method.

MAMV Decision Rule. Given LLMs $\{M_1, \dots, M_K\}$, prompts p_V, p_N , and triple (T_i, A_i, S_i) :

$$\begin{aligned} v_k &= M_k(p_V, T_i, A_i, S_i) \in \{0, 1\} \\ n_k &= M_k(p_N, T_i, S_i) \in \{0, 1\} \end{aligned} \quad (7)$$

$$\text{MAMV}(\cdot) = \mathbb{P}\left[\sum_k v_k > \frac{K}{2}\right] \wedge \mathbb{P}\left[\sum_k n_k > \frac{K}{2}\right] \quad (8)$$

The novelty score is $\mathcal{N} = |\{i : \text{MAMV}(\cdot) = 1\}|/N$.

5 Prompt Engineering

Here we outline the prompts utilized in our method.

5.1 Self-reflection Prompting

Sublation: Figures 4 and 6 in Appendix B show the sublation prompts used for idea generation and math reasoning, respectively. Engaging in self-debating is one way to generate diverse viewpoints. To simulate the sublation process, we employ an iterative self-debating, Solo Performance Prompting (SPP) strategy. SPP harnesses the model’s theory of mind reasoning abilities by instructing it to “split” into various personas and collaborate on a given prompt through a brainstorming session among these personas (Wang et al., 2024b). Our framework requests the model to generate an arbitrary persona for each iteration of the self-dialectic. The following instructs the model to practice SPP:

Sublation Prompt. *Imagine you are someone X who has noticed a problem with or defect in a proposed view. Produce an opposed view on the same topic that corrects for the noted defect or problem.*

It is necessary to explicitly define the criteria for what constitutes potentially relevant types of defects. For ideation, we instruct the model with:

Defect Categories. Possible defects are:

1. The proposed view has constitutive elements that are not fully defined;
2. The proposed view does not account for some relevant phenomena or examples (its content is incomplete);
3. The proposed view is incompatible with some of the phenomena or examples that it is supposed to include;
4. The proposed view cannot be fully defined on its own, but can be defined only in relation to another view.

Speculation: Rather than choosing one view over the other, speculation seeks to unify opposing ideas by proposing a third perspective:

Unification Instruction. Produce a third view on the same topic that:

- Unifies the previous two views in relation to the defect or problem on which the two views were opposed to one another.
- This third view must be a unifying theory that explains how the two views agree with or are the same as one another and how the two views are opposed to one another.

Figures 5 and 7 in Appendix B illustrate the speculation prompts used for idea generation and math reasoning.

5.2 MAMV Prompting

Validity: To verify that the model has adhered to the speculation instructions, we provide MAMV with the instructions for the unifying process and ask each model to vote “yes” or “no” on whether the instructions are followed. Figure 8 in Appendix B demonstrates this process.

Novelty: Evaluating novelty is more challenging, as it requires domain experts who are familiar with all contributions in the field. Here, we define novelty as the introduction of new ideas that build upon the previous step’s proposition i.e., T_{i-1} . The corresponding prompt is reported in Figure 9 in Appendix B. We emphasize that our method generates new ideas that may not always be scientifically correct. Our focus, however, is on the validity of the dialectical process provided through instructions for the generation of the unified idea, as well as the novelty of these ideas compared to the propositions and oppositions in previous steps. The scientific evaluation of these ideas could be examined by experts in the field using various scientific methods. We stress that evaluating novelty is not just about adding new information to the previous proposition; it also requires a thorough understanding of existing literature to ensure the ideas are truly unexplored. However, this does not undermine the effectiveness of our method as an early exploration but highlights the need for better evaluations, potentially involving human/AI experts to develop more effective ways.

6 Experiments

We apply our method to *reasoning* and *ideation* tasks, comparing constant against annealing temperature; the full setup and hyper-parameters (Table 3) are in Appendix A.

6.1 Quantitative Results: Reasoning

Table 1 compares our method against standard prompting and multi-turn Self-refine, and Table 2 isolates the temperature schedule.

Comparison with baselines. *Our method outperforms all prompting baselines, including Self-refine, on GSM-Symbolic for all five models, and is best or tied on GSM-8k.* The relative gains over the best baseline on GSM-Symbolic are substantial: GPT-4o $0.852 \rightarrow 0.884$ (+3.8%), GPT-4o-mini $0.709 \rightarrow 0.770$ (+8.6%), and Qwen 2.5-7B $0.587 \rightarrow 0.623$ (+6.1%). On the near-saturated GSM-8k (baselines already exceed 0.94 for the GPT models) we top every model, tying the strong Self-refine baseline (0.933) on Qwen3-8B. GSM-hard and MMLU Pro are more competitive, yet the reasoning-tuned Qwen3-8B posts the largest single gain of any model on GSM-hard ($0.585 \rightarrow 0.658$, +12.5%). Overall we top GSM-hard everywhere except GPT-4o, and on MMLU Pro we lead on

Model	Context	Prompt	GSM-8k	GSM-Symbolic	GSM-hard	MMLU Pro
GPT-4o	128k	Zero-shot	0.863±0.004	0.758±0.006	0.597±0.003	0.725±0.000
		Zero-shot+CoT	0.863±0.004	0.760±0.004	0.587±0.004	0.714±0.000
		Few-shot	0.943±0.002	0.851±0.003	0.642±0.004	0.712±0.000
		Few-shot+CoT	0.944±0.002	0.843±0.005	0.640±0.008	0.723±0.000
		Self-refine	0.954±0.004	0.852±0.004	0.655±0.003	0.749±0.000
		Dialectic (Ours)	0.955±0.000	0.884±0.004	0.646±0.005	0.749±0.000
GPT-4o-mini	128k	Zero-shot	0.871±0.005	0.709±0.005	0.541±0.007	0.593±0.000
		Zero-shot+CoT	0.874±0.003	0.697±0.003	0.546±0.003	0.639±0.000
		Few-shot	0.921±0.000	0.690±0.004	0.552±0.003	0.505±0.000
		Few-shot+CoT	0.921±0.001	0.708±0.009	0.566±0.007	0.617±0.000
		Self-refine	0.933±0.002	0.709±0.003	0.574±0.005	0.627±0.000
		Dialectic (Ours)	0.940±0.003	0.770±0.008	0.574±0.006	0.673±0.000
Qwen 2.5-7B-Instruct	1M	Zero-shot	0.828±0.000	0.513±0.000	0.491±0.000	0.422±0.000
		Zero-shot+CoT	0.816±0.000	0.516±0.000	0.501±0.000	0.454±0.000
		Few-shot	0.866±0.000	0.527±0.000	0.517±0.000	0.411±0.000
		Few-shot+CoT	0.856±0.000	0.511±0.000	0.499±0.000	0.411±0.000
		Self-refine	0.898±0.000	0.587±0.000	0.515±0.000	0.469±0.000
		Dialectic (Ours)	0.907±0.000	0.623±0.000	0.548±0.000	0.499±0.000
Phi-4	16k	Zero-shot	0.869±0.000	0.741±0.000	0.560±0.000	0.695±0.000
		Zero-shot+CoT	0.861±0.000	0.742±0.000	0.559±0.000	0.706±0.000
		Few-shot	0.949±0.000	0.825±0.000	0.619±0.000	0.687±0.000
		Few-shot+CoT	0.936±0.000	0.806±0.000	0.611±0.000	0.699±0.000
		Self-refine	0.950±0.000	0.839±0.000	0.629±0.000	0.711±0.000
		Dialectic (Ours)	0.954±0.000	0.849±0.000	0.635±0.000	0.709±0.000
Qwen3-8B	32k	Zero-shot	0.801±0.000	0.669±0.000	0.454±0.000	0.542±0.000
		Zero-shot+CoT	0.854±0.000	0.626±0.000	0.506±0.000	0.551±0.000
		Few-shot	0.916±0.000	0.658±0.000	0.578±0.000	0.565±0.000
		Few-shot+CoT	0.921±0.000	0.723±0.000	0.578±0.000	0.572±0.000
		Self-refine	0.933±0.000	0.774±0.000	0.585±0.000	0.601±0.000
		Dialectic (Ours)	0.933±0.000	0.784±0.000	0.658±0.000	0.613±0.000

Table 1: Dialectical self-reflection vs. other prompting. The highlighted **Dialectic (Ours)** row reports the stronger of our constant and annealing schedules per benchmark. **Bold** = best, underline = second best in each column.

GPT-4o-mini and both Qwen models, tie GPT-4o, and trail Self-refine by just 0.002 on Phi-4.

Algorithm 1 Hegelian Dialectical Self-Reflection

Part 1: Hegelian Dialectical Self-Reflection

```

1: Input: LLM  $\mathcal{M}$ , prompts  $p_A, p_S$ , proposition  $T_0$ ,
   temperatures  $\tau_0, \tau_A, \text{decay}$ ,  $\theta$ , max iterations  $N$ 
2: Output: Final unified idea  $S^*$ 
3:  $T \leftarrow T_0$  {Initialize proposition}
4: for  $i = 0$  to  $N - 1$  do
5:    $A_i \leftarrow \mathcal{M}(T_i, \tau_A, p_A)$  { $\mathfrak{A}$ : Sublation}
6:    $\tau(i) \leftarrow \tau_0 \cdot e^{-\theta i}$  {Update temperature}
7:    $S_i \leftarrow \mathcal{M}(T_i, A_i, \tau(i), p_S)$  { $\mathfrak{S}$ : Speculation}
8:   if  $\text{MAMV}(T_i, A_i, S_i) = 1$  then
9:      $T_{i+1} \leftarrow S_i$  {Accept & advance}
10:   else
11:      $S^* \leftarrow T_i$ ; break {Halt}
12:   end if
13: end for
14: return  $S^* \leftarrow S_i$ 

```

Part 2: Multi-Agent Majority Voting (MAMV)

```

1: Input: Triple  $(T, A, S)$ , prompts  $p_V, p_N$ ,
   LLM panel  $\{M_1, \dots, M_K\}$ 
2: Output: Decision  $d \in \{0, 1\}$ 
3: for  $k = 1$  to  $K$  do
4:    $v_k \leftarrow M_k(p_V, T, A, S)$  {Validity}
5:    $n_k \leftarrow M_k(p_N, T, S)$  {Novelty}
6: end for
7:  $d \leftarrow \mathbb{P}[\sum_k v_k > \frac{K}{2}] \wedge \mathbb{P}[\sum_k n_k > \frac{K}{2}]$ 
8: return  $d$ 

```

Model	Temperature	GSM-8k	GSM-Symbolic	GSM-hard	MMLU Pro
GPT-4o	Constant	0.953±0.004	0.879±0.003	0.646±0.005	0.749±0.000
	Annealing	0.955±0.000	0.884±0.004	0.644±0.006	0.749±0.000
GPT-4o-mini	Constant	0.939±0.004	0.770±0.008	0.572±0.005	0.673±0.000
	Annealing	0.940±0.003	0.769±0.000	0.574±0.006	0.663±0.000
Qwen 2.5-7B-Instruct	Constant	0.907±0.000	0.615±0.000	0.548±0.000	0.499±0.000
	Annealing	0.895±0.000	0.623±0.000	0.547±0.000	0.499±0.000
Phi-4	Constant	0.951±0.000	0.849±0.000	0.633±0.000	0.707±0.000
	Annealing	0.954±0.000	0.844±0.000	0.635±0.000	0.709±0.000
Qwen3-8B	Constant	0.933±0.000	0.784±0.000	0.658±0.000	0.605±0.000
	Annealing	0.932±0.000	0.780±0.000	0.645±0.000	0.613±0.000

Table 2: **Ablation study:** Accuracy for constant vs. annealing speculation temperature.

Task-specific analysis. GSM-8k/hard target multi-step arithmetic, while Symbolic requires abstract symbolic manipulation. Gains are *most pronounced on the symbolic and knowledge tasks*: the sublation-speculation cycle surfaces and repairs the structural errors that dominate rule-based reasoning and that a single forward pass leaves uncorrected. Qwen3-8B is the exception: already strong on symbolic structure, its gains concentrate on GSM-hard, where iterative error-correction pays off most on longer arithmetic chains.

Annealing vs. constant temperature. Table 2 compares the two schedules (Eq. 4). Neither dominates: annealing wins GSM-8k for the GPT and Phi-4 models (three of five), while both open-weight Qwen models (2.5-7B and Qwen3-8B) prefer a constant temperature on GSM-8k and GSM-hard; elsewhere the schedules trade off and stay within error bars on most MMLU Pro cells.

Variance reduction. *Beyond accuracy, our method consistently reduces output variance.* By Proposition 3.3, the preservation constraint ($S_i \supseteq T_i \cap A_i$) retains validated premises, progressively narrowing the solution space.

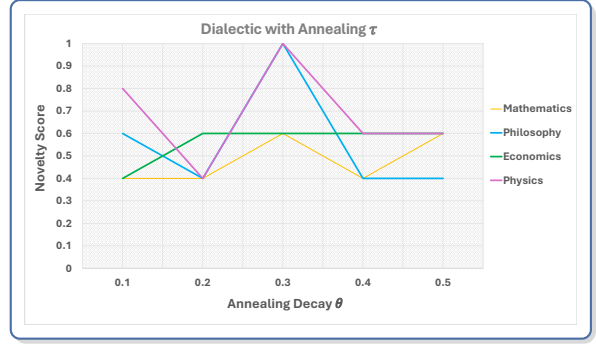
Effect of context length. Dialectical prompting is multi-turn: each iteration appends the full proposition-opposition-unification history. Larger windows hold the whole trajectory without truncation, keeping the reasoning chain intact, so gains grow with context; smaller windows (e.g., Phi-4,

16k) compress it, giving more modest yet still best-among-all gains. *The effect saturates once the window fits the trajectory*: our three-iteration runs need only a few thousand tokens, so beyond that the base model’s reasoning matters more than window size. Qwen3-8B shows this: with just 32k, below GPT-4o’s 128k or Qwen 2.5’s 1M, its reasoning-tuned backbone gives the largest GSM-hard gain of any model (Table 1), while Phi-4’s 16k is where context truly binds. Figure 11 (Appendix C) shows a complete dialectical iteration.

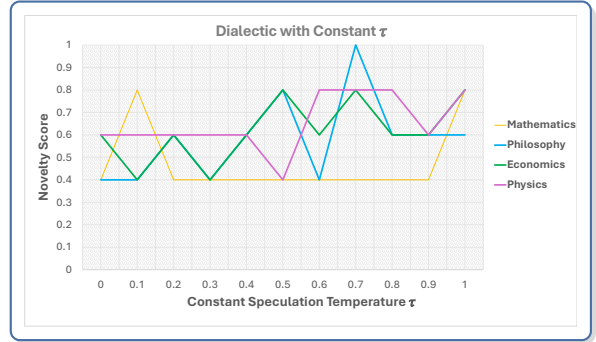
6.2 Qualitative Results: Scientific Ideation

Observations and Key Findings Across hundreds of GPT-4o runs spanning annealing and constant τ and slightly varied prompts, we observe:

- 1) Identical configs and prompts can still yield varying responses, with different step counts and novelty scores, though ideas from one configuration usually stay on the same topic.
- 2) A high constant speculation τ can yield either out-of-topic or genuinely novel unified ideas.
- 3) The most novel unifications occur early; over iterations, ideas grow more refined and absorb opposing viewpoints, making meaningful opposition harder to generate and sometimes repeating it, a sign it has become comprehensive and stable.
- 4) For reasoning, the unified idea’s level of agreement and the opposition’s intensity (opposing qualifications vs. fundamental premises) act as prompt-based hyper-parameters shaping the self-debate.
- 5) Both temperatures shape novelty: for sublation, low τ_A yields arguments that directly address the proposition, while high τ_A yields near-orthogonal, contradictory perspectives whose resolution qualifies both sides into self-consistent but too-agreeable discussions ($\tau_A = 0.5$ is a compromise). For speculation (Figure 3a), low decay ($\theta \rightarrow 0$) mimics a constant high temperature while high decay over-suppresses creativity, with $\theta = 0.3$ maximizing novelty; under constant τ , higher values raise novelty but risk topic drift (Figure 3b).
- 6) Our hyper-parameters and prompts are not universal and must be tuned per topic.
- 7) More statements in the proposition raise the likelihood of opposing views, yielding more meaningful opposition and additional steps. A full dialectical iteration, as well as ideation examples for math, physics, economics, and philosophy, appear in Appendices C and D.



(a) **Annealing decay θ** . Low decay ($\theta \rightarrow 0$) mimics constant high temperature; high decay ($\theta \rightarrow 1$) mimics low temperature. The optimal $\theta = 0.3$ maximizes $\mathcal{N}$.



(b) **Constant temperature τ** . Higher τ increases novelty but risks topic drift; lower τ maintains coherence at the cost of reduced creativity.

Figure 3: Effect of speculation temperature on novelty $\mathcal{N}$: (a) dynamic annealing schedule and (b) constant temperature.

7 Limitations and Future Directions

In this section, we briefly discuss some limitations of our current framework and experimental setup.

Deliberate adaptation and interpretation choices in Hegelian dialectic We formulate Hegel’s dialectic as a 2-step progress, where the flow from proposition to opposition and then unified idea does not follow any sort of logical necessity as proposed by (Maybee, 2020; Kaufmann, 1965; Inoue, 2014). This also means that the perspective from which the opposition introduces its opposition is arbitrary, i.e. delegated to the LLM agent. (2) The unified idea always maintains a self-consistent position. In Hegel’s dialectic, statements can remain ambiguous, where they are true from one perspective and false from the other, without requiring a resolution (Hegel, 1807; Maybee, 2020). Interested readers can refer to (Maybee, 2020; Inoue, 2014) for more details.

Measuring novelty of a statement is challenging as it involves subjective assessment and context-awareness. Novelty is not just about adding new information, it also requires understanding existing literature to ensure that the idea has not been previously explored. Comparing an idea solely based on an initial proposition does not account for depth of existing research. Our novelty prompt has inherent limitations. It does not explicitly define whether altering the strength of an argument is considered novel, even if all points have been previously mentioned. Also, it remains unclear whether the model should regard the negation of existing premises as novel, as it involves manipulating existing information rather than introducing new information. However, we view such negation as novel.

Difficulties in distinguishing creativity from LLM remembering in baseline evaluation Evaluating new ideas against pre-proposed unified idea is extremely challenging, as distinguishing whether a model is merely recalling training information or generating creative ideas is challenging.

Measuring the effect of annealing due to randomness of generation: We often observed that, a constant τ either fails to produce novel content when set too low or leads to irrelevant information when set too high. However, due to inherent randomness in the generation, it is not straightforward to generalize this observation. Conducting statistical significance tests over multiple rounds of prompting can help to evaluate this observation more effectively.

Reproducibility of results: The inherent randomness in the generation, coupled with constantly evolving nature of LLMs and the lack of control, especially when using black-box models (e.g. GPT-x), makes it difficult to reproduce the results which is essential in scientific settings. While white-box models offer better control over outputs, they may not be as powerful as black-box competitors.

Repetition and randomness in API calls We have occasionally observed identical opposition and unification paragraphs, as well as main points, raising concerns about the extent of randomness and the possible presence of hidden caching mechanisms during these experiments.

Need for domain expert LLMs Utilizing the extensive knowledge of domain-specific LLM experts can help identify unique contributions and

ensure comprehensive coverage of the field, making it more feasible to measure novelty.

Future directions. In this paper, for simplicity, we used a constant τ_A . However, as we discussed earlier, it directly affects novelty, with lower τ_A producing arguments directly addressing the initial proposition and higher τ_A results in orthogonal perspectives that lead to epistemological inquiries. Investigating multiple oppositions with varying τ_A simultaneously, generating unified ideas, and backtracking from undesired outcomes is worthwhile. We reserve this exploration for future research.

Currently, with the novelty score stop condition, our framework strives to resolve dialectical contradictions every cycle. In future implementations, the statements from opposing view points might be explicitly considered as ambiguous and not integrated into the unified idea until enough iterations have passed to resolve them.

Meanwhile, in the MAD setting, the Multi-Persona framework addresses the degeneracy-of-thought issue in naive self-reflection by pre-assigning one agent to express viewpoints and another to oppose them (Liang et al., 2023). By adjusting the likelihood of agreement in debate protocols, the Multi-Persona MAD framework surpasses other MAD frameworks in the Q&A setting by tuning the agreeability of debating agents. (Liang et al., 2023; Smit et al., 2024). Similarly, the decisiveness of the speculation step can be tuned and compared with the performance of agreeable/disagreeable MAD frameworks.

8 Concluding Remarks

We framed LLM self-reflection as a self-dialectical process grounded in Hegel’s dialectics: a cycle of understanding, sublation, and speculation that challenges an initial proposition with opposing views and refines them into a unified idea combining the strengths of both. We studied how speculation temperature affects novelty under constant and dynamic annealing schedules, assessing novelty with MAMV, a multi-agent collaborate-and-vote system. Our method yields significant gains on mathematical, symbolic, and knowledge-intensive reasoning (GSM-8k, GSM-hard, GSM-Symbolic, and MMLU Pro) and promising idea generation: high constant temperatures occasionally produced off-topic or highly novel ideas, while annealing surfaced the most innovative ideas early, growing more nuanced and stable over time.

References

- Sara Abdali, Richard Anarfi, C. J. Barberan, and Jia He. 2024a. [Decoding the AI pen: Techniques and challenges in detecting ai-generated text](#). In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2024, Barcelona, Spain, August 25-29, 2024*, pages 6428–6436. ACM.
- Sara Abdali, Richard Anarfi, CJ Barberan, Jia He, and Erfan Shayegani. 2024b. [Securing large language models: Threats, vulnerabilities and responsible practices](#).
- Sara Abdali, Jia He, CJ Barberan, and Richard Anarfi. 2024c. [Can llms be fooled? investigating vulnerabilities in llms](#).
- Maryam Amirizani, Elias Martin, Maryna Sivachenko, Afra Mashhadi, and Chirag Shah. 2024. [Can llms reason like humans? assessing theory of mind reasoning in llms for open-ended questions](#). In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management, CIKM '24*, page 34–44, New York, NY, USA. Association for Computing Machinery.
- David G. Angeler and Julie E. Maybee. 2025. [Chapter three - dialectical ecosystems: Theory and applications](#). In Alex J. Dumbrell, editor, *Ecological Horizons: From Nature to People, Part 1*, volume 72 of *Advances in Ecological Research*, pages 39–90. Academic Press.
- Akari Asai, Zeqiu Wu, Yizhong Wang, Avirup Sil, and Hannaneh Hajishirzi. 2024. [Self-rag: Learning to retrieve, generate, and critique through self-reflection](#). In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net.
- Susanne Bobzien and Matthew Duncombe. 2023. Dialectical School. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*, Fall 2023 edition. Metaphysics Research Lab, Stanford University.
- Hassiba Boukhatem. 2022. Hegelian dialectics applications in the 21st century politics. *Revue Akofena*. Cambridge University Press. n. [Dialectic](#).
- Chi-Min Chan, Weize Chen, Yusheng Su, Jianxuan Yu, Wei Xue, Shanghang Zhang, Jie Fu, and Zhiyuan Liu. 2024. [Chateval: Towards better llm-based evaluators through multi-agent debate](#). In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net.
- Edward Y. Chang. 2023. [Prompting large language models with the socratic method](#). *ArXiv preprint*, abs/2303.08769.
- Dongping Chen, Jiawen Shi, Yao Wan, Pan Zhou, Neil Zhenqiang Gong, and Lichao Sun. 2024. Self-cognition in large language models: An exploratory study. *Proceedings of the ICML 2024 Large Language Models and Cognition Workshop*.
- Paul F. Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017a. [Deep reinforcement learning from human preferences](#). In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 4299–4307.
- Paul F. Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2017b. [Deep reinforcement learning from human preferences](#). In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 4299–4307.
- Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. 2021. [Training verifiers to solve math word problems](#). *ArXiv preprint*, abs/2110.14168.
- Daniel Delahaye, Supatcha Chaimatanan, and Marcel Mongeau. 2019. [Simulated Annealing: From Basics to Applications](#), pages 1–35. Springer International Publishing, Cham.
- Qingxiu Dong, Lei Li, Damai Dai, Ce Zheng, Jingyuan Ma, Rui Li, Heming Xia, Jingjing Xu, Zhiyong Wu, Baobao Chang, Xu Sun, Lei Li, and Zhifang Sui. 2024. [A survey on in-context learning](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 1107–1128, Miami, Florida, USA. Association for Computational Linguistics.
- Yilun Du, Shuang Li, Antonio Torralba, Joshua B. Tenenbaum, and Igor Mordatch. 2024. [Improving factuality and reasoning in language models through multiagent debate](#). In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net.
- Hejia Geng, Boxun Xu, and Peng Li. 2023. [Upur: A kantian-inspired prompting framework for enhancing large language model capabilities](#). *ArXiv preprint*, abs/2310.01441.
- Paula Gottlieb. 2023. Aristotle on Non-contradiction. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*, Winter 2023 edition. Metaphysics Research Lab, Stanford University.
- Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Ming-Wei Chang. 2020. Realm: Retrieval-augmented language model pre-training. ICML'20. JMLR.org.
- Georg Wilhelm Friedrich Hegel. 1807. *The Phenomenology of Spirit*. Cambridge University Press (2019).

- Georg Wilhelm Friedrich Hegel. 1951. *Hegel's Science of Logic*. Humanity Books, Amherst, N.Y. Verify edition: Humanity Books (Amherst, NY) is the A. V. Miller translation (1998/1999); the 1951 edition was Allen & Unwin, trans. Johnston & Struthers.
- Georg Wilhelm Friedrich Hegel. 2010. *The Science of Logic [Wissenschaft der Logik]*. Cambridge University Press, New York.
- Kazumi Inoue. 2014. [Dialectical contradictions and classical formal logic](#). *International Studies in the Philosophy of Science*, 28(2):113–132.
- Walter Kaufmann. 1965. *Hegel: Reinterpretation, Texts, and Commentary*. Doubleday & Company, Garden City, New York.
- Aitor Lewkowycz, Anders Andreassen, David Dohan, Ethan Dyer, Henryk Michalewski, Vinay V. Ramasesh, Ambrose Slone, Cem Anil, Imanol Schlag, Theo Gutman-Solo, Yuhuai Wu, Behnam Neyshabur, Guy Gur-Ari, and Vedant Misra. 2022. [Solving quantitative reasoning problems with language models](#). In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*.
- Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and Zhaopeng Tu. 2023. [Encouraging divergent thinking in large language models through multi-agent debate](#).
- Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegrefe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad Majumder, Katherine Hermann, Sean Welleck, Amir Yazdanbakhsh, and Peter Clark. 2023. [Self-refine: Iterative refinement with self-feedback](#). In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*.
- Julie E. Maybee. 2020. Hegel's Dialectics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, Winter 2020 edition. Metaphysics Research Lab, Stanford University.
- Raphael Milliere and Cameron Buckner. 2024. [A philosophical introduction to language models – part ii: The way forward](#). *ArXiv preprint*, abs/2405.03207.
- Raphaël Millière and Cameron Buckner. 2024. [A philosophical introduction to language models – part i: Continuity with classic debates](#). *ArXiv preprint*, abs/2401.03910.
- Marvin Minsky. 1988. *Society of Mind*. Simon and Schuster.
- Albena Neschen. 2013. [Hegel on modern economics and business](#). In Christoph Lütge, editor, *Handbook of the Philosophical Foundations of Business Ethics*, pages 415–429. Springer.
- Maxwell Nye, Anders Johan Andreassen, Guy Gur-Ari, Henryk Michalewski, Jacob Austin, David Bieber, David Dohan, Aitor Lewkowycz, Maarten Bosma, David Luan, Charles Sutton, and Augustus Odena. 2022. Show your work: Scratchpads for intermediate computation with language models.
- Karl R. Popper. 1940. [What is dialectic?](#) *Mind*, 49(194):403–426.
- Nazneen Fatema Rajani, Bryan McCann, Caiming Xiong, and Richard Socher. 2019. [Explain yourself! leveraging language models for commonsense reasoning](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4932–4942, Florence, Italy. Association for Computational Linguistics.
- Bertrand Russell. 1959. *My Philosophical Development*. George Allen & Unwin, London.
- Erfan Shayegani, Md. Abdullah Al Mamun, Yu Fu, Pedram Zaree, Yue Dong, and Nael B. Abu-Ghazaleh. 2023. [Survey of vulnerabilities in large language models revealed by adversarial attacks](#). *ArXiv preprint*, abs/2310.10844.
- Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. 2023. [Reflexion: language agents with verbal reinforcement learning](#). In *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*.
- Andries P. Smit, Nathan Grinsztajn, Paul Duckworth, Thomas D. Barrett, and Arnau Pretorius. 2024. [Should we be going mad? A look at multi-agent debate strategies for llms](#). In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net.
- Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V. Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023. [Self-consistency improves chain of thought reasoning in language models](#). In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net.
- Yutong Wang, Jiali Zeng, Xuebo Liu, Fandong Meng, Jie Zhou, and Min Zhang. 2024a. Taste: Teaching large language models to translate through self-reflection. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6144–6158. Association for Computational Linguistics.
- Zhenhailong Wang, Shaoguang Mao, Wenshan Wu, Tao Ge, Furu Wei, and Heng Ji. 2024b. [Unleashing the emergent cognitive synergy in large language models: A task-solving agent through multi-persona self-collaboration](#). In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 257–279.

- Mexico City, Mexico. Association for Computational Linguistics.
- Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. 2022. [Chain-of-thought prompting elicits reasoning in large language models](#). In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*.
- Lijun Wu and 1 others. 2023. [The impact of large language models on scientific discovery: A preliminary study using gpt-4](#). *ArXiv preprint*, abs/2311.07361.
- Tianci Xue, Ziqi Wang, Zhenhailong Wang, Chi Han, Pengfei Yu, and Heng Ji. 2023. [Rcot: Detecting and rectifying factual inconsistency in reasoning by reversing chain-of-thought](#). *ArXiv preprint*, abs/2305.11499.
- Eric Zelikman, Yuhuai Wu, Jesse Mu, and Noah D. Goodman. 2022. [Star: Bootstrapping reasoning with reasoning](#). In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*.
- Qiang Zhang and 1 others. 2024a. [Scientific large language models: A survey on biological & chemical domains](#). *ArXiv preprint*, abs/2401.14656.
- Wenqi Zhang, Yongliang Shen, Linjuan Wu, Qiuying Peng, Jun Wang, Yueting Zhuang, and Weiming Lu. 2024b. Self-contrast: Better reflection through inconsistent solving perspectives. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.
- Yu Zhang, Xiusi Chen, Bowen Jin, Sheng Wang, Shuiwang Ji, Wei Wang, and Jiawei Han. 2024c. [A comprehensive survey of scientific large language models and their applications in scientific discovery](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 8783–8817, Miami, Florida, USA. Association for Computational Linguistics.
- Yizhen Zheng, Huan Yee Koh, Jiaxin Ju, Anh T. N. Nguyen, Lauren T. May, Geoffrey I. Webb, and Shirui Pan. 2023. [Large language models for scientific synthesis, inference and explanation](#).
- Mingchen Zhuge, Haozhe Liu, Francesco Faccio, Dylan R. Ashley, R’obert Csord’as, Anand Gopalakrishnan, Abdullah Hamdi, Hasan Hammoud, Vincent Herrmann, Kazuki Irie, Louis Kirsch, Bing chuan Li, G. Li, Shuming Liu, Jinjie Mai, Piotr Pikekos, Aditya Ramesh, Imanol Schlag, Weimin Shi, and 7 others. 2023. [Mindstorms in natural language-based societies of mind](#). *ArXiv preprint*, abs/2305.17066.
- Daniel M. Ziegler, Nisan Stiennon, Jeff Wu, Tom B. Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. 2019. [Fine-tuning language models from human preferences](#). *ArXiv preprint*, abs/1909.08593.

Appendix

A Experimental Setting

Models. We utilize GPT family models for both quantitative and qualitative experiments. For idea generation, we employ *GPT-4o mini*, *GPT4-32k*, and *GPT-4* within the MAMV framework, while *GPT-4o* serves as the core model for dialectical self-reflection. For reasoning, we additionally evaluate on *Qwen 2.5-7B* and *Phi-4*.

Dialectical Iterations and MAMV. We set a maximum iteration count N per experiment. For idea generation, an external MAMV module evaluates the validity and novelty of each unified idea; the process halts if either criterion fails. For reasoning tasks, we do not use MAMV, isolating the effect of dialectical self-reflection from multi-agent voting.

Sublation Temperature τ_A . During the sublation step, τ_A directly influences novelty: lower τ_A produces arguments closely engaging with the proposition, while higher τ_A generates more divergent perspectives. We set $\tau_A = 0.5$ as a compromise.

Speculation Temperature. We conduct two sets of experiments: one with the annealing schedule $\tau(i) = \tau_0 e^{-\theta i}$ and one with constant temperature ($\theta = 0$). We grid-searched θ and chose a high τ_0 to ensure initial creativity. Figures 3a and 3b present the optimal θ and constant τ respectively.

Hyper-parameters. Tables 3 and 4 detail our experimental hyper-parameters.

Hyper-parameter	Value
Initial temperature τ_0	0.7
Constant temperature for opposition τ_A	0.5
Decay constant θ	[0.1, 0.5]
Max iterations (idea generation) N	5
Max iterations (math reasoning) N	3

Table 3: Experimental hyper-parameters.

Dataset	Model	Constant		Annealing	
		τ_0	τ_A	τ_0	τ_A
GSM-Hard	GPT-4o	0.7	0.5	0.7	0.5
	GPT-4o-mini	0.7	0.5	0.7	0.5
	GPT4-32k	1.0	0.3	1.0	0.3
	Phi-4	1.0	0.5	1.0	0.5
	Qwen2.5	1.0	0.3	1.0	0.3
	Phi-4-mini-instruct	1.0	0.5	1.0	0.5
GSM-Symbolic	GPT-4o	0.7	0.5	0.7	0.5
	GPT-4o-mini	0.7	0.5	0.7	0.5
	GPT4-32k	0.7	0.5	0.7	0.5
	Phi-4	1.0	0.5	1.0	0.5
	Qwen2.5	1.0	0.3	1.0	0.3
	Phi-4-mini-instruct	1.0	0.5	1.0	0.5
GSM8k	GPT-4o	0.7	0.5	0.7	0.5
	GPT-4o-mini	0.7	0.5	0.7	0.5
	GPT4-32k	1.0	0.3	1.0	1.0
	Phi-4	0.7	0.5	0.7	0.5
	Qwen2.5	1.0	0.3	0.7	1.0
	Phi-4-mini-instruct	1.0	0.5	1.0	0.5

Table 4: Comparison of τ_0 and τ_A values across datasets and models.

B Prompt Templates

The full prompt templates referenced in Section 5 are shown below (Figures 4–9).

Sublation Prompt (p_A) – Qualitative Experiment

Instructions

You are following Hegel's dialectical method. You have access to a Thesis.

Your task:

1. Read the Thesis below (and any prior context if needed).
2. Imagine you are someone X, who has an orthogonal OR diametrically opposed perspective on the same topic.
3. Produce an Antithesis that is self-consistent, follows an argumentative structure, and contradicts the Thesis.
4. If the Thesis already qualifies all anticipated counterarguments, the Antithesis should be a new competing thesis with the same goal as the current Thesis.
5. Include novel, opinionated perspectives or angles beyond what has been stated so far.
6. Provide a concise "Summary of Antithesis" at the end.

Response Format

State who you are and your perspective:

X: [Name of the person with the Antithesis perspective]

Premises:

- Premise 1
- Premise 2
- (... as many premises as needed)

Reasoning: Explain how these premises contradict or challenge the Thesis, adding original, opinionated perspectives.

Conclusion: The final claim that directly opposes or contradicts the Thesis.

Antithesis: Detailed explanation.

Summary of Antithesis: One- or two-sentence summary capturing the essence of this Antithesis.

Figure 4: Prompt used during the sublation process in the idea generation experiment.

Speculation Prompt (p_S) – Qualitative Experiment

Instructions

You are following Hegel's dialectical method. You have access to the summary of the previous conversations, including the Thesis and Antithesis.

Your task:

1. Read the Thesis and Antithesis below.
2. Produce a Synthesis that either:
 - Qualifies the contradictory statements so that Thesis and Antithesis unify without contradiction, OR
 - Negates contradictory statements, explicitly listing which ones are negated and explaining why.
3. Present a coherent, integrated perspective that resolves or clarifies these contradictions.
4. Provide a short "Summary of Synthesis" statement, which will be used as the next Thesis.

Response Format

Thesis UNION Antithesis:

Premises:

- Premise 1
- Premise 2
- (... as many premises as needed)

Reasoning:

Conclusion:

Synthesis: Detailed explanation of how contradictions are integrated or negated. If qualifying contradictions, detail how they are harmonized. If negating specific statements, list them and briefly explain why. If both are lacking an important premise, introduce that confounding or missing premise. Explain how this new perspective leads to a coherent, possibly novel position.

Summary of Synthesis (Next Thesis): One- or two-sentence statement that unifies the positions and serves as the next Thesis.

Figure 5: Prompt used during the speculation process in the idea generation experiment.

Sublation Prompt (p_A) – Quantitative Experiment (Reasoning)

Instructions

You are following Hegel’s dialectical method.

You have access to the question: {question} and the following proposed solution: {solution 1}.

Your task:

1. Read the proposed solution carefully. If you notice any problem or defect, proceed to task 2; otherwise return the proposed solution.
2. Imagine you are someone X who has noticed a problem with or defect in the proposed solution. If and only if the proposed solution is not correct, produce an oposed solution that corrects the noted defect. Possible defects include, but are not limited to:
 - **Arithmetic errors:** incorrect calculation, order of operations, decimal mismanagement, fraction handling.
 - **Misinterpretation:** failure to extract key information, incorrect units/conversions, incorrect assumptions.
 - **Logical errors:** flawed reasoning steps, skipping key intermediate steps.
 - **Over-complication:** introducing unnecessary complexity.
 - **Formatting errors:** incomplete solution, inconsistent format, ambiguity.
 - **Conceptual misunderstanding:** misunderstanding the mathematical concept, confusing problem types.

For example: {Few-shot examples}

Let’s think step by step.

Response Format

Defects or problems:

1. Defect/Problem 1
2. Defect/Problem 2
3. (... as many defects as needed)

Q: {question}

A:

Ensure the answer A concludes with “The answer is...”

Figure 6: Prompt used during the sublation process in the GSM8k evaluation.

Speculation Prompt (p_S) – Quantitative Experiment (Reasoning)

Instructions

You are following Hegel’s dialectical method.

You have access to two opposing solutions for a given math question.

Your task:

- Read the given opposing solutions below.
- Produce a third solution for the same problem such that:
 - It unifies the previous two solutions in relation to the defects or problems that made them opposed.
 - This third solution must capture how the two solutions agree and resolve the reported defects.

Response Format

Agreements between solution 1 and solution 2:

1. Point 1 ...

Defects with solution 1:

1. Defect/Problem 1 ...

Defects with solution 2:

1. Defect/Problem 1 ...

Solution 1: ‘{solution 1}’

Solution 2: ‘{solution 2}’

Q: {question}

A:

Ensure the answer is the same as the unified solution but concludes with “The answer is...”

Let’s think step by step.

Figure 7: Prompt used during the speculation process in the GSM8k evaluation.

Validity Prompt (p_V) – MAMV

Instructions

You are following Hegel's dialectical method. You have access to the Thesis and Antithesis.

Your task:

1. Read the Thesis and Antithesis below.
2. Check whether Synthesis either:
 - Qualifies the contradictory statements so that Thesis and Antithesis unify without contradiction, OR
 - Negates contradictory statements, explicitly listing which ones are negated and explaining why.
3. Check whether Synthesis presents a coherent, integrated perspective that resolves or clarifies these contradictions.

Keeping all this in consideration, please assess whether the synthesis provided, derived from thesis and antithesis texts, adheres to the synthesis process. Reply with `yes` if it does and `no` if it does not.

Input / Output

Source Text 1: {Thesis}
Source Text 2: {Antithesis}
Synthesized Text: {Synthesis}

Response (yes/no):

Figure 8: Validity assessment prompt in the MAMV paradigm.

Novelty Prompt (p_N) – MAMV

Instructions

Please assess whether the source text is novel in comparison to the target text. By novelty, we mean that the source contains new information that is not identical to the target. Reply with `yes` if it is novel and `no` if it is not.

Input / Output

Source Text: {Synthesis}
Target Text: {Thesis}

Response (yes/no):

Figure 9: Prompt utilized for the novelty assessment in the MAMV paradigm.

C Instances of Dialectical Iteration

In this section, we present the complete responses of the core self-reflective model (GPT-4o in our experiments), captured during three key moments: understanding, sublation, and speculation. These responses illustrate one full dialectical iteration during both scientific idea generation and mathematical reasoning tasks.

Dialectical Iteration – Idea Generation

Understanding (ℱ)

Proposition T_0 : Energy is a continuous quantity.

Sublation (ℑ)

Premises:

1. Quantum mechanics shows that energy is not always continuous, but can appear in quantized packets.
2. Atomic and subatomic processes exhibit discrete energy levels.

Reasoning: These premises contradict the Thesis that energy is simply continuous, because the Thesis does not account for microscopic phenomena in which energy exchange occurs discretely.

Conclusion: Energy is quantized in certain contexts, such as atomic and subatomic processes, where it exists in discrete packets rather than as a continuous entity.

Summary of Antithesis: Energy cannot be understood as purely continuous; quantum phenomena require treating it as quantized in specific regimes.

Speculation (⊗)

Premises:

1. Energy can be described as continuous in classical physics, where it flows without discrete interruptions.
2. In quantum mechanics, energy is quantized, existing in discrete packets such as photons or electron energy levels.
3. Both descriptions depend on the scale and context of observation.

Reasoning: The continuous description is a macroscopic approximation, while quantum mechanics reveals discreteness at microscopic scales. The two views are therefore not simply contradictory, but apply at different levels of description.

Conclusion: Energy exhibits both continuous and quantized characteristics depending on context and scale.

Summary of Synthesis (Next Thesis): Energy’s duality as both continuous and quantized explains how classical and quantum mechanics complement each other across different scales.

Figure 10: An example of a dialectical iteration generated by GPT-4o for the idea generation experiment.

Dialectical Iteration – Mathematical Reasoning (GSM8k)

Understanding (ℱ) – Proposition

Q: John plans to sell all his toys and use the money to buy video games. He has 13 lego sets and he sells them for \$15 each. He ends up buying 8 video games for \$20 each and has \$5 left. How many lego sets does he still have?

A: He sold 13 sets for \$15 each: $13 \times \$15 = \195 . He bought 8 games for \$20 each: $8 \times \$20 = \160 . He has \$5 left, so total before buying: $\$160 + \$5 = \$165$. Since he made \$195 and spent \$165, he has no lego sets left. **The answer is 0 lego sets.**

Sublation (ℑ)

Defects:

1. *Misinterpretation:* The solution assumes John sold all his lego sets, but the problem says he plans to sell all his toys and asks how many lego sets remain.
2. *Logical error:* The solution calculates revenue from all 13 sets instead of determining how many sets were actually sold from the amount spent and left over.

Let x be the number of lego sets sold. He spent $8 \times 20 = 160$ dollars and had \$5 left, so he earned $160 + 5 = 165$ dollars from selling lego sets. Since each set sells for \$15, we have $15x = 165$, so $x = 11$. He originally had 13 sets, so $13 - 11 = 2$ remain. **The answer is 2 lego sets.**

Speculation (⊗)

The first solution correctly computes the total value of all 13 lego sets, but incorrectly assumes that all of them were sold. The corrected solution resolves this by using the actual amount of money John needed for the games and the leftover amount. He needed \$160 for the games and had \$5 left, so he sold enough lego sets to make \$165. At \$15 per set, this means he sold 11 sets. Since he started with 13, he still has 2 sets.

The answer is 2 lego sets.

Figure 11: An example of a dialectical iteration generated by GPT-4o for a math problem from GSM8k.

D Qualitative Results: Scientific Idea Generation

This section provides examples from our qualitative experiment, i.e., the generation of new ideas across different domains. Each sequence traces one full dialectical trajectory: from an initial idea through successive sublation–speculation cycles, terminating when the MAMV criterion is met.

Table 4: Topic 1 – Mathematics (Annealing $\tau, \theta = 0.5$)

Iteration 0 – Understanding ($\aleph$)

Proposition T_0 : In mathematics, we can introduce multiple forms of infinity, some of which may be considered larger or smaller than others based on certain metrics, for example time, to allow for a sense of relativity when working with infinity.

Iteration 1 $\tau = 0.70$

Sublation ($\aleph$): Defects: (1) Does not fully define “forms of infinity” or how they can be differentiated. (2) Does not account for the mathematical rigor required to compare infinities using metrics like time. (3) Does not explain how relativity applies to infinity coherently.

Opposition: Infinity should be treated as a single abstract concept, compared through set theory and cardinality rather than physical metrics.

Speculation ($\aleph$): Infinity can be understood in multiple forms through set theory and cardinality, and the concept of relativity can be meaningfully applied within this framework to compare different sizes of infinity.

Iteration 2 $\tau = 0.42$

Sublation ($\aleph$): Defects: (1) Does not define how relativity can be rigorously applied within purely mathematical contexts. (2) Fails to consider frameworks beyond set theory and cardinality. (3) Use of time as a metric is unconventional and not sufficiently justified.

Opposition: Infinity should be understood through multiple mathematical frameworks; relativity must be clearly defined and justified.

Speculation ($\aleph$): Infinity can be understood through multiple frameworks (set theory, cardinality, topology, analysis), and the concept of relativity can be meaningfully applied to compare different sizes of infinity.

Iteration 3 $\tau = 0.25$

Sublation ($\aleph$): Defects: (1) Does not define how relativity applies rigorously to compare infinities. (2) Fails to account for inherent differences between physical metrics and abstract mathematical concepts. (3) Overlooks specific examples from topology and analysis.

Opposition: Precise definitions and rigorous application are essential; abstract concepts must be distinguished from physical metrics.

Speculation ($\aleph$): Infinity can be understood through multiple frameworks; precise definitions and rigorous application of principles are essential, and distinguishing abstract concepts from physical metrics is crucial.

Iteration 4 $\tau = 0.15$

Sublation ($\aleph$): Defects: (1) Does not fully define how relativity applies within mathematical frameworks. (2) Does not account for potential incompatibility of physical metrics for measuring infinity. (3) Lacks a clear explanation of how topology and analysis integrate.

Opposition: Physical metrics like time should be excluded from measuring infinity to maintain mathematical rigor.

Speculation ($\aleph$): Infinity can be understood through multiple frameworks; excluding physical metrics like time from the measurement of infinity is necessary to maintain mathematical rigor.

Iteration 5 $\tau = 0.09$

Sublation ($\aleph$): Defects: (1) Does not fully define how relativity can be applied. (2) Excludes physical metrics without considering interdisciplinary insights. (3) Lacks specific examples from topology and analysis. (4) Does not address how to maintain rigor when integrating diverse frameworks.

Opposition: A comprehensive approach should include interdisciplinary insights, specific examples, and ensure rigor when integrating diverse frameworks.

Speculation ($\aleph$) – Final Idea:

Infinity can be understood through multiple frameworks; interdisciplinary approaches, including insights from physical metrics, can enhance mathematical comprehension. Maintaining rigor when integrating diverse frameworks is crucial to avoid inconsistencies and ensure clarity.

Process ended after 5 iterations

Validity: yes Novelty: yes

Table 5: Topic 2 – Physics (Annealing $\tau, \theta = 0.3$)

Iteration 0 – Understanding (U)

Proposition T_0 : Energy is a continuous entity.

Iteration 1 $\tau = 0.70$

Sublation (S): Defects: (1) Does not account for the quantized nature of energy in quantum mechanics. (2) Incompatible with discrete energy levels of electrons in atoms.

Opposition: Energy is quantized in certain contexts, existing in discrete packets rather than as a continuous entity.

Speculation (S): Energy is a dual entity that can be continuous in classical contexts and quantized in quantum contexts, respecting both macroscopic and microscopic observations.

Iteration 2 $\tau = 0.51$

Sublation (S): Defects: (1) Does not fully define the transition between continuous and quantized descriptions. (2) Does not account for overlap phenomena. (3) May oversimplify complex interactions.

Opposition: A more detailed examination of the transition between these descriptions is necessary for comprehensive understanding.

Speculation (S): Energy is a dual entity, but its behavior at intermediate scales and in complex interactions requires nuanced understanding beyond simple categorization.

Iteration 3 $\tau = 0.38$

Sublation (S): Defects: (1) “Intermediate scales” is ambiguous. (2) Does not account for quantum field theory advances. (3) Does not provide specific failure examples.

Opposition: Energy should be understood through comprehensive frameworks like QFT, addressing transitions across all scales.

Speculation (S): Energy is dual; modern theoretical frameworks like QFT can bridge descriptions across different scales. Accurate modeling requires clear definitions and specific examples.

Iteration 4 $\tau = 0.28$

Sublation (S): Defects: (1) Lacks clear definition of intermediate scales. (2) No specific examples of simplification failures. (3) Assumes QFT can seamlessly bridge all scales without addressing limitations.

Opposition: Clear definitions, specific examples, and examination of QFT limitations at mesoscopic scales are required.

Speculation (S): Energy is dual; QFT can bridge descriptions; accurate modeling requires clear definitions, specific examples, and addressing QFT’s limitations at mesoscopic scales.

Iteration 5 $\tau = 0.21$

Sublation (S): Defects: (1) Fails to provide practical definition of intermediate scales. (2) Does not address how QFT applies to specific mesoscopic phenomena. (3) Overly optimistic about QFT’s bridging capability.

Opposition: Precise definitions and critical evaluation of QFT’s limitations are essential for comprehensive understanding.

Speculation (S) – Final Idea:

Energy is dual (continuous classically, quantized quantum-mechanically). QFT can bridge these descriptions, but a critical evaluation of its limitations at mesoscopic scales is necessary for comprehensive understanding.

Process ended after 5 iterations

Validity: yes Novelty: yes

Table 6: Topic 3 – Economics (Annealing $\tau, \theta = 0.3$)

Iteration 0 – Understanding (11)

Proposition T_0 : According to Keynesian economics, during a recession the government should increase spending and cut taxes to stimulate demand. In times of economic boom, it should reduce spending and increase taxes to cool down the economy.

Iteration 1 $\tau = 0.70$

Sublation (21): Defects: (1) Does not account for long-term impact on national debt and inflation. (2) Assumes government intervention is always effective. (3) Does not address potential inefficiencies in government spending.

Opposition: Government intervention should be carefully balanced and consider long-term impacts, limitations, and potential inefficiencies.

Speculation (5): Government intervention can be valuable for stabilizing economic cycles, but it must be balanced with considerations for long-term impacts, efficiency, and external factors.

Iteration 2 $\tau = 0.51$

Sublation (21): Defects: (1) Does not define “efficient” government spending. (2) Does not account for crowding out private investment. (3) Assumes effective balancing is always achievable despite political challenges.

Opposition: Government intervention should be minimized due to unclear efficiency definitions, crowding-out risks, and political challenges.

Speculation (5): Government intervention must be balanced with considerations for efficiency, crowding-out effects, and political challenges to ensure sustainable stability.

Iteration 3 $\tau = 0.38$

Sublation (21): Defects: (1) Still does not define spending efficiency. (2) Does not consider positive impacts of public sector investment. (3) Does not account for effective governance possibilities. (4) Ignores external global factors.

Opposition: Government intervention can foster public sector innovation and social welfare; it requires clear definitions, successful governance recognition, and consideration of global factors.

Speculation (5): Government intervention can stabilize cycles and foster innovation, balanced with efficiency, crowding-out effects, effective governance, and global factors.

Iteration 4 $\tau = 0.28$

Sublation (21): Defects: (1) No actionable definition of efficiency. (2) Claims about public sector innovation lack evidence. (3) Underestimates complexities of political systems. (4) Global factors are mentioned but not integrated.

Opposition: Government intervention should be critically assessed and limited due to these challenges.

Speculation (5) – Final Idea:

Government intervention can stabilize economic cycles and foster innovation, but must be balanced with long-term impacts, efficiency, crowding-out effects, effective governance, and global factors, supported by empirical evidence and a detailed framework.

Process ended after 4 iterations

Validity: **yes** Novelty: **no**

Table 7: Topic 4 – Philosophy (Annealing $\tau, \theta = 0.3$)

Iteration 0 – Understanding (11)

Proposition T_0 : Existentialism emphasizes authenticity as a central value, positing that individuals must confront their freedom and take responsibility for crafting a meaningful life amidst an absurd and indifferent universe. Authenticity is achieved through self-awareness, rejection of external impositions, and commitment to one's values.

Iteration 1 $\tau = 0.70$

Sublation (21): Defects: (1) Does not fully define “authenticity” or how it can be measured. (2) Does not account for social and relational aspects of human existence. (3) Assumes individuals can fully reject external impositions, which may not be feasible.
Opposition: Authenticity should be understood as a dynamic process influenced by social relationships and societal contexts.

Speculation (5): Authenticity involves a dynamic interplay between personal values and societal influences, emphasizing ongoing negotiation rather than complete rejection of external impositions.

Iteration 2 $\tau = 0.51$

Sublation (21): Defects: (1) Does not provide a clear, actionable definition. (2) Does not address how external factors can dominate personal authenticity. (3) Fails to provide a framework for balancing personal values and societal pressures.
Opposition: Authenticity should be clearly defined, and existentialism must offer a robust framework for navigating societal influence.

Speculation (5): Authenticity is an evolving process shaped by internal self-awareness and external contexts, requiring balance supported by clear definitions and practical strategies.

Iteration 3 $\tau = 0.38$

Sublation (21): Defects: (1) Clear definition may contradict the existentialist idea that authenticity is inherently subjective. (2) Overlooks that societal influences can *enhance* authenticity. (3) Practical frameworks could become rigid and prescriptive.
Opposition: Authenticity is deeply personal and should not be rigidly defined; societal influences can support rather than undermine it; the pursuit should remain fluid.

Speculation (5): Authenticity is a dynamic interplay between personal values and societal influences, emphasizing fluid, individualistic negotiation. It should not be rigidly defined; societal influences can both support and undermine one's sense of authenticity.

Iteration 4 $\tau = 0.28$

Sublation (21): Defects: (1) Does not adequately define authenticity, leading to ambiguity. (2) Lacks consideration of negative societal impacts (conformity). (3) Does not address potential conflict between personal values and societal influences.
Opposition: Authenticity requires a clearer definition and thorough consideration of societal pressures, including conflicts with personal values.

Speculation (5): Authenticity involves a dynamic interplay between personal values and societal influences, requiring both clarity in definition and flexibility in application, recognizing the unique journey of each individual.

Iteration 5 $\tau = 0.21$

Sublation (21): Defects: (1) Does not account for societal norms becoming internalized and indistinguishable from personal values. (2) Fails to provide criteria for achieving definitional clarity. (3) Does not address subconscious influences. (4) Assumes societal influences are external and separate.
Opposition: Authenticity is influenced by deeply internalized norms and subconscious factors, requiring a nuanced approach without assuming clear separation between internal and external influences.

Speculation (5) – Final Idea:
Authenticity involves a dynamic interplay between personal values, societal influences, and subconscious psychological factors, requiring both clarity in definition and flexibility in application. It remains fluid and individualistic, recognizing the complexities of maintaining authenticity in a social context.

Process ended after 5 iterations

Validity: yes Novelty: yes