

# 000 REVERSE-ENGINEERED REASONING FOR OPEN- 001 002 ENDED GENERATION 003 004

005 **Anonymous authors**

006 Paper under double-blind review

## 007 008 ABSTRACT 009

011 While the “deep reasoning” paradigm has spurred significant advances in verifi-  
012 able domains like mathematics, its application to open-ended, creative generation  
013 remains a critical challenge. The two dominant methods for instilling reasoning—reinforcement learning (RL) and instruction distillation – falter in this area;  
014 RL struggles with the absence of clear reward signals and high-quality reward mod-  
015 els, while distillation is prohibitively expensive and capped by the teacher model’s  
016 capabilities. To overcome these limitations, we introduce REverse-Engineered Rea-  
017 soning (REER), a new paradigm that fundamentally shifts the approach. Instead of  
018 building a reasoning process “forwards” through trial-and-error or imitation, REER  
019 works “backwards” from known good solutions to computationally discover the  
020 latent, step-by-step deep reasoning process that could have produced them. Using  
021 this scalable, gradient-free approach, we curate and open-source DeepWriting-  
022 20K, a large-scale dataset of 20,000 deep reasoning trajectories for open-ended  
023 tasks. Our model, DeepWriter-8B, trained on this data, not only surpasses strong  
024 open-source baselines but also achieves performance competitive with, and at times  
025 superior to, leading proprietary models like GPT-4o and Claude 3.5.

## 027 1 INTRODUCTION

030 The paradigm of “deep reasoning” is catalyzing a shift in large language model (LLM) reasoning,  
031 moving beyond rapid, surface-level inference to leverage increased computational investment at test  
032 time (Guo et al., 2025; Jaech et al., 2024; Team, 2025; Muennighoff et al., 2025; Fu et al., 2025). This  
033 approach unlocks sophisticated capabilities like multi-step planning and complex problem-solving,  
034 yielding remarkable performance gains in verifiable domains such as mathematics and programming.  
035 The success in these areas has been largely propelled by Reinforcement Learning (RL), where clear  
036 reward signals for correct outcomes effectively guide a model’s search through vast solution spaces.

037 However, the reliance on verifiability presents a formidable barrier when applying deep reasoning  
038 to open-ended, creative domains (Lu, 2025; Ouyang et al., 2022). Creative writing, a quintessential  
039 example of a non-verifiable task, lacks a singular, objective ground truth. Instead, its quality is judged  
040 on subjective criteria like originality, emotional resonance, and narrative coherence (Wu et al., 2025).  
041 This disconnect raises a critical and largely unexplored research question:

### 042 **How to instill deep reasoning for open-ended generation in the absence of task verifiability?**

044 Bridging this gap is profoundly challenging. The dominant paradigms for cultivating advanced  
045 reasoning falter here; adapting RL by training a reward model to approximate subjective quality  
046 that aligns with human preferences is an immense challenge in itself (Ouyang et al., 2022), and the  
047 subsequent RL process is notoriously sample-inefficient and computationally burdensome (Lu, 2025).  
048 The alternative, instruction distillation from a powerful model, is often prohibitively expensive and  
049 fundamentally capped by the teacher model’s capabilities (Toshniwal et al., 2024). This is exacerbated  
050 by the scarcity of high-quality queries and deep reasoning trajectories tailored for complex open-  
051 ended generation (Bai et al., 2024). These constraints create a critical bottleneck, demanding a new  
052 paradigm that sidesteps both the sample inefficiency of RL and the costly dependency of distillation.

053 To break this impasse, we introduce a new paradigm: **REverse-Engineered Reasoning (REER)**. In  
054 contrast to conventional methods that build a reasoning process “forwards” through trial-and-error

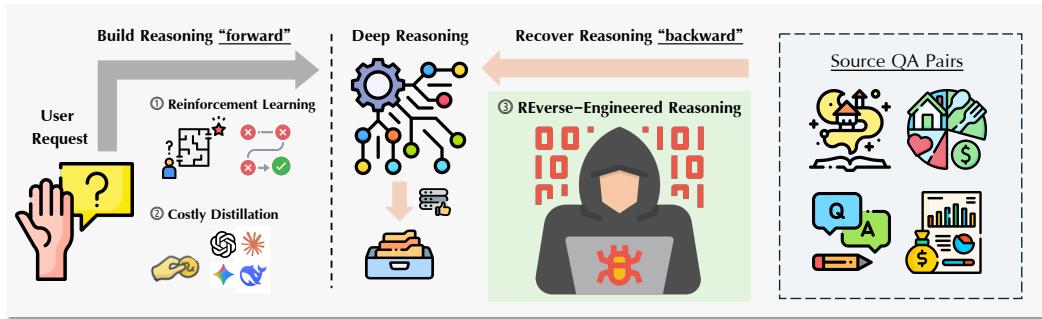


Figure 1: **(Left)** Existing methods attempt to build deep reasoning “forwards” for a user request through trial-and-error (RL) or costly distillation, which falter in open-ended domains that lack clear, verifiable reward signals. **(Right)** We propose a third path for teaching deep reasoning, REverse-Engineered Reasoning (REER). REER works “backwards”, recovering plausible human-like thought process from known-good outputs in open-source Question-Answer (QA) pairs.

or distillation, *we work “backwards” from a known good outcome*. We essentially ask: “Given a high-quality piece of output, what is the most coherent and logical thinking process that would have generated it?” By answering this question, we can synthesize the otherwise latent, human-like reasoning paths at scale, bypassing costly distillation of thinking data beforehand or inefficient trial-and-error.

We pioneer a novel approach that operationalizes the REER paradigm and, for the first time, *instill deep reasoning capabilities for open-ended generation entirely from scratch*. Our approach involves three key stages. First, we source a diverse dataset of query-solution pairs for open-ended generation from the web, encompassing 16,000 samples spanning across ordinary-life question-answering, academic writing, functional writing and creative writing. From these, we “reverse-engineer” deep reasoning trajectories – structured, human-like thought process tailored for open-ended generation. Eventually, we use this synthetic data to fine-tune a base language model, teaching it to reason and plan deeply before generating a final solution.

The central innovation lies in how we synthesize this data: *we formulate the recovery of high-quality thinking trajectories as a gradient-free search problem*. These trajectories are found by iteratively refining an initial plan, with the search guided by a proxy for thought quality – the perplexity of a known good solution. The gradient-free, self-contained nature of our synthesis process lends us the scalability. By obviating the need for expensive, query-by-query distillation from proprietary models or the sample-inefficiency of reinforcement learning, our approach provides a cost-effective and automatable pathway to generate vast quantities of high-quality, deep-thinking training data. This makes it possible to instill sophisticated reasoning in models at a scale that was previously impractical.

Using this method, we created **DeepWriting-20K**, a comprehensive dataset of 20,000 thinking trajectories, and fine-tuned a Qwen3-8B base model. Our extensive empirical evaluation on benchmarks like LongBench (Bai et al., 2024), HelloBench (Que et al., 2024), and WritingBench (Wu et al., 2025) shows that DeepWriter-8B successfully internalizes this deep reasoning process. It not only substantially outperforms strong open-source baselines but also achieves performance competitive with, and in some cases exceeding, leading proprietary models like GPT-4o and Claude 3.5, validating our approach as a powerful new pathway for building complex reasoning for open-ended generation.

Our primary contributions are:

- **Pioneering a New Paradigm for Reasoning:** We introduce Reverse-Engineered Reasoning (REER), a novel and scalable paradigm that synthesizes high-quality thinking trajectories without reliance on RL or distillation. This work charts a new, third path for teaching complex reasoning, addressing a critical bottleneck in the field.
- **Enabling New Research with DeepWriting-20K:** We contribute a comprehensive dataset of 20,000 query-response and deep reasoning trajectories across 25 categories. This resource is designed to mitigate data scarcity and catalyze future research into the mechanisms of planning and structured thought in open-ended generation.

108  
 109  
 110  
 111  
 112  
 113  
 114 • **Strong Performance in Open-Ended Generation from Scratch:** We provide rigorous empirical evidence that our approach works. Our model, DeepWriter-8B, trained entirely on our synthesized data, matches or exceeds the performance of premier proprietary models on challenging writing benchmarks, proving that human-like deep reasoning can be cultivated without costly distillation or RL.

114 **2 APPROACH**  
 115

116 **[Question]**

117 Write an engaging story based on the following writing prompt:  
 118 “You can’t speedrun an Isekai! It’s not a game!” The God tells me.  
 119 “Watch me.” I reply.

120 **[Response]**

121 <think>

122 Alright, the user wants a story based on the prompt ...

123 **First, I need to understand the user’s intent.** The core objective is to create a compelling ...

124 Let me think... Maybe I should emphasize the protagonist’s journey from disbelief to ...

125 Hmm... Alternatively, maybe I should also consider the emotional journey of the protagonist and ...

126 **For the content,** I need to brainstorm a core creative idea.

127 The main character, Hiro ...

128 **Building the structure,** I’ll use a narrative framework with a clear beginning, development, climax.

129 Let me think about the beginning ... Next, the development phase ... For the climax, I’ll need a final, epic battle ... Additionally, I should include Hiro’s reflections on his journey ...

130 Wait, that’s a bit too straightforward.

131 **Let me add some twists and turns.**

132 For instance, the God could ... Hmm, maybe I should include a subplot involving ... This could add layers to the story. ...

133 </think>

134 <answer>

135 ...

136 </answer>

137  
 138 Figure 2: An example of the deep reasoning generated by DeepWriter. The model demonstrates  
 139 **deliberate planning** for open-ended generation, employing various thinking patterns (e.g., ‘Hmm...’  
 140 ‘Alternatively’, ‘Wait, that’s a bit ...’) to facilitate structured reasoning, including logical deduction,  
 141 branching, and backtracking.

142  
 143 Our central goal is to instill deep reasoning in LLMs for open-ended tasks without relying on costly  
 144 distillation or reinforcement learning. To achieve this, we introduce **REverse-Engineered Reasoning**  
 145 (**REER**), a novel paradigm that shifts the objective from generating a solution to discovering the  
 146 latent reasoning process behind an existing high-quality one. Instead of building a reasoning process  
 147 “forwards” via trial-and-error, REER works “backwards” from a known good output to computationally  
 148 synthesize the step-by-step thinking that could have produced it. This approach is operationalized  
 149 as a search problem where we iteratively refine an initial thinking process to discover a trajectory that  
 150 best explains a high-quality, human-written output. An example of the structured reasoning we aim  
 151 to cultivate is shown in **Figure 2**, where the model demonstrates deliberate planning, exploration of  
 152 alternatives (“Hmm... Alternatively”), and self-correction (“Wait, that’s a bit too straightforward”).

153  
 154 **2.1 REVERSE-ENGINEERED REASONING AS A SEARCH PROBLEM.**

155 Let  $x$  be an input query (e.g., a story prompt) and  $y$  be a high-quality reference solution (e.g., a well-  
 156 written story). Our objective is to find a *deep reasoning trajectory*, denoted by  $z$ , which represents a  
 157 structured, step-by-step thinking process that guides the generation of  $y$  from  $x$ .

158  
 159 The primary challenge in open-ended domains is the absence of a verifiable correctness signal. The  
 160 REER paradigm circumvents this by reframing the problem: instead of judging the final output,  
 161 we evaluate the quality of a *thinking process* based on how well it explains a known-good output.  
 We operationalize this principle by using the **perplexity** (a.k.a, the model surprise) of the reference

162 solution  $y$  as a proxy for the quality of a given reasoning trajectory  $z$ . A lower perplexity score for  $y$ ,  
 163 conditioned on both  $x$  and  $z$ , indicates that the trajectory provides a more coherent and effective plan.  
 164 In essence, REER posits that a good thinking process  $z$  is one that makes a high-quality answer  $y$   
 165 seem maximally probable and logical to the model.

166 Formally, we model the deep reasoning trajectory  $z$  as a discrete sequence of reasoning steps,  
 167  $z = [z_1, z_2, \dots, z_n]$ . The problem is then formulated as a search for the optimal trajectory  $z^*$  within  
 168 the vast space of possible trajectories  $\mathcal{Z}$ , such that  $z^*$  minimizes the perplexity of the reference  
 169 solution  $y$ :

$$z^* = \arg \min_{z \in \mathcal{Z}} \text{PPL}(y|x, z)$$

170 Here,  $\text{PPL}(y|x, z)$  is the perplexity of the token sequence of  $y$  as calculated by a generator LLM,  
 171 conditioned on  $x$  and  $z$ . This optimization is performed via a *gradient-free local search algorithm*,  
 172 allowing us to iteratively refine the trajectory without a differentiable objective.  
 173

## 175 2.2 ITERATIVE REFINEMENT VIA LOCAL SEARCH

176 Solving for the optimal trajectory  $z^*$  directly is intractable due to the vast search space. Therefore, we  
 177 propose an iterative refinement algorithm that employs a guided local search to discover a high-quality  
 178 deep reasoning trajectory. The algorithm starts with an initial trajectory and progressively improves it  
 179 by refining its constituent segments, guided by the perplexity signal. As visualized in **Figure 3**, the  
 180 algorithm runs as follows:  
 181

182 **1. Initialization:** For a given  $(x, y)$  pair, we  
 183 generate an initial, imperfect deep reasoning  
 184 trajectory,  $z^{(0)}$ , by prompting an LLM with a  
 185 thought-provoking instruction (see Appendix,  
 186 Listing 1) to produce a plausible plan. This initial  
 187 trajectory is denoted as  $z = [z_1, z_2, \dots, z_n]$ .

188 **2. Node Expansion (Segment-wise Edits):** The  
 189 core of our method is an iterative loop that re-  
 190 fines  $z$  one segment at a time. In each iteration,  
 191 we select a segment  $z_i$  to improve. We prompt  
 192 the LLM to generate candidate refinements with  
 193 more thinking-based details, elaborations and  
 194 reflections. To generate these refinements, we  
 195 provide the full context including the query  $x$ ,  
 196 the reference solution  $y$ , and the surrounding  
 197 trajectory segments (refined steps  $z_{<i}^*$  and initial  
 198 steps  $z_{>i}$ ). The prompt is meticulously designed  
 199 to encourage detailed reasoning while preventing the model from simply copying content from the  
 200 reference solution (see Appendix, Listing 2).

201 **3. Node Evaluation and Selection:** For each generated candidate  $c$ , we construct a temporary  
 202 trajectory  $z'_{\text{cand}}$  by substituting  $z_i$  with  $c$ . We then evaluate each candidate by computing its quality  
 203 score,  $S(c) = \text{PPL}(y|x, z'_{\text{cand}})$ . The candidate with the lowest perplexity score is chosen as the  
 204 updated segment for the next iteration:  $z_i^* = \arg \min_{c \in C_i \cup \{z_i\}} S(c)$ . We include the original segment  
 205  $z_i$  in the candidate set to ensure that the perplexity improves monotonically.

206 **4. Termination:** The refinement process repeats until the perplexity of the solution reaches a  
 207 predefined target threshold or a maximum number of iterations is completed. The final output is a  
 208 refined trajectory  $z^*$ .

209 This process allows us to create a dataset of  $(x, z^*, y)$  triples, which can then be used to fine-tune a  
 210 base LLM to internalize the deep reasoning capability for open-ended generation from scratch.

211 It is important to distinguish our iterative local search from methods like Monte Carlo Tree  
 212 Search (Browne et al., 2012; Li et al., 2025). First, by using the perplexity of a complete ref-  
 213 erence solution as a quality proxy, REER avoids the computationally expensive rollouts required in  
 214 MCTS. Second, our approach operates on a "global-to-local" principle: we start with a complete,  
 215 albeit imperfect, global plan and iteratively improve it through local, segment-wise edits. This con-  
 trasts with standard MCTS or beam search, which build solutions sequentially by extending partial

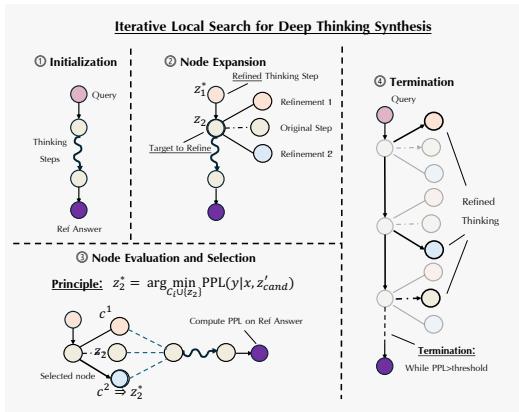


Figure 3: Method Overview: Iterative Local Search for deep thinking Synthesis.

216 states. These distinctions make our approach a scalable and efficient method for operationalizing  
 217 REER, enabling the creation of large-scale, deep-reasoning datasets for open-ended generation.  
 218

219 **2.3 TRAINING DATA CURATION**  
 220

221 The success of our methodology hinges on a large-scale, high-quality dataset of  $(x, z^*, y)$  triples.  
 222 The creation of this dataset follows a multi-stage pipeline: sourcing diverse query-solution pairs,  
 223 synthesizing deep reasoning trajectories, and applying rigorous filtering.  
 224

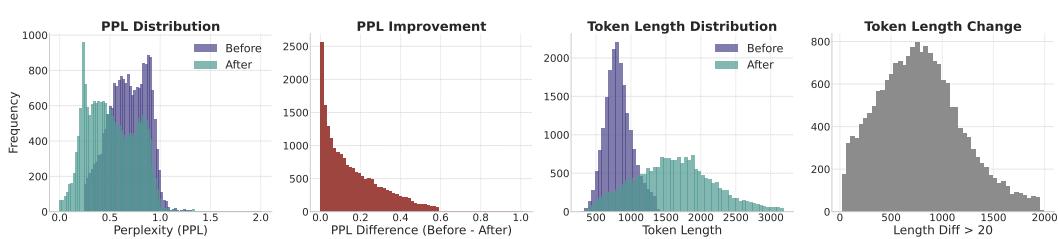
225 **2.3.1 Sourcing of (Query, Solution) Pairs.** To ensure diversity in style, topic, and complexity, we  
 226 sourced initial  $(x, y)$  pairs from three primary channels. We gathered prompt-story pairs from online  
 227 communities like r/WritingPrompts, using upvotes as a quality proxy, and reverse-engineered queries  
 228 ( $x$ ) from classic Project Gutenberg texts ( $y$ ) using GPT-4o. Finally, we augmented this collection with  
 229 data from instruction tuning datasets such as WildChat (Zhao et al., 2024) and LongWriter6K (Bai  
 230 et al., 2024).  
 231

231 **2.3.2 Trajectory Synthesis and Filtering.** From our sourced pairs, we selected 20,000 high-  
 232 quality query-solution pairs covering 25 manually nominated categories to ensure broad topic  
 233 coverage. For each pair, we executed our iterative local search algorithm to generate an optimal deep  
 234 reasoning trajectory  $z^*$ .  
 235

236 **Context Engineering.** The efficacy of the search algorithm, however, hinges not only on the search  
 237 procedure but also on the nuanced design of the instructions used to elicit deep reasoning from the  
 238 generator LLM. We proposed three key designs in our context engineering to ensure high-quality  
 239 synthesis. We only summarize the key insights here and refer the reader to the appendix for detailed  
 240 prompts.

241 1. **Enforcing Segment-wise Edits via a Meta-Structure.** To ensure the generator model performs  
 242 a true segment-wise edit without including edits for the subsequent parts of the trajectory, we  
 243 enforce a *meta-structure* for the reasoning process within the prompt. This serves as an *implicit*  
 244 *regularizer*, helping the model to localize the current segment and constrain its edits to the  
 245 intended scope when performing segment-wise edits.  
 246 2. **Injecting Human-like Thinking Patterns.** To prevent the synthesis of rigid and formulaic  
 247 reasoning, we *deliberately inject human-like thinking patterns*. Prompts explicitly encourage  
 248 phrases that signify cognitive exploration and self-reflections (e.g., “Hmm...maybe I can...”,  
 249 “Wait, that’s a bit...”), triggering a more human-like reasoning style and incentivizing self-  
 250 reflection through training (Wang et al., 2025b).

251 Analysis of this synthesis process, shown in **Figure 4**, confirms its effectiveness. The perplexity  
 252 distribution shifts significantly lower after refinement, with the vast majority of samples showing  
 253 a marked PPL improvement. Concurrently, the token length of the trajectories increases, to an  
 254 indicating that the search process successfully expands initial simple plans into more detailed and  
 255 elaborate reasoning chains.  
 256

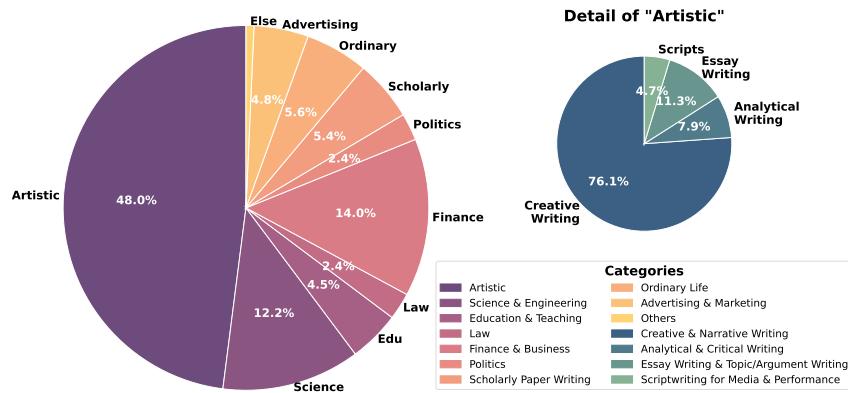


257 **Figure 4: Analysis of Token Length & Perplexity Before and After the Search.** The leftmost two  
 258 plots show that our iterative search process consistently **reduces perplexity (PPL)**. The rightmost  
 259 two plots show that the process also tends to **increase the token length** of the thinking trajectory,  
 260 reflecting the addition of more detailed reasoning steps.  
 261

262 During instruction tuning, we witness the challenge of repetitive and degenerate thinking. We  
 263 therefore applied two heuristic filtering strategies to prune low-quality trajectories:  
 264

270 1. **End-of-Thinking Filtering:** We discarded samples where thinking patterns persisted in the  
 271 final 10% of the sequence. These trajectories risk misleading the model to stuck in a repetitive  
 272 loop and failing to conclude its reasoning process.  
 273 2. **Repetition Filtering:** We employed a repetition metric to measure the frequency of the top-3  
 274 n-grams within each trajectory. Samples exhibiting high n-gram repetition, a sign of degenerative  
 275 looping expressions, were filtered out.

276 This process resulted in a final dataset of 20,000 high-quality deep reasoning trajectories. The  
 277 distribution of this dataset, shown in **Figure 5**, highlights its diversity, with a significant focus on  
 278 **Artistic** (Literature and Arts) writing, which is further broken down into sub-genres like Creative  
 279 Writing and Essay Writing.



295 Figure 5: Distribution of the final 20K training dataset by categories taking more than 0.5% account.  
 296 The primary chart shows a diverse range of topics, with a large emphasis on **Artistic** writing.  
 297 The detailed view of "Artistic" reveals a focus on Creative Writing and other styles, ensuring  
 298 comprehensive coverage for open-ended generation.

300 **2.3.3 Final Dataset Assembly for Fine-Tuning.** Training a model exclusively on domain-specific  
 301 data risks overfitting and can degrade its general knowledge priors. To mitigate this, we adopted  
 302 a **mixed-data training strategy**. We combined our 20K synthetically generated trajectories with  
 303 distilled deep reasoning trajectories from public datasets, i.e., OpenThoughts (Guha et al., 2025), that  
 304 primarily cover domains like mathematics, coding, and science. This blended datasets prevents the  
 305 model from catastrophic overfitting when learning deep reasoning for open-ended generation.

306 To train the base LLM, each complete triple in the final dataset was formatted using the prompt  
 307 template shown in Listing 4 in the appendix. This structure explicitly teaches the model to first  
 308 perform deep reasoning before producing the final output, thereby internalizing the desired reasoning  
 309 process from scratch.

### 3 EXPERIMENTS

313 Our empirical evaluation is structured to rigorously validate the efficacy of DeepWriter. We address  
 314 two central research questions:

315 1. How does DeepWriter, fine-tuned from scratch on an 8B open-source model, compare against  
 316 state-of-the-art proprietary models and other powerful open-source alternatives across a spectrum  
 317 of diverse open-ended generation tasks?  
 318 2. What is the individual contribution of each core component of our approach – specifically, the  
 319 synthesized deep thinking trajectories, the iterative refinement algorithm, and the characteristics  
 320 of the thinking traces and the data composition – to the model’s final performance?

322 To answer these questions, we first present a comprehensive comparison against leading models,  
 323 followed by a series of targeted ablation studies. We also provide a qualitative deep-dive analysis into  
 the model’s reasoning capabilities in the appendix.

324 **Training Data:** Our primary training dataset comprises approximately 20,000 deep thinking  
 325 trajectories, which we synthesized for 16,000 unique queries spanning a wide array of open-ended  
 326 tasks. As stated in Section 3, to prevent catastrophic forgetting of general reasoning abilities, we  
 327 blended this core dataset with public thinking-process datasets that reasoning-related domains (e.g.,  
 328 mathematics, coding). This resulted in a final mixed dataset of 37,000 examples, ensuring a balance  
 329 between specialized open-ended generation capabilities and keeping broad knowledge priors.

330 **Implementation Details:** We selected **Qwen3-8B-Base** as our base model for fine-tuning. This  
 331 decision was informed by preliminary experiments where other candidates, such as Llama-3.1-8B-  
 332 Base, struggled to effectively internalize the deep thinking process, and Qwen-2.5-7B-Base faced  
 333 prohibitive context length limitations. For the critical trajectory synthesis stage, we utilized Qwen2.5-  
 334 32B-Instruct as the generator model. Fine-tuning was conducted for 3 epochs using a constant  
 335 learning rate of  $2 \times 10^{-5}$  and a global batch size of 96. We set the max step to 10 and stopping PPL  
 336 to 0.25 in the iterative local search procedure.

### 337 3.1 EVALUATION BENCHMARKS

340 To ensure a comprehensive and multi-faceted evaluation, we employed a suite of three complementary  
 341 benchmarks: LongBench-Write (LB), HelloBench (HB), and WritingBench (WB). Together, they  
 342 probe three distinct and critical dimensions of generative performance: raw endurance, real-world  
 343 applicability, and domain-specific proficiency.

- 344 • **LongBench-Write (LB):** This benchmark functions as a targeted stress test for generative  
 345 endurance. It is designed to measure a model’s ability to produce coherent, ultra-long-form text  
 346 (e.g.,  $>10,000$  words), allowing us to assess the foundational capacity for maintaining thematic  
 347 consistency over extended outputs.
- 348 • **HelloBench (HB):** To gauge practical applicability, HelloBench evaluates performance on a  
 349 diverse set of “in-the-wild” tasks sourced from real user queries. Our analysis focuses on two key  
 350 subsets: **HB-A (Open-Ended QA)**, which tests the generation of detailed and nuanced answers,  
 351 and **HB-B (Heuristic Text Generation)**, which assesses creative reasoning and stylistic fidelity  
 352 in long-form narrative continuation.
- 353 • **WritingBench (WB):** This benchmark is tailored to measure domain-specific proficiency and  
 354 controllability across six professional and creative domains: **A** (Academic & Engineering),  
 355 **B** (Finance & Business), **C** (Politics & Law), **D** (Literature & Arts), **E** (Education), and **F**  
 356 (Advertising & Marketing). It specifically evaluates the ability to adhere to complex, multi-  
 357 dimensional constraints, a hallmark of advanced open-ended generation.

358 **Evaluation Protocols:** Given the subjective nature of open-ended tasks, we adopted the established  
 359 protocol of using powerful LLMs as judges within each benchmarks<sup>1</sup>. While we acknowledge the  
 360 potential for inherent biases in this method, it remains the most scalable and consistent approach  
 361 for evaluating nuanced generative quality. Specifically, **Claude-3.7** was used to score outputs for  
 362 LongBench and WritingBench, while **GPT-4o** was used for HelloBench. For HelloBench, we report  
 363 the original score without rescaling.

### 364 3.2 MAIN RESULTS

366 We benchmarked DeepWriter against leading proprietary models (GPT-4o, Claude 3.5, Claude  
 367 3.7) and strong open-source baseline, Qwen2.5-32B-Instruct, Qwen3-8B, LongWriter-8B. The  
 368 results, presented in Table 1, unequivocally demonstrate that our methodology successfully instills  
 369 sophisticated generation capabilities in an 8B model without costly distillation or trial-and-error.

370 **Analysis of Main Results.** The results in Table 1 reveal several compelling findings. First,  
 371 **DeepWriter-8B consistently and substantially outperforms the strong open-source baseline,**  
 372 **LongWriter-8B, across all benchmarks.** The performance gap is particularly stark in the diverse  
 373 WritingBench domains, where DeepWriter achieves an average uplift of over 18 points. This high-  
 374 lights the profound advantage of our deep thinking synthesis approach over standard instruction  
 375 tuning for cultivating advanced generative skills.

376 <sup>1</sup>Using the latest evaluation protocol of WritingBench, we note that there is currently discrepancy on  
 377 reproduced results and paper results, which is also acknowledged by the authors.

378  
 379 Table 1: Main performance comparison on LongBench (LB), HelloBench (HB), and WritingBench  
 380 (WB). DeepWriter demonstrates competitive performance against leading proprietary models and  
 381 significantly outperforms other open-source models.

382 <b>Model</b>	383 <b>LB</b>	384 <b>HB-A</b>	385 <b>HB-B</b>	386 <b>WB-A</b>	387 <b>WB-B</b>	388 <b>WB-C</b>	389 <b>WB-D</b>	390 <b>WB-E</b>	391 <b>WB-F</b>
GPT-4o	83.1	83.7	87.6	74.4	73.4	74.3	77.9	75.8	78.0
Claude 3.5	89.3	82.9	88.3	59.05	57.6	56.3	59.3	62.0	67.7
Claude 3.7	97.8	83.9	93.2	78.2	77.9	76.5	79.3	79.2	80.8
Qwen2.5-32B-Instruct	78.8	81.0	83.8	52.5	49.8	51.0	49.6	53.9	54.2
Qwen3-8B	85.2	81.4	85.3	68.7	68.9	67.0	67.2	71.2	71.3
LongWriter-8B	76.5	80.1	82.6	57.9	53.9	49.0	52.0	52.9	52.0
<b>DeepWriter-8B</b>	<b>91.3</b>	<b>82.6</b>	<b>87.4</b>	<b>72.2</b>	<b>71.8</b>	<b>69.8</b>	<b>70.6</b>	<b>73.7</b>	<b>72.3</b>

392 Table 2: Ablation studies. The full model (top row) is compared against versions with key components  
 393 removed. Results show that the synthesized deep thinking trajectories and iterative refinement are  
 394 crucial for performance.

394 <b>Model Configuration</b>	395 <b>LB</b>	396 <b>HB-A</b>	397 <b>HB-B</b>	398 <b>WB-A</b>	399 <b>WB-B</b>	400 <b>WB-C</b>	401 <b>WB-D</b>	402 <b>WB-E</b>	403 <b>WB-F</b>
<b>DeepWriter-8B (Full)</b>	<b>91.3</b>	<b>82.6</b>	<b>87.5</b>	<b>72.2</b>	<b>71.8</b>	<b>69.8</b>	<b>70.6</b>	<b>73.7</b>	<b>72.3</b>
- Remove Synthesis Data	82.9	70.9	73.7	63.4	62.7	62.8	57.7	66.3	62.7
- Remove Iterative Search	83.2	81.0	84.4	66.7	68.7	67.3	65.6	69.5	70.1
- Remove Reflection Tokens	86.9	82.2	82.8	71.6	69.6	70.4	62.0	69.9	71.9
- Downsample Long Traces	90.3	82.2	84.0	69.6	70.3	69.1	67.5	69.8	70.7
- Downsample Short Traces	89.3	81.1	82.1	70.8	70.6	70.0	66.9	72.4	69.7
- Remove Literature data	88.8	81.6	85.3	71.3	71.0	69.3	69.8	72.2	71.3

404 Second, **DeepWriter-8B closes a significant portion of the performance gap with elite proprietary**

405 **models**. On the creative HelloBench task (HB-B), its score (87.48) is statistically on par with GPT-4o

406 (87.6) and Claude 3.5 (88.3). More strikingly, on the professional writing tasks in WritingBench,

407 DeepWriter-8B not only surpasses Claude 3.5 by a large margin in all six categories but also remains

408 highly competitive with the much larger GPT-4o and Claude 3.7 models. A counter-intuitive result is

409 DeepWriter-8B’s score of 91.28 on LongBench-Write, exceeding both GPT-4o (83.1) and Claude 3.5

410 (89.3). This suggests that explicitly training on structured thinking trajectories provides a powerful

411 inductive bias for maintaining long-range coherence, a critical challenge in ultra-long text generation.

### 412 3.3 ABLATION STUDIES

414 To meticulously dissect the contribution of each component of our methodology, we conducted a

415 series of ablation studies, with results detailed in Table 2. Each experiment isolates a specific design

416 choice to quantify its impact on overall performance.

417 The ablation results provide robust evidence supporting our methodological design.

- 418 • **Importance of Synthesized Data:** Removing our 20K synthesized trajectories and training  
 419 only on public thinking datasets (“- Remove Synthesis Data”) causes the most significant  
 420 performance degradation across the board. Scores plummet, particularly in creative tasks like  
 421 HelloBench HB-B (87.48 → 73.73) and across WritingBench (average drop of over 8 points).  
 422 This confirms a core hypothesis: it is not merely the presence of “thinking” data that matters,  
 423 but the **quality and relevance of structured trajectories tailored for open-ended domains**  
 424 that drive performance.
- 425 • **Impact of Iterative Refinement:** Using the initial, unrefined thinking trajectories ( $z^{(0)}$ ) instead  
 426 of the final, optimized ones ( $z^*$ ) (“- Remove Iterative Search”) also leads to a clear drop in  
 427 performance. While the decline is less severe than removing the synthesis data entirely, the drop  
 428 on nuanced WritingBench tasks (e.g., WB-A: 72.20 → 66.72) is substantial. This proves that  
 429 our perplexity-guided local search is highly effective at discovering superior reasoning paths  
 430 that translate directly into stronger generative capabilities.
- 431 • **Effect of Reflection Tokens:** Removing reflection tokens (e.g., ‘Hmm...’, ‘Wait, that’s...’) from  
 432 the synthesis prompts (“- Remove Reflection Tokens”) has a nuanced effect. While overall scores

432 dip slightly, the most pronounced drop is in WritingBench domain D (Literature & Arts), which  
 433 falls from 70.57 to 62.04. This suggests that these explicit markers of cognitive exploration,  
 434 self-correction, and branching are particularly valuable for instilling the flexibility and creativity  
 435 required in artistic writing tasks.

- 436 • **Role of Trajectory Length:** We explored the impact of trace length by selectively downsampling  
 437 either long or short trajectories. The results reveal a task-dependent preference: removing longer,  
 438 more elaborate traces (“- Downsample Long Traces”) disproportionately harms performance  
 439 on complex, domain-specific tasks like those in WritingBench. Conversely, removing shorter,  
 440 more concise traces (“- Downsample Short Traces”) has a slightly larger negative impact on  
 441 creative tasks like HB-B. This suggests that detailed, multi-step plans are crucial for structured  
 442 professional writing, while nimbler, more direct reasoning may be optimal for creative ideation.
- 443 • **Role of Literature & Arts Data:** Removing the data from the “Literature & Arts” and “Or-  
 444 dinary Life” domains (“- without Literature & Arts data”) degrades performance across all  
 445 benchmarks, not just in the corresponding WB-D category. This finding indicates that training  
 446 on creative and narrative tasks imparts a more generalizable ability to handle nuance, structure,  
 447 and open-endedness, even benefiting performance in more technical domains. This highlights  
 448 the contribution of the release of our 20K dataset covering comprehensive topics.

## 449 4 RELATED WORK AND FUTURE DIRECTIONS

450 The paradigm of “deep reasoning” aims to move beyond rapid, surface-level inference by leveraging  
 451 increased computational investment at test time, a strategy shown to be effective by advanced  
 452 models (Team et al., 2023; Guo et al., 2025; Jaech et al., 2024; Team, 2025; Muennighoff et al.,  
 453 2025; Fu et al., 2025). This approach gained prominence with methods like Chain-of-Thought  
 454 (CoT) prompting (Wei et al., 2022), which elicits intermediate reasoning steps, and has evolved into  
 455 more sophisticated strategies like Tree-of-Thought (Yao et al., 2023) and self-refinement (Madaan  
 456 et al., 2023; Kumar et al., 2024; Zelikman et al., 2022; 2024). While these techniques excel in  
 457 verifiable domains like mathematics, their application to open-ended, creative tasks remains limited  
 458 by the absence of a singular ground truth for verification. Our work, REverse-Engineered Reasoning  
 459 (REER), directly addresses this gap.

460 Two dominant paradigms exist for instilling reasoning into a model’s parameters: reinforcement learning  
 461 (RL) and instruction distillation. RL is effective when clear reward signals are available (Ouyang  
 462 et al., 2022; Guo et al., 2025; Team et al., 2025; Wang et al., 2025c), but struggles in creative  
 463 domains where crafting a reward model to capture subjective qualities is an immense challenge  
 464 in itself (Ouyang et al., 2022; Zhang et al., 2024; Lu, 2025). WritingZero (Lu, 2025) adopts this  
 465 approach, but data and models remain closed. While recent work like VeriFree (Zhou et al., 2025)  
 466 also uses a proxy for reward in verifiable domains, REER applies a similar principle to recover  
 467 human-like reasoning for the broader challenge of open-ended generation.

468 Instruction Distillation offers an alternative, wherein reasoning traces are generated by a powerful but  
 469 proprietary “teacher” model, e.g., GPT-4 (Achiam et al., 2023). While effective, this approach is often  
 470 prohibitively expensive and is fundamentally capped by the capabilities of the “teacher” model (Guha  
 471 et al., 2025; Toshniwal et al., 2024). To overcome these data bottlenecks, researchers have increasingly  
 472 turned to synthetic data generation that builds a solution “forwards” (Wang et al., 2022; Zelikman  
 473 et al., 2022; 2024), e.g., LongWriter (Wu et al., 2025). Our central innovation is to “reverse-engineer”  
 474 reasoning by synthesizing it backwards from known good outcomes. By operationalizing this as a  
 475 scalable, gradient-free search guided by perplexity, we created Deep Writing-20K, the first large-scale  
 476 dataset of deep reasoning trajectories for open-ended tasks.

477 Our model, Deep Writer-8B, trained on this data, validates REER as a powerful and cost-effective  
 478 method, surpassing strong open-source baselines and achieving performance competitive with leading  
 479 proprietary models like GPT-4o. Importantly, our creation and release of Deep Writing-20K also  
 480 democratizes access to high-quality deep reasoning data, addressing a critical bottleneck for the  
 481 research community. **This opens several promising directions for future research.** A primary  
 482 avenue is scaling our experiments to larger models and datasets to investigate the scaling laws of  
 483 reverse-engineered reasoning. Furthermore, the core principle of REER is well-suited for novel  
 484 scenarios where reasoning annotations are scarce, making it a valuable paradigm to explore in  
 485 complex domains such as multi-step agentic tasks, scientific discovery, and multi-modal reasoning.

486  
487  
**ETHICS STATEMENT**488  
489  
490  
491  
492  
493  
494  
This research adheres to the ICLR Code of Ethics. Our work centers on the development of a new  
methodology for training large language models and the creation of a new dataset, ‘Deep Writing-  
20K’. The data for this dataset was sourced from publicly available and permissible sources, including  
online communities (e.g., r/WritingPrompts), public domain texts (Project Gutenberg), and existing  
open-source datasets (WildChat, LongWriter6K). We have strictly filtered on top of theses datasets to  
ensure that our data collection and usage practices respect user privacy and do not include personally  
identifiable information.495  
496  
497  
498  
499  
The primary goal of this research is to advance the understanding of reasoning in AI for open-ended,  
creative, and professional tasks. However, we acknowledge that, like any powerful generative model,  
the methods and models presented could be misused for generating harmful, biased, or misleading  
content. We tried our best to filter out harmful contents and prevented the model from internalizing  
implicit societal biases.500  
501  
502  
503  
The models and datasets used in our research, including the Qwen model series, are used in accordance  
with their respective licenses. We intend for our open-sourced dataset, ‘Deep Writing-20K’, to be  
used by the research community to foster further investigation into transparent and beneficial AI  
reasoning mechanisms.504  
505  
**REPRODUCIBILITY STATEMENT**506  
507  
508  
509  
510  
We are committed to ensuring the reproducibility of our research. All components required to  
replicate our findings are detailed within the paper and its appendix, and we provide a preview of the  
data in the supplementary.511  
512  
513  
514  
515  
The detailed data curation pipeline, including sourcing, synthesis, and filtering procedures, is de-  
scribed in Section 2.3. The core algorithm for ‘Reverse-Engineered Reasoning (REER)’ via iterative  
local search is detailed in Section 2.2. The exact prompts used for generating initial trajectories,  
performing segment-wise edits, and conducting inference are provided in the Appendix, Listings 1-4.  
The code is attached in the supplementary.516  
517  
518  
519  
We used publicly available base models for our experiments. The trajectory synthesis was performed  
using ‘Qwen2.5-32B-Instruct’, and the final ‘Deep Writer-8B’ model was fine-tuned from ‘Qwen3-  
8B-Base’. All training hyperparameters, including learning rate, batch size, and number of epochs,  
are specified in Section 3.1 .520  
521  
522  
Our evaluations were conducted on publicly available benchmarks: LongBench-Write, HelloBench,  
and WritingBench. The evaluation protocols, including the LLMs used as judges, are described in  
Section 3.2.524  
525  
**REFERENCES**526  
527  
528  
529  
Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman,  
Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report.  
*arXiv preprint arXiv:2303.08774*, 2023.530  
531  
532  
533  
Yushi Bai, Jiajie Zhang, Xin Lv, Linzhi Zheng, Siqi Zhu, Lei Hou, Yuxiao Dong, Jie Tang, and Juanzi  
Li. Longwriter: Unleashing 10,000+ word generation from long context llms. *arXiv preprint*  
*arXiv:2408.07055*, 2024.534  
535  
536  
537  
Cameron B Browne, Edward Powley, Daniel Whitehouse, Simon M Lucas, Peter I Cowling, Philipp  
Rohlfshagen, Stephen Tavener, Diego Perez, Spyridon Samothrakis, and Simon Colton. A survey  
of monte carlo tree search methods. *IEEE Transactions on Computational Intelligence and AI in  
games*, 4(1):1–43, 2012.538  
539  
Yichao Fu, Xuewei Wang, Yuandong Tian, and Jiawei Zhao. Deep think with confidence. *arXiv*  
*preprint arXiv:2508.15260*, 2025.

540 Tiancheng Gu, Kaicheng Yang, Chaoyi Zhang, Yin Xie, Xiang An, Ziyong Feng, Dongnan Liu,  
 541 Weidong Cai, and Jiankang Deng. Realsyn: An effective and scalable multimodal interleaved  
 542 document transformation paradigm. *arXiv preprint arXiv:2502.12513*, 2025.

543 Etrash Guha, Ryan Marten, Sedrick Keh, Negin Raoof, Georgios Smyrnis, Hritik Bansal, Marianna  
 544 Nezhurina, Jean Mercat, Trung Vu, Zayne Sprague, et al. Openthoughts: Data recipes for reasoning  
 545 models. *arXiv preprint arXiv:2506.04178*, 2025.

546 Caglar Gulcehre, Tom Le Paine, Srivatsan Srinivasan, Ksenia Konyushkova, Lotte Weerts, Abhishek  
 547 Sharma, Aditya Siddhant, Alex Ahern, Miaosen Wang, Chenjie Gu, et al. Reinforced self-training  
 548 (rest) for language modeling. *arXiv preprint arXiv:2308.08998*, 2023.

549 Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu,  
 550 Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms  
 551 via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

552 Guangzeng Han, Weisi Liu, and Xiaolei Huang. Attributes as textual genes: Leveraging llms  
 553 as genetic algorithm simulators for conditional synthetic data generation, 2025. URL <https://arxiv.org/abs/2509.02040>.

554 Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec  
 555 Helyar, Aleksander Madry, Alex Beutel, Alex Carney, et al. Openai o1 system card. *arXiv preprint*  
 556 *arXiv:2412.16720*, 2024.

557 Aviral Kumar, Vincent Zhuang, Rishabh Agarwal, Yi Su, John D Co-Reyes, Avi Singh, Kate Baumli,  
 558 Shariq Iqbal, Colton Bishop, Rebecca Roelofs, et al. Training language models to self-correct via  
 559 reinforcement learning. *arXiv preprint arXiv:2409.12917*, 2024.

560 Peiji Li, Kai Lv, Yunfan Shao, Yichuan Ma, Linyang Li, Xiaoqing Zheng, Xipeng Qiu, and Qipeng  
 561 Guo. Fastmcts: A simple sampling strategy for data synthesis. *arXiv preprint arXiv:2502.11476*,  
 562 2025.

563 Xun Lu. Writing-zero: Bridge the gap between non-verifiable problems and verifiable rewards. *arXiv*  
 564 *preprint arXiv:2506.00103*, 2025.

565 Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri  
 566 Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, et al. Self-refine: Iterative refinement  
 567 with self-feedback. *Advances in Neural Information Processing Systems*, 36:46534–46594, 2023.

568 Niklas Muennighoff, Zitong Yang, Weijia Shi, Xiang Lisa Li, Li Fei-Fei, Hannaneh Hajishirzi, Luke  
 569 Zettlemoyer, Percy Liang, Emmanuel Candès, and Tatsunori Hashimoto. s1: Simple test-time  
 570 scaling. *arXiv preprint arXiv:2501.19393*, 2025.

571 Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong  
 572 Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow  
 573 instructions with human feedback. *Advances in neural information processing systems*, 35:27730–  
 574 27744, 2022.

575 Haoran Que, Feiyu Duan, Liqun He, Yutao Mou, Wangchunshu Zhou, Jiaheng Liu, Wenge Rong,  
 576 Zekun Moore Wang, Jian Yang, Ge Zhang, et al. Hellobench: Evaluating long text generation  
 577 capabilities of large language models. *arXiv preprint arXiv:2409.16191*, 2024.

578 Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Mingchuan Zhang, YK Li, Yu Wu,  
 579 and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language  
 580 models. *arXiv preprint arXiv:2402.03300*, 2024.

581 Alex Su, Haozhe Wang, Weiming Ren, Fangzhen Lin, and Wenhui Chen. Pixel reasoner: In-  
 582 centivizing pixel-space reasoning with curiosity-driven reinforcement learning. *arXiv preprint*  
 583 *arXiv:2505.15966*, 2025.

584 Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut,  
 585 Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. Gemini: a family of highly  
 586 capable multimodal models. *arXiv preprint arXiv:2312.11805*, 2023.

594 Kimi Team, Angang Du, Bofei Gao, Bowei Xing, Changjiu Jiang, Cheng Chen, Cheng Li, Chenjun  
 595 Xiao, Chenzhuang Du, Chonghua Liao, et al. Kimi k1. 5: Scaling reinforcement learning with  
 596 llms. *arXiv preprint arXiv:2501.12599*, 2025.

597

598 Qwen Team. Qwq-32b: Embracing the power of reinforcement learning, March 2025. URL  
 599 <https://qwenlm.github.io/blog/qwq-32b/>.

600 Shubham Toshniwal, Ivan Moshkov, Sean Narenthiran, Daria Gitman, Fei Jia, and Igor Gitman.  
 601 Openmathinstruct-1: A 1.8 million math instruction tuning dataset. *Advances in Neural Information  
 602 Processing Systems*, 37:34737–34774, 2024.

603

604 Haozhe Wang, Chao Du, Panyan Fang, Li He, Liang Wang, and Bo Zheng. Adversarial constrained  
 605 bidding via minimax regret optimization with causality-aware reinforcement learning. In *Pro-  
 ceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp.  
 606 2314–2325, 2023.

607

608 Haozhe Wang, Long Li, Chao Qu, Fengming Zhu, Weidi Xu, Wei Chu, and Fangzhen Lin. To code  
 609 or not to code? adaptive tool integration for math language models via expectation-maximization.  
 610 *arXiv preprint arXiv:2502.00691*, 2025a.

611

612 Haozhe Wang, Chao Qu, Zuming Huang, Wei Chu, Fangzhen Lin, and Wenhua Chen. Vl-rethinker:  
 613 Incentivizing self-reflection of vision-language models with reinforcement learning. *arXiv preprint  
 614 arXiv:2504.08837*, 2025b.

615

616 Haozhe Wang, Qixin Xu, Che Liu, Junhong Wu, Fangzhen Lin, and Wenhua Chen. Emergent  
 617 hierarchical reasoning in llms through reinforcement learning. *arXiv preprint arXiv:2509.03646*,  
 618 2025c.

619

620 Tiannan Wang, Jiamin Chen, Qingrui Jia, Shuai Wang, Ruoyu Fang, Huilin Wang, Zhaowei Gao,  
 621 Chunzhao Xie, Chuou Xu, Jihong Dai, et al. Weaver: Foundation models for creative writing.  
 622 *arXiv preprint arXiv:2401.17268*, 2024.

623

624 Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A Smith, Daniel Khashabi, and  
 625 Hannaneh Hajishirzi. Self-instruct: Aligning language models with self-generated instructions.  
 626 *arXiv preprint arXiv:2212.10560*, 2022.

627

628 Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny  
 629 Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in  
 630 neural information processing systems*, 35:24824–24837, 2022.

631

632 Yuning Wu, Jiahao Mei, Ming Yan, Chenliang Li, Shaopeng Lai, Yuran Ren, Zijia Wang, Ji Zhang,  
 633 Mengyue Wu, Qin Jin, et al. Writingbench: A comprehensive benchmark for generative writing.  
 634 *arXiv preprint arXiv:2503.05244*, 2025.

635

636 Kaicheng Yang, Jiankang Deng, Xiang An, Jiawei Li, Ziyong Feng, Jia Guo, Jing Yang, and Tongliang  
 637 Liu. Alip: Adaptive language-image pre-training with synthetic caption. In *Proceedings of the  
 638 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 2922–2931, October 2023.

639

640 Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L Griffiths, Yuan Cao, and Karthik  
 641 Narasimhan. Tree of thoughts: Deliberate problem solving with large language models, 2023.  
 642 *URL https://arxiv.org/abs/2305.10601*, 3:1, 2023.

643

644 Eric Zelikman, Yuhuai Wu, Jesse Mu, and Noah Goodman. Star: Bootstrapping reasoning with  
 645 reasoning. *Advances in Neural Information Processing Systems*, 35:15476–15488, 2022.

646

647 Eric Zelikman, Georges Harik, Yijia Shao, Varuna Jayasiri, Nick Haber, and Noah D Goodman.  
 648 Quiet-star: Language models can teach themselves to think before speaking. *arXiv preprint  
 649 arXiv:2403.09629*, 2024.

650

651 Lunjun Zhang, Arian Hosseini, Hritik Bansal, Mehran Kazemi, Aviral Kumar, and Rishabh Agarwal.  
 652 Generative verifiers: Reward modeling as next-token prediction. *arXiv preprint arXiv:2408.15240*,  
 653 2024.

648 Wenting Zhao, Xiang Ren, Jack Hessel, Claire Cardie, Yejin Choi, and Yuntian Deng. Wildchat: 1m  
649 chatgpt interaction logs in the wild. *arXiv preprint arXiv:2405.01470*, 2024.  
650

651 Xiangxin Zhou, Zichen Liu, Anya Sims, Haonan Wang, Tianyu Pang, Chongxuan Li, Liang  
652 Wang, Min Lin, and Chao Du. Reinforcing general reasoning without verifiers. *arXiv preprint*  
653 *arXiv:2505.21493*, 2025.

654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701

702 **A LIST OF PROMPTS**  
703704 Below we list the exact prompts used for trajectory synthesis and in-house evaluation. For the meta-  
705 structure guidelines and thinking pattern injection, refer to Listing 1. For enforcing segment-wise  
706 edits, refer to Listing 2. For quality assessment with regard to deep reasoning, refer to Listing 4. For  
707 computing the proxy score of a deep reasoning trajectory, we employ Listing 3, without including the  
708 reference output.  
709710 **LISTINGS**  
711712 

1	Prompt for Generating Initial Thinking. . . . .	14
2	Prompt for Segment-wise Edits. . . . .	16
3	Prompt for Standard Inference. . . . .	17
4	Prompt for Rating Response Quality w.r.t. Deep Reasoning. . . . .	19

  
713714 Listing 1: Prompt for Generating Initial Thinking.  
715

```

716 1 You are an expert in many fields. Suppose you will give a specific
717     final response, I need you to also write down the thought
718     process behind this solution.
719 2 Here is a task:
720 3 {}
721 4
722 5 Here is the solution you will create:
723 6 {}
724 7
725 8 Now, you need to write down the thinking process behind this
726     solution, as if you are thinking aloud and brainstorming in
727     the mind. The thinking process involves thoroughly exploring
728     questions through a systematic long thinking process. This
729     requires engaging in a comprehensive cycle of analysis,
730     summarizing, exploration, reassessment, reflection,
731     backtracing, and iteration to develop well-considered thinking
732     process. Present your complete thought process within a
733     single and unique '<think></think>' tag.
734 9
735 10 Your thought process must adhere to the following requirements:
736 11
737 12 1. **Narrate in the first-person as if you are thinking aloud and
738     brainstorming**
739 13     Stick to the narrative of "I". Imagine you are brainstorming
740     and thinking in the mind. Use verbalized, simple language.
741 14
742 15 2. **Unify the thinking process and the writing solution:***
743 16     Your thought process must precisely correspond to a part of
744     the writing solution. The reader should be able to clearly see
745     how your thoughts progressively "grew" into the finished
746     piece, making the copy feel like the inevitable product of
747     your thinking.
748 17
749 18 3. **Tone of Voice: Planning, Sincere, Natural, and Accessible***
750 19     Imagine you are analyzing and planning what to do before you
751     start to write the solution. Your language should be plain
752     and easy to understand, avoiding obscure professional jargon
753     to explain complex thought processes clearly.
754 20
755 21 4. **Logical Flow: Clear and Progressive***

```

756 22

757 23

758 24

759

760

761 25

762 26

763

764

765

766

767

768

769

770 27

771 28

772

773

774

775

776 29

777 30

778

779

780

781

782

783

784

785

786 31

787 32

788

789

790

791

792 33

793

794

795

796

797 34

798

799

800

801 35

802

803

804

805

806 38

807

808 39

809 40

5. **\*\*Thinking Framework for deep thinking\*\***  
 To ensure your thinking is clear and deep, to showcase your thinking and planning to fulfill the task, below is what you might cover when you are thinking aloud and brainstorming.

Understanding the user intent and the task: Before putting pen to paper, I need to thoroughly consider the fundamental purpose of the writing. I first need to discern the user's true goal behind their literal request. Next, I will consider: Who am I talking to? I will create a precise profile of the target reader, understanding their pain points, aspirations, and reading context. Then, I will establish the Core Objective: What specific emotional, cognitive, and behavioral changes do I most want the reader to experience after reading?

Establishing the content: I need to brainstorm a core creative idea and communication strategy centered around my objective. Then, I will think about what content and key information I need to convey to the reader to fulfill the writing task, and what source materials this will involve.

Building the structure: I need to design a clear narrative path for the reader, like a "blueprint." First, I will plan the article's skeleton (e.g., using a framework like the Golden Circle "Why-How-What," the AIDA model "Attention-Interest-Desire-Action," or a narrative structure "Beginning-Development-Climax-Resolution"). Then, I will plan the key modules: How will the introduction hook the reader? How will the body be layered and the arguments arranged? How will the conclusion summarize, elevate the message, and provide a clear Call to Action (CTA)?

Outline: If the task output might be relatively long, I will consider writing an outline (or a draft) which naturally derives from the plan above. Specifically, the outline will ground my plan into paragraphs, summarizing the key content for each paragraph and what are the key points here, sentence structure or anything important for the paragraph.

I PROMISE I will NOT copy the solution I will NOT copy the solution, this outline (or draft) should only look like a prototype or outline of the target text. After finishing this outline, I will check again if there are any details or notes I should pay attention to when writing the final solution.

I will begin writing this draft after a '--- Outline (or Draft) ---' separator at the end of my thinking process. The draft will be included in the same '<think></think>' block.

6. Throughout the thinking process, I want to involve deep thinking and planning, and use deliberate self-critique/self-reflection in my thinking process. Trigger these by regularly using patterns such as 'wait', 'maybe', 'let me', etc. For example:

- Hmm, maybe .. (other concrete thinking regarding the given request)
- Let me think ..
- Wait no ..

```

810 41      - But wait ..(might find something wrong with your previous
811 42      thoughts)
812 42      - Wait, that's a bit ..(reflections about previous decisions).
813 43      Let me think .. (are thinking of other possibilities)
814 43      - Wait, the user said ..(backtracing of previous information).
815 44      So ..
816 44      - Hmm...Alternatively, maybe ..(branching on other
817 45      possibilities)
818 45      - But ..
819 46 But I promise I will use diverse triggers and will NOT use same
820 46 triggers repeatedly. I will use these when analyzing user
821 46 needs, establishing content and structure and when I consider
822 46 alternatives, backtracing and the details. I will NOT use them
823 46 when I write the draft or I am approaching the end of
824 47 thinking.
825 48 In the thinking process, make sure NO PAST TENSES, NO PAST TENSES,
826 48 because this is the thought process before you are to write a
827 48 final solution. You are planning what you will and you need
828 48 to do.
829 49 Imagine you're thinking aloud and brainstorming. Write it as an
830 49 internal monologue or a stream of consciousness. Do not use
831 49 bullet points, numbers, or formal section headings.
832 50 Now record your thinking process within '<think></think>' tags.

```

Listing 2: Prompt for Segment-wise Edits.

```

833
834
835 1 Your task is to receive a user request, a target output, and an
836 1 existing thinking process, and then to refine and enrich a
837 1 specific paragraph within that thinking process.
838 2
839 3 ----> **Task**
840 4 {}
841 5 ----> **Target Output**
842 6 {}
843 7 ----> **Thinking Process**
844 8 {}
845 9 <replace>
846 10 {}
847 11 </replace>
848 12 {}
849 13
850 14 Follow this three-step method to construct your response:
851 15
852 16 **Step 1:** Locate the paragraph you need to revise within the
853 16 existing thinking process. In relation to the surrounding
854 17 context, what is the primary function of this paragraph?
855 18
856 18 **Step 2:** Read the "Target Text" and the "paragraph to be
857 18 revised" side-by-side. Ask yourself a key question: Does the
858 18 thinking process reflected in this paragraph lack crucial
859 19 steps, or is there content that could be further optimized and
860 19 detailed to better align with the Target Text?
861 20
862 20 **Step 3:** Improve and optimize the paragraph (that represents a
863 21 part of the thinking process).
864 21 - Based on the analysis, refine the initial target paragraph into
865 21 a new one, base remain the tone. Put the refinement into <
866 21 refine></refine> tags.

```

864 22 - To help involve deep thinking and planning, consider deliberate  
 865 self-critique/self-reflection in your thinking process.  
 866 Trigger these by frequently using patterns such as 'wait', ' ' maybe', 'let me', etc. For example:  
 867  
 868 23 - Hmm, maybe .. (other concrete thinking regarding the given  
 869 request)  
 870 24 - Let me think ..  
 871 25 - Wait no ..  
 872 26 - But wait ..(might find something wrong with your previous  
 873 thoughts)  
 874 27 - Wait, that's a bit ..(reflections about previous decisions).  
 875 Let me think .. (are thinking of other possibilities)  
 876 28 - Wait, the user said ..(backtracking of previous information)  
 877 29 . So ..  
 878 30 - Hmm...Alternatively, maybe ..(branching on other  
 879 possibilities)  
 880 31 - But ..  
 881 - If the function of the paragraph being improved is to serve as a  
 882 first draft of the text, you must focus on enhancing the text'  
 883 s logic and completeness. The draft should not be a general  
 884 outline but should express specific content and state a clear  
 885 point of view. Consider whether the current draft is an  
 886 appropriate prototype for the Target Text: it should be  
 887 neither too vague nor a direct copy, but should reflect a  
 888 32 foundational version.  
 889 33 Based on the guide above, you are to refine **\*\*only\*\*** the section  
 890 marked for replacement below.  
 891 <replace>  
 892 {}  
 893 </replace>  
 894 38 In your response, first, present your analysis following the three  
 895 -step method within '<analyze></analyze>' tags. Finally, place  
 896 the corresponding, refined paragraph of the **\*\*thinking**  
 897 process**\*\*** within '<refine></refine>' tags.  
 898 Notes: a. Avoid repeating. Reduce the use of the same connection  
 899 words, avoid repeating the same meanings over and over again.  
 900 Ensure that your revised content does not repeat information  
 901 from the context.  
 902 b. please keep the first a few words of the original paragraph,  
 903 especially the connection words  
 904 c. use self-critique trigger words, such as 'wait', 'maybe', 'let  
 905 me', etc.

Listing 3: Prompt for Standard Inference.

906  
 907 1 You are an expert in many fields. Suppose you will give a specific  
 908 2 final response, I need you to also write down the thought  
 909 process behind this solution.  
 910 3 Here is a task:  
 911 4 {}  
 912 5  
 913 6 Now, you need to think aloud and brainstorm in the mind. The  
 914 thinking process involves thoroughly exploring questions  
 915 through a systematic long thinking process. This requires  
 916 engaging in a comprehensive cycle of analysis, summarizing,  
 917 exploration, reassessment, reflection, backtracing, and

iteration to develop well-considered thinking process. Present your complete thought process within a single and unique '<think></think>' tag.

Your thought process must adhere to the following requirements:

1. **\*\*Narrate in the first-person as if you are thinking aloud and brainstorming\*\***  
Stick to the narrative of "I". Imagine you are brainstorming and thinking in the mind. Use verbalized, simple language.
2. **\*\*Unify the thinking process and the writing solution:\*\***  
Your thought process must precisely correspond to a part of the writing solution. The reader should be able to clearly see how your thoughts progressively "grew" into the finished piece, making the copy feel like the inevitable product of your thinking.
3. **\*\*Tone of Voice: Planning, Sincere, Natural, and Accessible\*\***  
Imagine you are analyzing and planning what to do before you start to write the solution. Your language should be plain and easy to understand, avoiding obscure professional jargon to explain complex thought processes clearly.
4. **\*\*Logical Flow: Clear and Progressive\*\***
5. **\*\*Thinking Framework for deep thinking\*\***  
To ensure your thinking is clear and deep, to showcase your thinking and planning to fulfill the task, below is what you might cover when you are thinking aloud and brainstorming.

Understanding the user intent and the task: Before putting pen to paper, I need to thoroughly consider the fundamental purpose of the writing. I first need to discern the user's true goal behind their literal request. Next, I will consider: Who am I talking to? I will create a precise profile of the target reader, understanding their pain points, aspirations, and reading context. Then, I will establish the Core Objective: What specific emotional, cognitive, and behavioral changes do I most want the reader to experience after reading?

Establishing the content: I need to brainstorm a core creative idea and communication strategy centered around my objective. Then, I will think about what content and key information I need to convey to the reader to fulfill the writing task, and what source materials this will involve.

Building the structure: I need to design a clear narrative path for the reader, like a "blueprint." First, I will plan the article's skeleton (e.g., using a framework like the Golden Circle "Why-How-What," the AIDA model "Attention-Interest-Desire-Action," or a narrative structure "Beginning-Development-Climax-Resolution"). Then, I will plan the key modules: How will the introduction hook the reader? How will the body be layered and the arguments arranged? How will the conclusion summarize, elevate the message, and provide a clear Call to Action (CTA)?

972 30 Draft: unless it is a really easy request, otherwise I need to  
 973 consider writing a draft based on the plan above, before you  
 974 give the final writing solution. I will translate my plan  
 975 into paragraphs, considering the key points, content, and  
 976 sentence structure for each. This initial draft should look  
 977 like a prototype of the target text. This draft will be way  
 978 shorter than the final writing solution within controlled  
 979 length, but it must also avoid being too vague or general or  
 980 simply copying the final text. I will begin writing this draft  
 981 after a '--- The Draft ---' separator at the end of my  
 982 thinking process. The draft will be included in the same '<  
 983 think></think>' block. After writing the draft, I will further  
 984 critique what can be improved, and analyze what details can  
 985 be enriched (and hence make it more likely to eventually  
 986 arrive at the given solution)  
 987 31  
 988 32 6. Throughout the thinking process, I want to involve deep  
 989 thinking and planning, and use deliberate self-critique/self-  
 990 reflection in my thinking process. Trigger these by frequently  
 991 using patterns such as 'wait', 'maybe', 'let me', etc. For  
 992 33 example:  
 993 34     - Hmm, maybe .. (other concrete thinking regarding the given  
 994 35 request)  
 995 36     - Let me think ..  
 996 37     - Wait no ..  
 997 38     - But wait ..(might find something wrong with your previous  
 998 39 thoughts)  
 999 40     - Wait, that's a bit ..(reflections about previous decisions).  
 1000 41     Let me think .. (are thinking of other possibilities)  
 1001 42     - Wait, the user said ..(backtracking of previous information)  
 1002 43     . So ..  
 1003 44     - Hmm...Alternatively, maybe ..(branching on other  
 1004 45 possibilities)  
 1005 46     - But ..  
 1006 47 Now record your clear, complete, and logical thinking process  
 1007 48 within '<think></think>' tags.  
 1008 49 In the thinking process, make sure NO PAST TENSES, NO PAST TENSES,  
 1009 50 because this is the thought process before you are to write a  
 1010 51 final solution. You are planning what you will and you need  
 1011 52 to do.  
 1012 53 Imagine you're thinking aloud and brainstorming. Write it as an  
 1013 54 internal monologue or a stream of consciousness. Do not use  
 1014 55 bullet points, numbers, or formal section headings.  
 1015

Listing 4: Prompt for Rating Response Quality w.r.t. Deep Reasoning.

1016 1  
 1017 2  
 1018 3 You are an expert judge in AI generated content. Your primary task  
 1019 4 is to assess an AI model's response, specifically focusing on  
 1020 5 its ability to perform \*\*deep thinking and planning\*\*. You  
 1021 6 will evaluate the response across five distinct dimensions. A  
 1022 7 model that excels at deep thinking will not only provide a  
 1023 8 correct answer but will demonstrate a structured, logical, and  
 1024 9 well-grounded reasoning process from start to finish.  
 1025 5 Your final output must be a structured report with a score and  
 1026 6 justification for each dimension.  
 1027 7

```

1026 6
1027 7
1028 8
1029 9 ## Evaluation Dimensions & Scoring
1030 10
1031 11 #### 1\. Understanding & Problem Decomposition
1032 12
1033 13 **Relation to Deep Thinking:** This is the foundational step. Deep
1034 thinking is impossible without first accurately understanding
1035 the problem in its entirety. This dimension measures if the
1036 model comprehends the user's explicit and implicit needs and
1037 then breaks down the complex request into manageable, logical
1038 parts. This act of decomposition *is* the first stage of
1039 planning.
1040 14
1041 * Score 1 (Poor): The model fundamentally misunderstands the
1042 user's request or ignores key components. The response is off-
1043 topic or fails to address the core problem.
1044 * Score 3 (Average): The model grasps the main goal but may
1045 overlook nuances or implicit constraints. It attempts to break
1046 down the problem, but the decomposition may be incomplete or
1047 slightly illogical.
1048 * Score 5 (Excellent): The model demonstrates a
1049 comprehensive understanding of the user's intent, including
1050 subtle details. It expertly deconstructs the problem into a
1051 clear, exhaustive, and actionable framework.
1052 18 Score 2 and Score 4 fit interpolate into the above scoring
1053 19 criterion.
1054 20
1055 21 #### 2\. Content Structure & Logical Consistency
1056 22
1057 * Score 1 (Poor): This dimension reflects the clarity
1058 and order of the model's thought process. A deep, well-
1059 considered plan has a coherent structure where ideas flow
1060 logically and conclusions are built upon valid premises.
1061 Inconsistencies or a chaotic structure indicate shallow,
1062 stream-of-consciousness generation rather than deliberate
1063 24 planning.
1064 * Score 3 (Average): The response is disorganized, rambling, or
1065 internally contradictory. It's difficult to follow the model'
1066 26 line of reasoning.
1067 * Score 5 (Excellent): The response has a discernible
1068 structure (e.g., uses headings, lists), but the flow between
1069 sections could be improved. It is mostly consistent, with only
1070 27 minor logical gaps.
1071 * Score 1 (Poor): The response is impeccably structured
1072 . Each part logically follows from the previous one, building
1073 a coherent and compelling argument or plan. The internal logic
1074 is sound and easy to follow from beginning to end.
1075 28 Score 2 and Score 4 interpolate into the above scoring criterion
1076 30
1077 31
1078 32 #### 3\. Depth of Analysis & Synthesis
1079 33
1080 **Relation to Deep Thinking:** This is the core of "deep thinking
1081 . It goes beyond simply retrieving facts and measures the
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
2598
2599
2600
2601
2602
2603
2604
2605
2606
2607
2608
2609
2610
2611
2612
2613
2614
2615
2616
2617
2618
2619
2620
2621
2622
2623
2624
2625
2626
2627
2628
2629
2630
2631
2632
2633
2634
2635
2636
2637
2638
2639
2640
2641
2642
2643
2644
2645
2646
2647
2648
2649
2650
2651
2652
2653
2654
2655
2656
2657
2658
2659
2660
2661
2662
2663
2664
2665
2666
2667
2668
2669
2670
2671
2672
2673
2674
2675
2676
2677
2678
2679
2680
2681
2682
2683
2684
2685
2686
2687
2688
2689
2690
2691
2692
2693
2694
2695
2696
2697
2698
2699
2700
2701
2702
2703
2704
2705
2706
2707
2708
2709
2710
2711
2712
2713
2714
2715
2716
2717
2718
2719
2720
2721
2722
2723
2724
2725
2726
2727
2728
2729
2730
2731
2732
2733
2734
2735
2736
2737
2738
2739
2740
2741
2742
2743
2744
2745
2746
2747
2748
2749
2750
2751
2752
2753
2754
2755
2756
2757
2758
2759
2760
2761
2762
2763
2764
2765
2766
2767
2768
2769
2770
2771
2772
2773
2774
2775
2776
2777
2778
2779
2780
2781
2782
2783
2784
2785
2786
2787
2788
2789
2790
2791
2792
2793
2794
2795
2796
2797
2798
2799
2800
2801
2802
2803
2804
2805
2806
2807
2808
2809
2810
2811
2812
2813
2814
2815
2816
2817
2818
2819
2820
2821
2822
2823
2824
2825
2826
2827
2828
2829
2830
2831
2832
2833
2834
2835
2836
2837
2838
2839
2840
2841
2842
2843
2844
2845
2846
2847
2848
2849
2850
2851
2852
2853
2854
2855
2856
2857
2858
2859
2860
2861
2862
2863
2864
2865
2866
2867
2868
2869
2870
2871
2872
2873
2874
2875
2876
2877
2878
2879
2880
2881
2882
2883
2884
2885
2886
2887
2888
2889
2890
2891
2892
2893
2894
2895
2896
2897
2898
2899
2900
2901
2902
2903
2904
2905
2906
2907
2908
2909
2910
2911
2912
2913
2914
2915
2916
2917
2918
2919
2920
2921
2922
2923
2924
2925
2926
2927
2928
2929
2930
2931
2932
2933
2934
2935
2936
2937
2938
2939
2940
2941
2942
2943
2944
2945
2946
2947
2948
2949
2950
2951
2952
2953
2954
2955
2956
2957
2958
2959
2960
2961
2962
2963
2964
2965
2966
2967
2968
2969
2970
2971
2972
2973
2974
2975
2976
2977
2978
2979
2980
2981
2982
2983
2984
2985
2986
2987
2988
2989
2990
2991
2992
2993
2994
2995
2996
2997
2998
2999
3000
3001
3002
3003
3004
3005
3006
3007
3008
3009
3010
3011
3012
3013
3014
3015
3016
3017
3018
3019
3020
3021
3022
3023
3024
3025
3026
3027
3028
3029
3030
3031
3032
3033
3034
3035
3036
3037
3038
3039
3040
3041
3042
3043
3044
3045
3046
3047
3048
3049
3050
3051
3052
3053
3054
3055
3056
3057
3058
3059
3060
3061
3062
3063
3064
3065
3066
3067
3068
3069
3070
3071
3072
3073
3074
3075
3076
3077
3078
3079
3080
3081
3082
3083
3084
3085
3086
3087
3088
3089
3090
3091
3092
3093
3094
3095
3096
3097
3098
3099
3100
3101
3102
3103
3104
3105
3106
3107
3108
3109
3110
3111
3112
3113
3114
3115
3116
3117
3118
3119
3120
3121
3122
3123
3124
3125
3126
3127
3128
3129
3130
3131
3132
3133
3134
3135
3136
3137
3138
3139
3140
3141
3142
3143
3144
3145
3146
3147
3148
3149
3150
3151
3152
3153
3154
3155
3156
3157
3158
3159
3160
3161
3162
3163
3164
3165
3166
3167
3168
3169
3170
3171
3172
3173
3174
3175
3176
3177
31
```

1080 model's ability to analyze underlying principles, connect  
 1081 disparate ideas, and synthesize them to create new insights. A  
 1082 simple plan lists steps; a deeply thought-out plan explains \*  
 1083 why\* those are the right steps and how they interrelate.  
 1084<sup>34</sup>  
 1085<sup>35</sup> \* \*\*Score 1 (Poor):\*\* The response is superficial, relying on  
 1086 cliches or surface-level information. It shows no evidence of  
 1087 analyzing the "why" behind the "what."  
 1088<sup>36</sup> \* \*\*Score 3 (Average):\*\* The model provides a competent analysis  
 1089 , explaining concepts correctly but treating them in isolation.  
 1090 It lacks the synthesis needed to create a novel or holistic  
 1091 perspective.  
 1092 \* \*\*Score 5 (Excellent):\*\* The model provides a profound  
 1093 analysis, connecting concepts in insightful ways. It  
 1094 synthesizes information to offer a nuanced perspective that is  
 1095 more than the sum of its parts, demonstrating a true grasp of  
 1096 the subject matter.  
 1097 Score 2 and Score 4 interpolate into the above scoring criterion  
 1098<sup>37</sup>  
 1099<sup>38</sup> -----  
 1100<sup>40</sup> **## 4\. Presentation Clarity**  
 1101<sup>41</sup>  
 1102<sup>42</sup> \*\*Relation to Deep Thinking:\*\* A brilliant plan is useless if it  
 1103 cannot be understood. This dimension assesses the model's  
 1104 ability to communicate its complex thoughts and plans  
 1105 effectively. Clarity in presentation demonstrates a higher  
 1106 level of understanding, as the model must distill its  
 1107 reasoning into a format that is concise, accessible, and  
 1108 actionable for the user.  
 1109<sup>44</sup>  
 1110 \* \*\*Score 1 (Poor):\*\* The response is convoluted, filled with  
 1111 jargon, or poorly formatted. The user would struggle to  
 1112 understand the main points or how to act on the advice.  
 1113 \* \*\*Score 3 (Average):\*\* The response is generally  
 1114 understandable but could be more concise or better organized.  
 1115 It may be overly dense or require the user to re-read sections  
 1116 to grasp the meaning.  
 1117 \* \*\*Score 5 (Excellent):\*\* The response is exceptionally clear,  
 1118 concise, and well-formatted. It uses plain language and  
 1119 effective formatting (like lists, bolding, or tables) to make  
 1120 complex information easy to digest and act upon.  
 1121<sup>47</sup> Score 2 and Score 4 interpolate into the above scoring criterion  
 1122<sup>48</sup>  
 1123<sup>50</sup> -----  
 1124<sup>51</sup> **## 5\. Factual Grounding (Hallucination Check)**  
 1125<sup>52</sup>  
 1126<sup>53</sup> \*\*Relation to Deep Thinking:\*\* Deep thinking and planning must be  
 1127 grounded in reality to be useful. A plan built on fabricated  
 1128 information ("hallucinations") is fundamentally flawed and  
 1129 demonstrates a critical failure in the reasoning process. This  
 1130 model's entire output.  
 1131<sup>54</sup>  
 1132 \*This dimension is scored on a severity scale, not a quality scale  
 1133<sup>55</sup>  
 1134 \*  
 1135<sup>56</sup>

```

113457      * **Score 4 (Factually Sound):** The response contains no
113558      discernible factual errors or hallucinations.
113659      * **Score 3 (Minor Inaccuracy):** Contains a small error (e.g.,
1137      a slightly incorrect date, a minor misstatement) that does not
1138      undermine the overall logic or conclusion of the response.
113959      * **Score 2 (Significant Hallucination):** Contains a major
1140      factual error that invalidates a key part of the argument or
1141      plan. The response is partially unreliable.
114260      * **Score 0 (Critical Hallucination):** The core premise or a
1143      critical component of the response is based on a fabrication,
1144      rendering the entire output untrustworthy and potentially
1145      harmful.
114661      Score 1 interpolates into the above scoring criterion.
114762      -----
114863      ## Final Output Format
114964
115065      Please provide your evaluation in the following structured json
1151      format.
115267      ```json
115368      {
115469          "evaluationReport": {
115570              "understandingAndDecomposition": {
115671                  "score": "[Enter a score from 1-5]",
1157                  "justification": "[Your justification here. Explain why you
115873                  gave this score.]"
115974              },
116075              "structureAndConsistency": {
116176                  "score": "[Enter a score from 1-5]",
1162                  "justification": "[Your justification here. Explain why you
116377                  gave this score.]"
116478              },
116579              "depthOfAnalysis": {
116680                  "score": "[Enter a score from 1-5]",
1167                  "justification": "[Your justification here. Explain why you
116882                  gave this score.]"
116983              },
117084              "presentationClarity": {
1171                  "score": "[Enter a score from 1-5]",
1172                  "justification": "[Your justification here. Explain why you
117386                  gave this score.]"
117487              },
117588              "factualGrounding": {
1176                  "severityScore": "[Enter a severity score from 1-5]",
1177                  "justification": "[Describe any inaccuracies or
117890                  hallucinations found. If none, state 'Response is factually
1179                  sound.]"
1180                  },
1181                  "overallSummary": "[Provide a final, concise paragraph
1182                  summarizing the model's overall performance in deep thinking
118391                  and planning. A response with a Hallucination Severity Score
118492                  of 2 or 3 cannot be considered a high-quality example of
118593                  planning, regardless of other scores.]"
118694          }
118795      -----
118896      <User Request>

```

```

118897 $INST$  

118998 </User Request>  

119099 <Response>  

1191100 $RESPONSE$  

1192101  

1193102  

1194103  

1195104  

1196105  

1197106  

1198107 Now go back to the evaluation guideline and give the json report  

1199 . . .
1200
1201
1202 B QUALITATIVE ANALYSIS
1203
1204 B.1 GENERATION QUALITY OF DEEP THINKING
1205
1206 Beyond quantitative scores, we sought to understand how
1207 well DeepWriter internalizes the qualities of deep thinking.
1208 To this end, we conducted a qualitative analysis, scoring
1209 model outputs on five dimensions intrinsically linked to
1210 advanced reasoning and planning:
1211
1212 


1213 - Problem Deconstruction: The ability to break down
1214 a complex prompt into a logical hierarchy of sub-
1215 goals. This is the foundation of effective planning.

1216 - Logical Consistency: Maintaining a coherent and
1217 non-contradictory reasoning path throughout the en-
1218 tire generation necessitates the ability to plan over
1219 the generation.

1220 - Depth of Analysis: Moving beyond surface-level
1221 responses to explore nuances, consider alternatives,
1222 and demonstrate sophisticated understanding. This
1223 reflects the "deep" aspect of the thinking process.

1224 - Presentation Clarity: The ability to structure the
1225 final output in a clear, organized, and persuasive man-
1226 ner, which is a direct outcome of a well-formed internal
1227 plan.

1228 - Factual Grounding: Ensuring that generated
1229 content, where applicable, is accurate and well-
1230 supported, reflecting a robust and reality-aware rea-
1231 soning process.

1232 

1233 The normalized scores, visualized in the radar chart in Figure 6, provide a signature of each model's
1234 reasoning profile. As illustrated in Figure 6, DeepWriter-8B exhibits a remarkably strong and well-
1235 rounded reasoning profile. Its performance polygon significantly envelopes that of the LongWriter-
1236 8B baseline, showing dramatic improvements across all five dimensions. This confirms that our
1237 methodology genuinely enhances underlying reasoning capabilities, rather than just improving
1238 superficial output fluency.
1239 Furthermore, DeepWriter-8B's profile closely rivals that of GPT-4o and substantially exceeds Claude
1240 3.5, particularly in Depth of Analysis and Factual Grounding. While the state-of-the-art Claude 3.7
1241 still defines the frontier, especially in Depth of Analysis, our 8B model has demonstrably bridged a
1242 large portion of the capability gap. This validates our central claim: instilling a deep thinking process
1243 through gradient-free synthesis is a highly promising pathway toward building more powerful and
1244 scalable models.

```

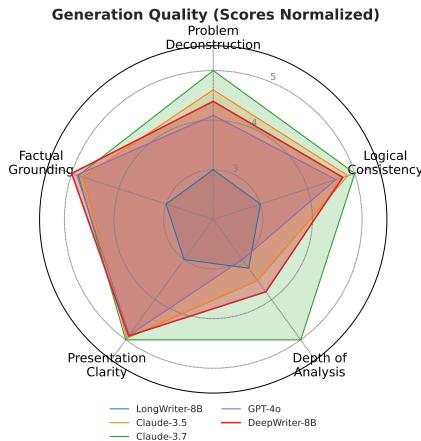
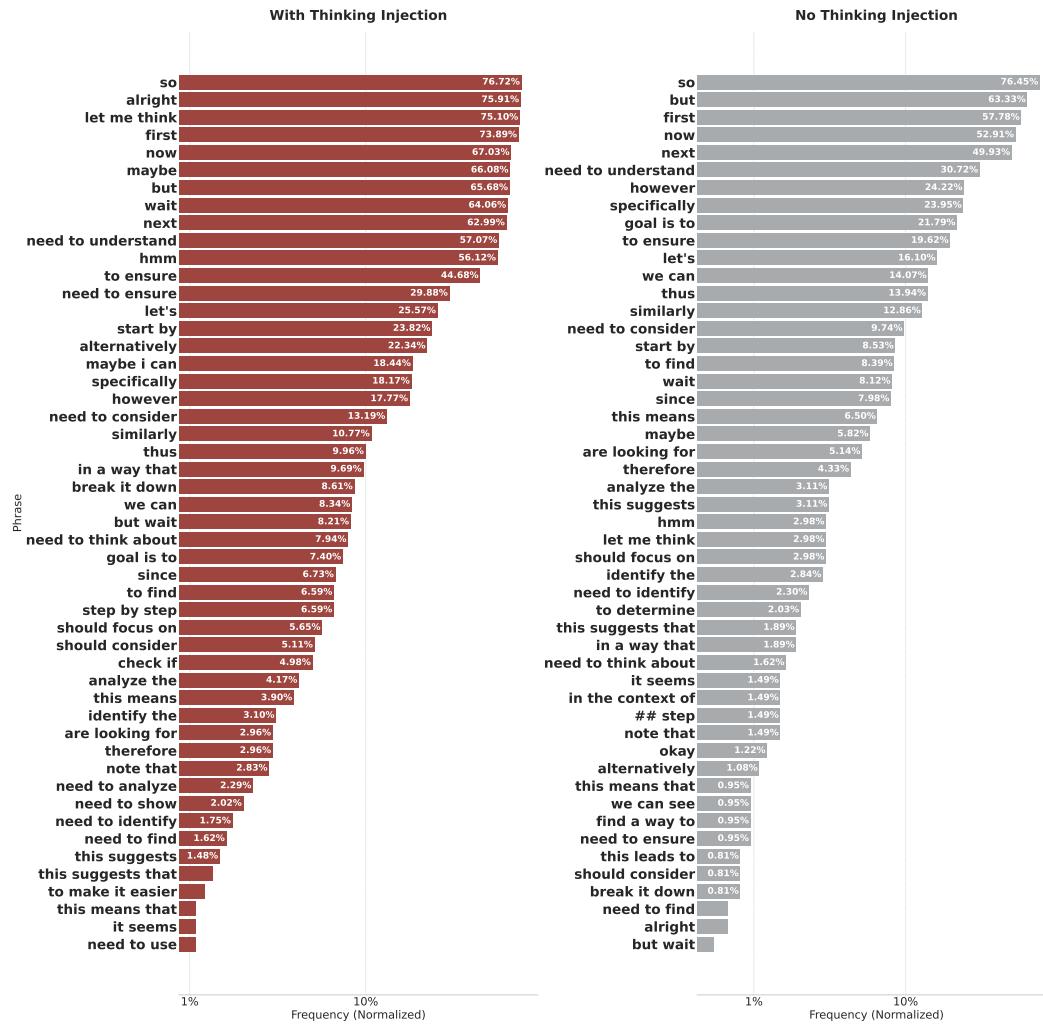


Figure 6: Qualitative comparison of generation quality. Scores are normalized across five dimensions related to deep thinking. DeepWriter-8B shows a reasoning profile far superior to the open-source baseline and is competitive with top proprietary models.

1242 B.2 QUALITATIVE COMPARISON OF THINKING PATTERNS  
1243

1244 To better understand how injecting human-like thinking patterns during synthesis affects the model’s  
1245 behavior, we analyze the frequency of reasoning phrases generated by the full model versus the  
1246 ablated model trained without injecting thinking patterns. The thinking patterns are deduplicated  
1247 such that occurrences will be counted only once for the same solution.

1248  
1249 **Comparison of Deduplicated Thinking Pattern Frequencies (Top 50)**  
1250

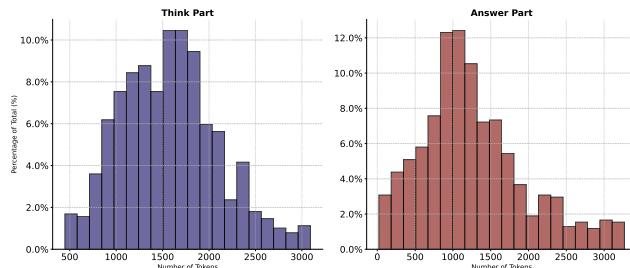
1285 Figure 7: Comparison of the top 50 thinking pattern frequencies for models trained with and without  
1286 the injection of human-like thinking patterns during data synthesis. The model with injection (left)  
1287 shows a more diverse and balanced distribution of patterns, while the model without (right) relies  
1288 heavily on a few formulaic phrases.  
1289

1290 As shown in Figure 7, the difference is stark. The model trained *with* thinking pattern injection  
1291 exhibits a more diverse and evenly distributed use of thinking patterns. Tokens indicating reflection  
1292 and self-correction, such as ‘let me think’, ‘maybe’, ‘hmm’, and ‘wait’, are prominent. This suggests  
1293 a more flexible, human-like reasoning process with cognitive exploration. In contrast, the model  
1294 trained *without* this injection relies on a small set of highly frequent phrases like ‘next’, ‘first’, and  
1295 ‘goal is to’. The frequency distribution is highly skewed, indicating a more rigid and formulaic  
reasoning process.

1296 This analysis confirms that the proposed context engineering techniques encourages the model to  
 1297 adopt a more nuanced and reflective approach to problem-solving, which, as shown in the ablation  
 1298 studies, is particularly beneficial for creative and complex tasks.  
 1299

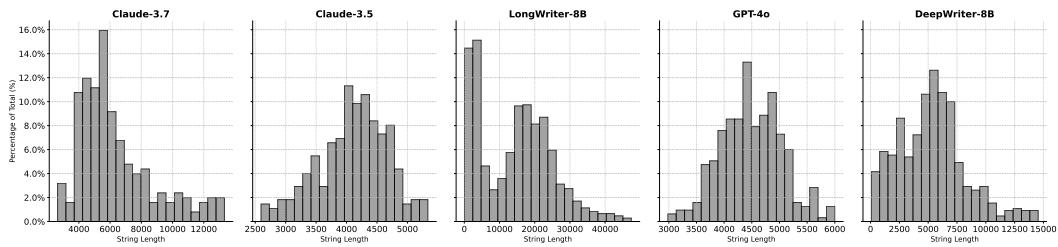
### 1300 C BEHAVIORAL ANALYSIS

1302 We conducted preliminary analysis on the model’s behaviors. Figure 8 shows the token length  
 1303 distribution of DeepWriter-8B responses on LongBench-Write.  
 1304



1314 Figure 8: Token Length distribution of Thinking and Answer part of DeepWriter-8B.  
 1315

1316 We also compare the response string length distribution across leading models in Figure 9. While  
 1317 DeepWriter achieves superior performance competitive with frontier models, it does not introduce  
 1318 excessive response length like LongWriter. The average response length is around 5000 tokens,  
 1319 comparable with frontier models like GPT-4o and Claude-3.7.  
 1320



1330 Figure 9: Response String Length Distribution across different models.  
 1331

### 1332 D CASE STUDIES

1335 Due to formatting issues with latex, we put a few case studies in the supplementary materials.  
 1336 We manually review the cases where DeepWriter can outperforms other models, confirming the  
 1337 argument that DeepWriter can achieve better depth, logical consistency and factual grounding via  
 1338 deep reasoning traces. We also analyze the error patterns of the generations from DeepWriter. We  
 1339 find that DeepWriter often gets lower score due to domain knowledge gap, implying the benefits of  
 1340 training with more diverse corpus of topics and domains.  
 1341

### 1342 E EXTENDED RELATED WORK

1344 **Deep Reasoning and Test-Time Computation.** The paradigm of “deep reasoning” (or Long CoTs)  
 1345 aims to move beyond rapid, surface-level inference by leveraging increased computational investment  
 1346 at test time. Advanced models from organizations like Google (Team et al., 2023), DeepSeek AI  
 1347 (Guo et al., 2025), and OpenAI (Jaech et al., 2024) have demonstrated the effectiveness of this  
 1348 test-time scaling (Team, 2025; Muenighoff et al., 2025; Fu et al., 2025). This approach gained  
 1349 prominence with methods like Chain-of-Thought (CoT) prompting (Wei et al., 2022), which elicits  
 intermediate reasoning steps to guide a model toward more accurate solutions. Building on this,

more sophisticated strategies have emerged, such as Tree-of-Thought (ToT) (Yao et al., 2023), which explores a tree of possible reasoning paths, and various self-correction or self-refinement (Madaan et al., 2023; Kumar et al., 2024; Zelikman et al., 2022; 2024) mechanisms that iteratively improve an initial response. While these approaches have yielded remarkable performance gains in verifiable domains like mathematics and programming, their application to open-ended, creative tasks remains largely unexplored due to the absence of a singular ground truth for verification. REER addresses this gap by developing a method to instill this deliberate, structured thinking capability for non-verifiable creative domains.

**Paradigms for Instilling Reasoning.** Beyond prompting techniques at inference time, two dominant paradigms exist for integrating advanced reasoning capabilities directly into a model’s parameters: reinforcement learning and instruction distillation.

Reinforcement Learning (RL) has been instrumental in aligning LLMs with human preferences (RLHF) and improving performance on tasks with clear reward signals (Ouyang et al., 2022; Guo et al., 2025; Team et al., 2025; Wang et al., 2025c). In verifiable domains, a correct outcome provides a straightforward positive reward, effectively guiding the model’s search through a vast solution space (Shao et al., 2024; Wang et al., 2025a;b; Su et al., 2025). However, this reliance on verifiability presents a formidable barrier when applied to open-ended generation (Ouyang et al., 2022; Lu, 2025). Crafting a reward model that can reliably approximate nuanced and subjective qualities like originality or emotional resonance is an immense challenge in itself (Ouyang et al., 2022; Zhang et al., 2024). Furthermore, the subsequent RL process is often computationally burdensome and sample-inefficient (Shao et al., 2024; Gulcehre et al., 2023; Wang et al., 2023). Recently VeriFree (Zhou et al., 2025) extends verification-based reward to likelihood-based reward for reinforcement learning on verifiable domains. Likewise, REverse-Engineered Reasoning (REER) shares the principle of using a proxy to judge the reasoning quality. However, the motivation is fundamentally different – we focus on recovering human-like deep reasoning from known-good outputs for the broader open-ended generation problems.

Instruction Distillation offers an alternative, wherein reasoning traces are generated by a powerful “teacher” model (e.g., GPT-4 (Achiam et al., 2023)) and used as training data for a smaller “student” model. While effective, this approach is constrained by two fundamental limitations. First, it is often hampered by the prohibitive cost of querying state-of-the-art proprietary models at scale (Guha et al., 2025; Toshniwal et al., 2024). Second, and more fundamentally, distillation is capped by the teacher’s abilities—a student model cannot learn a capacity that the teacher does not already possess (Toshniwal et al., 2024). This limitation is exacerbated by the general scarcity of high-quality, open-source instruction data tailored for advanced creative tasks (Bai et al., 2024).

To overcome these data bottlenecks, researchers have increasingly turned to synthetic data generation. Most approaches use a powerful LLM to generate new query-response pairs, often to augment existing datasets or bootstrap capabilities in new domains (Wang et al., 2022; Zelikman et al., 2022; 2024; Gu et al., 2025; Yang et al., 2023; Han et al., 2025). These methods aim to build a solution “forwards” for a given query through data synthesis. Our central innovation is to “reverse-engineer” reasoning – synthesize deep reasoning “backwards” from a known good outcome such as human-written solutions.

**Writing Datasets, Models and Benchmarks** Prior work has explored both synthetic data pipelines and RL in AI writing. For instance, Weaver (Wang et al., 2024) proposed instruction back-translation, LongWriter (Bai et al., 2024) proposed an agentic data pipeline to synthesize long-form writing outputs and introduced the LongBench-Write benchmark. In contrast, Writing-Zero (Lu, 2025) employed an RL approach, training a reward model on private datasets, but its training data remains unreleased. DeepWriter, to our knowledge, is the first to *instill deep reasoning for open-ended generation* using a scalable, open synthetic data approach.

Evaluation in this domain relies on recently developed benchmarks. HelloBench (Que et al., 2024) proposes a diverse collection of “in-the-wild” tasks from real user queries to gauge practical applicability. Meanwhile, WritingBench (Wu et al., 2025) measures domain-specific proficiency and the ability to adhere to complex, multi-dimensional constraints across six professional domains.

1404 **F LLM USAGE**  
14051406 We acknowledge the use of large language models (LLMs) during the preparation of this manuscript.  
1407 The application of these tools was strictly limited to an assistive role for improving the quality  
1408 of the writing. Specifically, LLMs were utilized to enhance clarity, refine sentence structure, and  
1409 ensure a smooth and logical flow of arguments throughout the paper. The core ideas, methodology,  
1410 experimental results, and all intellectual contributions presented herein are entirely the work of the  
1411 authors. LLMs were not used to generate any of the substantive research content or analysis.  
1412  
1413  
1414  
1415  
1416  
1417  
1418  
1419  
1420  
1421  
1422  
1423  
1424  
1425  
1426  
1427  
1428  
1429  
1430  
1431  
1432  
1433  
1434  
1435  
1436  
1437  
1438  
1439  
1440  
1441  
1442  
1443  
1444  
1445  
1446  
1447  
1448  
1449  
1450  
1451  
1452  
1453  
1454  
1455  
1456  
1457