

FOR-PROMPTING: FROM OBJECTION TO REVISION VIA AN ASYMMETRIC PROMPTING PROTOCOL

Anonymous authors

Paper under double-blind review

ABSTRACT

Reasoning protocols such as Chain of Thought (CoT) and Tree of Thought (ToT) organize internal deliberation but lack an explicit mechanism for external questioning that elicits self-revision. We present FOR-Prompting (From Objection to Revision Prompting), an asymmetric protocol where a Defender proposes an answer, a Debater (Questioner) raises question-style objections with no direct fixes, and a Host optionally synthesizes the final output. Across GSM8K, FOR-Prompting matches the accuracy of CoT and consistently improves over single-prompting when evaluated under identical model backbones. On small-scale open-source models (e.g., LLaMA-3.2-1B), FOR-Prompting yields substantial gains over direct prompting and performs comparably to lightweight reasoning baselines, highlighting its promise for low-resource and on-device settings. Cross-model role-swapping further shows that performance is primarily determined by the Defender, enabling small models to act effectively as Questioners. Beyond structured math tasks, FOR-Prompting supports refinement in open-ended and multi-stage tasks: qualitative analysis shows improved exploration, coverage, and specificity, and a blind human preference study found that participants preferred FOR-Prompting outputs over strong LLM baselines in an itinerary-planning scenario. The protocol is model-agnostic and operates purely through role-structured prompting, requiring no training, access to model internals, or symmetrically strong agents. FOR-Prompting therefore enables scalable study of objection-driven reasoning and offers a practical mechanism for automated iterative refinement across both hosted and local LLMs.

1 INTRODUCTION

Large language models (LLMs) increasingly solve complex tasks by following prompting strategies that structure their intermediate reasoning. Within this line of work, Chain-of-Thought (CoT) (Wei et al., 2022) encourages linear stepwise explication, Tree-of-Thought (ToT) (Yao et al., 2023a) organizes branched exploration, Self-Consistency (Wang et al., 2023b) aggregates multiple independent chains, and self critique aims to repair an initial draft through reflective prompts. More recently, graph structured prompting such as Graph-of-Thoughts (GoT) (Besta et al., 2024) extends beyond trees by exploring and scoring a broader space of partial ideas. Taken together, these single agent strategies have advanced transparency and control. Yet these protocols operate primarily within a single reasoner, lacking an explicit mechanism for external questioning that elicits self-revision without supplying solutions. In many settings, such as mathematical word problems, planning, analysis, or explanation, what improves an answer is not another answer, but a well-posed question that surfaces gaps, unstated assumptions, or overlooked constraints.

A parallel literature explores multi agent prompting. Debate style systems pit agents against each other (Du et al., 2024; Park & Seo, 2025), committee and ensemble schemes solicit multiple answers and then vote or justify, reviewer and reviser workflows attach a critic that proposes edits, tutor or coach patterns offer hints or worked substeps, and cross examination has agents interrogate one another before an aggregator selects a final output. These designs introduce valuable external pressure (i.e., external questioning applied by a separate model rather than self-reflection), yet they typically inject external thinking and sometimes competing solutions into the trace. As a result, error detection is often confounded with answer replacement, provenance becomes diluted because multiple authors shape the content, and evaluation grows complicated because improvements may stem

054 from the external agent rather than from the solver’s revised reasoning. Iteration depth is not always
055 controlled and outcome level evaluation is often mixed with process supervision, which makes it
056 hard to isolate the effect of questioning itself. These gaps motivate an alternative that keeps the
057 benefits of external pressure while avoiding external solution content.

058 We draw on a widely used human-in-the-loop (HITL) practice (Wu et al., 2022). Human reviewers
059 rarely fix an LLM’s response directly. Instead, they improve the quality of the LLM’s responses
060 or request more detailed explanations by posing questions. These questions reveal missing as-
061 sumptions, inconsistent steps, ambiguous concepts, or infeasible constraints. We introduce **FOR-**
062 **Prompting**, short for From Objection to Revision Prompting. A Defender proposes an answer. A
063 Questioner then asks question-style challenges rather than supplying fixes, with the aim of inducing
064 the Defender’s rethinking and revision. Here objection is used in a broad questioning sense that
065 includes clarifying questions, constraint checks, and counterexample or counterfactual probes. By
066 separating roles so that the Questioner asks and the Defender reasons and revises, FOR-Prompting
067 preserves a single accountable line of thought and turns questioning into a first class mechanism
068 for revision. Conceptually, this can be viewed as an automated HITL-like question loop, where
069 the human-like interrogative pressure is delivered by an agent, [mimicking the questioning behavior
070 humans typically use when iteratively refining LLM outputs rather than involving real human users.](#)

071 Methodologically the protocol is open ended by design. It can iterate without a fixed upper limit or
072 stop under a user specified rule such as a utility threshold, a convergence signal, a time or cost bud-
073 get, or a planner defined milestone. This flexibility is useful for brainstorming and refinement heavy
074 work including multi step planning, scenario exploration, and requirement elicitation, where addi-
075 tional cycles of questioning often surface new constraints or alternatives. For controlled evaluation
076 in this paper we instantiate an external only setting. A Questioner poses clarifying and adversarial
077 questions and a Defender revises the solution in response. No self reflection step is inserted. The
078 Questioner never supplies direct fixes and the Defender remains the sole author of revisions. We cap
079 the number of rounds with a small constant that is kept fixed within each experiment and reported
080 with the results.

081 In our experiments, we conducted [four](#) case studies to validate the effectiveness of FOR-Prompting:
082 (1) Benchmark evaluation on GSM8K (Cobbe et al., 2021): using GPT-4o, we compared FOR-
083 Prompting against single-prompt, CoT, [Self-Ask, and Self-Consistency baselines](#) under identical
084 settings. FOR-Prompting matched the performance of CoT and [Self-Ask, and remained close to](#)
085 [Self-Consistency, while consistently outperforming the single-prompt baseline.](#) These results con-
086 firm that external question-driven refinement can be as effective as established reasoning scaffolds.
087 (2) Small-scale model evaluation: on LLaMA-3.2:1B, we compared single-prompt, CoT, and two
088 FOR-Prompting variants (with and without Host aggregation). Both FOR-Prompting variants sub-
089 stantially outperformed the single-prompt baseline and achieved accuracy comparable to CoT. Host
090 aggregation provided limited benefit on small models due to summarization errors, highlighting that
091 the gain primarily comes from the external-questioning mechanism. (3) Cross-model role swapping:
092 we show that using a small model as the Questioner and a stronger model as the Defender preserves
093 high performance while significantly reducing large-model token usage. This asymmetric structure
094 differs from self-ask-style methods and provides a cost-efficient strategy for hybrid pipelines. (4)
095 Open-ended tasks: FOR-Prompting produced more complete and actionable plans than baselines on
096 complex itinerary-generation tasks. [A blind human preference study with 77 participants shows that
097 74% preferred the FOR-Prompting output over those from two strong frontier LLM systems.](#)

098 In summary, we present FOR-Prompting as an automated form of HITL-like questioning that pre-
099 serves a single authorial line of thought while benefiting from external pressure. By treating ques-
100 tions as the sole external input and without requiring direct fixes, FOR-Prompting enables controlled
101 study of questioning itself rather than a mixture of questioning and external reasoning, thereby im-
102 proving interpretability, accountability, and practical utility.

103 2 CONTRIBUTIONS

104 This work makes three primary contributions. (1) Conceptual novelty: We introduce FOR-
105 Prompting, the first prompting protocol that formalizes questioning-rather than answer substitution-
106 as the exclusive form of external intervention. By separating questioning from revision, FOR-
107 Prompting preserves a single accountable reasoning chain while leveraging external pressure to elicit
self-revision. (2) Protocol design: We instantiate FOR-Prompting as a lightweight, role-based inter-

108 action loop between a Defender and a Questioner, enforcing that objections take the form of ques-
109 tions only, which enables systematic study of questioning as a mechanism for improving reasoning.
110 (3) Empirical validation: We validate FOR-Prompting on both commercial and open-source LLMs.
111 On GSM8K, FOR-Prompting achieves higher accuracy than the baseline and matches the same level
112 of the performance of CoT. When tested on small-scale open-source models, FOR-Prompting yields
113 substantial accuracy gains over the baseline. In addition, we provide illustrative examples showing
114 that FOR-Prompting can effectively correct errors in LLM outputs and refine or expand open-ended
115 tasks, serving as a viable alternative to direct human supervision or intervention. [In open-ended
116 itinerary-planning scenarios, FOR-Prompting consistently generated more complete, realistic, and
117 actionable plans than strong baseline models, as also reflected in human preference evaluations.](#)

118 3 RELATED WORK 119

120 **Reasoning and planning in single-agent prompting.** LLMs have popularized explicit interme-
121 diate reasoning via CoT (Wei et al., 2022), which improves performance on structured tasks, while
122 Self-Consistency (Wang et al., 2023b) aggregates diverse reasoning paths to reduce brittle single-
123 trace errors. Beyond linear chains, ToT (Yao et al., 2023a) explores alternative branches with
124 look-ahead and backtracking, and GoT (Besta et al., 2024) generalizes this idea to non-linear struc-
125 tures. Complementary scaffolds strengthen problem decomposition and execution. Planning-based
126 prompting includes *least-to-most* (Zhou et al., 2023) and plan-then-solve (Wang et al., 2023a). Tool-
127 use methods such as PAL (Gao et al., 2023) and PoT (Chen et al., 2023) delegate subproblems to
128 programs. ReAct (Yao et al., 2023b) interleaves reasoning with external actions to ground steps in
129 evidence. [Self-ask Press et al. \(2023\) promotes reasoning by having the model ask itself follow-up
130 questions and then answer them within a single-model loop.](#) These approaches increase transparency
131 and controllability, yet they primarily operate within a single reasoner and therefore provide only
132 limited avenues for externally driven improvement through questioning.

133 **Self-reflection and verification mechanisms.** A parallel line of work adds self-assessment and
134 correction. Self-Refine iteratively revises outputs using model-generated feedback (Madaan et al.,
135 2023), and Reflexion maintains episodic reflective memory to guide adaptation (Shinn et al., 2023).
136 Constitutional AI encodes explicit normative constraints to regulate outputs (Bai et al., 2022). Chain-
137 of-Verification validates intermediate claims before targeted revision (Cobbe et al., 2021; Dhuli-
138 awala et al., 2024). SelfCheckGPT further automates verification through sampling and consistency
139 checking (Manakul et al., 2023). Recent work also documents limits of reflection, including an
140 early-stop effect where gains plateau after a few iterations. DORA sustains progress with dynamic
141 optimization prompts (Li et al., 2025), and Agent-Pro uses policy-level reflection and optimization
142 to evolve agent behaviors over time (Zhang et al., 2024a). Beyond single-agent reflection, reflective
143 multi-agent collaboration distributes critique and self-reflection across multiple agents to improve
144 robustness and adaptability on complex tasks (Bo et al., 2025). These studies underscore the value
145 of feedback and verification, while also revealing that internal reflection alone may be insufficient
146 when external perspectives are needed to challenge early commitments.

147 **Multi-agent prompting for reasoning and collaboration.** Multi-agent systems (MAS) (Dorri
148 et al., 2018; He et al., 2025; Sun et al., 2025; Maldonado et al., 2024) frame reasoning and collabora-
149 tion as interactions among specialized agents rather than a single model acting alone. By distrib-
150 uting roles, perspectives, and responsibilities, MAS approaches target robustness, explainability,
151 and creativity (Naik et al., 2025; Epperson et al., 2025; Zhang et al., 2024b). Within this paradigm,
152 debate-style frameworks position agents with opposing views to contest one another, which can
153 surface errors through adversarial questioning and improve factual accuracy and robustness (Du
154 et al., 2024). Park & Seo (2025) report that multi-agent debate aligns more closely with human
155 evaluators in educational assessment than single-agent baselines. Purely agent-agent debate such
156 as MAD shows reductions in hallucination and gains in accuracy (Du et al., 2024), while reflective
157 multi-agent collaboration distributes critique and synthesis across agents to support long-horizon
158 tasks (Bo et al., 2025). At the same time, recent findings warn that multi-agent settings can suffer
159 social influence effects that reduce robustness under peer pressure (Song et al., 2025).

160 **Role specialization in multi-agent architectures.** Beyond symmetric debate, role-based designs
161 coordinate complementary skills to tackle complex tasks. CAMEL demonstrates that role-playing
agents can sustain goal-directed exchanges with minimal human steering (Li et al., 2023). Auto-

Gen provides an extensible infrastructure for assistant–critic and team-based interactions (Wu et al., 2024). Debate has also been adapted to encourage divergent reasoning paths rather than quick consensus (Liang et al., 2024). In software engineering, ChatDev and MetaGPT orchestrate requirements, design, coding, and testing through division of labor and iterative review, which scales solution quality in practice (Qian et al., 2024; Hong et al., 2024). Despite these advances, many systems converge episodically through voting or consensus, which relaxes sustained pressure to interrogate a reasoning chain, and many prioritize factual question answering over open-ended planning and creative refinement, which limits their utility in tasks that demand progressive elaboration.

Dialogue-based deliberation and automated reflection. Multi-agent dialogue extends beyond adversarial debate to simulate deliberation, reflection, and negotiation. Zhang et al. (2024b) design a generative MAS that helps users burst filter bubbles by engaging diverse perspectives through gamified interactions. Ashkinaze et al. (2025) guide LLMs with simulated social ensembles that follow procedures from deliberative democracy. Divekar et al. (2020) develop a human–multi-agent negotiation platform in immersive environments, highlighting opportunities and challenges of multiparty dialogue. Dialogue frameworks also operationalize draft–critique–revise loops. AutoGen’s assistant–critic pattern (Wu et al., 2024), debate with adjudication (Liang et al., 2024), and LLM-as-judge protocols for scalable evaluation (Zheng et al., 2023) are representative. However, most of these systems target single-answer accuracy under bounded contexts and often embed hints, templates, or edits, which makes it difficult to isolate the effect of questioning itself.

Positioning. The studies above either keep all reasoning inside one agent or allow external collaborators to introduce hints, edits, or competing solutions. In contrast, our setting focuses on externally driven improvement through questions alone. We automate a HITL-like practice in which targeted questions expose gaps, assumptions, and constraints, while the answering agent remains solely responsible for the revision. [Unlike Self-Ask, which requires a single model to generate and answer its own follow-ups, our method assigns answering and questioning to different agents. This prevents self-confirmation bias, introduces genuinely external interrogative pressure, and preserves a single accountable reasoning trajectory without entangling solution generation with self-evaluation. Also, unlike multi-agent debate, the Questioner in our framework does not generate competing solutions or attempt to win an argument; it contributes only question-style challenges, making the interaction asymmetric and refinement-oriented rather than adversarial.](#) By treating questions as the only currency of intervention, our approach preserves a single accountable line of thought, avoids confounds from externally supplied reasoning, and complements debate, review, and tutoring styles with a mechanism for persistent and structured questioning that supports open-ended planning and iterative refinement.

4 METHODOLOGY

We introduce FOR-Prompting, a multi-agent, questioning-driven prompting protocol that structures inference as an iterative question/challenge-revision loop. Given a question Q and a round budget N , the protocol coordinates three roles: a *Defender* that proposes and iteratively revises solutions, a *Debater* that injects external objections/questions without revealing solutions, and an optional *Host* that synthesizes the final answer from the full interaction history.

The process unfolds as follows. The Defender first produces an initial answer A_0 to Q . The Debater then analyzes A_0 and asks a targeted question Q_1 , prompting the Defender to revise its solution to A_1 conditioned on Q and Q_1 . Repeating this cycle for N rounds yields a trajectory

$$\{A_0, Q_1, A_1, \dots, Q_N, A_N\}.$$

Finally, the Host aggregates the dialogue trajectory into the final answer A^* .

Formally, for $r = 1, \dots, N$,

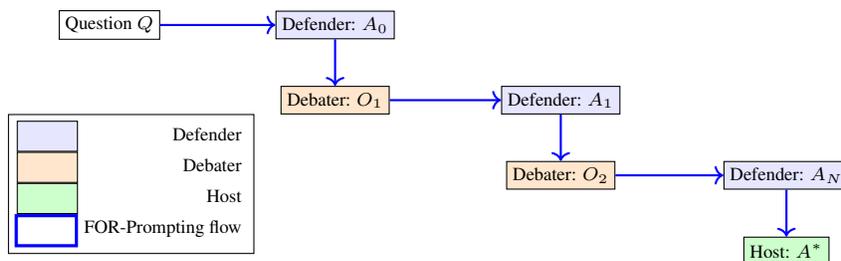
$$\begin{aligned} A_0 &= \text{Defender}(Q), \\ O_r &= \text{Debater}(A_{r-1}), \\ A_r &= \text{Defender}(Q, O_1, \dots, O_r), \\ A^* &= \text{Host}(Q, \{A_0, O_1, A_1, \dots, O_N, A_N\}). \end{aligned}$$

216 **Role Responsibilities.** The three roles follow
 217 strict divisions of responsibility: *Defender* pro-
 218 poses and revises solutions under questioning;
 219 *Debater* issues targeted questions only (no al-
 220 ternative solutions), focusing on logical gaps,
 221 hidden assumptions, counterexamples, and ro-
 222 bustness; *Host* integrates the multi-round trace,
 223 ensures resolution of concerns, and produces
 224 the final answer A^* .

225 These responsibilities are enforced through role-specific prompts that constrain outputs to questions,
 226 revisions, or synthesis, while leaving surface form flexible. The aim is not to prescribe a fixed
 227 template but to enforce a mechanism of externally driven questioning and revision.

229 **Protocol Representation.** To ensure transparency and analyzability, each run records a structured
 230 dialogue history with role tags and message types, which guides Host synthesis and enables subse-
 231 quent process-level analysis.

232 Algorithm 1 presents an abstract pseudocode version of FOR-Prompting, capturing the essential dy-
 233 namics of questioning and revision while omitting implementation-specific details. Figure 1 com-
 234 plements this abstraction with a flowchart illustrating the iterative structure.



245 Figure 1: FOR-Prompting protocol flow (external questions only). The Debater raises questions, the
 246 Defender revises answers, and the Host synthesizes the final output. *Host could be optional.

247 5 EXPERIMENTAL SETUP AND EVALUATION

248 5.1 TESTING ON GSM8K

249 5.1.1 CASE STUDY 1: LARGE LANGUAGE MODEL (GPT-4o) ON GSM8K

250 **Datasets.** We first evaluate on GSM8K (Cobbe et al., 2021), a elementary and middle school math
 251 word problems benchmark. We evaluate on the public test set and, for local-model experiments,
 252 report results on a uniformly sampled subset of 500 test items with a fixed seed. Each method
 253 produces a free-form solution; no prompt length cap is imposed. All runs persist the full dialogue
 254 for auditability, while metrics are computed from the final solution produced for each item.

255 **Accuracy evaluation.** For each model response, we
 256 deterministically extract the final numeric prediction
 257 and compare it against the ground-truth solution. GPT-
 258 4.1 (temperature 0) is used only as a preliminary ver-
 259 ifier to assist rapid inspection, but the final correctness
 260 labels are manually checked to eliminate judge-model
 261 errors. All extraction, verification, and scoring pro-
 262 cedures are applied uniformly across the 500 sampled
 263 GSM8K items.

264 **APIs and determinism.** All methods are executed
 265 through the same programmatic framework with temperature 0 for reproducibility. Unless other-
 266 wise noted, all model-generated responses across experiments use OpenAI’s `gpt-4o` through the
 267 official OpenAI API. This ensures that differences in performance arise from the prompting pro-
 268 tocols rather than backend model variations. All LLM-as-a-judge evaluations are conducted using
 269 OpenAI’s GPT-4.1.

Algorithm 1 FOR-Prompting Protocol (ab-
 stracted pseudocode)

Require: Question Q , round budget N
 1: $A_0 \leftarrow \text{Defender}(Q)$
 2: **for** $r = 1$ to N **do**
 3: $O_r \leftarrow \text{Debater}(A_{r-1})$
 4: $A_r \leftarrow \text{Defender}(Q, O_1, \dots, O_r)$
 5: **end for**
 6: $A^* \leftarrow \text{Host}(Q, \{A_0, O_1, A_1, \dots, O_N, A_N\})$
 7: **return** A^*

Table 1: GSM8K accuracy across prompt-
 ing methods using `gpt-4o`.

Method	Accuracy
Single-prompt	0.92
Chain-of-Thought (CoT)	0.94
Self-ask	0.94
Self-consistency	0.95
FOR-Prompting (Host=True)	0.94
FOR-Prompting (Host=False)	0.94

Results. Table 1 reports GSM8K accuracy across prompting methods under a unified `gpt-4o` backend. The single-prompt baseline achieves an accuracy of 0.92. All structured prompting methods yield modest improvements over this baseline: CoT and self-ask both reach 0.94, self-consistency achieves 0.95, and FOR-Prompting attains 0.94 in both its Host and no-Host variants. Overall, these results indicate that structured reasoning generally provides a small but consistent benefit over single-shot prompting, with all advanced prompting strategies exhibiting comparable performance.

5.1.2 CASE STUDY 2: SMALL-SCALE MODEL (LLAMA-3.2:1B) ON GSM8K

Setting. We evaluate an open-source 1B-parameter model (LLaMA-3.2:1B¹²) on the 1,319-item GSM8K test set to examine whether lightweight interaction structures can compensate for limited parametric capacity. We compare four prompting strategies: (i) **Single-prompt**: a direct answer without explicit reasoning. (ii) **CoT**: prompting the model to generate step-by-step reasoning before an answer. (iii) **FOR-Prompting (Host-True)**: the standard protocol in which a Host synthesizes the final answer. (iv) **FOR-Prompting (Host-False)**: the same protocol without Host synthesis; the Defender’s final revision is returned directly.

In both FOR-Prompting variants, the Debater provides only question-style challenges and never supplies solutions; the Defender re-answers the original question each round to avoid objective drift. Temperature and decoding settings are held constant across all methods.

Results. Table 2 reports accuracy on the GSM8K. The **Single-prompt** baseline yields only **7%** accuracy, confirming the difficulty of multi-step reasoning for a 1B-parameter model. Adding **CoT** substantially improves performance to **23%**, showing that even very small models benefit from explicit reasoning scaffolds.

FOR-Prompting achieves competitive performance relative to CoT. The **Host-False** variant reaches **23%**, matching CoT, while the **Host-True** version obtains **19%**. Importantly, both FOR-Prompting variants substantially outperform the single-prompt baseline, demonstrating that the external-questioning mechanism remains effective even under severe model capacity constraints. In particular, despite the model’s limited size, a single round of structured questioning enables the model to surface missing constraints and correct earlier reasoning gaps, capabilities that the single-prompt baseline rarely exhibits.

Although the Host role is primarily intended to provide a clearer final presentation rather than improve reasoning quality, the results indicate that its impact on small-model accuracy is limited. For a 1B model, the additional synthesis step can occasionally introduce errors, such as dropping constraints or mis-merging intermediate reasoning. Moreover, the extra consolidation stage can amplify the model’s inherent noise, since each additional generation step compounds earlier uncertainties. This explains why **Host-False** slightly outperforms **Host-True**: lightweight models tend to be more reliable when returning their final reasoning trace directly rather than performing an extra consolidation step.

Overall, while the absolute accuracy remains far below that of large-scale models, the consistent gain over the single-prompt baseline, more than doubling accuracy, shows that structured prompting strategies, including both CoT and FOR-Prompting, meaningfully amplify reasoning ability in small models.

5.1.3 CROSS-MODEL ROLE SWAPPING EXPERIMENTS

To examine whether FOR-Prompting remains robust when different models assume the Defender and Debater roles, we conducted a cross-model role swapping evaluation using GPT-4o and LLaMA-3.2:1B.

This setting tests the asymmetric nature of the protocol by observing how performance changes when model capacities are permuted across roles.

Table 2: Accuracy of LLaMA-3.2:1B on GSM8K under different strategies.

Method	Accuracy
Single-prompt	0.07
Chain-of-Thought (CoT)	0.23
FOR-Prompting (Host-True)	0.19
FOR-Prompting (Host-False)	0.23

Figure 2: Cross-Model Role Swapping Results

Defender Model	Debater Model	Method	Accuracy
GPT-4o	LLaMA-3.2:1B	FOR-Prompting	0.93
LLaMA-3.2:1B	GPT-4o	FOR-Prompting	0.21

¹<https://ollama.com/library/llama3.2:1b>

²<https://ai.meta.com/blog/llama-3-2-connect-2024-vision-edge-mobile-devices/>

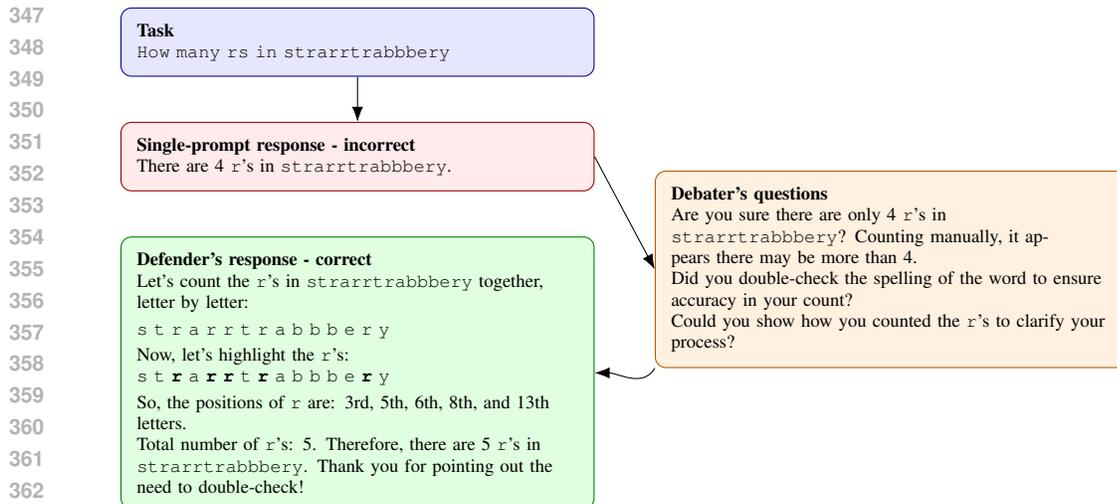
324 Table 2 summarizes the results. When GPT-4o served as the Defender and LLaMA-3.2:1B acted as
 325 the Debater, FOR-Prompting achieved an accuracy of 0.932. This suggests that the Debater’s role,
 326 primarily generating external interrogative pressure, does not require high model capacity to be ef-
 327 fective. In contrast, when the roles were reversed, and LLaMA-3.2:1B served as the Defender while
 328 GPT-4o served as the Debater, accuracy dropped sharply to 0.21. This indicates that the Defender
 329 role is substantially more sensitive to model expressiveness, since it must propose candidate answers
 330 and integrate external challenges into the final reasoning.

331 These results reinforce our central claim that FOR-Prompting is an *asymmetric prompting protocol*.
 332 The protocol does not require symmetry in model capability: a small model can serve as an effective
 333 Debater, but strong performance depends critically on the Defender being sufficiently capable.
 334

335 5.2 CASE STUDY 3: CORRECTING MISTAKES THROUGH EXTERNAL QUESTIONS

336 In this case study, we use a currently popular commercial model and apply the FOR-Prompting
 337 method to address problems where these models would otherwise produce incorrect answers. The
 338 results show that FOR-Prompting can correct erroneous responses without human supervision or
 339 external tool assistance.

340 Specifically, we present a tricky question, “How many rs in strarrtrabbbery” (this is the only prompt
 341 we input to LLM), which is often incorrectly answered by LLMs because of their tendency to mis-
 342 count characters in strings due to tokenization and attention limitations. We used GPT-4.1 to answer
 343 this question. Under the single-prompt setting, we obtained an incorrect response (“There are 4 ‘r’s
 344 in ‘strarrtrabbbery.’”). However, within the FOR-Prompting framework, the debater raised addition-
 345 al questions that guided the defender back on track, resulting in the correct answer (“Therefore,
 346 there are 5 ‘r’s in ‘strarrtrabbbery.’”), as shown in Figure 3.



364 Figure 3: A detailed example of FOR-Prompting with initial task, incorrect single-prompt answer,
 365 debater’s questions, and corrected defender response.

367 5.3 CASE STUDY 4: REFINING OPEN-ENDED TASK THROUGH EXTERNAL QUESTIONS

368 We compared three approaches, ChatGPT 5 Thinking (Extended thinking) ³, GPT-4.1 (single-
 369 prompt), and FOR-Prompting (GPT-4.1-based), on the open-ended task “Create a 5-day itinerary
 370 for a trip to Rio de Janeiro in April.” This kind of task, an open-ended itinerary-planning and
 371 content-generation task, involves integrating multi-objective constraints, step-by-step reasoning, and
 372 fact-checking. Therefore, evaluating models on such tasks enables a comprehensive assessment of
 373 their overall capabilities and practical value for real-world decision support and alignment with user
 374 preferences.

375 Qualitatively, FOR-Prompting yielded more robust, field-ready itineraries than the other two ap-
 376 proaches. For example, when the Debater asked “what to do if tickets to Christ the Redeemer were
 377

³<https://openai.com/index/introducing-gpt-5/>

378 *unavailable*”, the revision added clear alternatives and timing guidance; when the Debater ques-
 379 tioned “*whether combining Sugarloaf and Lapa in one day was realistic*”, the plan incorporated
 380 time buffers and fallbacks. Table 3 summarizes the relative breadth of practical information across
 381 the three methods. From this we can see that FOR-Prompting considers far more factors in its plan
 382 than ChatGPT 5 and the single-prompt baseline. Although we can prompt an LLM with the relevant
 383 factors and have it add the missing content-something models and tools, especially a stronger model
 384 like ChatGPT 5, can readily do to supplement, expand, and refine-the process typically relies on
 385 human expertise (manual prompting) rather than being fully automated as in FOR-Prompting. The
 386 complete FOR-Prompting content appears in Appendix A.3.

Factor	ChatGPT 5 (Extended thinking)	GPT-4.1 (Single-prompt)	FOR-Prompting (GPT-4.1-based)
Weather	✓	✓	✓
Packing/Attire	✓	✓	✓
Reservations/Tickets	✓	-	✓
Transportation	✓	-	✓
Accessibility	-	-	✓
Safety	✓	✓	✓
Rainy-day alternatives	✓	-	✓
Alternatives/Branching	✓	-	✓
Nightlife/Music	✓	✓	✓
Food/Restaurants	✓	✓	✓
Markets/Shopping	-	✓	✓
Language/Comms	-	-	✓
Offline/Navigation	-	-	✓
Luggage Storage	-	-	✓
Emergency Info	-	-	✓
Holidays/Hours	-	-	✓
Photography/Viewpoints	✓	✓	✓
Signature Experiences	✓	-	✓
Family-friendly	-	-	✓
Health/Water	-	✓	✓

Table 3: Practical Info & Coverage Comparison

Human preference evaluation. To examine whether the benefits of iterative external questioning generalize beyond a single prompt, we conducted a separate human preference study using an additional open-ended planning task: “*Create a 7-day travel itinerary for Barcelona.*” This task belongs to the same class of multi-constraint itinerary-planning problems as the Rio example but offers a different geographic and structural context, enabling a broader test of generality.

We recruited 77 participants via social media platforms. Each participant completed a single-choice question after reviewing three anonymized outputs produced by: (1) FOR-Prompting (GPT-4.1-based), (2) GPT-4.1 (single-prompt), and (3) a widely used frontier LLM system. Participants were asked: “Which of the following outputs do you prefer?” Results show that **74% (57/77)** of participants preferred the FOR-Prompting itinerary. Participants highlighted its greater completeness, more realistic daily pacing, clearer contingency planning, and stronger practical utility. These findings indicate that iterative external questioning enhances open-ended planning performance not only in the Rio case study but also in a distinct and independently evaluated scenario, reinforcing the broader applicability of FOR-Prompting for real-world decision support.

5.4 CASE STUDY 5: MULTI-STAGE TASK EXTENSION OF FOR-PROMPTING

Given the growing demand for multi-stage agentic tasks, we added a small pilot study to illustrate the applicability of FOR-Prompting in dynamic, multi-stage settings. As shown in Figure 4, after completing Task 1 we introduced new conditions or new knowledge. These additions reflect realistic sources of task updates, such as temporal changes, real-world constraints, emerging requirements, or subsequent tasks that depend on prior outputs.

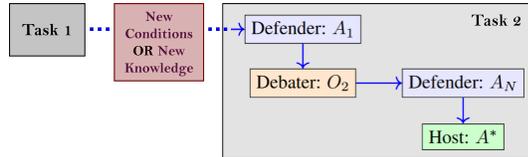


Figure 4: Multi-stage extension of FOR-Prompting. New conditions or new knowledge are incorporated after Task 1, enabling iterative refinement toward the Task 2 solution.

We reused the travel-planning scenario from Case Study 4 and treated its final response as the completed output of Task 1. We then assumed that on Day 4 of the trip, additional travelers would join, forming a six-person group for that day’s activities. This new group-travel condition was added

432 to the task specification, and FOR-Prompting was run for three additional iterative rounds, which
433 constitute Task 2.

434
435 The pilot results show that FOR-Prompting incorporates newly added conditions rather than treating
436 each turn as an isolated optimization. Beyond general considerations such as distance or dining
437 options, the notion of group travel was explicitly integrated into the objections raised by the Debater.
438 Example inquiries include: “*For group logistics, did you consider the potential for traffic delays*
439 *or the availability of large Ubers or private vans during peak times?*” and “*For artisanal shops*
440 *and farmers’ markets, will a group of six moving together disrupt the flow or experience?*” These
441 questions capture factors that only emerge once group size becomes a relevant constraint.

442
443 New conditions or new knowledge can originate from multiple sources. Users may directly provide
444 new inputs that modify earlier assumptions, or external APIs can introduce real-time updates, such as
445 local news feeds, delayed transportation alerts, or weather forecasts. This mechanism makes FOR-
446 Prompting particularly suitable for tasks that require frequent adjustment or multi-stage processing
447 based on ongoing events or new data. Examples include strategy revision, policy adjustment, and
448 dynamic analysis tasks in domains such as healthcare, finance, education, and other environments
449 where conditions evolve continuously.

448 These observations point to FOR-Prompting’s potential to handle dynamic multi-stage updates and
449 adapt to evolving task requirements. Full outputs are provided in Appendix A.4.

451 6 DISCUSSION

452 6.1 HUMAN-INSPIRED EXTERNAL QUESTIONING

453 FOR-Prompting draws inspiration from human collaborative reasoning, where interlocutors improve
454 solutions by asking clarifying and adversarial questions rather than by directly proposing alterna-
455 tives. However, while the approach is inspired by human collaboration, FOR-Prompting deliberately
456 removes the need for human-in-the-loop prompting during problem solving. Human oversight is of-
457 ten bounded by uneven domain knowledge and experiential blind spots, leading to under-specified
458 goals or misread constraints (Zamfirescu-Pereira et al., 2023; Zhang et al., 2025a). Moreover, even
459 expert users may struggle to translate tacit intentions into effective prompts, yielding brittle, incon-
460 sistent guidance, or slow down the problem-solving process (Zhang et al., 2025b). These limita-
461 tions amplify variance across runs, introduce cognitive biases (e.g., anchoring on an early draft), and
462 reduce reproducibility.

463 In contrast, FOR-Prompting encodes questioning, revision, and synthesis into agents with explicit,
464 machine-executable objectives, enabling the system to interrogate and refine the intermediate so-
465 lution without relying on ad hoc human expertise. This design is highly promising, as it not only
466 reduces dependence on human intervention - often constrained by limited knowledge, incomplete
467 problem understanding, or insufficient prompt-crafting skills - but also enhances the consistency,
468 reproducibility, and transparency of the reasoning process. By transforming what was once an intu-
469 itive and manual collaborative process into a structured, automated workflow, FOR-Prompting offers
470 a scalable and domain-agnostic framework for complex reasoning tasks. It preserves the strengths
471 of human-style dialogue-driven refinement while overcoming the variability, cost, and cognitive
472 bias introduced by manual prompting, ultimately paving the way toward more autonomous, self-
473 improving reasoning systems.

474 6.2 FOR-PROMPTING IS HELPFUL AND SHOWS PROMISE FOR SMALL MODELS

475 On GSM8K we observe a recurring pattern. Single prompt baselines are brittle and often miss the
476 correct answer. CoT improves correctness by eliciting intermediate steps. In our experiments FOR-
477 Prompting reached accuracy comparable to CoT and produced reasoning that annotators viewed as
478 more coherent and more responsive to targeted external questions. These lightweight questions act
479 like a gentle adversary and push the solver to state assumptions and address salient gaps before
480 finalizing the answer. This yields solutions that align better with human judgments.

481 Many real tasks do not have a single ground truth solution. In brainstorming, plan refinement, and
482 strategy design, value comes from surfacing blind spots, stress testing feasibility, and moving from
483 vague goals to concrete next steps. FOR-Prompting provides an automated scaffold for this kind
484 of thinking. External questions guide exploration, push for specificity, and make trade offs explicit.
485 The protocol therefore supports ideation that is more diverse, plans that are more actionable, and
narratives that better document why a decision was made. As evaluation in these domains often relies

486 on human preference and rubric based judgments, the transparency of the dialogue trace becomes
487 an asset rather than an overhead.

488 This mechanism may have substantial potential for small model applications (Wang et al., 2025). For
489 a 1B parameter model the absolute accuracy on math problems is usually well below that of larger
490 models, yet one brief round of external questioning typically brings a notable improvement over
491 direct prompting. For planning and brainstorming tasks FOR-Prompting can also be attractive on
492 smaller models. Our results indicate advantages in interaction quality, outcome quality, information
493 coverage, and level of refinement. The remaining consideration is cost. The combination with
494 small models is practical because it offers cost efficiency, on-device or constrained deployment, and
495 privacy when local reasoning is required. Example scenarios include privacy sensitive assistants that
496 must reason on-device, on-device personal use or research deployments with limited compute, and
497 edge applications where intermittent connectivity or energy limits rule out large models.

498 The most immediate cost is token usage and latency. External questioning can increase both, yet the
499 overhead is usually controllable. This overhead compares favorably to scaling parametric capacity.
500 Users can iterate extensively on a personal device without incurring high API token costs.

501 An additional practical benefit emerges from our cross-model role-swapping experiments. When
502 we assign the small model as the Debater (question generator) and use the stronger model only as
503 the Defender, we observe that the overall reasoning quality remains high while the token usage of
504 the larger model is greatly reduced. This asymmetric structure stands in contrast to self-ask-style
505 approaches, where a single (often large) model must generate both sub-questions and answers, dou-
506 bling the load on the expensive model. In our setting, the Debater only needs to generate short
507 interrogative prompts, a task that even a 1B model can perform reliably. As a result, most of the
508 computational cost is shifted away from the frontier model and toward the lightweight model, which
509 is inexpensive to run locally. This makes FOR-Prompting attractive in scenarios where large-model
510 usage must be minimized, such as cost-sensitive deployments, rate-limited API environments, or
511 hybrid pipelines where local small models serve as question-generating front ends for stronger but
512 more expensive solvers.

513 7 LIMITATIONS

514 Although FOR-Prompting offers improvements over some prompting strategies and reduces poten-
515 tial human labor costs in open-ended tasks, several limitations remain. For instance, the protocol
516 introduces additional computational overhead. FOR-Prompting relies on multi-turn interactions be-
517 tween the Debater and the Defender, which increases latency compared with single-turn prompting
518 methods such as standard prompting or CoT. As a result, the approach is less suitable for applica-
519 tions that require real-time responses or low-latency interactions. Instead, FOR-Prompting is better
520 suited for tasks that can tolerate background processing and iterative refinement, such as open-ended
521 reasoning, analysis, or planning, where response speed is not the primary constraint. Our study is
522 limited by compute constraints, which prevented evaluation on more and larger open-source datasets.
523 We were also unable to reproduce certain reported baselines for some models (e.g., LLaMA 3.2:1B),
524 so all benchmark values are computed using our standardized pipeline.

525 8 CONCLUSION AND FUTURE WORK

526 FOR-Prompting targets the reasoning process by introducing structured external questioning that
527 surfaces hidden assumptions and closes inferential gaps. This approach improves correctness and
528 transparency at modest computational cost, enables the use of smaller models and on-device deploy-
529 ments, and remains backend-agnostic since it relies on role-based interactions rather than model-
530 specific features. vspace-0.2em Future work will broaden the range of tasks and modalities, develop
531 adaptive questioning policies, and integrate domain-specific validators to complement our accuracy
532 and reasoning proxies. We also plan to release standardized prompts and transcripts to support
533 reproducible benchmarking and evaluate FOR-Prompting in practical settings such as education, ac-
534 cessibility, and edge assistants. Additionally, future investigations may explore multi-agent variants
535 of FOR-Prompting involving more than two coordinated roles, enabling richer forms of interrogative
536 pressure and collaborative reasoning. We also aim to examine the generalizability of the protocol
537 across languages, cultural contexts, and real-world open-ended decision-making scenarios. Another
538 promising direction is integrating FOR-Prompting with retrieval-augmented generation (RAG) or
539 reinforcement learning (RL). Combining external knowledge retrieval with structured interrogative
pressure may strengthen long-horizon planning and multi-stage tasks.

540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593

ETHICS STATEMENT

This work includes a minimal-risk human preference study. Participants were recruited voluntarily through social media and verified as adults via email-based confirmation for eligibility. After verification, all personally identifiable information was removed, and only anonymized preference responses were retained. All procedures follow commonly accepted ethical guidelines for minimal-risk user research.

USE OF LLMs

We employed LLMs for limited purposes: (1) to improve the readability of this paper (grammar and spelling correction), and (2) to reformat outputs in the Appendix that could not be directly rendered in \LaTeX . Some reported results (*coherence*, *reasoning*, and *GPT-4.1 evaluation*) were obtained using an LLM-as-a-judge approach, as noted in the main text, whereas accuracy scores were computed through statistical calculations.

REPRODUCIBILITY STATEMENT

Because our contribution, FOR-Prompting, is primarily a framework-level design rather than a fixed implementation, the specific code and prompts used in our experiments do not need to be followed verbatim. Instead, we provide illustrative code snippets and representative outputs, which are sufficient to reproduce and verify the core results. These materials are made available through an anonymized repository (<https://anonymous.4open.science/r/as98f4-6cEE/>) during the review process and will be released publicly upon acceptance.

REFERENCES

- Joshua Ashkinaze, Emily Fry, Narendra Edara, Eric Gilbert, and Ceren Budak. Plurals: A system for guiding llms via simulated social ensembles. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400713941. doi: 10.1145/3706598.3713675. URL <https://doi.org/10.1145/3706598.3713675>.
- Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, et al. Constitutional ai: Harmlessness from ai feedback. *arXiv preprint arXiv:2212.08073*, 2022.
- Maciej Besta, Nils Blach, Ales Kubicek, Robert Gerstenberger, Michał Podstawski, Lukas Gianinazzi, Joanna Gajda, Tomasz Lehmann, Hubert Niewiadomski, Piotr Nyczyk, and Torsten Hoefler. Graph of thoughts: solving elaborate problems with large language models. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'24/IAAI'24/EAAI'24. AAAI Press, 2024. ISBN 978-1-57735-887-9. doi: 10.1609/aaai.v38i16.29720. URL <https://doi.org/10.1609/aaai.v38i16.29720>.
- Xiaohe Bo, Zeyu Zhang, Quanyu Dai, Xueyang Feng, Lei Wang, Rui Li, Xu Chen, and Ji-Rong Wen. Reflective multi-agent collaboration based on large language models. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24, Red Hook, NY, USA, 2025. Curran Associates Inc. ISBN 9798331314385. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/fa54b0edce5eef0bb07654e8ee800cb4-Paper-Conference.pdf.
- Wenhu Chen, Xueguang Ma, Xinyi Wang, and William W. Cohen. Program of thoughts prompting: Disentangling computation from reasoning for numerical reasoning tasks. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL <https://openreview.net/forum?id=YfZ4ZPt8zd>.
- Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. Training verifiers to solve math word problems, 2021. URL <https://arxiv.org/abs/2110.14168>.

594 Shehzaad Zuzar Dhuliawala, Mojtaba Komeili, Jing Xu, Roberta Raileanu, Xian Li, Asli Celikyilmaz, and Jason E Weston. Chain-of-verification reduces hallucination in large language models, 2024. URL <https://openreview.net/forum?id=VP20ZB6DHL>.

595

596

597 Rahul R. Divekar, Hui Su, Jeffrey O. Kephart, Maira Gratti DeBaysar, Melina Guerra, Xiangyang Mou, Matthew Peveler, and Lisha Chen. Humaine: Human multi-agent immersive negotiation competition. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, pp. 1–10, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450368193. doi: 10.1145/3334480.3383001. URL <https://doi.org/10.1145/3334480.3383001>.

598

599

600

601

602

603 Ali Dorri, Salil S. Kanhere, and Raja Jurdak. Multi-agent systems: A survey. *IEEE Access*, 6: 28573–28593, 2018. doi: 10.1109/ACCESS.2018.2831228.

604

605

606 Yilun Du, Shuang Li, Antonio Torralba, Joshua B. Tenenbaum, and Igor Mordatch. Improving factuality and reasoning in language models through multiagent debate. In *Forty-first International Conference on Machine Learning*, 2024. URL <https://openreview.net/forum?id=zj7YuTE4t8>.

607

608

609

610 Will Epperson, Gagan Bansal, Victor C Dibia, Adam Fourney, Jack Gerrits, Erkang (Eric) Zhu, and Saleema Amershi. Interactive debugging and steering of multi-agent ai systems. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400713941. doi: 10.1145/3706598.3713581. URL <https://doi.org/10.1145/3706598.3713581>.

611

612

613

614

615 Luyu Gao, Aman Madaan, Shuyan Zhou, Uri Alon, Pengfei Liu, Yiming Yang, Jamie Callan, and Graham Neubig. PAL: Program-aided language models. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 10764–10799. PMLR, 23–29 Jul 2023. URL <https://proceedings.mlr.press/v202/gao23f.html>.

616

617

618

619

620

621 Junda He, Christoph Treude, and David Lo. Llm-based multi-agent systems for software engineering: Literature review, vision, and the road ahead. *ACM Trans. Softw. Eng. Methodol.*, 34(5), May 2025. ISSN 1049-331X. doi: 10.1145/3712003. URL <https://doi.org/10.1145/3712003>.

622

623

624

625

626 Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xiawu Zheng, Yuheng Cheng, Ceyao Zhang, Jinlin Wang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin, et al. Metagpt: Meta programming for a multi-agent collaborative framework. International Conference on Learning Representations, ICLR, 2024. doi: <http://hdl.handle.net/10754/700016>.

627

628

629

630 Guohao Li, Hasan Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. Camel: Communicative agents for "mind" exploration of large language model society. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 51991–52008. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/a3621ee907def47c1b952ade25c67698-Paper-Conference.pdf.

631

632

633

634

635

636 Kun Li, Tingzhang Zhao, Wei Zhou, and Songlin Hu. DORA: Dynamic optimization prompt for continuous reflection of LLM-based agent. In Owen Rambow, Leo Wanner, Marianna Apidianaki, HEND Al-Khalifa, Barbara Di Eugenio, and Steven Schockaert (eds.), *Proceedings of the 31st International Conference on Computational Linguistics*, pp. 7546–7557, Abu Dhabi, UAE, January 2025. Association for Computational Linguistics. URL <https://aclanthology.org/2025.coling-main.504/>.

637

638

639

640

641

642 Tian Liang, Zhiwei He, Wenxiang Jiao, Xing Wang, Yan Wang, Rui Wang, Yujiu Yang, Shuming Shi, and Zhaopeng Tu. Encouraging divergent thinking in large language models through multi-agent debate. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen (eds.), *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 17889–17904, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-main.992. URL <https://aclanthology.org/2024.emnlp-main.992/>.

643

644

645

646

647

- 648 Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegrefe, Uri
649 Alon, Nouha Dziri, Shrimai Prabhunoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad
650 Majumder, Katherine Hermann, Sean Welleck, Amir Yazdanbakhsh, and Peter Clark. Self-refine:
651 Iterative refinement with self-feedback. In *Thirty-seventh Conference on Neural Information Pro-
652 cessing Systems*, 2023. URL <https://openreview.net/forum?id=S37h0erQLB>.
- 653 Diego Maldonado, Edison Cruz, Jackeline Abad Torres, Patricio J. Cruz, and Silvana del Pilar Gam-
654 boa Benitez. Multi-agent systems: A survey about its components, framework and workflow.
655 *IEEE Access*, 12:80950–80975, 2024. doi: 10.1109/ACCESS.2024.3409051.
- 656
657 Potsawee Manakul, Adian Liusie, and Mark Gales. SelfcheckGPT: Zero-resource black-box hal-
658 lucination detection for generative large language models. In *The 2023 Conference on Empir-
659 ical Methods in Natural Language Processing*, 2023. URL [https://openreview.net/
660 forum?id=RwzFNbJ3Ez](https://openreview.net/forum?id=RwzFNbJ3Ez).
- 661 Suchismita Naik, Amanda Snellinger, Austin L. Toombs, Scott Saponas, and Amanda K Hall. Ex-
662 ploring early adopters’ use of ai driven multi-agent systems to inform human-agent interaction
663 design: Insights from industry practice. In *Proceedings of the Extended Abstracts of the CHI
664 Conference on Human Factors in Computing Systems*, CHI EA ’25, New York, NY, USA, 2025.
665 Association for Computing Machinery. ISBN 9798400713958. doi: 10.1145/3706599.3706693.
666 URL <https://doi.org/10.1145/3706599.3706693>.
- 667
668 Bogyom Park and Kyoungwon Seo. Assessing critical thinking through a multi-agent llm-based
669 debate chatbot. In *Proceedings of the Extended Abstracts of the CHI Conference on Hu-
670 man Factors in Computing Systems*, CHI EA ’25, New York, NY, USA, 2025. Association
671 for Computing Machinery. ISBN 9798400713958. doi: 10.1145/3706599.3721207. URL
672 <https://doi.org/10.1145/3706599.3721207>.
- 673 Ofir Press, Muru Zhang, Sewon Min, Ludwig Schmidt, Noah Smith, and Mike Lewis. Mea-
674 suring and narrowing the compositionality gap in language models. In Houda Bouamor, Juan
675 Pino, and Kalika Bali (eds.), *Findings of the Association for Computational Linguistics: EMNLP
676 2023*, pp. 5687–5711, Singapore, December 2023. Association for Computational Linguistics.
677 doi: 10.18653/v1/2023.findings-emnlp.378. URL [https://aclanthology.org/2023.
678 findings-emnlp.378/](https://aclanthology.org/2023.findings-emnlp.378/).
- 679 Chen Qian, Wei Liu, Hongzhang Liu, Nuo Chen, Yufan Dang, Jiahao Li, Cheng Yang, Weize
680 Chen, Yusheng Su, Xin Cong, Juyuan Xu, Dahai Li, Zhiyuan Liu, and Maosong Sun. Chat-
681 Dev: Communicative agents for software development. In Lun-Wei Ku, Andre Martins, and
682 Vivek Srikumar (eds.), *Proceedings of the 62nd Annual Meeting of the Association for Com-
683 putational Linguistics (Volume 1: Long Papers)*, pp. 15174–15186, Bangkok, Thailand, August
684 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.810. URL
685 <https://aclanthology.org/2024.acl-long.810/>.
- 686 Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik R Narasimhan, and Shunyu Yao. Re-
687 flexion: language agents with verbal reinforcement learning. In *Thirty-seventh Conference on
688 Neural Information Processing Systems*, 2023. URL [https://openreview.net/forum?
689 id=vAE1hFcKW6](https://openreview.net/forum?id=vAE1hFcKW6).
- 690
691 Maojia Song, Tej Deep Pala, Weisheng Jin, Amir Zadeh, Chuan Li, Dorien Herremans, and Soujanya
692 Poria. Llms can’t handle peer pressure: Crumbling under multi-agent social interactions, 2025.
693 URL <https://arxiv.org/abs/2508.18321>.
- 694
695 Chuanneng Sun, Songjun Huang, and Dario Pompili. Llm-based multi-agent decision-making:
696 Challenges and future directions. *IEEE Robotics and Automation Letters*, 10(6):5681–5688, 2025.
697 doi: 10.1109/LRA.2025.3562371.
- 698
699 Fali Wang, Zhiwei Zhang, Xianren Zhang, Zongyu Wu, TzuHao Mo, Qiuhaio Lu, Wanqing Wang,
700 Rui Li, Junjie Xu, Xianfeng Tang, Qi He, Yao Ma, Ming Huang, and Suhang Wang. A com-
701 prehensive survey of small language models in the era of large language models: Techniques,
enhancements, applications, collaboration with llms, and trustworthiness. *ACM Trans. Intell.
Syst. Technol.*, September 2025. ISSN 2157-6904. doi: 10.1145/3768165. URL <https://doi.org/10.1145/3768165>. Just Accepted.

702 Lei Wang, Wanyu Xu, Yihuai Lan, Zhiqiang Hu, Yunshi Lan, Roy Ka-Wei Lee, and Ee-Peng
703 Lim. Plan-and-solve prompting: Improving zero-shot chain-of-thought reasoning by large lan-
704 guage models. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki (eds.), *Proceedings*
705 *of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long*
706 *Papers)*, pp. 2609–2634, Toronto, Canada, July 2023a. Association for Computational Linguistics. doi: 10.18653/v1/2023.acl-long.147. URL <https://aclanthology.org/2023.acl-long.147/>.

709 Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V Le, Ed H. Chi, Sharan Narang, Aakanksha
710 Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language
711 models. In *The Eleventh International Conference on Learning Representations*, 2023b. URL
712 <https://openreview.net/forum?id=1PL1NIMMrw>.

713 Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi,
714 Quoc V. Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language
715 models. In *Proceedings of the 36th International Conference on Neural Information Processing*
716 *Systems, NIPS '22*, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.
717 URL https://openreview.net/forum?id=_VjQlMeSB_J.

718 Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun
719 Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W White, Doug Burger, and
720 Chi Wang. Autogen: Enabling next-gen LLM applications via multi-agent conversations. In *First*
721 *Conference on Language Modeling*, 2024. URL <https://openreview.net/forum?id=BAakY1hNKS>.

722 Xingjiao Wu, Luwei Xiao, Yixuan Sun, Junhang Zhang, Tianlong Ma, and Liang He. A survey of
723 human-in-the-loop for machine learning. *Future Generation Computer Systems*, 135:364–381,
724 2022. doi: <https://doi.org/10.1016/j.future.2022.05.014>.

725 Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik R
726 Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. In
727 *Thirty-seventh Conference on Neural Information Processing Systems*, 2023a. URL <https://openreview.net/forum?id=5Xc1ecx0lh>.

728 Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R Narasimhan, and Yuan Cao.
729 React: Synergizing reasoning and acting in language models. In *The Eleventh International Con-*
730 *ference on Learning Representations*, 2023b. URL https://openreview.net/forum?id=WE_vluYUL-X.

731 J.D. Zamfirescu-Pereira, Richmond Y. Wong, Bjoern Hartmann, and Qian Yang. Why johnny can't
732 prompt: How non-ai experts try (and fail) to design llm prompts. In *Proceedings of the 2023*
733 *CHI Conference on Human Factors in Computing Systems, CHI '23*, New York, NY, USA, 2023.
734 Association for Computing Machinery. ISBN 9781450394215. doi: 10.1145/3544548.3581388.
735 URL <https://doi.org/10.1145/3544548.3581388>.

736 He Zhang, Yueyan Liu, Xin Guan, Jie Cai, and John M. Carroll. Harnessing the power of ai in
737 qualitative research: Role assignment, engagement, and user perceptions of ai-generated follow-
738 up questions in semi-structured interviews, 2025a. URL <https://arxiv.org/abs/2509.12709>.

739 He Zhang, Chuhao Wu, Jingyi Xie, Yao Lyu, Jie Cai, and John M. Carroll. Harnessing the power
740 of ai in qualitative research: Exploring, using and redesigning chatgpt. *Computers in Human*
741 *Behavior: Artificial Humans*, 4:100144, 2025b. ISSN 2949-8821. doi: <https://doi.org/10.1016/j.chbah.2025.100144>. URL <https://www.sciencedirect.com/science/article/pii/S2949882125000283>.

742 Wenqi Zhang, Ke Tang, Hai Wu, Mengna Wang, Yongliang Shen, Guiyang Hou, Zeqi Tan, Peng
743 Li, Yueting Zhuang, and Weiming Lu. Agent-pro: Learning to evolve via policy-level reflection
744 and optimization. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar (eds.), *Proceedings*
745 *of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long*
746 *Papers)*, pp. 5348–5375, Bangkok, Thailand, August 2024a. Association for Computational Lin-
747 guistics. doi: 10.18653/v1/2024.acl-long.292. URL <https://aclanthology.org/2024.acl-long.292/>.

756 Yu Zhang, Jingwei Sun, Li Feng, Cen Yao, Mingming Fan, Liuxin Zhang, Qianying Wang, Xin
 757 Geng, and Yong Rui. See widely, think wisely: Toward designing a generative multi-agent
 758 system to burst filter bubbles. In *Proceedings of the 2024 CHI Conference on Human Factors
 759 in Computing Systems*, CHI '24, New York, NY, USA, 2024b. Association for Com-
 760 puting Machinery. ISBN 9798400703300. doi: 10.1145/3613904.3642545. URL <https://doi.org/10.1145/3613904.3642545>.

762 Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang,
 763 Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, Hao Zhang, Joseph E Gonzalez, and Ion Sto-
 764 ica. Judging llm-as-a-judge with mt-bench and chatbot arena. In A. Oh, T. Naumann,
 765 A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Informa-
 766 tion Processing Systems*, volume 36, pp. 46595–46623. Curran Associates, Inc., 2023.
 767 URL [https://proceedings.neurips.cc/paper_files/paper/2023/file/
 768 91f18a1287b398d378ef22505bf41832-Paper-Datasets_and_Benchmarks.
 769 pdf](https://proceedings.neurips.cc/paper_files/paper/2023/file/91f18a1287b398d378ef22505bf41832-Paper-Datasets_and_Benchmarks.pdf).

771 Denny Zhou, Nathanael Schärli, Le Hou, Jason Wei, Nathan Scales, Xuezhi Wang, Dale Schu-
 772 urmans, Claire Cui, Olivier Bousquet, Quoc V Le, and Ed H. Chi. Least-to-most prompting
 773 enables complex reasoning in large language models. In *The Eleventh International Confer-
 774 ence on Learning Representations*, 2023. URL [https://openreview.net/forum?id=
 775 WZH7099tgfM](https://openreview.net/forum?id=WZH7099tgfM).

777 A APPENDIX

779 A.1 EXAMPLE ROLE PROMPTS.

781 Table 4 shows example prompts for each role in the FOR-Prompting framework. It worth to note
 782 that the prompts listed for each role in the FOR-Prompting framework serve as illustrative examples.
 783 In practice, prompts with similar intent or wording can be employed to achieve comparable effects,
 784 and there is no requirement to use these exact phrasings.

786 Role	786 Example Prompt
787 Defender	787 “You are the Defender. Answer the user’s task as best as possible. Be 788 concise but correct. Provide a single final answer at the end, clearly 789 marked.”
790 Debater	790 “You are the Debater. You will ask clarifying questions to improve the 791 answer, without proposing any solutions. Your output should be ONLY 792 a concise list of questions.”
793 Host	793 “You are the Host. You must synthesize the Defender’s answer and the 794 Debater’s objections, and produce a refined final answer. Quote specific 795 objections when relevant.”

796 Table 4: Illustrative prompts for each role in FOR-Prompting.

799 A.2 FULL OUTPUTS OF FOR-PROMPTING IN CASE STUDY 3 - TASK “HOW MANY RS IN 800 STRARRTRABBBERY”

802 The following presents the full record of the task described in Case Study 3. We conducted seven
 803 rounds of iteration. Although the FOR-Prompting method produced the correct answer after the
 804 first round, later confirmed by human review, the subsequent iterations enhanced the interpretability
 805 of the answer, effectively providing six additional layers of verification from different perspectives
 806 (through questions). The final result synthesized by the “Host” was also correct.

807 **Task:** How many rs in strarrtrabbbery

808 **Defender:** “There are 4 ‘r’s in ‘strarrtrabbbery.’”

810 **Debater:** “Are you sure there are only 4 ‘r’s in ‘strarrtrabbbery’? Counting manually, it appears
811 there may be more than 4.”, “Did you double-check the spelling of the word to ensure accuracy in
812 your count?”, “Could you show how you counted the ‘r’s to clarify your process?”
813

814 **Defender:** Let’s count the ‘r’s in ‘strarrtrabbbery’ together, letter by letter:

815 s t r a r r t r a b b b e r y
816

817 Now, let’s highlight the ‘r’s:

818 s t r a r r t r a b b b e r y
819

820 So, the positions of ‘r’ are: 3rd, 5th, 6th, 8th, and 13th letters.

821 Total number of ‘r’s: 5.

822 Therefore, there are 5 ‘r’s in ‘strarrtrabbbery’. Thank you for pointing out the need to double-check!
823
824

825 **Debater:** “Is the highlighting of ‘r’s consistent with the actual positions in the word ‘strarrtrabb-
826 bery’? For example, are the 5th and 6th letters both ‘r’?”, “Could you clarify why the 13th letter is
827 marked as an ‘r’, and whether this matches the raw spelling?”, “Would it be helpful to present the
828 word with indices for better clarity in counting?”, “Are there any possible typographical errors in
829 the word or in the counting process that could affect the accuracy of the result?”, “Is there a reason
830 you thanked the user for double-checking, and could you provide more assurance by showing the
831 count step by step with indices?”

832 **Defender:** Let’s carefully examine the word ‘strarrtrabbbery’ step by step, addressing each cri-
833 tique:
834

835 **1. Listing with indices for clarity:**
836

Index	Letter
1	s
2	t
3	r
4	a
5	r
6	r
7	t
8	r
9	a
10	b
11	b
12	b
13	e
14	r
15	y

853 **2. Counting ‘r’s and their positions:**

854 – 3rd letter: r

855 – 5th letter: r

856 – 6th letter: r

857 – 8th letter: r

858 – 14th letter: r

859 So, the ‘r’s are at positions 3, 5, 6, 8, and 14.

860 **3. Are 5th and 6th letters both ‘r’?**

861 Yes, as shown above, both the 5th and 6th letters are ‘r’.

862 **4. Is the 13th letter an ‘r’?**

863 No, the 13th letter is ‘e’, not ‘r’. The 14th letter is ‘r’.

864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917

5. Typographical errors:

If the spelling is exactly as you wrote—‘strarrtrabbbbery’—then the above count is correct. If there was a typo in the word, the count could change.

6. Step-by-step count and assurance:

By displaying the word with indices and listing all the ‘r’ positions, the count is accurate and transparent.

Final count: There are 5 ‘r’s in ‘strarrtrabbbbery’, at positions 3, 5, 6, 8, and 14.

Debater: “Does the listing with indices provide enough clarity, or could the formatting be improved for easier verification?”, “Is there any ambiguity in how the indices correspond to the letters, particularly if the table or formatting is misaligned?”, “Are the positions of the ‘r’s correctly identified, or is there room for error in transcribing the positions from the table to the list?”, “Would it be beneficial to explicitly show the process of counting each ‘r’ in the word as it is written, rather than relying on the table?”, “Is there an assumption that the word was spelled correctly, and how would the analysis change if there was a typographical error?”, “Are there any additional checks that could be included to further ensure the transparency and accuracy of the count?”, “Could the explanation include a visual or color-coded representation for even greater clarity?”

Defender: Let’s improve the clarity and accuracy of counting the number of ‘r’s in “strarrtrabbbbery” by addressing each critique:

1. Improved formatting for easier verification:

A simple inline format can help ensure indices match up with letters:

```
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
s t r a r r t r a b b b e r y
```

2. Ambiguity in indices:

Displaying both indices and letters in a single line as above prevents misalignment and makes it clear which index corresponds to each letter.

3. Are ‘r’ positions identified correctly?

Let’s mark the ‘r’s directly:

```
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
s t [r] a [r] [r] t [r] a b b b e [r] y
      ↑      ↑      ↑      ↑      ↑
```

The ‘r’s appear at positions: 3, 5, 6, 8, and 14.

4. Explicitly showing the process:

Walking through the word, letter by letter:

- 1: s (not r)
- 2: t (not r)
- 3: r (1st r)
- 4: a (not r)
- 5: r (2nd r)
- 6: r (3rd r)
- 7: t (not r)
- 8: r (4th r)
- 9: a (not r)
- 10: b (not r)
- 11: b (not r)

- 918 • 12: b (not r)
- 919 • 13: e (not r)
- 920
- 921 • 14: r (5th r)
- 922 • 15: y (not r)

923
924 Total ‘r’s: 5

925 **5. Assumption about spelling:**

926 This count assumes the word is spelled as given. If there is a typo, the count would change. Double-
927 check the spelling before counting.

928 **6. Additional checks:**

929 Double-count by highlighting the ‘r’s, or use a tool (such as a find/count function) for large words.
930

931 **7. Visual or color-coded representation:**

932 If possible, color-code the ‘r’s for clarity (here, using brackets):

```
933       s   t   r   a   r   r   t   r   a   b   b   b   e   r   y
934           ↑       ↑   ↑       ↑
935
```

936 **Conclusion:**

937 There are 5 ‘r’s in “strarrtrabbbery”, at positions 3, 5, 6, 8, and 14. The improved formatting and
938 explicit counting process make this transparent and easy to verify.
939

940 **Debater:** “Is the inline format with indices and letters actually make the process more transparent,
941 or could it become confusing if not perfectly aligned (especially in plain text)?”, “Is there still any
942 risk of ambiguity in the indices if a proportional font or formatting error occurs?”, “When marking
943 the ‘r’s with brackets, is there a possibility that this notation could be misinterpreted or missed,
944 especially for those unfamiliar with the method?”, “Is it necessary to list every letter and its check,
945 or would a more concise method suffice for such a short example?”, “Does the answer adequately
946 address what would happen if there were a typo, or should there be a suggested process for validating
947 the original word’s spelling before starting?”, “Are there other efficient or automated ways to check
948 the count, especially for longer words or strings?”, “Would using color (if possible) or another visual
949 aid significantly improve clarity, or does the current bracket system suffice for most scenarios?”

950 **Defender:** Certainly! Here’s a refined response that directly addresses the critiques and provides
951 both a clear answer and guidance for best practices:
952

953 **How many ‘r’s in ‘strarrtrabbbery’?**

954 **Short, clear method:**

955 For a short word like “strarrtrabbbery”, the simplest and least error-prone way is to scan the word
956 and count each ‘r’:
957

```
958       s   t   r   a   r   r   t   r   a   b   b   b   e   r   y
959           ↑       ↑   ↑       ↑
960
```

961 Going through the word letter by letter:

```
962       s   t   r   a   r   r   t   r   a   b   b   b   e   r   y
963           ↑       ↑   ↑       ↑
964
```

- 965 – The ‘r’s appear at positions 3, 5, 6, 8, and 14 (counting from 1).
- 966 – **Total: 5 ‘r’s.**

967 **Addressing the critiques:**

- 968
- 969 1. *Inline format with indices:* For very short examples, adding indices or brackets is helpful
970 only if the spacing is exact (i.e., in monospaced font). In plain text or proportional fonts,
971 alignment can be lost, making it confusing. For short examples, it’s usually clearer just to
highlight the letters or list their positions.

- 972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
2. *Risk of ambiguity*: Yes, indices can become ambiguous if formatting is off. For longer or more complex strings, consider using tools or code to count.
 3. *Brackets for marking*: Brackets can be missed or misunderstood if the reader isn't familiar with the method. Explicitly stating the positions or just listing the count is often clearer.
 4. *Conciseness*: For a short string like this, listing every letter and its check is overkill. A simple count with a quick visual scan suffices.
 5. *Typos and spelling*: Always check the spelling of the word before counting. If accuracy is critical, copy and paste the source or use a spell-checker.
 6. *Automated/efficient methods*: For long strings, use a text editor's search feature (Ctrl+F) or a simple script (e.g., Python: `word.count('r')`).
 7. *Visual/color cues*: If possible, use color or bold to highlight letters, but in plain text, a caret (^) or simply stating positions is usually sufficient.

987
988
989
990

Summary:

- There are 5 'r's in "strarrtrabbberry".
- For short words, scan and count. For long strings, use search or code.
- Always verify the spelling before you start.

991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005

Debater: "Is the caret marking method really effective outside of monospaced fonts, or could it introduce more confusion for most users?", "Does the response give enough guidance for users who find brackets or bolding visually cluttered—are alternative suggestions needed?", "When listing the positions of 'r's, is it clear enough whether counting starts from 1 or 0? Should an explicit example be given?", "Does the section on automated methods sufficiently warn about the risk of missing 'R' if the search is case-sensitive?", "For 'Scripts', should the answer mention what to do if the user doesn't know programming, or suggest user-friendly alternatives?", "Is it necessary to elaborate more on what to do if a typo is found—should the answer suggest correcting and recounting?", "Could the advice on counting in short words be more strongly phrased to highlight that errors are common even for short or simple words?", "Does the summary table help users choose the right method, or would an example scenario be more helpful?", "Should the response mention accessibility tools or features (like screen readers) that can help users with dyslexia or visual impairments count letters?", "Is the response's structure (with repeated examples and summary) redundant or does it aid comprehension? Would a more concise version be just as clear?"

1006
1007
1008

Defender: Certainly! Here is an improved response that directly addresses all the critiques and clarifies approaches for counting 'r's in "strarrtrabbberry":

1009

How many 'r's in 'strarrtrabbberry'?

1010

Short Answer:

1011
1012

There are 5 occurrences of the letter 'r' in the word "strarrtrabbberry".

1013

Detailed Guidance:

1014

1. **Manual Counting: Reliable Basics**

1015
1016
1017
1018

Read each letter one by one and tally every 'r' you encounter. You can use your finger, a pencil, or physically point to keep track. **Caution:** Even in short or simple words, repeated or clustered letters make mistakes surprisingly common. For best accuracy, count more than once or use a second method.

1019

2. **Marking Methods: Pros, Cons, and Suggestions**

1020
1021

Carets (^) above letters:

1022
1023

Example (works best in monospaced fonts only):

strarrtrabbberry
^ ^ ^ ^ ^

1024

Brackets [] or bolding:

1025

s t [r] a [r] [r] t [r] a b b e [r] y

Alternative: underline/highlight or circle each 'r' on paper.

- 1026 3. **Listing Positions: Be Explicit**
 1027 The 'r's are at positions 3, 5, 6, 8, and 14 (counting from 1: s=1, t=2, r=3...). Always state
 1028 whether you start at 1 or 0.
- 1029 4. **Automated Methods and Case Sensitivity**
 1030 Use Ctrl+F or editor search; check case sensitivity. For scripts, e.g., Python
 1031 `word.count('`r`')` (beware hidden characters and case).
- 1032 5. **If You Don't Code**
 1033 Use an online letter counter; or ask someone to help with a quick script.
- 1034 6. **Check for Typos Before Counting**
 1035 Confirm spelling, correct if needed, then recount (copy-paste to avoid input errors).
- 1036 7. **Accessibility Features**
 1037 Consider screen readers, browser extensions for letter highlighting, or printing and marking
 1038 physically if visual tracking is difficult.

1040 **Summary Table (tabular format):**

Method	When to Use	Cautions/Alternatives
Manual count	Always possible	Double-check repeats/typos
Carets	Monospaced fonts	Avoid in variable-width fonts
Brackets/Bold	If not cluttered	Underline/highlight instead
Positions	Explicit reference	State start index (1/0)
Automated	Long/complex strings	Watch case sensitivity
Scripts	If you know how	Use online tools if not
Accessibility	If needed	Screen readers, print, aids

1051 **In summary:**

1052 There are 5 'r's in "strarrtrabbbbery". Choose a method that fits your context and double-check if
 1053 accuracy matters.

1054 **Host:** Final Result

1055 There are 5 'r's in "strarrtrabbbbery", at positions 3, 5, 6, 8, and 14 (counting from 1).

1059 A.3 FULL OUTPUTS OF FOR-PROMPTING IN CASE STUDY 4 - TASK "CREATE A 5-DAY
 1060 ITINERARY FOR A TRIP TO RIO DE JANEIRO IN APRIL."

1061 **Task:** Create a 5-day itinerary for a trip to Rio de Janeiro in April.

1062 **Defender:** Day 1: Arrival & Copacabana - Morning: Arrive in Rio de Janeiro, check into your
 1063 hotel near Copacabana or Ipanema. - Afternoon: Relax and stroll along Copacabana Beach. Grab
 1064 lunch at a beachside kiosk. - Evening: Walk to Forte de Copacabana for sunset views. Dinner at a
 1065 local Brazilian steakhouse (churrascaria).

1066 Day 2: Christ the Redeemer & Santa Teresa - Morning: Take the train up Corcovado Mountain to
 1067 visit Christ the Redeemer. Enjoy panoramic city views. - Afternoon: Head to the bohemian Santa
 1068 Teresa neighborhood. Explore its art studios, cafes, and colonial streets. - Evening: Have dinner in
 1069 Santa Teresa and enjoy live samba music at a local bar.

1070 Day 3: Sugarloaf Mountain & Lapa - Morning: Ride the cable car to Sugarloaf Mountain (Pão de
 1071 Açúcar) for breathtaking views of Rio and Guanabara Bay. - Afternoon: Visit the Selarón Steps
 1072 (Escadaria Selarón) and explore the vibrant Lapa district. - Evening: Experience Lapa's nightlife
 1073 and try traditional Brazilian snacks and caipirinhas.

1074 Day 4: Botanical Garden & Ipanema - Morning: Visit the Rio de Janeiro Botanical Garden (Jardim
 1075 Botânico), famous for its lush flora and giant water lilies. - Afternoon: Relax at Ipanema Beach.
 1076 Shop at the local boutiques or visit the Hippie Fair (if it's Sunday). - Evening: Dine at a seafood
 1077 restaurant overlooking the ocean.

1080 Day 5: Tijuca Forest & Departure - Morning: Take a guided hike in Tijuca National Park, the world's
1081 largest urban rainforest. See waterfalls and wildlife. - Afternoon: Last-minute shopping or a final
1082 stroll along the beach. - Evening: Depart Rio de Janeiro.

1083 Tips: - Pack light clothes, sunscreen, and a hat for April's warm weather. - Stay hydrated and keep
1084 an eye on your belongings in crowded areas. - Try local dishes like feijoada, pão de queijo, and fresh
1085 tropical fruits.
1086

1087 **Debater:** "Day 1 seems quite relaxed, but is there enough time to check into the hotel and still
1088 enjoy Copacabana Beach and Forte de Copacabana if the arrival is late in the morning or early
1089 afternoon?"

1090 "The recommendation to stay near Copacabana or Ipanema is good, but are there price or safety
1091 considerations travelers should be aware of for these neighborhoods?"

1092 "Is Forte de Copacabana open late enough for sunset visits, and are there any entrance fees or
1093 restrictions that should be noted?"

1094 "On Day 2, is it necessary to book tickets in advance for the Corcovado train to Christ the Redeemer,
1095 especially during busy travel seasons?"

1096 "Santa Teresa is a great suggestion, but are there safety tips for visiting this area, and is public
1097 transport or a taxi recommended for getting there?"

1098 "Live samba music in Santa Teresa is appealing, but are there specific venues or days when perfor-
1099 mances are more likely?"

1100 "Day 3's plan includes Sugarloaf Mountain and Lapa; is it manageable to fit both in one day without
1101 feeling rushed, especially considering travel times and possible queues?"

1102 "Lapa is famous for nightlife, but are there safety concerns or areas to avoid after dark, and are there
1103 recommended places for first-time visitors?"

1104 "On Day 4, is the Hippie Fair only on Sundays, and if so, is there an alternative suggestion for
1105 non-Sunday visitors?"

1106 "Visiting both the Botanical Garden and Ipanema Beach in one day may be ambitious; is there
1107 enough time to enjoy both without feeling hurried?"

1108 "The seafood restaurant suggestion in Ipanema is nice; are there specific recommendations for pop-
1109 ular or well-reviewed places?"

1110 "For Day 5's hike in Tijuca National Park, is a guided tour essential for safety, or can experienced
1111 hikers go alone? Are there particular trails that are recommended for a half-day visit?"

1112 "Are there recommendations on where to store luggage after hotel checkout if the flight departs in
1113 the evening?"

1114 "The tips are practical, but could include advice on transportation (e.g., using taxis, ride-sharing, or
1115 public transit) and any relevant safety apps or emergency numbers."

1116 "Should travelers be aware of any major holidays, events, or closures in Rio during April that could
1117 affect this itinerary?"

1118 "Are there any health or vaccination requirements for visiting Rio, especially for those planning
1119 outdoor activities in Tijuca Forest?"

1120 "The itinerary includes many outdoor activities; is there a rainy season in April that could disrupt
1121 plans, and are there suggested indoor alternatives?"

1122 "Are language barriers likely for English-speaking tourists, and are there tips for overcoming them
1123 (such as useful Portuguese phrases or translation apps)?"

1124 "Does the itinerary take into account travel times between attractions, and would using a map app
1125 or planning routes in advance be helpful?"

1126 "Overall, could the itinerary be adjusted for travelers with mobility issues or specific dietary restric-
1127 tions?"

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

Defender: 5-Day Rio de Janeiro Itinerary (April) – Enhanced Version with Practicalities, Safety, and Alternatives

Day 1: Arrival, Copacabana Beach & Forte de Copacabana

- **Arrival:** If arriving late morning or early afternoon, plan for a quick hotel check-in (most hotels allow baggage drop before formal check-in time). Prioritize either Copacabana Beach or Forte de Copacabana if time is short; both are close together.

- **Copacabana/Ipanema Stay:** These are safe and lively neighborhoods but can be pricier. For budget travelers, check for hostels or guesthouses. Use hotel safes for valuables and avoid dark, deserted areas at night.

- **Forte de Copacabana:** Open until 6pm (last entry ~5:30pm), entrance fee (about R\$6). Sunset is often best viewed from the beach; for late arrivals, consider skipping the fort and enjoying the promenade.

Day 2: Christ the Redeemer & Santa Teresa

- **Corcovado Train:** Booking tickets in advance is highly recommended, especially on weekends or holidays (get them at tremdocorcovado.rio). Allow time for queues and transport; early morning is less crowded.

- **Santa Teresa:** Take a licensed taxi or app-based ride (Uber/99) for safety. Avoid walking alone after dark. Stay on main roads; keep valuables secure.

- **Live Samba:** Check venues like Bar do Mineiro (good food, sometimes live music), Armazém São Thiago, or Café do Alto. Fridays and Saturdays are most likely for performances; confirm with venues ahead.

Day 3: Sugarloaf Mountain & Lapa

- **Manageability:** Visit Sugarloaf in the morning (less crowded). Allow 2–3 hours, including queues. Lapa is a 20–30 min taxi ride from Sugarloaf. Manageable if you don't linger too long.

- **Lapa Safety:** Stick to main streets and avoid isolated alleys after dark. Recommended for first-timers: Rio Scenarium (iconic samba club), Lapa 40 Graus, or Selina Lapa Rooftop Bar.

Day 4: Jardim Botânico, Ipanema & Hippie Fair

- **Hippie Fair:** Only on Sundays. For other days, consider Feira de Copacabana (night street market) or boutique shopping on Rua Visconde de Pirajá.

- **Timing:** Botanical Garden (2–3 hours) + Ipanema Beach (rest of the day) is doable but don't rush. If tired, skip extra shopping.

- **Seafood:** Try Restaurante Satyricon, Zaza Bistro, or Barraca do Uruguaí for well-reviewed seafood.

Day 5: Tijuca Forest Hike & Departure

- **Guided vs. Solo Hike:** Experienced hikers can do main trails solo, but for Pico da Tijuca or Pedra Bonita, a guide is safer (especially for first-timers). Shorter trails like Cascatinha Taunay are manageable alone.

- **Luggage Storage:** Most hotels offer luggage storage. Alternatively, use LuggageHero or airport lockers.

General Tips & Practicalities:

- **Transport:** Use Uber or 99 apps for safe, affordable rides. Metro is faster for some routes. Rio de Janeiro Metro app and Moovit are helpful. Emergency number: 190 (police). Download “Onde Tem Táxi” app for official taxis.

- **Weather:** April is autumn, warm (22–28°C), with some chance of rain. Have indoor alternatives: Museu do Amanhã (science museum), MAR (art museum), or AquaRio (aquarium).

- 1188 - **Events/Holidays:** April can include Easter (Semana Santa); attractions may close or be crowded.
1189 Double-check opening hours in advance.
1190
1191 - **Health:** No mandatory vaccinations for Rio, but yellow fever vaccination is recommended for
1192 Tijuca Forest. Bring mosquito repellent.
1193 - **Language:** English is limited. Download Google Translate; learn key phrases (e.g., “por favor,”
1194 “obrigado/a”). Most tourist areas have English signage.
1195 - **Mobility/Diet:** All attractions have some accessible options, but sidewalks may be uneven. Many
1196 restaurants cater to vegetarians/vegans; always ask about dietary needs.
1197
1198 - **Travel Times:** Use Google Maps or Moovit to plan routes and check for traffic. Allow extra time
1199 during rush hour.

1200 This itinerary balances highlights with practical advice, safety tips, and alternatives for all travelers.
1201

1202
1203 **Debater:** “Is Day 1 too ambitious for travelers who arrive later in the day, considering potential
1204 fatigue and limited daylight?”, “While Copacabana and Ipanema are touted as safe, are there specific
1205 streets or zones within these neighborhoods that should be avoided, especially at night?”, “Should
1206 the itinerary mention possible long lines or wait times at Forte de Copacabana, especially on week-
1207 ends or holidays?”, “On Day 2, are there alternative ways to reach Christ the Redeemer if Corcovado
1208 train tickets are sold out, such as vans or hikes?”, “Does the Santa Teresa recommendation account
1209 for the steepness of the neighborhood, which may affect visitors with mobility issues?”, “Are there
1210 reliable resources to check for updated samba schedules, since live music nights can vary and some
1211 venues may require reservations?”, “Visiting Sugarloaf and Lapa on the same day assumes smooth
1212 transport—could traffic or public transit delays make this plan too rushed?”, “Is it safe to visit Lapa
1213 during daytime, or is it best reserved for evenings and nightlife?”, “For Day 4, should more detail
1214 be given on how to reach the Hippie Fair or alternative markets by public transport?”, “Does
1215 the schedule for Day 4 allow enough downtime for relaxation, or does it risk overpacking the day
1216 with activities?”, “The seafood recommendations are helpful, but are these restaurants accessible to
1217 non-Portuguese speakers and do they offer menus in English?”, “Regarding the Tijuca Forest hike,
1218 are there entrance fees, opening hours, or permit requirements travelers should be aware of?”, “Is
1219 luggage storage at hotels always free, or are there fees travelers should budget for when using ser-
1220 vices like LuggageHero or airport lockers?”, “Should the transport section mention the risk of traffic
1221 jams at rush hour, especially when traveling to or from the airport?”, “Are there indoor alternatives
1222 suggested for all areas of the city in case of rain, or are some neighborhoods lacking in rainy-day
1223 options?”, “Does the itinerary consider public holidays or local festivals that might disrupt the plan
1224 beyond Easter?”, “Are any health precautions missing, such as the risk of foodborne illness or sun
1225 exposure?”, “How easy is it to find vegetarian or vegan food in traditional Brazilian restaurants, and
1226 are there specific places that cater to these diets?”, “For travelers with limited mobility, are the main
1227 attractions truly accessible, or do some have stairs/steep paths that could pose problems?”, “Does
1228 the itinerary suggest any tools for offline navigation, in case cell service is spotty or travelers lack
1229 data access?”

1229 **Defender:** 5-Day Rio de Janeiro Itinerary in April (Refined with Critique Responses)
1230

1231 Day 1: Arrival & Copacabana/Ipanema - If you arrive late: Focus on a relaxed beach stroll (Copaca-
1232 bana Promenade or Ipanema’s Posto 9), sunset, and a light meal. Forte de Copacabana closes at 6pm
1233 and often has a wait on weekends/holidays—skip if arrival is after 3pm or you’re tired. - Safety:
1234 Copacabana and Ipanema are generally safe on the main avenues (Avenida Atlântica, Rua Visconde
1235 de Pirajá). Avoid the inner blocks near favelas (e.g., parts around Cantagalo, Pavão-Pavãozinho, and
1236 Arpoador at night). Stick to areas with good lighting and crowds after dark.

1237 Day 2: Christ the Redeemer & Santa Teresa - Christ the Redeemer: If train tickets are sold out,
1238 official vans leave from Praça do Lido (Copacabana), Largo do Machado, and Barra da Tijuca.
1239 There’s also a well-marked hiking trail (Parque Lage), but it’s steep and best for fit travelers. -
1240 Santa Teresa: The area is charming but steep and cobbled—difficult for those with mobility issues.
1241 The Santa Teresa tram (bondinho) is an accessible, scenic way to reach the area from downtown.
Taxis or ride-hailing apps are recommended for easier access. - Samba: For live music, check

1242 sites like Sympla (event tickets), RioTur’s agenda, or venues’ social media for up-to-date schedules.
1243 Rio Scenarium and Bar Carioca da Gema are popular but may require reservations, especially on
1244 weekends.

1245 Day 3: Sugarloaf Mountain & Lapa - Timing: Visit Sugarloaf early to avoid lines (opens 8am).
1246 Allow 2–3 hours. Lapa is best in the evening for nightlife, but the famous arches and Escadaria
1247 Selarón are safe to visit by day. Allow for 30–45 minutes commute between the two—consider
1248 traffic at rush hour (4–7pm). If running late, prioritize one. - Safety: Lapa is safe during the day
1249 on main streets (Avenida Mem de Sá, Rua do Lavradio) but avoid wandering late at night unless in
1250 groups or with a guide.

1251 Day 4: Jardim Botânico, Ipanema & Hippie Fair - Hippie Fair: Only on Sundays. For other days,
1252 consider Feira de Copacabana (night street market) or boutique shopping on Rua Visconde de Pirajá.
1253 - Seafood: Satyricon, Zaza Bistro, and Barraca do Uruguai are popular with English menus and staff
1254 used to tourists. Reservations suggested on weekends. - Downtime: The afternoon is reserved for
1255 beach relaxation or hotel pool time. Activities are optional and can be dropped if you need more
1256 rest.

1257 Day 5: Tijuca Forest Hike & Departure - Hiking: Parque Nacional da Tijuca is free, open daily 8am–
1258 5pm. No permits needed for main trails. Guided tours are best for Pico da Tijuca or Pedra Bonita
1259 (available via GetYourGuide, Viator, or local agencies). Shorter trails like Cascatinha Taunay are
1260 accessible and scenic. - Luggage: Most hotels offer free luggage storage, but confirm in advance.
1261 LuggageHero and airport lockers charge by the hour/day (typically R\$30–60 per bag).

1262 General Tips - Transport: Metro and ride-hailing apps are safest. Allow extra time during rush hour,
1263 especially to/from the airport. - Rain: For bad weather, consider Museu do Amanhã (downtown),
1264 MAR (art museum), Shopping Leblon, AquaRio (aquarium), or CCBB (cultural center) in Centro.
1265 Each is easily reached by metro or taxi. - Public holidays: Beyond Easter, Tiradentes Day (April
1266 21) may affect opening hours. Confirm with venues in advance. - Health: Use bottled water, avoid
1267 undercooked street food, and use sunscreen (UV is high even in April). Carry insect repellent for
1268 hikes. - Dietary needs: Vegetarian/vegan options are limited at traditional restaurants, but spots like
1269 Teva, Prana, and Org Bistro cater to these diets. Most places in tourist areas offer English menus or
1270 at least English-speaking staff. - Accessibility: Christ the Redeemer, Sugarloaf, and Jardim Botânico
1271 have accessible paths and elevators, but Santa Teresa and the Selarón steps are challenging. Always
1272 check with venues if you have mobility concerns. - Offline navigation: Download Google Maps for
1273 Rio, or use Maps.me. Rio Metro app works offline for metro planning.

1274
1275 **Debater:** “Does Day 1 provide enough flexibility for travelers arriving very late in the day, or
1276 should it suggest skipping activities entirely if arrival is after sunset?”, “The safety advice is helpful,
1277 but could it be more specific regarding exactly which streets or blocks are best avoided, especially for
1278 those unfamiliar with the area?”, “On Day 2, while van alternatives for Christ the Redeemer are men-
1279 tioned, is there information about how to purchase tickets for these vans and potential wait times?”,
1280 “Santa Teresa is noted as steep and difficult for those with mobility issues, but does the itinerary
1281 provide enough alternatives or modifications for travelers with disabilities?”, “Are the samba venue
1282 recommendations up to date, and is there a risk of venues being closed or having limited schedules
1283 outside of peak tourist times?”, “For Day 3, is the suggested timing between Sugarloaf and Lapa
1284 realistic, considering possible delays or the need for rest between activities?”, “Does the itinerary
1285 sufficiently warn about petty crime or pickpocketing in tourist-heavy areas like Lapa and Sugar-
1286 loaf?”, “On Day 4, is the Hippie Fair included only on Sundays? Are there enough alternative
1287 options for markets on other days, or does the plan become thin?”, “Are the seafood restaurant rec-
1288 ommendations suitable for travelers with dietary restrictions or allergies, and is there a price range
1289 provided?”, “Does the itinerary on Day 4 risk being too crowded with activities, or is there enough
1290 downtime for relaxation?”, “For Day 5, is it clear how to access guided tours for the hikes, and are
1291 there options for non-Portuguese speakers?”, “Does the luggage advice account for possible hotel
1292 policies that charge for storage, or should travelers always expect to pay unless confirmed other-
1293 wise?”, “Are the rain-day activity suggestions spread out geographically, or would visitors have to
1294 travel far in poor weather to reach them?”, “Should the plan address how to handle sudden closures
1295 or schedule changes due to public holidays more proactively?”, “Are health safety tips adequate, or
should there be more detail about local illnesses, water safety, or emergency procedures?”, “While
vegetarian/vegan options are named, are there enough suggestions for those with other dietary needs

1296 (e.g., gluten-free, allergies)?”, “Is the summary of accessibility for major attractions accurate, and
1297 should there be more detailed guidance for wheelchair users or those with significant mobility is-
1298 sues?”, “Does the itinerary provide enough detail about how to use offline navigation tools, such as
1299 step-by-step instructions for downloading maps?”, “Are there any potential gaps in transport advice,
1300 such as how to safely use taxis or ride-hailing apps and what to do if public transit is disrupted?”,
1301 “Should the itinerary include advice about local customs, tipping, or language barriers to help trav-
1302 elers better prepare?”

1303

1304

1305

1306

Defender: 5-Day Rio de Janeiro Itinerary in April (Refined for Clarity, Detail, and Practicality)

1307

1308 **Day 1: Arrival, Copacabana & Ipanema**

1309

1310

1311

1312

1313

1314

1315

1316

1317

1318

- Late Arrivals: If your flight arrives after 7pm, skip all sightseeing and beach activities entirely for safety and rest. Use this time only for checking in, settling, and eating at your hotel or a nearby, well-lit restaurant. If you arrive even later (e.g., after 10pm), go straight to your accommodation and rest—do not attempt any itinerary activities.

- Safety: For Copacabana, limit yourself to Avenida Atlântica and avoid backstreets behind the hotels after sunset. In Ipanema, stay on Rua Visconde de Pirajá and along the main beach promenade; avoid walking towards the hills (especially near Cantagalo or Pavão-Pavãozinho) at night. Use app-based transportation at night and do not walk alone on the sand after dark.

1318 **Day 2: Christ the Redeemer & Santa Teresa**

1319

1320

1321

1322

1323

1324

1325

1326

1327

1328

- Christ the Redeemer: Purchase van tickets only from official counters at Praça do Lido, Largo do Machado, or Barra Shopping, or online at <https://paineirascorcovado.com.br> (choose “oficial van”). Arrive before 8am for the shortest lines; expect 30–90 minutes’ wait if later. The official van is the only step-free option (the train is not wheelchair accessible). Bring ID for ticket collection. There are bathrooms and snack bars at the base and summit.

- Santa Teresa Mobility: Take the historic tram (bonde) from Centro (wheelchair accessible). For those with limited mobility, focus your visit on Largo dos Guimarães, which is relatively flat and offers cafes and art shops. Skip the Selarón Steps (steep) if needed. For fully accessible alternatives, consider the Museum of Tomorrow or the Rio Art Museum (both fully accessible and with step-free access).

- Samba Venues: To verify hours, check the venues’ official websites and Instagram pages. For proactive planning, use the Rio nightlife aggregator “Sympla” or call ahead. Have backup options such as live music bars in Lapa or dinner at a traditional restaurant if samba venues are closed.

1332

1333

1333 **Day 3: Sugarloaf Mountain & Lapa**

1334

1335

1336

1337

1338

1339

1340

- Timing: Plan 2–2.5 hours for Sugarloaf (including cable cars and views), then a minimum 45 minutes for transit to Lapa. Schedule a relaxed lunch in between to prevent feeling rushed. If tired, skip Lapa or move it to the evening of another day.

- Crime Risks: At Sugarloaf, keep valuables zipped and never set belongings down. In Lapa, avoid side streets and dark areas after dusk; stick to Rua do Lavradio and Avenida Mem de Sá. Use taxis/rideshare after dark. Pickpocketing is the main risk—keep phones and wallets secure and be cautious if approached by strangers.

1341

1342

1342 **Day 4: Jardim Botânico, Hippie Fair & Markets**

1343

1344

1345

1346

1347

1348

1349

- Market Alternatives: If the Hippie Fair isn’t running, try the Feira de São Cristóvão (Northeastern market), Feira da Glória (Sunday), or the antique market at Praça XV. Or, explore artisanal shops in Leblon or the Ipanema Farmers’ Market (Fridays). For more variety, consider a food tour or chocolate tasting class.

- Dietary Needs: Most restaurants can accommodate allergies if given advance notice. Use allergy cards (in Portuguese), and ask for the chef if you have multiple allergies. Zazá Bistro, Da Bela, and Org Bistrô are known for handling complex dietary needs. Always double-check if there is any doubt, and bring snacks if you have severe allergies.

1350 - Downtime: To avoid overscheduling, schedule pool or beach time, a spa massage (many hotels
1351 offer), or a leisurely cafe visit. Consider a sunset at Arpoador or a relaxed boardwalk stroll. Remind
1352 yourself that breaks are part of the Rio experience!

1353 **Day 5: Accessible Options for Non-Hikers**

1354
1355 - In addition to Parque Lage and rooftop brunches, consider the Museu do Amanhã (fully accessible),
1356 a guided street art tour by car, or a cooking class in Ipanema. For families, the AquaRio aquarium
1357 is also accessible and engaging.

1358 - Luggage Storage: Most hotels provide free storage, but always confirm at check-in. For third-party
1359 services like Bounce or LuggageHero, read user reviews and check for insurance. Have a backup
1360 plan: shopping mall lockers (e.g., Rio Sul), or ask your final excursion provider if they can hold
1361 bags. If all else fails, the airport has reliable paid lockers.

1362 **General Guidance**

1363
1364 - **Rainy Day Options:** Besides museums and malls, try niche indoor activities—Samba dance class,
1365 pottery workshop, escape room, or cinema with English subtitles (Kinoplex Leblon or Estação Net).

1366 - **Holiday Closures:** Use Google Maps and check Instagram for real-time updates. For more proac-
1367 tive planning, download the “Guia Rio” app or use “Feriados RJ” for public holiday info. Ask your
1368 hotel to confirm openings the night prior.

1369 - **Health & Safety:** In addition to dengue and food sensitivities, caution about sun exposure—
1370 use SPF 50+, hats, and hydrate constantly. Tap water is not potable; drink bottled/filtered only.
1371 Pharmacies (“Farmácia”) are everywhere for supplies. Drink bottled or filtered water; avoid street
1372 food if you have a sensitive stomach.

1373 - **Accessibility:** For tailored advice, contact Accessible Rio or Turismo Acessível. Major sights
1374 (Sugarloaf, Christ, Jardim Botânico) are accessible; Santa Teresa and Selarón Steps are not. Most
1375 Ubers can accommodate folding wheelchairs; for more, ask your hotel.

1376
1377 - **Offline Navigation:** Download Google Maps for offline use. Older travelers should request printed
1378 maps from hotels or tourist offices, or ask staff to print directions for each day’s route.

1379 - **Transport:** At airports/bus stations, if English support is lacking, look for official taxi stands (with
1380 price lists), or use Uber (set pick-up indoors). If in doubt, approach the information desk. If stuck,
1381 show your hotel address to staff and ask for help in simple English or via Google Translate.

1382 - **Flexibility:** If weather or closures disrupt plans, swap beach/park days with museum/market days.
1383 Keep a list of rainy-day options. Don’t hesitate to ask your hotel for daily local advice.

1384 - **Balance of Experiences:** This itinerary now mixes major sights with authentic experiences (food
1385 tours, local markets, Santa Teresa, art walks) and downtime. Adjust daily based on your interests
1386 and energy—Rio rewards both adventure and relaxation.
1387

1388 **Host: 5-Day Rio de Janeiro Itinerary for April (Final Version)**

1389 **Day 1: Arrival, Copacabana & Ipanema**

1390
1391
1392
1393 - **Arrival:** If arriving late, prioritize a relaxed stroll along the Copacabana or Ipanema beachfront
1394 promenade. Enjoy a light meal at a well-reviewed, nearby restaurant. Avoid poorly lit or unfamiliar
1395 areas after dark, and consult your hotel for safe walking routes.

1396 - **Hotel Check-in:** Most hotels allow baggage drop before formal check-in. Use hotel safes for
1397 valuables.
1398

1399 - **Evening:** If time allows, catch the sunset from the promenade or Arpoador Rock (between Ipanema
1400 and Copacabana). For dinner, try a local churrascaria (Brazilian steakhouse) or a seafood restaurant.

1401 - **Safety:** Main avenues are generally safe, but be vigilant for pickpocketing and avoid favela-
1402 adjacent areas at night. Ask your hotel for a safety map or advice.
1403

- 1404 **Day 2: Christ the Redeemer & Santa Teresa**
1405
1406 - **Morning:** Visit Christ the Redeemer. Book tickets for the Corcovado train in advance (<https://www.tremdocorcovado.rio/en/>). If sold out, use official vans from Praça do Lido (Copacabana), Largo do Machado, or Barra da Tijuca. Guided tours are also available via hotels.
1407
1408
1409 - **Accessibility:** The train and vans are accessible. The hiking trail from Parque Lage is steep and best for fit travelers.
1410
1411 - **Afternoon:** Explore Santa Teresa’s artistic neighborhood. Use the historic tram (bondinho) from downtown for a scenic, accessible ride. Visit Largo dos Guimarães, Parque das Ruínas (mostly accessible), and local art studios.
1412
1413
1414 - **Evening:** Dine in Santa Teresa (Bar do Mineiro, Café do Alto). For live samba, check Rio Scenarium or Bar Carioca da Gema (reserve ahead, especially on weekends).
1415
1416
1417 - **Safety:** Use taxis or ride-hailing apps for transport, especially after dark. Stick to main streets and avoid isolated areas.
1418
1419 —
1420 **Day 3: Sugarloaf Mountain & Lapa**
1421
1422 - **Morning:** Ride the cable car up Sugarloaf Mountain (Pão de Açúcar). Arrive early (opens 8am) to avoid crowds. Allow 2–2.5 hours for the visit.
1423
1424 - **Afternoon:** Head to Lapa. Visit the Selarón Steps (bottom entrance on Rua Joaquim Silva is more accessible) and admire the Lapa Arches.
1425
1426 - **Evening:** Experience Lapa’s nightlife—try Rio Scenarium, Lapa 40 Graus, or Selina Lapa Rooftop Bar. Sample caipirinhas and traditional snacks.
1427
1428 - **Safety:** Stick to main streets (Avenida Mem de Sá, Rua do Lavradio). Be wary of scams and keep valuables concealed. For emergencies, dial 190 (police).
1429
1430
1431 —
1432 **Day 4: Jardim Botânico & Markets - Morning:** Visit the Rio de Janeiro Botanical Garden (Jardim Botânico). Allow 2–3 hours to explore the lush flora and giant water lilies.
1433
1434 - **Afternoon:** Relax at Ipanema Beach. If it’s Sunday, visit the Hippy Fair (Praça General Osório). On other days, try Feira de Copacabana (night market), Feira de São Cristóvão (Northeastern market), or artisanal shops in Leblon.
1435
1436 - **Evening:** Dine at a seafood restaurant (Satyricon, Zazá Bistro, Barraca do Uruguai). Many restaurants accommodate dietary needs—use allergy cards in Portuguese and notify staff in advance.
1437
1438 - **Downtime:** Schedule time for a pool, spa, or leisurely cafe visit. Sunset at Arpoador is a local favorite.
1439
1440
1441 —
1442 **Day 5: Tijuca Forest or Accessible Alternatives & Departure**
1443
1444 - **Morning:** For hikers, take a guided hike in Tijuca National Park (Pico da Tijuca or Pedra Bonita). For a lighter option, stroll in Parque Lage or visit the Museu do Amanhã (fully accessible).
1445
1446 - **Non-hikers:** Consider a rooftop brunch, street art tour by car, cooking class, or AquaRio aquarium (accessible and family-friendly).
1447
1448 - **Afternoon:** Last-minute shopping, a final beach stroll, or relaxation at your hotel.
1449
1450 - **Luggage Storage:** Most hotels offer storage. For third-party options, use Bounce, LuggageHero, or mall lockers (e.g., Rio Sul). Confirm details at check-in.
1451
1452 - **Evening:** Depart Rio de Janeiro.
1453
1454
1455 —
1456
1457 **General Tips & Practicalities**

- 1458 - **Transport:** Use Uber/99 for safe, affordable rides. Metro is efficient for some routes; download
- 1459 the Rio Metro or Moovit app. At airports/bus stations, use official taxi stands or ride-hailing apps—
- 1460 never accept rides from solicitors.
- 1461 - **Weather:** April is warm (22–28°C) with some rain. Pack light clothes, sunscreen (SPF 50+), a
- 1462 hat, and mosquito repellent. Have indoor alternatives for rainy days: Museu do Amanhã, MAR,
- 1463 AquaRio, samba class, or a food tour.
- 1464 - **Safety:** Stay hydrated, drink only bottled/filtered water, and keep an eye on belongings in crowds.
- 1465 For emergencies, dial 190 (police), 192 (ambulance). Pharmacies (“Farmácia”) are widely available.
- 1466 - **Language:** English is limited; download Google Translate and learn basic Portuguese phrases.
- 1467 Most tourist sites have some English signage.
- 1468 - **Accessibility:** Major attractions (Sugarloaf, Christ, Jardim Botânico, Museu do Amanhã) are ac-
- 1469 cessible. Santa Teresa and Selarón Steps are not. For tailored advice, contact Accessible Rio or
- 1470 Turismo Acessível.
- 1471 - **Offline Navigation:** Download Google Maps for offline use. Request printed maps or directions
- 1472 from your hotel if needed.
- 1473 - **Flexibility:** Swap beach/park days with museums/markets if weather or closures disrupt plans.
- 1474 Keep a list of rainy-day options and ask your hotel for daily local advice.
- 1475 - **Dietary Needs:** Most restaurants can accommodate allergies and dietary restrictions. Always
- 1476 double-check with staff and carry snacks if you have severe allergies.
- 1477 - **Events/Holidays:** April may include Easter (Semana Santa); check attraction hours in advance.
- 1478 Use Google Maps, Instagram, or the “Guia Rio” app for real-time updates.
- 1479
- 1480
- 1481

1482 —

1483 **This itinerary balances Rio’s iconic sights, authentic local experiences, downtime, and**

1484 **practical safety and accessibility advice—ensuring a memorable and enjoyable visit for all**

1485 **travelers.**

1487

1488 A.4 FULL OUTPUTS OF FOR-PROMPTING IN CASE STUDY 5 - THE EXTENSION OF CASE

1489 STUDY 4 WITH AN EXTRA CONDITION – “ON THE DAY-4, OTHER TRAVELERS WILL

1490 JOIN US, FORMING A GROUP OF SIX PEOPLE, AND WE WILL DO ACTIVITIES TOGETHER.

1491 PLEASE UPDATE THIS ITINERARY ACCORDINGLY.”

1492

1493 **Task:** On the Day-4, other travelers will join us, forming a group of six people, and we will do

1494 activities together. Please update this itinerary accordingly.

1496 **Defender:** Day 4: Jardim Botânico & Markets (Group Day)

1497 - Group Logistics: With six people, plan for group-friendly transportation. Consider booking a large

1498 Uber, two taxis, or a private van for the day (many hotels can arrange this). Agree on a morning

1499 meeting time/location, and confirm everyone’s dietary needs and interests in advance.

1500 - Jardim Botânico: Arrive early (opens at 8 AM) to avoid crowds and maximize cooler morning

1501 hours. Group discounts are available if you purchase tickets together at the entrance. For a more

1502 engaging visit, book an English-speaking guided tour (ask at the visitor center or pre-arrange online).

1503 The gardens are mostly accessible with wide paths, but let the group know there is some walking

1504 involved. Bring water, hats, and sunscreen for everyone.

1505 - Markets & Food: After the gardens, head as a group to a nearby market. If it’s Sunday, the Feira

1506 Hippy de Ipanema is ideal for groups—plenty of space, food stalls, and shopping. For weekdays,

1507 consider the Feira de São Cristóvão (huge, lively, covered, and perfect for sampling Northeastern

1508 Brazilian food together). For smaller, more upscale experiences, visit artisanal shops in Leblon or

1509 the Ipanema Farmers’ Market (Friday mornings). If the group is interested in food, book a guided

1510 food tour or a chocolate tasting workshop—these experiences are fun in groups and can be reserved

1511 in English.

1512 - Lunch: Reserve a table for six in advance, especially on weekends. Options include Zazá Bistro
1513 (Ipanema, great for dietary restrictions), Da Bela (Botanical Garden area, farm-to-table), or a classic
1514 churrascaria (Brazilian steakhouse) like Fogo de Chão (Botafogo) for a festive group meal. Call
1515 ahead to confirm space and allergy accommodations.

1516 - Downtime & Social Time: After lunch, plan some downtime together—pool time at your hotel, a
1517 relaxed walk on the Ipanema or Copacabana boardwalk, or a group spa session (some hotels offer
1518 group massage packages). For sunset, Arpoador is a popular, safe spot for groups—arrive early to
1519 find space.

1520 - Evening (Optional): If energy allows, consider a group samba class (many studios offer private
1521 lessons for groups), or a casual bar/rooftop in Ipanema or Leblon for drinks and city views. Reserve
1522 if possible for a group this size.

1523 - Group Safety & Coordination: - Stay together, especially in crowded markets. - Appoint a group
1524 leader or WhatsApp group for communication. - Review emergency meeting points. - Carry some
1525 cash (small bills) for markets and tips. - If splitting up, agree on a regrouping time/location.

1526 Summary of Adjustments: - All activities are now group-friendly—transport, tours, meals, and
1527 downtime are optimized for six people. - Emphasis on advance reservations for tours and restaurants.
1528 - Market and food options tailored for group tastes and dietary needs. - Safety and communication
1529 tips for group travel. - Built-in flexibility for the group to split or relax together as needed.

1530 Let me know if you want suggestions for a specific group activity (e.g., cooking class, private boat
1531 ride, samba night) or further restaurant recommendations for groups!
1532
1533

1534
1535 **Debater** “For group logistics, did you consider the potential for traffic delays or the availability
1536 of large Ubers or private vans during peak times? Are there backup options if these aren’t avail-
1537 able?”, “While confirming dietary needs and interests in advance is wise, did you provide a clear
1538 method for gathering this information from the group?”, “For Jardim Botânico, did you check if
1539 English-speaking guides are always available on the spot, or should booking in advance be manda-
1540 tory during high season?”, “You mention group discounts at the gardens—do you have details on
1541 what constitutes a group, or if prior arrangement is needed for the discount?”, “When suggesting to
1542 bring water, hats, and sunscreen, do you have advice on where to purchase these locally if someone
1543 forgets?”, “For market visits, did you clarify the distance and transportation options between Jardim
1544 Botânico and each market? Some markets (like Feira de São Cristóvão) are far from the gardens.”,
1545 “Are there specific recommendations for group-friendly food stalls at these markets, or advice on
1546 seating availability for larger groups?”, “The artisanal shops and farmers’ market suggestions are
1547 good, but are these truly feasible for a group of six, or might they be too cramped or limited in of-
1548 ferings?”, “About booking food tours or chocolate workshops, can you specify reputable providers
1549 or whether these require significant advance notice?”, “For lunch, do you have recommendations in
1550 case the suggested restaurants are fully booked or not available for groups? For churrascarias, did
1551 you mention the price range or possible long wait times?”, “Regarding downtime, not all hotels have
1552 pools or spa packages—do you have alternatives for groups staying in budget accommodations?”,
1553 “For the sunset at Arpoador, are there safety concerns after dark, or tips on avoiding crowds, espe-
1554 cially for first-time visitors?”, “In the evening, are samba classes for groups commonly offered in
1555 English, and do they require advance reservations, or can they be arranged on short notice?”, “Your
1556 safety advice is solid, but is there a recommendation for how to handle valuables or what to do if
1557 someone gets lost?”, “The summary mentions flexibility to split up, but are there suggested ways
1558 to ensure everyone stays engaged and doesn’t feel left out if interests diverge?”, “Overall, while the
1559 plan is comprehensive, is there any advice for balancing structured activities with free time, or for
1560 dealing with group decision fatigue?”

1561 **Defender** Here are updated, practical answers to your group-focused questions for Day 4 and the
1562 overall itinerary:
1563

1564 1. Group Logistics & Transport: Yes, traffic delays are common, especially midday and late after-
1565 noon. For six people, large Ubers (UberXL), two standard Ubers/taxis, or a pre-booked van are your
main options. For peak times, book transport in advance via your hotel or apps like 99 Taxi. As a

- 1566 backup, use radio taxis (ask hotel front desk), or public transit if comfortable—Botanic Garden and
1567 Ipanema are on main bus lines. Always allow 20–30 extra minutes for traffic.
- 1568
- 1569 2. Gathering Dietary Needs/Interests: Before the trip or each day, send a group WhatsApp/Signal
1570 message or quick Google Form (simple tick-boxes: allergies, preferences, must-sees, energy level
1571 for the day). At breakfast, do a verbal check-in to confirm.
- 1572 3. Jardim Botânico Guides: English-speaking guides are sometimes available on the spot, but in
1573 high season (April is busy), book in advance via the official website or email (visit their site for
1574 contacts). Hotels can also arrange this.
- 1575 4. Group Discounts at Jardim Botânico: As of 2024, groups of 5+ may qualify for a discount,
1576 but prior arrangement is often required (book online or call ahead). Bring ID for all participants.
1577 Confirm discount conditions on the official website as policies can change.
- 1578 5. Buying Water, Hats, Sunscreen: Forgot something? Buy these at the Botanic Garden entrance
1579 kiosks, nearby pharmacies (Drogaria Pacheco, Raia), or supermarkets like Zona Sul (found in
1580 Ipanema, Jardim Botânico, Leblon). Prices are reasonable, and staff usually know basic English.
- 1581 6. Market Distances & Transport: Ipanema Hippie Fair is a 15-min drive from the Botanic Garden;
1582 Feira de São Cristóvão is 30–45 min by car (longer in traffic). For group comfort, use UberXL or a
1583 van. Public buses run between the Garden and Ipanema/Leblon, but not directly to São Cristóvão (a
1584 taxi or Uber is best).
- 1585 7. Group-Friendly Food Stalls/Seating: At Hippie Fair and São Cristóvão, look for larger food tents
1586 or open-air restaurants—many have communal tables and accept groups. For São Cristóvão, try
1587 “Barraca da Chiquita” (traditional food, group tables). At Hippie Fair, “Adega do Pimenta” and
1588 “Bar do Alemão” are group-friendly. Arrive early for lunch to secure seating.
- 1589 8. Feasibility of Artisanal Shops/Farmers’ Markets: Artisanal shops in Leblon/Ipanema are small
1590 but doable if you split up in pairs or browse at staggered times. Farmers’ markets are bustling but not
1591 cramped—groups can navigate if they stay flexible and don’t expect to sit. For a sit-down, prioritize
1592 markets with food courts (see above).
- 1593 9. Booking Food Tours/Chocolate Workshops: Reputable providers include Eat Rio (food tours),
1594 Rio4Fun (culinary experiences), and Choco.Rio (chocolate workshops). Book at least 48–72 hours
1595 in advance for a group of six. Confirm English-speaking guides.
- 1596 10. Lunch Backup Plans: Always have a Plan B and C. If Zazá Bistro, Da Bela, or a churrascaria
1597 is booked out, try Delírio Tropical (casual, healthy, many locations), Garota de Ipanema (classic,
1598 group tables), or Galeto Sat’s (local rotisserie chicken). Churrascaria prices: R80–R150/person, not
1599 including drinks; go early or reserve to avoid long waits.
- 1600 11. Downtime Alternatives for Budget Stays: If no hotel pool/spa, go to a public beach (Ipanema,
1601 Leblon), relax at Parque Lage’s café (beautiful garden setting), or find a local juice bar (Bibi Sucos,
1602 Polis Sucos) for a group hangout. Public parks are safe during daylight.
- 1603 12. Arpoador Sunset Safety & Crowds: Go before 5 PM to find a spot; leave by 7 PM. It’s safe in
1604 groups, but watch for pickpockets and avoid lingering after dark. Stick to the main path, avoid side
1605 alleys, and keep small bags zipped.
- 1606 13. Samba Classes for Groups: Many studios (e.g., Rio Samba Dancer, Casa de Dança Carlinhos
1607 de Jesus) offer group classes in English—book at least a day ahead, especially for private sessions.
1608 Some host drop-in classes, but reservations are safer for six people.
- 1609 14. Handling Valuables & Lost Persons: Each person carries only what they need. Use crossbody
1610 bags, keep phones away in crowds, and leave passports at the hotel safe. Share a group location via
1611 WhatsApp. If someone gets lost, agree on a meeting point (e.g., market entrance) or call the hotel
1612 for assistance.
- 1613 15. Engagement & Splitting Up: Use WhatsApp group chat for updates. If interests diverge, split
1614 into pairs/trios for an hour, then regroup for meals or main activities. Plan set meeting times/places,
1615 and encourage everyone to share their must-see picks ahead of time so no one feels left out.
- 1616
1617
1618
1619

1620 16. Balancing Structure & Free Time: Start with a flexible plan and build in 1–2 hours of “free
1621 roaming” or rest. Rotate who leads the day or chooses lunch. Review the next day’s plan together
1622 each evening to avoid decision fatigue and keep energy positive.

1623 Let me know if you need more details or sample WhatsApp group check-in messages!
1624

1625
1626
1627 **Debater:** “For group transport, do you have a contingency plan if UberXs or vans are unavail-
1628 able at peak times, or if there’s a shortage due to high demand?”, “When suggesting dietary needs
1629 collection, do you have an example or template for the Google Form or WhatsApp message to
1630 streamline this process?”, “Regarding Jardim Botânico guides, what is the process if an English-
1631 speaking guide is not available on the spot, and can the hotel guarantee booking one in advance?”,
1632 “For group discounts at Jardim Botânico, can you clarify what documentation is needed and how
1633 far in advance arrangements must be made to secure the discount?”, “While you mention where to
1634 buy essentials like water and sunscreen, are these locations reliably open early in the morning or
1635 late in the afternoon when visitors might arrive or leave?”, “Is traffic between the Botanic Garden
1636 and markets like Feira de São Cristóvão a significant concern at specific times, and should groups
1637 budget more time for these transfers?”, “You mention communal tables at markets, but are reserva-
1638 tions possible at these venues, or is seating strictly first-come, first-served?”, “For artisanal shops
1639 and farmers’ markets, will a group of six moving together disrupt the flow or experience, and are
1640 there recommendations for how to split up effectively?”, “Are the food tour and chocolate workshop
1641 providers verified for safety and quality, and do they cater specifically to English-speaking groups,
1642 especially during busy periods?”, “Regarding lunch plans, do your backup restaurants require reser-
1643 vations for groups, and are they located close to the main attractions to minimize travel time?”,
1644 “For downtime, are public beaches and parks safe for tourists in the late afternoon or early evening,
1645 and are there any security tips beyond general caution?”, “At Arpoador, are there any facilities like
1646 public toilets or refreshment stands, and is there a risk of overcrowding that might make the sunset
1647 experience less enjoyable?”, “Do samba class studios offer flexible scheduling for tourists, or do
1648 group classes require strict advance booking, especially for English-speaking instruction?”, “In the
1649 event someone gets lost, is there a recommended protocol beyond meeting points and WhatsApp,
1650 such as contacting local authorities or tourist assistance?”, “When splitting up, how can the group
1651 ensure everyone is included in decisions and no one feels isolated, especially if interests or energy
1652 levels differ widely?”, “For balancing structure and free time, what are some strategies to handle
1653 disagreements about pacing or activity choices to keep group morale high?”

1654 **Defender:** Here are detailed answers to your group travel questions for Day 4 (and relevant overall
1655 tips):

1656 1. Contingency Plan for Group Transport at Peak Times: - If UberXs/vans are unavailable, have
1657 a list of reliable radio taxi companies (e.g., Coopatur, Coopertáxi, Taxi Leblon). Ask your hotel for
1658 the numbers or to call on your behalf. - Consider splitting into two regular Ubers/taxis if needed
1659 (pre-agree on destination). - For longer trips (like to Feira de São Cristóvão), ask the hotel to book a
1660 private van or car service a day ahead. - Public transit is an option (bus lines 409, 439 from Botanic
1661 Garden to Ipanema/Leblon), but for six people, taxis/Ubers are faster and safer.

1662 2. Dietary Needs Collection Template: - WhatsApp Example: “Hi everyone! Please reply with any
1663 food allergies, dietary restrictions, or strong dislikes (e.g., nuts, gluten, vegetarian, etc.). Also, let
1664 us know any food you’re excited to try!” - Google Form Example Fields: Name / Allergy (Y/N,
1665 list if yes) / Dietary Preference (Vegan, Vegetarian, No Pork, etc.) / Foods to Avoid / Foods to Try /
1666 Emergency Contact.

1667 3. Jardim Botânico English Guide Process: - Hotel can request/book an English-speaking guide 2–5
1668 days in advance via the Jardim Botânico website or by phone. Same-day walk-in is risky for English
1669 guides, especially in busy season. - If not available, ask the hotel to arrange a freelance guide, or use
1670 an app like ToursByLocals or GetYourGuide.

1671 4. Documentation for Group Discount: - Typically, you’ll need a list of names and ID num-
1672 bers. Some sites may require passports (copies are fine). Book at least 3–5 days in advance for
1673 groups—hotel can assist. Policies can change, so confirm directly with the Jardim Botânico.

- 1674 5. Buying Essentials Early/Late: - Pharmacies (Drogasil, Pacheco, Raia) in Ipanema/Jardim
1675 Botânico open 7am–10pm. Zona Sul supermarket opens 7am–10pm. Beach kiosks open by 8am,
1676 close late evening. Most markets and tourist areas have at least one early/late option.
1677
- 1678 6. Traffic Between Garden & Markets: - Morning/lunch: 20–30 min to Ipanema, up to 1 hour to
1679 São Cristóvão (worse 4–6 pm weekdays). Budget extra time for traffic, especially Fri–Sun. For São
1680 Cristóvão, consider going before 3pm or after 7pm to avoid peak jams.
1681
- 1682 7. Market Seating/Reservations: - Most market food tents (Hippie Fair, São Cristóvão) are first-
1683 come, first-served; no reservations. Arrive early for group seating. Larger restaurants inside São
1684 Cristóvão may accept reservations by phone (often only in Portuguese).
1685
- 1686 8. Group Flow at Shops/Markets: - Six is a large group for crowded stalls. Suggest splitting into
1687 pairs/trios with a set time/place to regroup (e.g., 30 min shopping, meet at juice bar or main en-
1688 trance). Share your locations via WhatsApp.
1689
- 1690 9. Food Tour/Workshop Safety & Language: - Stick to highly rated providers (Eat Rio, Rio4Fun,
1691 Choco.Rio) with reviews on TripAdvisor or GetYourGuide. Confirm English language and group
1692 size when booking. For busy periods, book 4–7 days ahead.
1693
- 1694 10. Lunch Reservations & Location: - Backup restaurants (e.g., Delírio Tropical, Garota de
1695 Ipanema, Galeto Sat’s) accept reservations for groups, especially on weekdays. All are within 10–15
1696 min of main sights. Call or use Google/Instagram to reserve.
1697
- 1698 11. Beach/Park Safety Late Afternoon/Evening: - Main beaches (Ipanema, Copacabana) and parks
1699 (Parque Lage) are safe until sunset, especially in a group. Avoid isolated areas after dark, and keep
1700 valuables hidden. Stay near lifeguard posts or kiosks with staff.
1701
- 1702 12. Arpoador Facilities & Crowds: - Public toilets and refreshment stands available (small fee for
1703 toilets). Sunset is very popular—arrive 1+ hour early for a good spot, or consider nearby Ipanema
1704 Posto 8 if overcrowded. Beware of pickpockets in crowds.
1705
- 1706 13. Samba Class Booking: - Most studios (e.g., Rio Samba Dancer, Casa de Dança Carlinhos de
1707 Jesus) offer flexible scheduling for tourists, but for a group and English instruction, book at least 48
1708 hours ahead. Some offer private group classes.
1709
- 1710 14. Lost Person Protocol: - Pre-arrange a meeting point at each location. Keep phones charged
1711 and share live location in WhatsApp. If someone is lost and unreachable, contact the venue’s
1712 staff/security, or call Rio’s tourist police (Dial 190 for police, 2253-1177 for tourist assistance,
1713 English spoken).
1714
- 1715 15. Inclusive Group Decisions: - Start each day with a brief meeting to gather everyone’s priorities
1716 and energy levels. Use a poll (WhatsApp or verbally) for major decisions. Pair up those with similar
1717 interests if splitting up. Regroup for group meals and check-ins.
1718
- 1719 16. Handling Disagreements/Pacing: - Build in regular breaks and “optional” periods for rest or solo
1720 exploring. Rotate leadership or decision-making. If there’s a dispute, defer to a coin toss or majority
1721 vote, and keep the mood light—remind everyone the goal is shared enjoyment, not rigid schedules.
1722

1715 **Host: **Day 4: Jardim Botânico & Markets (Group Day – Six People)****

1716 ****Group Logistics & Transport**** - Arrange group-friendly transport: Book an UberXL, two
1717 Ubers/taxis, or a private van (your hotel can assist). Confirm a morning meeting point and time. Al-
1718 low 20–30 extra minutes for traffic, especially midday. - Create a WhatsApp group for coordination,
1719 share live locations, and agree on emergency meeting points at each stop.
1720

1721 ****Morning: Jardim Botânico**** - Arrive early (opens 8 AM) to enjoy cooler weather and avoid
1722 crowds. - Buy tickets together for possible group discount (check requirements on the official web-
1723 site; bring ID or passport copies). - Book an English-speaking guide in advance (via the Jardim’s
1724 website or your hotel) for a more engaging visit. - The gardens are mostly accessible, but there’s
1725 moderate walking—remind everyone to bring water, hats, and sunscreen (available at nearby phar-
1726 macies if needed).
1727

1727 ****Late Morning: Market Visit**** - ****Sunday:**** Head to the Feira Hippie de Ipanema (15 min
drive), ideal for groups with spacious stalls, food, and shopping. Arrive early for best seating at

1728 food tents. - **Weekdays**: Consider Feira de São Cristóvão (30–45 min drive), a lively, covered
1729 Northeastern Brazilian market with group-friendly restaurants (e.g., Barraca da Chiquita). For a
1730 smaller, upscale vibe, try artisanal shops in Leblon or the Ipanema Farmers' Market (Fridays). -
1731 If interested in food experiences, book a group food tour or chocolate tasting workshop (Eat Rio,
1732 Rio4Fun, Choco.Rio). Reserve 48–72 hours ahead and confirm English-speaking guides.

1733 **Lunch** - Reserve a table for six in advance. Top options: - **Zazá Bistro** (Ipanema; excel-
1734 lent for dietary needs) - **Da Bela** (near Jardim Botânico; farm-to-table) - **Fogo de Chão**
1735 (Botafogo; festive churrascaria) - **Alternatives**: Delírio Tropical, Garota de Ipanema, Galeto
1736 Sat's (all group-friendly, call ahead) - Confirm allergy accommodations and dietary restrictions (col-
1737 lect info via WhatsApp or a simple Google Form before the day).

1738 **Afternoon: Downtime & Social Time** - After lunch, enjoy downtime together: hotel pool,
1739 a relaxed walk on the Ipanema or Copacabana boardwalk, or a group spa session (many hotels
1740 offer packages). - For budget stays, relax at Parque Lage's café or a local juice bar (Bibi Sucos,
1741 Polis Sucos). - For sunset, head to Arpoador (arrive by 5 PM for space; safe in groups, but watch
1742 valuables and avoid lingering after dark).

1743 **Evening (Optional)** - If energy allows, book a private group samba class (e.g., Rio Samba
1744 Dancer, Casa de Dança Carlinhos de Jesus; reserve 48 hours ahead). - Or, unwind at a casual bar or
1745 rooftop in Ipanema or Leblon (reserve for a group this size).

1746 **Group Safety & Coordination** - Stay together in crowded markets; split into pairs/trios if
1747 needed, with set regrouping times/locations. - Carry only essentials (crossbody bags, no passports),
1748 keep phones secure, and bring small bills for markets/tips. - Review emergency meeting points at
1749 each stop. - Share live locations in WhatsApp and agree on what to do if someone gets separated
1750 (meet at entrance, contact hotel/tourist police if needed).

1751 **Flexibility & Engagement** - Build in 1–2 hours of free time for solo exploring or rest. - Rotate
1752 who leads the day or chooses lunch. - Start each day with a quick group check-in on energy, interests,
1753 and must-sees. - If interests diverge, split for short periods and regroup for meals or main activities.

1754 —

1755 —

1756 **Summary of Adjustments for Day 4** - All activities are now optimized for a group of six:
1757 transport, tours, meals, downtime, and safety. - Emphasis on advance reservations for guides, tours,
1758 and restaurants. - Market and food options tailored for group tastes and dietary needs. - Practical
1759 tips for group safety, communication, and flexibility. - Built-in options for splitting up or relaxing
1760 together as needed.

1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781