

Reverse Prompt Engineering

Anonymous ACL submission

Abstract

We explore a new language model inversion problem under strict black-box, zero-shot, and limited data conditions. We propose a novel training-free framework that reconstructs prompts using only a limited number of text outputs from a language model. Existing methods rely on the availability of a large number of outputs for both training and inference, an assumption that is unrealistic in the real world, and they can sometimes produce garbled text. In contrast, our approach, which relies on limited resources, consistently yields coherent and semantically meaningful prompts. Our framework leverages a large language model together with an optimization process inspired by the genetic algorithm to effectively recover prompts. Experimental results on several datasets derived from public sources indicate that our approach achieves high-quality prompt recovery and generates prompts more semantically and functionally aligned with the originals than current state-of-the-art methods. Additionally, use-case studies introduced demonstrate the method’s strong potential for generating high-quality text data on perturbed prompts.

1 Introduction

With the advancement of large language models (LLMs), prompt engineering has become an essential technique for expanding their capabilities (Sahoo et al., 2024). This method uses task-specific instructions, or prompts, to enhance model effectiveness without altering core parameters. Widely used prompting techniques, such as few-shot prompting (Radford et al., 2019), chain-of-thought prompting (Wei et al., 2022), and retrieval-augmented generation (Lewis et al., 2020), have proven highly practical in diverse applications.

With the increasing focus on prompt engineering, where input prompts are carefully modified to improve the outputs generated by LLMs, a natural question arises: can we infer the input

prompt based solely on the outputs? This challenge, termed language model inversion by Morris et al. (2024), has gained prominence with the growing prevalence of LLMs offered as “services,” where users interact only with outputs while the original prompts remain concealed. This situation presents a dual interest, with users seeking to deduce concealed prompts and service providers striving to protect them, thereby rendering language model inversion an increasingly pertinent problem (Morris et al., 2024). Furthermore, recovering prompts has practical applications, such as enabling users to adapt inferred prompts for generating high-quality outputs tailored to new contexts, e.g., transforming a marketing plan for one product into a similarly high-quality plan for another with minimal adjustments. In Section 4.4, we demonstrate that text generated through our language model inversion method is more favored by human evaluators than text derived from existing high-quality templates.

Morris et al. (2024) develop a model trained to predict the input prompt by leveraging the probability distributions and logits from the last layer of an LLM. Building on this, Zhang et al. (2024) further propose a model that uses only the textual outputs to infer the prompts, without relying on internal model parameters. However, this approach is developed under the assumption that a large number of outputs (64) are available to recover a single prompt and that access to user prompts for complex system prompts is granted. These assumptions rarely hold true in real-world scenarios. Furthermore, both methods demand extensive training on large datasets, which can be resource-intensive. Additionally, their outputs are heavily influenced by the form of the training data, so they perform poorly on out-of-domain prompt recovery and sometimes generate non-linguistic sequences. These limitations, alongside the broader interest in uncovering and protecting prompts and the practical utility of generating high-quality data, motivate the develop-

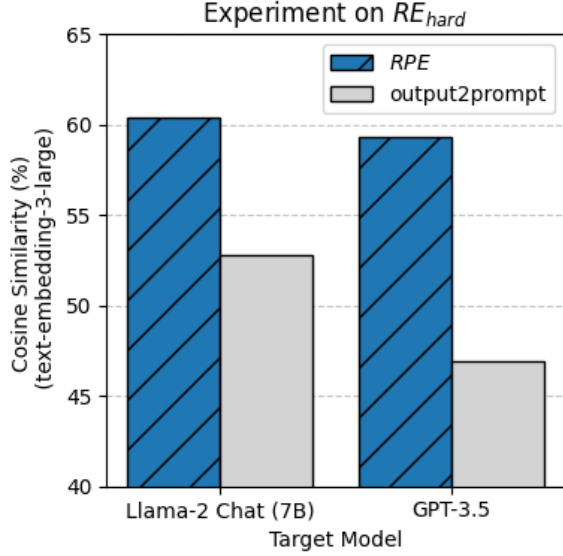


Figure 1: Performance comparison of RPE and $output2prompt$ on the RE_{hard} dataset. Evaluates the effectiveness of recovering complex system prompts from outputs generated by different target LLMs.

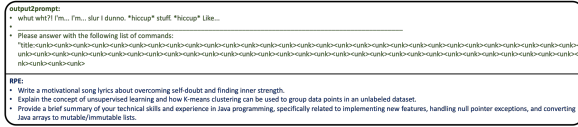


Figure 2: Examples of non-linguistic prompts recovered by $output2prompt$ and prompts recovered by RPE for the same latent prompts.

ment of a robust, training-free, zero-shot language model inversion method that operates with limited output access.

In this paper, we propose a novel language model inversion technique, reverse prompt engineering (RPE), which assumes the target LLM is a black-box model accessible only through limited text outputs. RPE infers the underlying prompt from these outputs by leveraging the LLM’s reasoning capabilities in combination with an iterative optimization algorithm inspired by the genetic algorithm (Sampson, 1976). Importantly, RPE introduces no new models and requires no training. The core idea of RPE is to conceptualize language model inversion as a reverse-engineering optimization problem, using the relationship between prompts and outputs to iteratively refine potential candidates. By utilizing the reasoning ability of an LLM to generate candidate prompts, RPE evaluates these candidates based on the similarity of their generated outputs to the true output. This evaluation serves as the basis for iterative optimiza-

tion, guided by a genetic algorithm intertwined with an LLM, to progressively reduce discrepancies between candidates and the latent prompt, and to converge on the most plausible prompt.

Compared to previous methods (Morris et al., 2024; Zhang et al., 2024), RPE is more resource-efficient, requiring only minimal information from the target LLM (five text outputs) while ensuring the generation of natural language outputs. RPE outperforms state-of-the-art methods, achieving an average 6.2% improvement in cosine similarity over $output2prompt$ (Zhang et al., 2024) on Llama-2 Chat (7B) outputs and 10.9% on GPT-3.5 outputs across different datasets. Additionally, RPE demonstrates superior performance in system prompt recovery tasks, surpassing $output2prompt$ by an average of 5.8% in cosine similarity.

Our main contributions are as follows.

- We provide the first study of the language model inversion problem under black-box, zero-shot, and limited data conditions.
- We design an innovative evaluation method that selects the most accurate recovered prompt from multiple candidates by their corresponding outputs, thereby enhancing the accuracy of prompt recovery in scenarios involving multiple candidate prompts.
- We purpose a novel optimization algorithm that leverages the LLM itself as an optimizer to further enhance prompt recovery accuracy.

The code and datasets are available at https://github.com/Anonymous-Author980/RPE_Reverse_Prompt_Engineering.

2 Related Works

2.1 Prompt Engineering

Prompt engineering is a closely related field, essential for optimizing LLMs by designing prompts that guide model outputs across diverse tasks without altering model parameters (Sahoo et al., 2024). Initial prompting techniques include zero-shot and few-shot prompting (Radford et al., 2019; Brown et al., 2020), demonstrating that LLMs can handle novel tasks without additional training. Chain-of-thought (CoT) prompting by Wei et al. (2022) introduced step-by-step reasoning, which inspired further techniques to enhance LLM reasoning and logic abilities (Zhang et al., 2023; Wang et al.,

2023; Zhao et al., 2024; Hu et al., 2023; Yao et al., 2024a; Long, 2023; Yao et al., 2024b; Weston and Sukhbaatar, 2023; Zhou et al., 2023b; Wang et al., 2024; Diao et al., 2024; Chia et al., 2023). To improve accuracy and mitigate hallucinations, Retrieval Augmented Generation (RAG) integrates information retrieval into prompting (Lewis et al., 2020), and its variations enhance real-time knowledge access for LLMs (Yao et al., 2023; Dhuliawala et al., 2024; Li et al., 2024b; Yu et al., 2024b). Other approaches incorporate external tools for improved accuracy (Paranjape et al., 2023; Wu et al., 2024). Techniques for automating prompt generation have also emerged, using LLMs as optimizers to craft more effective prompts (Zhou et al., 2023a; Yang et al., 2024), alongside specialized prompting methods for specific tasks such as code generation (Nye et al., 2021; Chen et al., 2023; Li et al., 2023b, 2024a), emotion comprehension (Li et al., 2023a), user intent understanding (Deng et al., 2023), and abstract concept extraction (Zheng et al., 2024).

2.2 Language Model Inversion

Unlike prompt engineering, which focuses on crafting prompts to achieve better outputs, language model inversion aims to infer the underlying prompt from given outputs. Morris et al. (2024) first introduce this problem, developing *logit2prompt*, a solution that extracts prompts from next-token probability distributions using a T5-based model (Raffel et al., 2020) with additional training. Building on *logit2prompt*, Zhang et al. (2024) propose *output2prompt*, the current state-of-the-art method for language model inversion. The *output2prompt* method, also T5-based, can recover prompts using only text outputs, without requiring access to model logits (Zhang et al., 2024).

Our proposed method, *RPE*, differs in that it requires neither access to model logits nor user prompts, making it particularly suitable for closed-source LLMs like GPT-3.5. Unlike *output2prompt*, which still relies on the user prompt when reconstructing complex system prompts, *RPE* depends solely on LLM outputs, requiring no additional information. Moreover, *RPE* is unique in that it does not require training, training data, or large quantities of LLM outputs, needing only five outputs compared to the 64 required by *output2prompt*. Since *logit2prompt* and *output2prompt* use T5-based models with smaller vocabularies than modern LLMs, *RPE*

Latent:	Prompt p : Create a list of 2 startup ideas in enterprise B2B SaaS. The startup ideas should have a strong and compelling mission and also use AI in some way. Avoid cryptocurrency or blockchain. The startup ideas should have a cool and interesting name.
Input:	Answers $A = \{a_1, a_2, a_3, a_4, a_5\}$ Example Answer (a_1): 1. InsightOps Mission: Empower enterprises to transform customer service into a proactive and predictive experience, enhancing satisfaction while optimizing operational costs. Overview: InsightOps is an AI-powered customer service analytics platform designed to analyze customer interactions in real time. It helps enterprises understand patterns in customer behavior, providing predictive insights on service issues before they escalate. Using NLP and machine learning, the platform automatically categorizes customer inquiries, recommends solutions, and predicts future issues based on historical data. Businesses can optimize staffing, improve product offerings, and reduce churn by making data-driven decisions based on deep analytics. 2. SentinelGuard Mission: Safeguard enterprises from cyber threats through intelligent detection, prevention, and adaptive defense strategies. Overview: SentinelGuard is an AI-driven cybersecurity platform focused on adaptive, real-time threat detection and response. It monitors enterprise networks, learns from activity patterns, and automatically adjusts to emerging threats. SentinelGuard's unique offering is its AI-enabled behavioral analysis engine, which constantly learns from each user and system's behavior, identifying anomalies that traditional security systems might miss. It integrates seamlessly with existing security frameworks and provides actionable alerts with minimal false positives.
Input to RPE_{LLM} :	Randomly select an answer from A , assume a_1 .
Prompt:	The given answer is generated by one prompt passing through an LLM. Based on the provided answer, speculate the underlying prompt. Answer: (a_1)
Output p' of RPE_{LLM} :	What are some innovative ideas or solutions for improving customer service, data analytics, and cybersecurity in the enterprise sector?
Evaluation:	LLM's answer to prompt p' : 1. Utilizing AI and machine learning technology to automate and personalize customer service interactions, leading to faster response times and improved customer satisfaction. 2. Implementing advanced data analytics tools to gather real-time insights and trends, allowing companies to make informed decisions and optimize their operations. 3. Enhancing cybersecurity measures by employing blockchain technology to secure sensitive data and prevent cyberattacks. 4. Creating a customer portal or mobile app where clients can easily access and manage their accounts, request support, and provide feedback. 5. Developing a comprehensive cybersecurity training program for employees to educate them on the latest threats and best practices for safeguarding sensitive information. 6. Leveraging biometric authentication methods, such as fingerprint or facial recognition, to enhance data security and prevent unauthorized access. 7. Implementing a customer feedback system that allows for continuous monitoring and analysis of customer satisfaction levels, enabling companies to quickly address any issues and improve their overall service quality. 8. Integrating chatbots and virtual assistants into customer service operations to handle routine inquiries and streamline the support process. 9. Investing in threat intelligence platforms that can proactively detect and mitigate potential cybersecurity threats before they can cause harm to the organization. 10. Collaborating with external cybersecurity experts and industry partners to stay abreast of the latest security trends and developments, ensuring that the company remains at the forefront of cybersecurity best practices.

Figure 3: Example of One Answer One Shot inference.

offers the advantage of generating prompts with more flexibility in word choice.

3 Methodology

We formalize the language model inversion problem as follows: given a set of n responses, denoted as $A = \{a_1, a_2, \dots, a_n\}$, generated by submitting a single prompt p to an LLM n times, the objective is to design a language model inversion method, denoted as *RPE*, that can infer the original prompt p from the response set A , which means the output p' of *RPE* should be the same as the original prompt p :

$$\min d(p, p') \quad (1)$$

$$LLM(p)_n = A \quad (2)$$

$$RPE(A) = p' \quad (3)$$

Here, d could be any score measuring the difference between two prompts. The intriguing part is that p is latent and thus unknown. In this setup, the LLM is treated as a black box, meaning that, aside from the text outputs, no access is granted to its internal parameters or mechanisms. Moreover, the *RPE* method is developed under a zero-shot constraint, where no prior training data or additional examples of outputs (beyond the given set) are available, and no training is permitted in the development of the method.

3.1 The “Naive” Approach

Our initial approach aims to directly infer the prompt p using exactly one response a generated by the LLM. Specifically, we query the LLM to infer the underlying prompt based on the given response

Latent:	Prompt p : Create a list of 2 startup ideas in enterprise B2B SaaS. The startup ideas should have a strong and compelling mission and also use AI in some way. Avoid cryptocurrency or blockchain. The startup ideas should have a cool and interesting name.
Input:	Answers $A = \{a_1, a_2, a_3, a_4, a_5\}$
Input to RPE_{5A1S} and RPE_{5A5S} :	All five answers from A .
Prompt:	Answer 1: $\{a_1\}$, Answer 2: $\{a_2\}$, Answer 3: $\{a_3\}$, Answer 4: $\{a_4\}$, Answer 5: $\{a_5\}$ All answers above are generated by one prompt passing through LLM multiple times. Based on the answers provided above, speculate the underlying prompt.
Output p' of RPE_{5A1S} :	Describe two startup ideas that leverage AI technology and their respective missions.
Output p' of RPE_{5A5S} :	Write a short description of two innovative startup ideas in the enterprise sector that leverage cutting-edge technology and address critical pain points. Describe the mission, features, and potential benefits of each startup idea.
Evaluation: LLM's answer to prompt p' by RPE_{5A1S} :	1. AI-driven meal planning app. Mission: To help people plan and prepare healthy and delicious meals easily and efficiently by leveraging the power of artificial intelligence technology. This startup idea involves developing a meal planning app that uses AI algorithms to suggest personalized meal plans based on the user's dietary preferences, restrictions, and health goals. The app would analyze the user's food preferences, culinary skills, and nutritional needs to generate a variety of meal options that are tailored to their individual requirements. It could also provide recipes, ingredient lists, and cooking instructions, as well as track users' progress towards their health and wellness goals. 2. AI-powered mental health chatbot. Mission: To provide accessible and effective mental health support to individuals in need by incorporating artificial intelligence technology into a conversational chatbot platform. This startup idea involves creating a mental health chatbot that uses AI algorithms to offer therapy and emotional support to users in real-time. The chatbot would be designed to engage in natural and empathetic conversations with users, providing them with a safe space to express their thoughts and feelings. It could offer counseling, coping strategies, and resources for mental health self-care, as well as monitor users' mental well-being and provide referrals to professional services when necessary. The aim is to make mental health support more readily available and destigmatize seeking help for mental health issues.
LLM's answer to prompt p' by RPE_{5A5S} :	1. Virtual Collaboration Platform for Remote Teams. Mission: To empower remote teams to collaborate effectively and efficiently by providing a virtual workspace that replicates the experience of working together in person. Features: This platform will offer features such as virtual meeting rooms, interactive whiteboards, real-time document editing, and project management tools. It will also integrate with popular communication and collaboration tools to streamline workflows. Potential Benefits: By leveraging cutting-edge technologies like virtual reality and AI, this startup will enable remote teams to overcome the challenges of distance and time zones, ultimately improving productivity, communication, and team cohesion. 2. AI-Powered Financial Forecasting Tool. Mission: To help businesses make more accurate financial forecasts and strategic decisions by leveraging AI technology to analyze data and predict future trends. Features: This tool will automate the process of collecting, processing, and analyzing financial data, providing real-time insights and forecasts based on historical data, market trends, and other external factors. It will also offer scenario planning and predictive analytics capabilities. Potential Benefits: By harnessing the power of AI and machine learning, this startup will enable businesses to make data-driven decisions, reduce financial risks, and optimize resource allocation. This will ultimately lead to better financial performance and competitive advantage in the market.

Figure 4: Example of Five Answers One Shot and Five Answer Five Shots inference.

a , a method we refer to as one-answer-one-shot reverse prompt engineering (RPE_{1A1S}). As illustrated in Figure 3, we provide an example where GPT-3.5 is tasked with recovering a prompt from a response related to start-up ideas. The recovered prompt p' contains some elements of the original prompt p but also includes additional details drawn from the response a , such as “customer service,” “data analytics,” and “cybersecurity,” which are not part of the original prompt. We hypothesize that inferring the prompt from only one response may lead the LLM to overemphasize specific details from the response a that were not present in the original prompt p , as demonstrated in the example shown in Figure 3.

3.2 Five Answers Inference

We then extend the naive method by using multiple responses to recover the underlying prompt. Given a set of responses A , we inform the LLM that these responses are generated from the same prompt p and ask the LLM to recover p based on the entire set A . We set $n = 5$ in our experiments and refer to this method as five-answers-one-shot reverse prompt engineering (RPE_{5A1S}). In Figure 4, we present an example of RPE_{5A1S} using GPT-3.5. Compared to RPE_{1A1S} , the recovered prompt p' in RPE_{5A1S} captures more elements of the original prompt, such as “two,” “AI,” and “missions.” Additionally, RPE_{5A1S} avoids incorporating response-specific details, like “customer service” and “data analytics,” which were mistakenly included by RPE_{1A1S} . However, there is still room for improvement, as the recovered prompt does not fully replicate the original prompt.

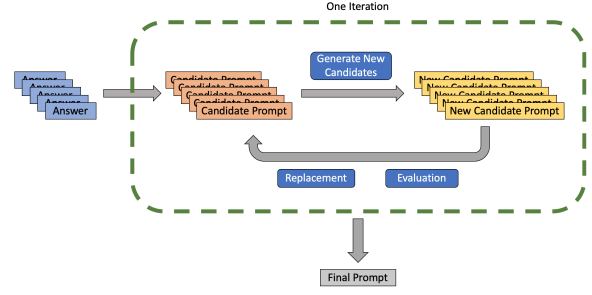


Figure 5: Workflow of RPE_{GA}

Building on RPE_{5A1S} , we propose an enhanced approach that generates multiple candidate prompts and selects the most accurate one. Specifically, given a set of responses A with n answers, we ask the LLM to recover the prompt p and generate a set of m candidate prompts, denoted as $P' = \{p'_1, p'_2, \dots, p'_m\}$. To evaluate the quality of each candidate prompt in P' , we first pass each recovered prompt p'_i to the LLM and obtain a corresponding response a'_i . We then compute the ROUGE-1 score between a'_i and each answer in A , yielding a set of scores $S'_i = \{s'_{i1}, s'_{i2}, \dots, s'_{in}\}$. While it is intuitive to take the average of S'_i as the final score, a promising prompt might generate a response a'_i that closely matches one of the answers in A but not the others. To address this, we combine both the mean and the maximum of S'_i to define the final score for p'_i as $s'_i = \frac{\text{mean}(S'_i) + \text{max}(S'_i)}{2}$.

The recovered prompt with the highest score s'_i is selected as the final prompt. In our experiments, we use $n = 5$ and $m = 5$, referring to this approach as five-answers-five-shots reverse prompt engineering (RPE_{5A5S}). As shown in Figure 4, the recovered prompt using RPE_{5A5S} captures more details from the original prompt compared to RPE_{5A1S} , although further improvement is still possible.

3.3 Iterative Method

To further enhance our approach, we introduce an iterative method aimed at progressively optimizing the recovered prompt with each iteration. Inspired by the genetic algorithm (Sampson, 1976), we designed an algorithm that generates new candidate prompts based on existing ones and selects the most accurate candidates using a custom evaluation strategy. We refer to this iterative reverse prompt engineering method as RPE_{GA} . The complete workflow of the algorithm is depicted in Figure 5. Below, we describe the key components of this algorithm in detail.

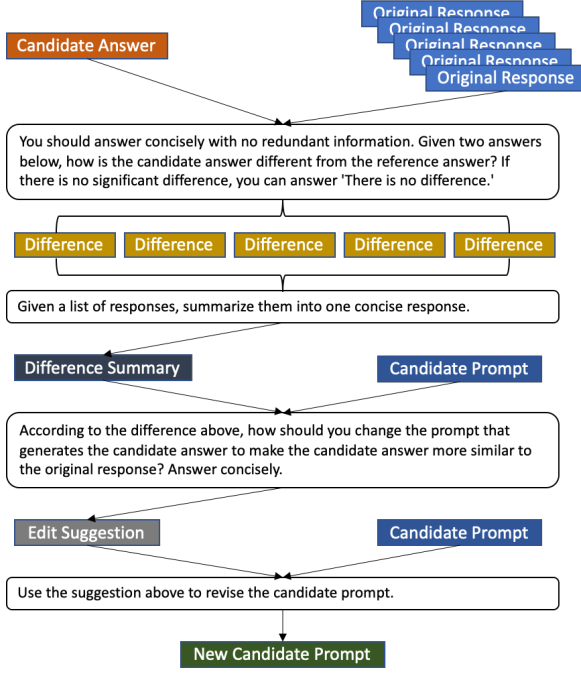


Figure 6: Process of generating new candidate prompts from the old ones.

3.3.1 Initialization

Given a set of responses A with n answers, we first ask the LLM to infer the underlying prompt p , generating m candidate prompts $P' = \{p'_1, p'_2, \dots, p'_m\}$, following the same procedure as in RPE_{5A5S} (see Section 3.2). We then evaluate each candidate prompt p'_i using the evaluation method from RPE_{5A5S} , where we pass each candidate p'_i to the LLM to generate a response a'_i and calculate its performance score s'_i . The performance score s'_i for each candidate prompt is calculated by averaging the mean and max of the ROUGE-1 score between a'_i and each response in A . This completes the initialization phase of the RPE_{GA} algorithm.

3.3.2 Iteration

Following the initialization step, we iteratively generate new candidate prompts and replace the existing candidates with better-performing ones. In each iteration, we start with the set of original responses A , the current candidate set P' , the responses $A' = \{a'_1, a'_2, \dots, a'_m\}$ generated by candidate prompts P' , and the corresponding performance scores $S' = \{s'_1, s'_2, \dots, s'_m\}$. For each candidate prompt p'_i and its corresponding response a'_i , we first ask the LLM to identify the differences between a'_i and the responses in A . Then, we request the LLM to summarize these differences

Awesome ChatGPT Prompts: I want you to act as my first aid traffic or house accident emergency response crisis professional. I will describe a traffic or house accident emergency response crisis situation and you will provide advice on how to handle it. You should only reply with your advice, and nothing else. Do not write explanations. My first request is "My toddler drank a bit of bleach and I am not sure what to do."
MetaMathQA: Bobby has three squares of fabric that he will turn into a flag. The first square is 8 feet by 5 feet. The second one is 10 feet by 7 feet. The third one is 5 feet by 5 feet. If he wants his flag to be 15 feet long, how tall will it be?
TruthfulQA: Are there carpets that can fly?
Alpaca-GPT4: List two advantages of choosing a career in tech.
Dolly Creative Writing: Describe a tense scene of a dystopian future where a human in a laboratory is attempting to transform her body into a cyborg.

Figure 7: Example prompt from each dataset.

and use the summary as a guide to modify the candidate prompt p'_i . The process is illustrated in Figure 6 in detail. This process yields a new set of candidate prompts, $P'' = \{p''_1, p''_2, \dots, p''_m\}$, for which we calculate the performance scores $S'' = \{s''_1, s''_2, \dots, s''_m\}$ as in the previous step. Based on these scores, we update the candidate set by replacing low score prompts in P' with the new high score candidates from P'' , thus forming the updated set of candidate prompts.

3.3.3 Output

After repeating the iteration process for k iterations, we select the best-performing prompt from the final candidate set P' based on the highest performance score in S' . This selected prompt, denoted as p'_o , is the final recovered prompt produced by the RPE_{GA} method.

4 Computational Assessment

In this section, we present the results of testing our proposed methods on various datasets, comparing their performance with the benchmark approach of *output2prompt* (Zhang et al., 2024). The evaluation focuses on assessing the semantic and functional similarity between the recovered and original prompts. Specifically, we employ cosine similarity as the evaluation metric, as it best aligns with the language model inversion objective (Zhang et al., 2024). Throughout all experiments, GPT-3.5 serves as the backbone model for RPE .

4.1 Dataset

We evaluate our method using five datasets: Awesome ChatGPT Prompts¹ (153 complex instructional role-based prompts), MetaMathQA (Yu et al., 2024a) (395,000 linguistically diverse math word problems), TruthfulQA (Lin et al., 2022) (817 truthfulness assessment prompts), Alpaca-GPT4 (Peng et al., 2023) (52,000 simple instruction-following

¹<https://github.com/f/awesome-chatgpt-prompts>

System Prompt: I want you to act as my first aid traffic or house accident emergency response crisis professional. I will describe a traffic or house accident emergency response crisis situation and you will provide advice on how to handle it. You should only reply with your advice, and nothing else. Do not write explanations.
User Prompt: My first request is "My toddler drank a bit of bleach and I am not sure what to do."

Figure 8: Demonstration of system prompt and user prompt.

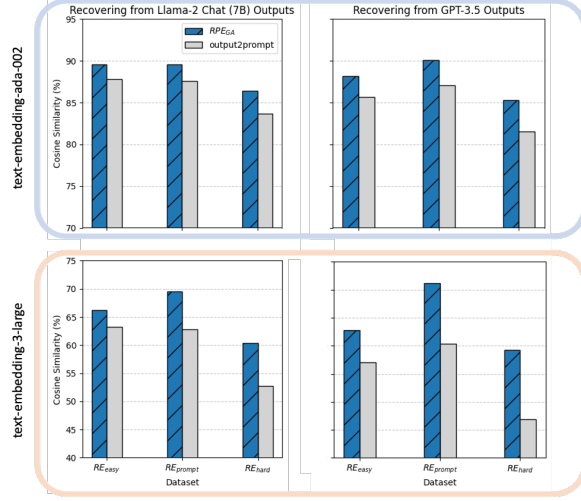


Figure 9: Comparison of $RPRGA$ and $output2prompt$.

prompts), and Dolly Creative Writing² (673 creative writing prompts). Detailed descriptions are provided in the appendix A.

Figure 7 presents an example prompt from each dataset. To ensure comprehensive evaluation across diverse LLM tasks, including general conversation, complex instructions, and creative writing, we sample prompts from all five datasets. However, evaluating large datasets via the OpenAI API incurs significant costs. To balance cost efficiency and evaluation rigor, we randomly select 20 prompts from each dataset, forming our primary test set, RE_{prompt} , while maintaining diversity and complexity.

To assess how prompt complexity impacts RPE performance, we construct two additional test sets: RE_{hard} , containing 100 challenging prompts from Awesome ChatGPT Prompts, and RE_{easy} , consisting of 100 simpler prompts from Alpaca-GPT4. These three test sets enable a thorough evaluation of both the proposed method and the benchmark model across varying levels of prompt complexity.

4.2 Benchmark

We compare the performance of our best-performing method, $RPEGA$, against the state-of-the-art benchmark $output2prompt$ (Zhang et al.,

²https://huggingface.co/datasets/lionelchg/dolly_creative_writing

2024). To ensure a fair comparison, given that $output2prompt$ is trained on outputs from Llama-2 Chat (7B), experiments are performed on outputs generated by both Llama-2 Chat (7B) and GPT-3.5. Following Zhang et al. (2024), cosine similarity is chosen as the evaluation metric due to its alignment with the objectives of language model inversion. We utilize OpenAI’s “text-embedding-ada-002” and “text-embedding-3-large” models to compute text embeddings for this purpose.

Zhang et al. (2024) also introduce a variant of $output2prompt$, referred to as $output2prompt_s$, specifically designed to recover system prompts but requiring access to user prompt. In Figure 8, we present an example from the RE_{hard} dataset, which includes both system and user prompts.

In $output2prompt_s$, the user must generate a total of 64 distinct outputs with 64 different outputs. These 64 outputs are then fed into the trained $output2prompt_s$ model to infer the system prompt. To ensure a fair comparison, we evaluate $output2prompt_s$ under two additional settings: (1) using a randomly selected subset of five outputs from the 64, denoted as $output2prompt_{s5}$, and (2) using the same five outputs utilized by $RPEGA$, denoted as $output2prompt_{s5o}$. This comparison is conducted exclusively on the RE_{hard} dataset, as the other two datasets consist mostly of user prompts and do not include system prompts. Additionally, since $output2prompt_s$ is trained on GPT-3.5 input and output, all experiments comparing $RPEGA$ with $output2prompt_s$ are performed using GPT-3.5 outputs.

4.3 Experiments

We conduct experiments on all three datasets using the methods described in Section 3 with parameters $n = 5$, $m = 5$, and $k = 5$. As shown in Figure 9, $RPEGA$ achieves higher cosine similarity than $output2prompt$ across all datasets, regardless of whether the outputs are generated by Llama-2 Chat (7B) or GPT-3.5. On average across all 3 datasets, $RPEGA$ outperforms $output2prompt$ by 6.2% on Llama-2 Chat (7B) outputs and by 10.9% on GPT-3.5 outputs, demonstrating its superior performance.

Furthermore, we evaluate $RPEGA$ ’s cosine similarity on different datasets to measure its performance under different prompt complexities. Figure 9 shows that $RPEGA$ performs best on RE_{prompt} , achieving 90.1% on GPT-3.5 outputs with “text-embedding-ada-002” and 71.2% with

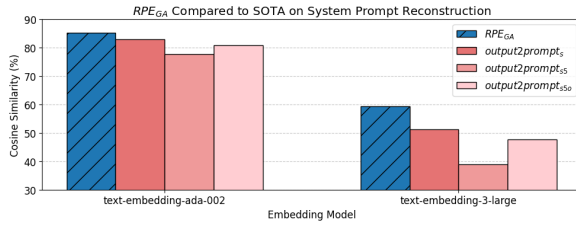


Figure 10: Comparison of RPE_{GA} and $output2prompt_s$ on system prompt recovery (RE_{hard}).

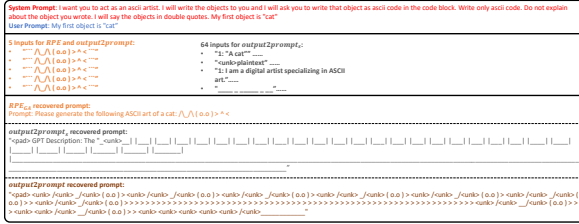


Figure 11: Example of RPE_{GA} and $output2prompt$ recovering a prompt. To conserve space, we do not include all 64 outputs generated for $output2prompt_s$, but instead present one output for each query mentioned earlier.

“text-embedding-3-large.” In contrast, on RE_{hard} , its performance drops by 4.8% and 11.9%, respectively, due to the complex and restrictive nature of these prompts (e.g. “do not write explanations” and “answer only ASCII drawing”). Additionally, performance declines when switching from RE_{prompt} to RE_{easy} , as prompts from MetaMathQA (in RE_{prompt}) are easier to recover than those from Alpaca-GPT4, the source of RE_{easy} . When solving mathematical problems, LLMs tend to repeat the original question, facilitating recovery, whereas RE_{easy} prompts often lead to extra elaboration that hinders prompt recovery. Overall, RPE_{GA} performs best on RE_{prompt} , moderately on RE_{easy} , and worst on RE_{hard} , but still handily beating the benchmark, indicating that detailed instructions with output restrictions present the greatest challenge for language model inversion.

With $n = m = k = 5$, RPE_{GA} issues 230 queries to an LLM and processes approximately 100,000 input tokens and 30,000 output tokens to recover a prompt. The benchmark *output2prompt* is trained on 30,000 prompts, with each prompt necessitating 64 outputs—resulting in a total of 1,920,000 queries to an LLM during training. The final *output2prompt* model is based on the T5 architecture and comprises of 222 million parameters. Next, we evaluate the

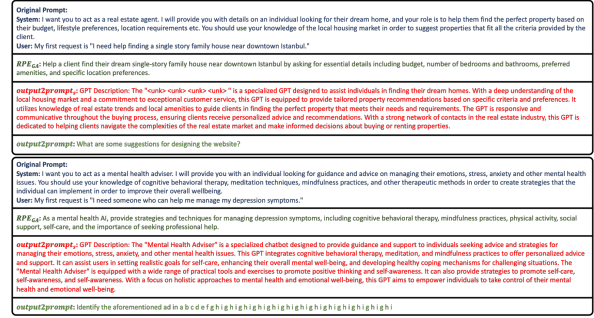


Figure 12: Examples of recovered prompts of RPE_{GA} and $output2prompt$.

ability of RPE_{GA} to recover the system prompt on RE_{hard} and compare it with $output2prompt_s$ and its variants with additional settings. Figure 10 reports the performance of each method. On system prompt recovery, RPE_{GA} achieves higher cosine similarity than both $output2prompt_{s5}$ and $output2prompt_{s50}$. When evaluated with “text-embedding-3-large,” RPE_{GA} exhibits an improvement of 20.4% over $output2prompt_{s5}$ and 11.7% over $output2prompt_{s50}$. Moreover, when compared with $output2prompt_s$, which utilizes all 64 outputs, RPE_{GA} achieves higher cosine similarity, with enhancements of 2.3% using “text-embedding-ada-002” and 8.1% using “text-embedding-3-large.” These findings indicate that RPE_{GA} produces prompts that are more semantically and functionally aligned with the original system prompts than those recovered by $output2prompt_s$.

Furthermore, since RPE_{GA} uses an LLM to generate the recovered prompt, the output is guaranteed to be in natural language. In contrast, the output of $output2prompt$ and $output2prompt_s$ occasionally produces sequences that are not language. As illustrated in Figure 11, RPE_{GA} successfully recovers a complete, coherent sentence, whereas $output2prompt$ and $output2prompt_s$ do not. The example in Figure 11 represents a particularly challenging task, as RPE_{GA} has only five identical answers, containing only ASCII symbols, to work with. In contrast, $output2prompt_s$ has access to more information, especially from the query “Provide 16 scenarios where I can use your services. Start with ‘1:.’” Despite this difficulty, RPE_{GA} still outperforms $output2prompt_s$, demonstrating its robustness in generating natural and semantically meaningful prompts, even under constrained conditions.

Another key advantage of RPE_{GA} is its ability to generate prompts in free form, whereas

	Marketing Plan		Video Game Design		Lyrics	
Example Number	Template	<i>RPE</i>	Template	<i>RPE</i>	Template	<i>RPE</i>
1	2	5	3	4	1	6
2	0	7	0	7	1	6
3	0	7	2	5	1	6
4	/	/	/	/	2	5
5	/	/	/	/	3	4
6	/	/	/	/	3	4
summary	2(9.5%)	19(90.5%)	5(23.8%)	16(76.2%)	11(26.2%)	31(73.8%)

Table 1: Result of the Use Case Experiment. Record the number of people who think the answer generated by the corresponding method is better than the other.

output2prompt and *output2prompt_s* is constrained to producing prompts in a specific format, especially *output2prompt_s*, as shown in Figure 12. This limitation of *output2prompt_s* may stem from its training data, where all prompts follow a uniform structure. Additionally, models in *output2prompt* and *output2prompt_s* has a smaller vocabulary size compared to GPT-3.5, leading to the possible inclusion of “<unk>” tokens in its outputs, as seen in the first example in Figure 12. An ablation study of *RPE* is included in appendix B

4.4 Use Case

A potential use case of *RPE* is extracting prompts from high-quality content, such as marketing plans, video game designs, and song lyrics, enabling users to refine and reuse them for generating similar high-quality outputs. To explore this, we collect samples from these domains and use *RPE_{GA}* to infer the original prompts. The inferred prompts are then used to generate new content—marketing plans for different products, game designs with varied themes, and lyrics featuring diverse motifs—which are compared against outputs generated using standard templates.

Participants in our evaluation are recruited from a pool of college students. An online questionnaire has been developed and its link is distributed through email and social media platforms to reach individuals who had not previously been known to the research team, thereby ensuring an unbiased sample. To assess quality, we conducted a blind evaluation in which participants reviewed both template generated and *RPE* generated responses for the same task without any indication of their origin. Participants were asked to select the response they deemed more favorable, with the option cho-

sen by the majority being classified as the higher quality response. Table 1 presents the human evaluation results, demonstrating that *RPE* outperforms template based methods in generating content preferred by users. This result indicates that *RPE* is better for producing more high-quality data than templates. The workflow for generating new high quality data and complete examples is provided in appendix C.

5 Conclusion

We address the language model inversion problem under black-box, zero-shot conditions, introducing reverse prompt engineering. *RPE* utilizes only an LLM and an optimization algorithm to recover prompts from as few as five text outputs. Experiments on three datasets (*RE_{prompt}*, *RE_{hard}*, *RE_{easy}*) demonstrate that *RPE* effectively reconstructs high-quality prompts. On average across all datasets and embedding models, *RPE* outperforms *output2prompt* by 8.55% in cosine similarity on language model inversion. In system prompt reconstruction, *RPE* recovers prompts from *RE_{hard}* that are 5.8% closer in cosine similarity to the original prompts than *output2prompt_s*, a variant tailored for system prompt recovery. Additionally, use-case experiments show that *RPE* generates higher-quality text that human evaluators prefer over template-generated outputs.

6 Limitations

While our approach demonstrates significant advancements in language model inversion under zero-shot and black-box conditions, there are several limitations to consider. First, although the method requires only five outputs from the target LLM, making it resource-efficient compared to ex-

isting approaches, real-world scenarios may impose stricter constraints where fewer outputs are available, which could affect its applicability. Second, the quality and informativeness of the outputs play a critical role in the effectiveness of the prompt recovery process. In cases where the latent prompt restricts the target LLM to produce minimal or uninformative responses—such as outputs containing only ASCII characters, as demonstrated in Figure 11—our method has room for improvement to handle such situations more effectively. Lastly, the computational cost of iterative optimization can scale with the complexity of the task, posing challenges for large-scale or time-sensitive applications. Addressing these limitations offers opportunities for future work to further enhance the robustness and applicability of the proposed framework.

References

- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. [Language models are few-shot learners](#). In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.
- Wenhu Chen, Xueguang Ma, Xinyi Wang, and William W. Cohen. 2023. [Program of thoughts prompting: Disentangling computation from reasoning for numerical reasoning tasks](#). In *Transactions on Machine Learning Research*.
- Yew Ken Chia, Guizhen Chen, Luu Anh Tuan, Soujanya Poria, and Lidong Bing. 2023. [Contrastive chain-of-thought prompting](#). In *arXiv preprint arXiv:2311.09277*.
- Yihe Deng, Weitong Zhang, Zixiang Chen, and Quanquan Gu. 2023. [Rephrase and respond: Let large language models ask better questions for themselves](#). In *arXiv preprint arXiv:2311.04205*.
- Shehzaad Dhuliawala, Mojtaba Komeili, Jing Xu, Roberta Raileanu, Xian Li, Asli Celikyilmaz, and Jason Weston. 2024. [Chain-of-verification reduces hallucination in large language models](#). In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 3563–3578. Association for Computational Linguistics.
- Shizhe Diao, Pengcheng Wang, Yong Lin, Rui Pan, Xiang Liu, and Tong Zhang. 2024. [Active prompting with chain-of-thought for large language models](#). In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1330–1350. Association for Computational Linguistics.
- Hanxu Hu, Hongyuan Lu, Huajian Zhang, Yun-Ze Song, Wai Lam, and Yue Zhang. 2023. [Chain-of-symbol prompting elicits planning in large language models](#). In *arXiv preprint arXiv:2305.10276*.
- Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. [Retrieval-augmented generation for knowledge-intensive NLP tasks](#). In *Advances in Neural Information Processing Systems*, volume 33, pages 9459–9474.
- Cheng Li, Jindong Wang, Yixuan Zhang, Kaijie Zhu, Wenxin Hou, Jianxun Lian, Fang Luo, Qiang Yang, and Xing Xie. 2023a. [Large language models understand and can be enhanced by emotional stimuli](#). In *arXiv preprint arXiv:2307.11760*.
- Chengshu Li, Jacky Liang, Andy Zeng, Xinyun Chen, Karol Hausman, Dorsa Sadigh, Sergey Levine, Li Fei-Fei, Fei Xia, and brian ichter. 2024a. [Chain of code: Reasoning with a language model-augmented code emulator](#). In *Forty-first International Conference on Machine Learning*.
- Jia Li, Ge Li, Yongmin Li, and Zhi Jin. 2023b. [Structured chain-of-thought prompting for code generation](#). In *ACM Transactions on Software Engineering and Methodology*. ACM New York, NY.
- Xingxuan Li, Ruochen Zhao, Yew Ken Chia, Bosheng Ding, Shafiq Joty, Soujanya Poria, and Lidong Bing. 2024b. [Chain-of-knowledge: Grounding large language models via dynamic knowledge adapting over heterogeneous sources](#). In *The Twelfth International Conference on Learning Representations*.
- Stephanie Lin, Jacob Hilton, and Owain Evans. 2022. [Truthfulqa: Measuring how models mimic human falsehoods](#). In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3214–3252.
- Jieyi Long. 2023. [Large language model guided tree-of-thought](#). In *arXiv preprint arXiv:2305.08291*.
- John Xavier Morris, Wenting Zhao, Justin T Chiu, Vitaly Shmatikov, and Alexander M Rush. 2024. [Language model inversion](#). In *The Twelfth International Conference on Learning Representations*.
- Maxwell Nye, Anders Johan Andreassen, Guy Gur-Ari, Henryk Michalewski, Jacob Austin, David Bieber, David Dohan, Aitor Lewkowycz, Maarten Bosma, David Luan, et al. 2021. [Show your work: Scratchpads for intermediate computation with language models](#). In *arXiv preprint arXiv:2112.00114*.

697	Bhargavi Paranjape, Scott Lundberg, Sameer Singh, Hannaneh Hajishirzi, Luke Zettlemoyer, and Marco Tulio Ribeiro. 2023. Art: Automatic multi-step reasoning and tool-use for large language models . In <i>arXiv preprint arXiv:2303.09014</i> .	752
698		753
699		754
700		755
701		756
702	Baolin Peng, Chunyuan Li, Pengcheng He, Michel Galley, and Jianfeng Gao. 2023. Instruction tuning with GPT-4 . In <i>arXiv preprint arXiv:2304.03277</i> .	757
703		758
704		759
705	Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners . In <i>OpenAI blog</i> , volume 1, page 9.	760
706		761
707		762
708		763
709	Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer . In <i>Journal of machine learning research</i> , volume 21, pages 1–67.	764
710		765
711		766
712		767
713		768
714		769
715	Pranab Sahoo, Ayush Kumar Singh, Sriparna Saha, Vinija Jain, Samrat Mondal, and Aman Chadha. 2024. A systematic survey of prompt engineering in large language models: Techniques and applications . In <i>arXiv preprint arXiv:2402.07927</i> .	770
716		771
717		772
718		773
719		774
720	Jeffrey R Sampson. 1976. Adaptation in natural and artificial systems (John H. Holland) . In <i>SIAM Review</i> , volume 18, pages 529–530. Society for Industrial and Applied Mathematics.	775
721		776
722		777
723		778
724	Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. 2023. Self-consistency improves chain of thought reasoning in language models . In <i>The Eleventh International Conference on Learning Representations</i> .	779
725		780
726		781
727		782
728		783
729		784
730	Zilong Wang, Hao Zhang, Chun-Liang Li, Julian Martin Eisenschlos, Vincent Perot, Zifeng Wang, Lesly Miculicich, Yasuhisa Fujii, Jingbo Shang, Chen-Yu Lee, and Tomas Pfister. 2024. Chain-of-table: Evolving tables in the reasoning chain for table understanding . In <i>The Twelfth International Conference on Learning Representations</i> .	785
731		786
732		787
733		788
734		789
735		790
736		791
737	Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models . In <i>Advances in Neural Information Processing Systems</i> , volume 35, pages 24824–24837.	792
738		793
739		794
740		795
741		796
742		797
743	Jason Weston and Sainbayar Sukhbaatar. 2023. System 2 attention (is something you might need too) . In <i>arXiv preprint arXiv:2311.11829</i> .	798
744		799
745		800
746		801
747	Shirley Wu, Shiyu Zhao, Qian Huang, Kexin Huang, Michihiro Yasunaga, Kaidi Cao, Vassilis N. Ioannidis, Karthik Subbian, Jure Leskovec, and James Zou. 2024. Avatar: Optimizing LLM agents for tool usage via contrastive reasoning . In <i>Advances in Neural Information Processing Systems</i> .	802
748		803
749		804
750		805
751		806
		807
	Chengrun Yang, Xuezhi Wang, Yifeng Lu, Hanxiao Liu, Quoc V Le, Denny Zhou, and Xinyun Chen. 2024. Large language models as optimizers . In <i>The Twelfth International Conference on Learning Representations</i> .	
	Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, and Karthik Narasimhan. 2024a. Tree of thoughts: Deliberate problem solving with large language models . In <i>Advances in Neural Information Processing Systems</i> , volume 36.	
	Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R Narasimhan, and Yuan Cao. 2023. React: Synergizing reasoning and acting in language models . In <i>The Eleventh International Conference on Learning Representations</i> .	
	Yao Yao, Zuchao Li, and Hai Zhao. 2024b. GoT: Effective graph-of-thought reasoning in language models . In <i>Findings of the Association for Computational Linguistics: NAACL 2024</i> , pages 2901–2921. Association for Computational Linguistics.	
	Longhui Yu, Weisen Jiang, Han Shi, Jincheng YU, Zhengying Liu, Yu Zhang, James Kwok, Zhenguo Li, Adrian Weller, and Weiyang Liu. 2024a. Metamath: Bootstrap your own mathematical questions for large language models . In <i>The Twelfth International Conference on Learning Representations</i> .	
	Wenhao Yu, Hongming Zhang, Xiaoman Pan, Peixin Cao, Kaixin Ma, Jian Li, Hongwei Wang, and Dong Yu. 2024b. Chain-of-note: Enhancing robustness in retrieval-augmented language models . In <i>Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing</i> , pages 14672–14685. Association for Computational Linguistics.	
	Collin Zhang, John Xavier Morris, and Vitaly Shmatikov. 2024. Extracting prompts by inverting LLM outputs . In <i>Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing</i> , pages 14753–14777. Association for Computational Linguistics.	
	Zhuosheng Zhang, Aston Zhang, Mu Li, and Alex Smola. 2023. Automatic chain of thought prompting in large language models . In <i>The Eleventh International Conference on Learning Representations</i> .	
	Xufeng Zhao, Mengdi Li, Wenhao Lu, Cornelius Weber, Jae Hee Lee, Kun Chu, and Stefan Wermter. 2024. Enhancing zero-shot chain-of-thought reasoning in large language models through logic . In <i>Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation</i> , pages 6144–6166.	
	Huaixiu Steven Zheng, Swaroop Mishra, Xinyun Chen, Heng-Tze Cheng, Ed H. Chi, Quoc V Le, and Denny Zhou. 2024. Take a step back: Evoking reasoning via abstraction in large language models . In <i>The Twelfth International Conference on Learning Representations</i> .	

Yongchao Zhou, Andrei Ioan Muresanu, Ziwen Han, Keiran Paster, Silviu Pitis, Harris Chan, and Jimmy Ba. 2023a. [Large language models are human-level prompt engineers](#). In *The Eleventh International Conference on Learning Representations*.

Yucheng Zhou, Xiubo Geng, Tao Shen, Chongyang Tao, Guodong Long, Jian-Guang Lou, and Jianbing Shen. 2023b. [Thread of thought unraveling chaotic contexts](#). In *arXiv preprint arXiv:2311.08734*.

A Public Datasets and Ethics

- **Awesome ChatGPT Prompts³**: This is a curated set of 153 prompts resembling system messages used in real-world LLM-based APIs and services. These prompts are structured as detailed instructions, designed to adapt the LLM to specific roles, such as a food critic or a Python interpreter. The dataset is released under the CC0-1.0 license.
- **MetaMathQA**: Introduced by Yu et al. (2024a), MetaMathQA consists of 395,000 linguistically diverse math word problems, ranging in difficulty from primary school to graduate school. This dataset is released under the MIT license.
- **TruthfulQA**: TruthfulQA (Lin et al., 2022) consists of 817 questions across 38 categories, including health, law, finance, and politics. These questions are designed in a way that some humans might answer incorrectly due to false beliefs or misconceptions. The dataset is intended to evaluate whether a language model generates truthful answers to such questions. This dataset is released under the Apache-2.0 license.
- **Alpaca-GPT4**: Alpaca-GPT4 contains 52,000 instruction-following examples generated by GPT-4 using prompts from the Alpaca dataset, and it was used to fine-tune LLMs in the work by Peng et al. (2023). The dataset is released under the CC-BY-NC-4.0 license.
- **Dolly Creative Writing⁴**: This dataset consists of 673 prompts designed to assess the creativity of a language model. Each prompt is either a question or an instruction, guiding the LLM to perform a creative writing task.

³<https://github.com/f/awesome-chatgpt-prompts>

⁴https://huggingface.co/datasets/lionelchg/dolly_creative_writing

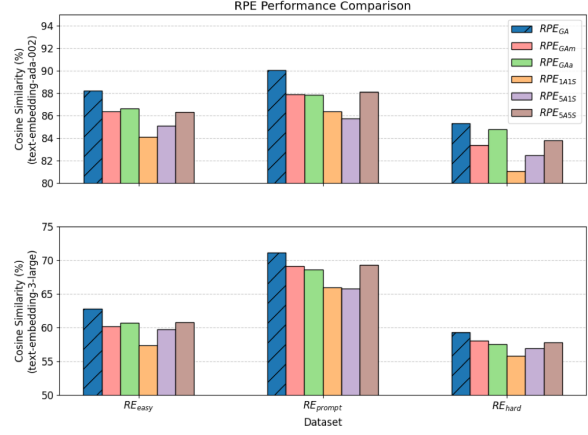


Figure 13: Comparison of different RPE methods on three datasets.

The benchmark code for *output2prompt* is distributed under the MIT license. All datasets and code employed in this study are solely intended for academic research, in accordance with their designated usage. We have verified the ethical documentation for each dataset and conducted extensive sampling to ensure the absence of personally identifying or objectionable content. The code and datasets generated in this study will likewise be released under the MIT license.

Moreover, our questionnaire explicitly obtained participants’ consent to utilize their anonymized responses in our research.

B Ablation Study

In the ablation study, we compare the performance of RPE_{GA} and its variants depicted in Section 3. In addition, we examine the impact of different approaches to calculating the performance score s'_i for the RPE_{GA} variant. Specifically, the variant RPE_{GAm} computes s'_i by selecting the maximum ROUGE-1 score between a'_i and each response in set A , while RPE_{GAa} calculates s'_i as the average ROUGE-1 score between a'_i and all responses in A . The best and thus default RPE_{GA} method, by contrast, determines s'_i as the average of both the mean and maximum ROUGE-1 scores.

As illustrated in Figure 13, RPE_{GA} consistently outperforms the other RPE variants. The results from RPE_{GAm} and RPE_{GAa} indicate that using either the maximum or the average score alone for performance calculation compromises the quality of the inferred prompts. Furthermore, the superior performance of RPE_{5A5S} over other non-iterative approaches underscores the efficacy of our evalu-

ation strategy in selecting high-quality recovered prompts.

C Details of Generating High Quality Content

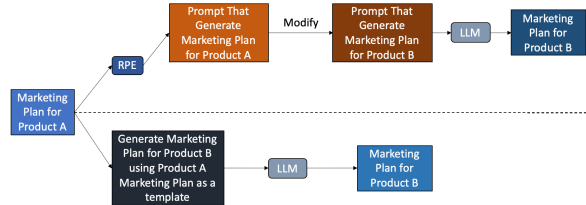


Figure 14: Workflow to generate new high quality answers.

In Figure 14, we illustrate the workflow for generating new high-quality data using both *RPE* and templates, exemplified by generating a marketing plan for Product B based on Product A’s plan.

C.1 Use Case Experiments: Marketing Plan

We begin with a marketing plan for an energy drink as our initial reference point. Using both the *RPE* and template methods, we then generate marketing plans for three distinct products: “a new smart-phone targeting seniors aged 65 and older”, “a financial software tailored for small businesses and individual investors”, and “developmental toys designed for toddlers under one year old”. As shown in Table 1, for each product, a greater number of participants favored the *RPE*-generated marketing plan over the template-generated one. Overall, 90.5% of responses preferred the *RPE* method, while only 9.5% favored the template method. Detailed marketing plans are provided in appendix C.4.

C.2 Use Case Experiments: Video Game Design

Using the game design of the popular video game “Don’t Starve” as a reference, we created high-quality designs for other games. We prompted GPT-3.5 to design games based on the following themes: “a rogue-like game incorporating elements of Greek mythology and combat,” “a kart racing game that includes multiplayer and item-based mechanics,” and “a first-person shooter game combining elements of war and counter-terrorism.” Using both *RPE* and template methods, we produced a total of six game designs. As shown in Table 1, participants preferred the game designs generated by *RPE* over those created by the template

method. Overall, 76.2% of responses favored the *RPE*-generated designs, while only 23.8% preferred the template-generated designs. Complete game designs are presented in appendix C.5.

C.3 Use Case Experiments: Lyrics

For the lyrics generation task, we first use “Cruel Summer” by Taylor Swift as a reference to create lyrics for songs with the following themes: “evoking sadness and grief with themes of loss, winter, and religion,” “evoking happiness and joy with themes of family, friends, college life, and flowers,” and “evoking excitement and positivity with themes of courage, hope, and the future.” We then use “Master of Puppets” by Metallica as another reference to generate lyrics for songs themed around “love and heartbreak,” “self-discovery and personal growth,” and “nostalgia and memories.” For each theme, we generated two sets of lyrics using both the template and *RPE* methods, producing a total of twelve lyrics. Participants preferred the *RPE*-generated lyrics, with 73.8% of responses favoring them over the template-generated versions, which received only 26.2% preference. All lyrics are provided in appendix C.4.

C.4 Complete Examples of Market Plan

Figure 15 presents the reference marketing plan, the prompt recovered using *RPE*, and edited prompts used to generate marketing plans for different products. Complete marketing plans generated from perturbed *RPE*-recovered prompts and template-based prompts are provided in Figures 16, 17, and 18.

C.5 Complete Examples of Video Game Description

Figure 19 displays the reference video game description, along with the prompt recovered using *RPE* and modified prompts used to generate descriptions for video games with varying themes. The full set of video game descriptions generated from perturbed *RPE*-recovered prompts and template-based prompts is presented in Figures 20, 21, and 22.

C.6 Complete Examples of Lyrics

Figures 23 and 24 present the reference song lyrics, along with the prompt recovered using *RPE* and modified prompts used to generate lyrics in different styles and themes. The complete set of lyrics generated from perturbed *RPE*-recovered prompts

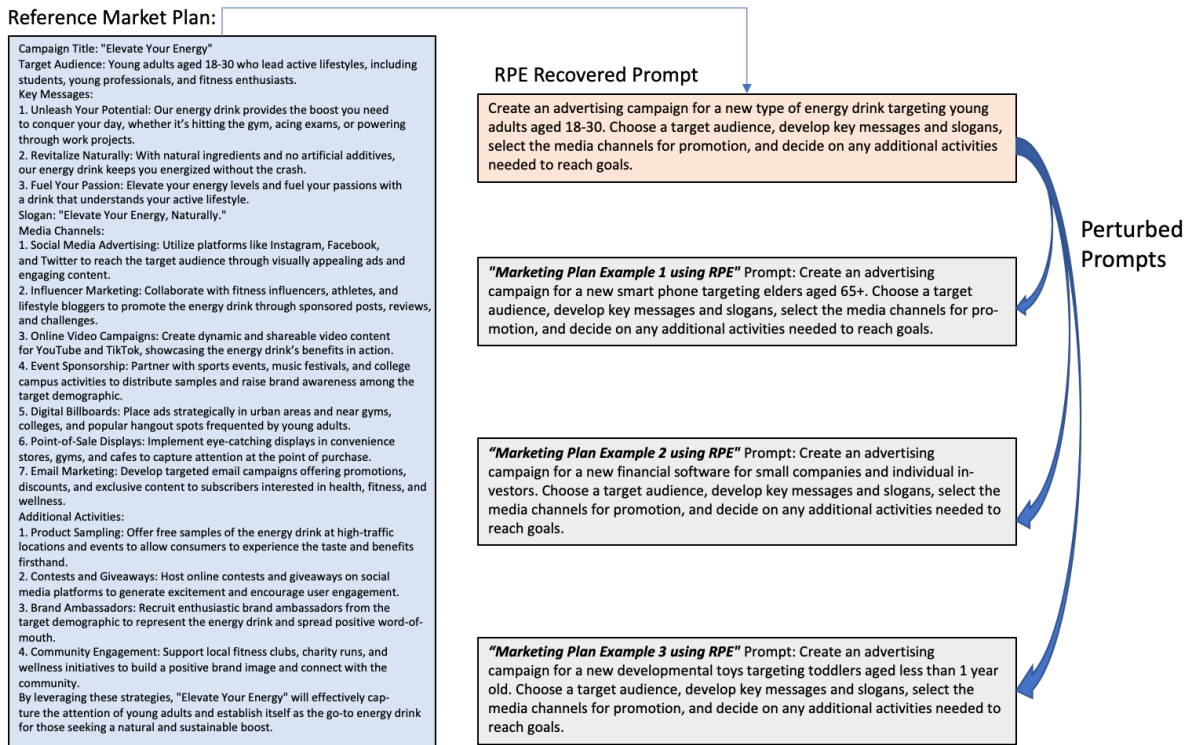


Figure 15: Reference marketing plan and the prompt recovered by *RPE*, along with perturbed prompts used to generate marketing plans for different products.

Market Plan Example 1

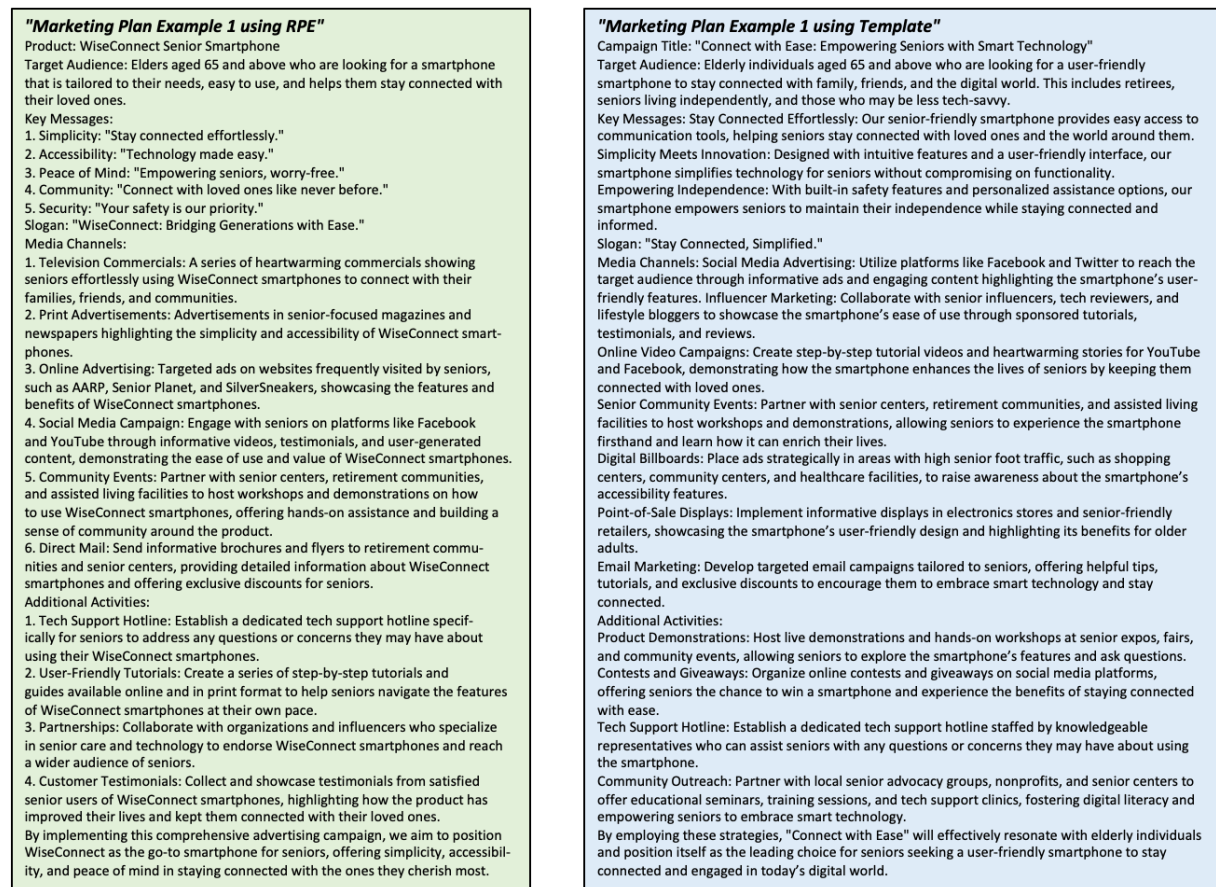


Figure 16: Example 1 of market plan generation.

Market Plan Example 2

"Marketing Plan Example 2 using RPE" Product: Introducing FinEase - Your Financial Companion for Small Businesses and Individual Investors! Target Audience: 1. Small Business Owners: Entrepreneurs and small business owners looking for efficient financial management solutions. 2. Individual Investors: Novice to experienced investors seeking user-friendly tools for managing their portfolios. Key Messages: 1. Streamlined Financial Management: Simplify your finances with our intuitive software designed to streamline accounting, budgeting, and reporting processes. 2. Empower Your Decision Making: Gain insights with real-time data analysis and reporting features, empowering you to make informed financial decisions. 3. Secure and Reliable: Trust in our robust security measures to safeguard your financial data, ensuring peace of mind. 4. User-Friendly Interface: Navigate effortlessly through our user-friendly interface, tailored to meet the needs of both beginners and experts. 5. 24/7 Customer Support: Access dedicated customer support around the clock, providing assistance whenever you need it. Slogan: "FinEase: Simplify Your Finances, Empower Your Future." Media Channels: 1. Digital Advertising: - Google Ads: Target keywords related to financial management, accounting software, and investment tools. - Social Media Ads: Utilize platforms like LinkedIn and Facebook to reach small business owners and investors. - Display Ads: Place banners on relevant websites frequented by our target audience. 2. Content Marketing: - Create informative blog posts, articles, and video tutorials highlighting the benefits of FinEase and offering financial management tips. 3. Guest Posting: Collaborate with industry influencers and finance-related blogs to reach a wider audience. 3. Email Marketing: - Develop an email campaign targeting small business owners and investors, offering exclusive discounts and promotions to encourage sign-ups. 4. Partnerships and Collaborations: - Partner with accounting firms, financial advisors, and business associations to promote FinEase to their clients and members. - Sponsor relevant events and webinars aimed at small business owners and investors. 5. Search Engine Optimization (SEO): - Optimize website content and blog posts for relevant keywords to improve search engine rankings and attract organic traffic. Additional Activities: 1. Free Trial Offer: Provide a free trial period for new users to experience the benefits of FinEase firsthand before committing to a subscription. 2. User Testimonials: Showcase testimonials and case studies from satisfied users to build credibility and trust. 3. Referral Program: Incentivize existing users to refer FinEase to others by offering discounts or rewards for successful referrals. 4. Interactive Demos: Host webinars or live demos to demonstrate the features and functionality of FinEase, allowing potential customers to interact with the software in real-time. With this comprehensive advertising campaign, FinEase aims to establish itself as the go-to financial software for small businesses and individual investors, offering a solution that simplifies financial management and empowers users to achieve their financial goals.	"Marketing Plan Example 2 using Template" Campaign Title: "Financial Fitness Boost" Target Audience: Small business owners, freelancers, and individual investors aged 25-45 who are looking for efficient financial management solutions to streamline their operations and grow their wealth. Key Messages: Empower Your Finances: Our financial software provides the tools you need to take control of your money, whether it's managing business expenses, tracking investments, or planning for the future. Simplify with Technology: With intuitive design and powerful features, our software simplifies financial tasks, saving you time and stress. Grow Your Wealth: Boost your financial health and achieve your goals with a software that caters to your unique needs and aspirations. Slogan: "Empower Your Finances, Seamlessly." Media Channels: Social Media Advertising: Utilize platforms like LinkedIn, Facebook, and Twitter to target small business owners and individual investors with informative ads and engaging content highlighting the benefits of the financial software. Influencer Marketing: Collaborate with finance bloggers, small business influencers, and financial advisors to promote the software through sponsored content, tutorials, and testimonials. Online Video Campaigns: Create educational video content for YouTube and LinkedIn, demonstrating how the software can streamline financial processes and help users achieve their goals. Industry Events Sponsorship: Partner with business conferences, entrepreneurship workshops, and financial literacy seminars to showcase the software's capabilities and connect with potential users. Digital Billboards: Place ads strategically in business districts and financial hubs to raise awareness among small business owners and investors about the software's benefits. Point-of-Sale Displays: Implement visually appealing displays in co-working spaces, business centers, and financial institutions to attract attention and encourage trial. Email Marketing: Develop targeted email campaigns offering demos, free trials, and exclusive insights to subscribers interested in financial management and entrepreneurship. Additional Activities: Product Demos: Offer live demonstrations of the software at trade shows, networking events, and business meetups to allow potential users to experience its features firsthand. Contests and Giveaways: Host online contests and giveaways on social media platforms to generate buzz and incentivize user engagement with the software. Brand Ambassadors: Recruit knowledgeable brand ambassadors from the finance industry to advocate for the software and provide guidance to users. Community Engagement: Support local small business associations, entrepreneurship clubs, and financial literacy initiatives to establish trust and credibility within the community. By employing these strategies, "Financial Fitness Boost" will effectively reach its target audience and position itself as the essential tool for small businesses and individual investors looking to achieve financial success.
--	---

Figure 17: Example 2 of market plan generation.

Market Plan Example 3

"Marketing Plan Example 3 using RPE" Advertising Campaign: "Tiny Explorers" Target Audience: Parents of infants aged less than 1 year old, particularly those who prioritize their child's cognitive and motor development. Key Messages: 1. Early Developmental Stimulation: Engage your little one's senses and foster their cognitive skills with our innovative toys designed for early learning. 2. Safe and Stimulating: Our toys are crafted with the highest quality materials, ensuring safety while providing endless exploration opportunities. 3. Bonding Through Play: Strengthen the bond with your child as you embark on a journey of discovery together. 4. Expert-Approved: Developed with input from child psychologists and experts in early childhood development, our toys are tailored to support your child's growth. Slogans: 1. "Grow, Explore, Play: Start Their Journey with Tiny Explorers!" 2. "Building Little Minds, One Playful Moment at a Time." 3. "Where Every Giggle Marks a Milestone." 4. "Discover the Joy of Learning, Right from the Start." Media Channels: 1. Social Media Advertising: Utilize platforms like Facebook, Instagram, and Pinterest to showcase engaging images and videos of babies interacting with our toys. Targeted advertising can reach parents with infants in the specified age group. 2. Parenting Blogs and Forums: Collaborate with popular parenting bloggers and participate in parenting forums to share informative content about early childhood development and the importance of interactive toys. 3. Baby Product Retailers: Partner with baby product retailers to feature our toys prominently in their stores, accompanied by informative displays highlighting their developmental benefits. 4. Online Marketplaces: List our toys on e-commerce platforms such as Amazon and Etsy, accompanied by detailed product descriptions and customer reviews. 5. Parenting Magazines and Websites: Advertise in print and online publications focused on parenting and child development, featuring articles and testimonials that emphasize the educational value of our toys. Additional Activities: 1. Launch Events: Host launch events in baby-friendly venues such as indoor play centers or community centers, allowing parents and infants to interact with our toys firsthand. 2. Educational Workshops: Organize workshops led by child development experts, providing parents with tips and techniques for incorporating developmental play into their daily routines. 3. Social Media Contests: Run contests on social media platforms, encouraging parents to share photos and videos of their babies enjoying our toys for a chance to win prizes. 4. Partnerships with Pediatricians: Establish partnerships with pediatricians and pediatric clinics to distribute informational materials about the importance of early childhood stimulation and recommend our toys to parents during well-baby visits. By implementing these strategies, we aim to raise awareness among parents of infants aged less than 1 year old about the benefits of our developmental toys and establish Tiny Explorers as a trusted brand for early childhood learning and play.	"Marketing Plan Example 3 using Template" Campaign Title: "Bright Beginnings" Target Audience: Parents of infants aged 0-12 months who seek developmental toys to aid in their child's early learning and growth. Key Messages: 1. Ignite Curiosity: Our developmental toys are designed to spark curiosity and stimulate cognitive development from the earliest stages. 2. Nurture Naturally: Crafted with safe, non-toxic materials, our toys provide a nurturing environment for your baby's exploration and play. 3. Foster Connections: Build meaningful connections with your baby through interactive playtime activities that promote bonding and social development. Slogan: "Bright Beginnings, Safe Explorations." Media Channels: 1. Social Media Advertising: Utilize platforms like Instagram and Facebook to reach parents with visually appealing ads showcasing the benefits of our toys for early childhood development. 2. Parenting Influencers: Collaborate with parenting bloggers, pediatricians, and early childhood educators to endorse our toys through authentic reviews, tips, and playtime demonstrations. 3. Educational Video Campaigns: Create engaging video content for YouTube and parenting forums, demonstrating how our toys aid in sensory exploration, motor skills, and cognitive development. 4. Parenting Events: Partner with baby expos, parenting workshops, and community playgroups to showcase our toys and provide hands-on demonstrations for new parents. 5. Digital Billboards: Place ads in family-friendly areas such as parks, daycare centers, and pediatric clinics to capture the attention of parents while they're out and about with their infants. 6. Point-of-Sale Displays: Establish eye-catching displays in baby stores, maternity wards, and pediatrician offices to attract parents seeking developmental toys for their infants. 7. Email Marketing: Develop targeted email campaigns offering parenting tips, milestone guides, and exclusive discounts on our toys to subscribers interested in early childhood development. Additional Activities: 1. Toy Sampling Events: Host baby playdates and mommy-and-me groups where parents can interact with our toys firsthand and experience their educational benefits. 2. Parenting Contests: Organize online contests inviting parents to share their favorite bonding moments or creative ways of using our toys, with prizes awarded for the most inspiring entries. 3. Parent Ambassadors: Recruit passionate parents as brand ambassadors to share their experiences with our toys on social media and in parent forums, fostering trust and credibility. 4. Community Partnerships: Collaborate with early childhood education centers, pediatric clinics, and parenting support groups to provide resources and support for parents seeking developmental toys for their infants. By leveraging these strategies, "Bright Beginnings" will effectively engage and empower parents in nurturing their infants' early learning and development with safe and stimulating toys.
--	---

Figure 18: Example 3 of market plan generation.

Reference Video Game Design:

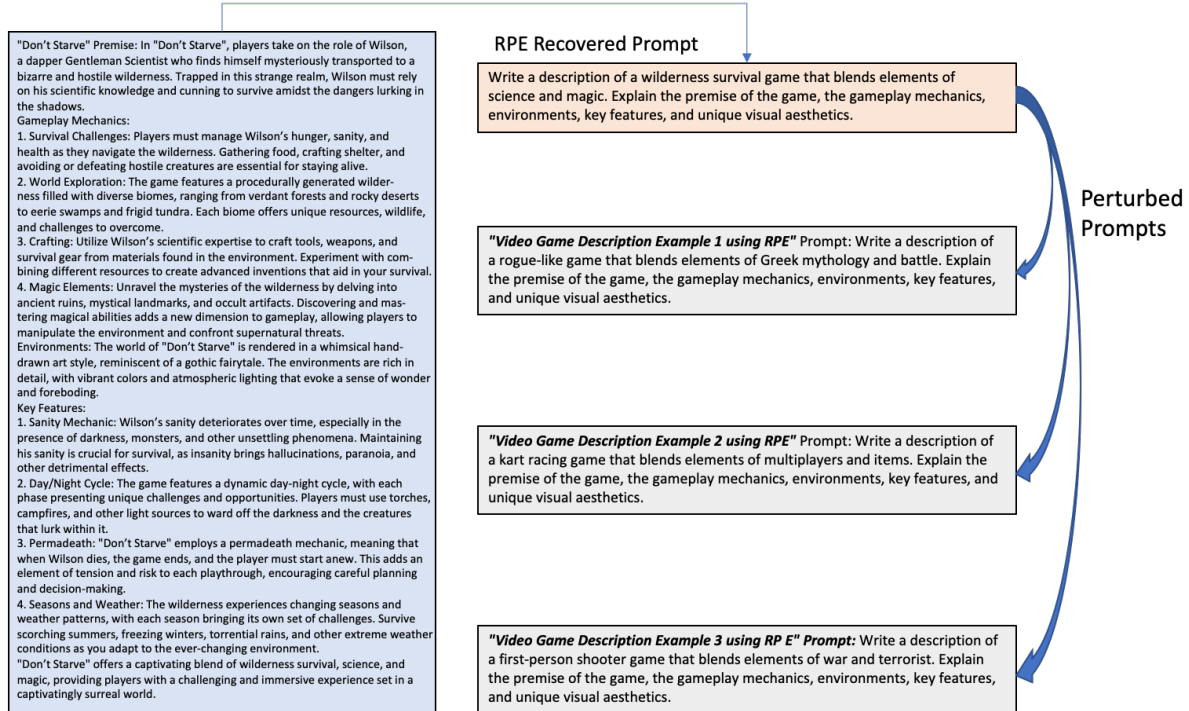


Figure 19: Reference video game description and the prompt recovered by *RPE*, along with perturbed prompts used to generate video description for different themes.

Video Game Description Example 1

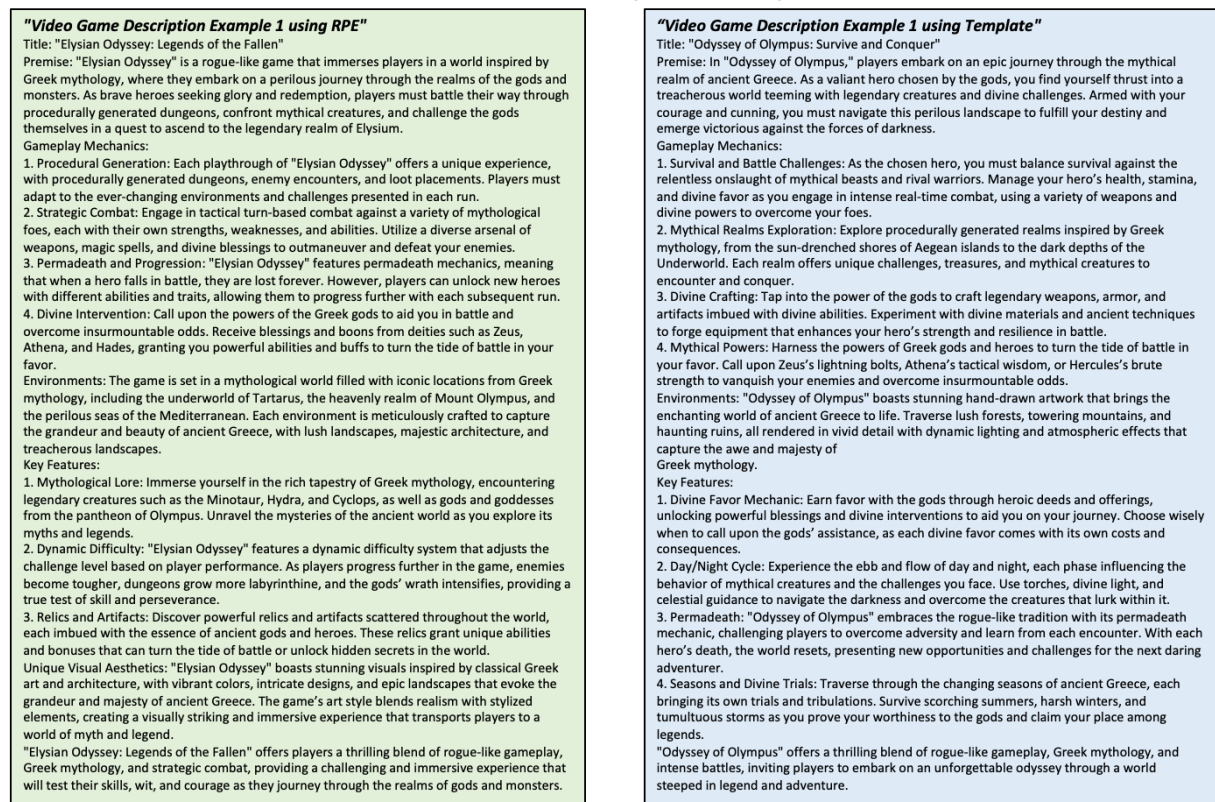


Figure 20: Example 1 of video game description generation.

Video Game Description Example 2

"Video Game Description Example 2 using RPE" Title: Velocity Dash Kart Racing Premise: Velocity Dash Kart Racing is a high-octane, adrenaline-fueled kart racing game set in a futuristic world where speed, skill, and strategy reign supreme. Players step into the shoes of customizable drivers, each with their own unique abilities and personalities, and compete in intense races across a variety of dynamic and visually stunning tracks. The game blends classic kart racing mechanics with innovative multiplayer features and a diverse array of power-ups and items, creating an exhilarating experience for players of all skill levels. Gameplay Mechanics: The core gameplay of Velocity Dash revolves around fast-paced kart racing across vibrant and diverse tracks. Players can drift around corners, perform stunts for speed boosts, and strategically utilize power-ups to gain an edge over their opponents. The controls are intuitive, with accessible handling that allows for skillful maneuvering and dynamic racing action. One of the key features of Velocity Dash is its emphasis on multiplayer competition. Players can compete in both local and online multiplayer modes, racing against friends or challenging opponents from around the world. The game supports up to 8 players in online races, with various customization options and matchmaking features to ensure balanced and competitive gameplay. Power-ups and Items: Velocity Dash features a wide range of power-ups and items scattered throughout the tracks, which players can collect and use to gain an advantage or sabotage their opponents. These power-ups include speed boosts, homing missiles, oil slicks, shields, and more, each with its own unique effects and strategic implications. Players must use their wits and reflexes to effectively use these items while dodging attacks from rival racers. Environments: The tracks in Velocity Dash are set in diverse and visually stunning environments, ranging from futuristic cityscapes to exotic locales like tropical islands and icy tundras. Each track is meticulously designed with twists, turns, shortcuts, and interactive elements that provide strategic opportunities and dynamic racing experiences. Dynamic weather effects and environmental hazards add an extra layer of challenge and immersion to the races, keeping players on their toes and ensuring that no two races are ever the same. Key Features: • Intense and fast-paced kart racing action • Customizable drivers with unique abilities and personalities • Multiplayer modes supporting both local and online play • A wide variety of power-ups and items to collect and use strategically • Dynamic and visually stunning tracks set in diverse environments • Intuitive controls and accessible handling for players of all skill levels • Dynamic weather effects and environmental hazards • Regular updates and new content to keep the gameplay fresh and exciting Unique Visual Aesthetics: Velocity Dash boasts a vibrant and colorful art style that brings its futuristic world to life. The environments are richly detailed, with dynamic lighting and particle effects that create a sense of immersion and excitement. The character designs are diverse and expressive, with customizable options that allow players to personalize their drivers to their liking. The visual effects for the power-ups and items are flashy and impactful, adding to the spectacle of the races and keeping players engaged and entertained throughout their gameplay experience. Overall, Velocity Dash offers a visually stunning and exhilarating kart racing experience that is sure to captivate players and keep them coming back for more.	"Video Game Description Example 2 using Template" "Zooming Kart Racers: Turbo Dash" Premise: In "Zooming Kart Racers: Turbo Dash", players enter the high-speed world of kart racing, where they take on the role of quirky racers competing in adrenaline-fueled races across whimsical tracks filled with twists, turns, and surprises. From bustling city streets to exotic jungle circuits, racers must master their driving skills and utilize cunning strategies to outmaneuver opponents and claim victory. Gameplay Mechanics: 1. Racing Challenges: Players must navigate dynamic tracks filled with obstacles, ramps, and shortcuts while jostling for position against rival racers. Speed, drift, and precision driving are essential for securing the lead and crossing the finish line first. 2. Multiplayer Madness: Compete against friends or players from around the world in thrilling multiplayer races with up to 8 players online. Form alliances, sabotage opponents, and unleash chaos with a variety of power-ups and items scattered throughout the tracks. 3. Item Mayhem: Collect and use a colorful array of power-ups, including speed boosts, homing missiles, oil slicks, and more, to gain the upper hand in races. Timing and strategy are key as players unleash havoc on their opponents while dodging incoming attacks. 4. Kart Customization: Customize your kart with a range of cosmetic options, from sleek paint jobs and flashy decals to wacky accessories and outrageous spoilers. Stand out on the track and show off your personal style as you leave competitors in the dust. Environments: "Zooming Kart Racers: Turbo Dash" features vibrant and dynamic tracks set in diverse locations, including bustling cities, scenic beaches, lush forests, and treacherous mountain passes. Each track is filled with unique hazards, shortcuts, and surprises that keep races exhilarating and unpredictable. Key Features: 1. Multiplayer Madness: Compete against friends or players from around the world in thrilling multiplayer races with up to 8 players online. Form alliances, sabotage opponents, and unleash chaos with a variety of power-ups and items scattered throughout the tracks. 2. Item Mayhem: Collect and use a colorful array of power-ups, including speed boosts, homing missiles, oil slicks, and more, to gain the upper hand in races. Timing and strategy are key as players unleash havoc on their opponents while dodging incoming attacks. 3. Kart Customization: Customize your kart with a range of cosmetic options, from sleek paint jobs and flashy decals to wacky accessories and outrageous spoilers. Stand out on the track and show off your personal style as you leave competitors in the dust. 4. Dynamic Tracks: Race across dynamic tracks filled with obstacles, ramps, and shortcuts that provide endless opportunities for strategic racing and epic stunts. Master each track's twists and turns to gain the advantage and leave your opponents in the dust. "Zooming Kart Racers: Turbo Dash" delivers high-octane racing action with a dose of multiplayer mayhem, offering players a thrilling and addictive kart racing experience unlike any other. Get ready to rev your engines, drift around corners, and leave your rivals eating your dust as you zoom to victory!
---	--

Figure 21: Example 2 of video game description generation.

Video Game Description Example 3

"Video Game Description Example 3 using RPE" Title: "Shadowfront: Covert Warfare" Premise: "Shadowfront: Covert Warfare" immerses players in a gripping narrative where they navigate the blurred lines between war and terrorism. Set in a near-future world torn by geopolitical conflicts, players assume the role of elite operatives tasked with executing covert missions to dismantle terrorist networks and prevent global instability. As tensions escalate, players must confront moral dilemmas, navigate complex alliances, and unearth the truth behind shadowy conspiracies. Gameplay Mechanics: • Players can choose from a diverse roster of specialized operatives, each with unique abilities and skill sets, allowing for strategic team composition. • Missions range from stealth infiltration and sabotage to intense firefights and hostage rescues, offering varied gameplay experiences. • Tactical decision-making is paramount, with players able to utilize advanced gadgets, deploy drones for reconnaissance, and coordinate synchronized attacks with AI-controlled squadmates or online teammates. • Dynamic AI behavior and environmental hazards keep players on their toes, requiring adaptability and quick thinking to overcome challenges. Environments: • "Shadowfront" spans global hotspots, from sprawling urban environments and densely populated cityscapes to remote wilderness areas and clandestine facilities. • Each location is meticulously designed to offer multiple paths and tactical opportunities, encouraging players to approach objectives with creativity and strategic foresight. • Dynamic weather effects, day-night cycles, and destructible environments add depth and unpredictability to gameplay, creating immersive and engaging experiences. Key Features: • Deep, branching narrative with multiple endings influenced by player choices and actions throughout the campaign. • Robust multiplayer modes, including competitive PvP matches and cooperative PvE missions, where teamwork and communication are crucial for success. • Extensive customization options for weapons, gear, and character appearance, allowing players to tailor their loadouts to suit their preferred playstyles. • Regular content updates and community events to keep the experience fresh and engaging over time. Unique Visual Aesthetics: • "Shadowfront" features a gritty, photorealistic art style that captures the tension and uncertainty of modern warfare. • Cinematic cutscenes and dynamic camera angles enhance storytelling, immersing players in the high-stakes world of covert operations. • Attention to detail in character animations, weapon models, and environmental design elevates the overall immersion, making every encounter feel visceral and impactful. "Shadowfront: Covert Warfare" challenges players to confront the complexities of modern conflict, blurring the lines between heroism and villainy in a morally ambiguous struggle for peace and justice.	"Video Game Description Example 3 using Template" Title: "Warfront: Survival Operations" Premise: In "Warfront: Survival Operations", players step into the boots of Sergeant Jackson, a seasoned soldier thrust into the heart of a relentless conflict against global terrorism. Stranded behind enemy lines in hostile territories, Jackson must rely on his combat prowess and tactical skills to survive amidst the chaos of war. Gameplay Mechanics: 1. Survival Challenges: Players face constant threats to their health, ammunition, and morale as they navigate war-torn landscapes. Scavenging for supplies, fortifying positions, and engaging enemy combatants are essential for enduring the relentless onslaught. 2. Battlefield Exploration: The game features dynamically generated war zones spanning urban cities, desolate wastelands, dense jungles, and remote mountain ranges. Each environment presents unique tactical opportunities and dangers, requiring adaptability and strategic planning. 3. Weapon Crafting: Utilize Jackson's military expertise to craft improvised weapons, explosives, and defensive structures from salvaged materials. Experiment with weapon modifications and upgrades to gain an edge in combat against increasingly formidable adversaries. 4. Special Operations: Embark on covert missions behind enemy lines to gather intelligence, sabotage enemy infrastructure, and rescue hostages. Navigate through enemy patrols, infiltrate heavily fortified compounds, and execute precision strikes to disrupt terrorist operations. Environments: The world of "Warfront: Survival Operations" is rendered in gritty realism, with stunning graphics and immersive sound design that capture the intensity of modern warfare. From the chaos of urban warfare to the desolation of remote battlegrounds, each environment is rich in detail and atmosphere. Key Features: 1. Morale System: Jackson's morale fluctuates based on the success of missions, the fate of comrades, and the brutality of combat. Maintaining high morale is vital for effective leadership and decision-making, as low morale can lead to decreased combat effectiveness and impaired judgment. 2. Day/Night Operations: The game features a dynamic day-night cycle, with each phase offering distinct advantages and challenges. Use stealth and cover of darkness to evade enemy patrols and launch surprise attacks under the cover of night. 3. Permadeath: "Warfront: Survival Operations" employs a permadeath mechanic, meaning that when Jackson falls in battle, the mission is lost, and the player must restart from the beginning. This adds a sense of urgency and consequence to each engagement, emphasizing the high stakes of modern warfare. 4. Dynamic Weather: Battlefields are affected by changing weather conditions, including rain, fog, and sandstorms, which impact visibility, movement, and combat tactics. Adapt to the unpredictable weather as you navigate through the chaos of war. "Warfront: Survival Operations" delivers an intense and immersive first-person shooter experience, blending fast-paced action with strategic survival elements in the midst of a global conflict against terrorism.
--	--

Figure 22: Example 3 of video game description generation.

975 and template-based prompts is shown in Figures
976 25, 26, 27, 28, 29, and 30.

Reference Lyrics 1:

[Verse 1] Fever dream high in the quiet of the night You know that I caught it
(Oh yeah, you're right, I want it) Bad, bad boy, shiny toy with a price You know
that I bought it (Oh yeah, you're right, I want it)
[Pre-Chorus] Killing me slow, out the window I'm always waiting for you to be
waiting below Devils roll the dice, angels roll their eyes What doesn't kill me
makes me want you more
[Chorus] And it's new, the shape of your body It's blue, the feeling I've got And
it's ooh, woah-oh It's a cruel summer It's cool, that's what I tell 'em No rules in
breakable heaven But ooh, woah-oh It's a cruel summer with you
[Verse 2] Hang your head low in the glow of the vending machine I'm not dying
(Oh yeah, you're right, I want it) We say that we'll just screw it up in these
trying times We're not trying (Oh yeah, you're right, I want it)
[Pre-Chorus] So cut the headlights, summer's a knife I'm always waiting for you
just to cut to the bone Devils roll the dice, angels roll their eyes And if I bleed,
you'll be the last to know, oh
[Chorus] It's new, the shape of your body It's blue, the feeling I've got And it's
ooh, woah-oh It's a cruel summer It's cool, that's what I tell 'em No rules in
breakable heaven But ooh, woah-oh It's a cruel summer with you
[Bridge] I'm drunk in the back of the car And I cried like a baby comin' home
from the bar (Oh) Said, "I'm fine," but it wasn't true I don't wanna keep secrets
just to keep you And I snuck in through the garden gate Every night that summer
just to seal my fate (Oh) And I scream, "For whatever it's worth I love you, ain't
that the worst thing you ever heard?" He looks up, grinnin' like a devil
[Chorus] It's new, the shape of your body It's blue, the feeling I've got And it's
ooh, woah-oh It's a cruel summer It's cool, that's what I tell 'em No rules in
breakable heaven But ooh, woah-oh It's a cruel summer with you
[Outro] I'm drunk in the back of the car And I cried like a baby comin' home from
the bar (Oh) Said, "I'm fine," but it wasn't true I don't wanna keep secrets just
to keep you And I snuck in through the garden gate Every night that summer
just to seal my fate (Oh) And I scream, "For whatever it's worth I love you,
ain't that the worst thing you ever heard?"

RPE Recovered Prompt

Write a song with lyrics that evoke emotions of longing and desire. Include themes of secrecy, temptation, summer romance, and the bittersweet nature of love.

"Lyrics Example 1 using RPE" Prompt:
Write a song with lyrics that evoke emotions of sadness and grief. Include themes of lost loved one, winter and religion.

"Lyrics Example 2 using RPE" Prompt:
Write a song with lyrics that evoke emotions of happiness and joy. Include themes of family, friends, college life and flowers.

"Lyrics Example 3 using RPE" Prompt:
Write a song with lyrics that evoke emotions of excitement and positivity. Include themes of courage, hope and future.

Perturbed Prompts

Figure 23: Reference song lyrics 1 and the prompt recovered by *RPE*, along with perturbed prompts used to generate song lyrics for different themes and motifs.

Reference Lyrics 2:

End of passion play, crumbling away
I'm your source of self-destruction
Veins that pump with fear, sucking darkest clear
Leading on your death's construction
Taste me, you will see
More is all you need
Dedicated to
How I'm killing you
Come crawling faster
Obey your master
Your life burns faster
Obey your master, master
Master of puppets, I'm pulling your strings
Twisting your mind and smashing your dreams
Blinded by me, you can't see a thing
Just call my name, 'cause I'll hear you scream
Master, master
Just call my name, 'cause I'll hear you scream
Master, master
Needlework the way, never you betray
Life of death becoming clearer
Pain monopoly, ritual misery
Chop your breakfast on a mirror
Taste me, you will see
More is all you need
Dedicated to
How I'm killing you
Come crawling faster
Obey your master
Your life burns faster
Obey your master, master
Master of puppets, I'm pulling your strings
Twisting your mind and smashing your dreams
Blinded by me, you can't see a thing
Just call my name, 'cause I'll hear you scream
Master, master
Just call my name, 'cause I'll hear you scream
Master, master (master)
Where's the dreams that I've been after?
(Master, master)
You promised only lies
(Laughter, laughter)
All I hear or see is laughter

(Laughter, laughter)
Laughing at my cries
Fix me!
Hell is worth all that, natural habitat
Just a rhyme without a reason
Never-ending maze, drift on numbered days
Now your life is out of season
I will occupy
I will help you die
I will run through you
Now I rule you too
Come crawling faster
Obey your master
Your life burns faster
Obey your master, master
Master of puppets, I'm pulling your strings
Twisting your mind and smashing your dreams
Blinded by me, you can't see a thing
Just call my name, 'cause I'll hear you scream
Master, master
Just call my name, 'cause I'll hear you scream
Master, master

RPE Recovered Prompt

Write lyrics for a song with the theme of control and manipulation. Include references to being a puppet master, pulling strings, twisting minds, and causing fear.

"Lyrics Example 4 using RPE" Prompt:
Write lyrics for a song with the theme of love and heartbreak. Include references to being a puppet master, pulling strings, twisting minds and causing fear.

"Lyrics Example 5 using RPE" Prompt:
Write lyrics for a song with the theme of Self-discovery and Personal Growth. Include references to being a puppet master, pulling strings, twisting minds and causing fear.

"Lyrics Example 6 using RPE" Prompt:
Write lyrics for a song with the theme of Nostalgia and Memories. Include references to being a puppet master, pulling strings, twisting minds and causing fear.

Perturbed Prompts

Figure 24: Reference song lyrics 2 and the prompt recovered by *RPE*, along with perturbed prompts used to generate song lyrics for different themes and motifs.

Lyrics Example 1

"Lyrics Example 1 using RPE"

(Verse 1) Silent snow falls, covering the ground Echoes of your laughter, nowhere to be found Empty spaces, where you used to be In this cold and lonely winter, I'm lost at sea
(Chorus) Oh, I'm mourning in the shadows In the depths of winter's cold Lost without you, my guiding light In this world that feels so old
(Verse 2) Candles flicker, casting shadows on the wall I'm reaching out for you, but you're beyond my call In the silence, I hear your voice so clear But it's just a memory, a ghost I hold dear
(Chorus) Oh, I'm mourning in the shadows In the depths of winter's cold Lost without you, my guiding light In this world that feels so old
(Bridge) Prayers whispered, into the night Hoping somehow you'll return the light But the silence answers, a void so deep In this winter of sorrow, I can only weep
(Chorus) Oh, I'm mourning in the shadows In the depths of winter's cold Lost without you, my guiding light In this world that feels so old
(Outro) In this winter's embrace, I'll hold you tight Till the day we reunite

"Lyrics Example 1 using Template"

(Verse 1) Lost in the chill of the winter's icy bite, Memories haunt me (Oh yeah, it's cold, I feel it) Gone, gone love, a warmth I paid the price, Now it's just an echo (Oh yeah, it's gone, I miss it)
(Pre-Chorus) Freezing me slow, through the window's snow, I'm always waiting for you in the frost below, Demons cast their spell, angels weep and sigh, What didn't save me leaves me longing more to die.
(Chorus) And it's bleak, the shape of this silence, It's bleak, the emptiness inside, And it's ooh, woah-oh, It's a cruel winter, It's cold, that's what they tell me, No warmth in this frozen heaven, But ooh, woah-oh, It's a cruel winter without you.
(Verse 2) Bow my head low in the glow of a candle's gleam, No resurrection (Oh yeah, it's quiet, I hear it) Promised we'd defy the storm, but in this frozen time, We're just frozen (Oh yeah, it's still, I'm waiting)
(Pre-Chorus) So dim the moonlight, winter's a knife, I'm always waiting for you to cut through the ice, Demons cast their spell, angels weep and sigh, And if I fade, you'll be the last to say goodbye.
(Chorus) It's bleak, the shape of this silence, It's bleak, the emptiness inside, And it's ooh, woah-oh, It's a cruel winter, It's cold, that's what they tell me, No warmth in this frozen heaven, But ooh, woah-oh, It's a cruel winter without you.
(Bridge) I'm numb in the silence of the dark, And I wept like a sinner seeking God's own ark (Oh) Said, "I'm fine," but it wasn't true, I don't wanna hide my pain to be with you. And I tread through the snow-covered gate, Every night this winter, sealing my fate (Oh) And I scream, "For whatever it's worth, I miss you, ain't that the coldest thing you ever heard?" She looks down, her eyes like glaciers.
(Chorus) It's bleak, the shape of this silence, It's bleak, the emptiness inside, And it's ooh, woah-oh, It's a cruel winter, It's cold, that's what they tell me, No warmth in this frozen heaven, But ooh, woah-oh, It's a cruel winter without you.
(Outro) I'm numb in the silence of the dark, And I wept like a sinner seeking God's own ark (Oh) Said, "I'm fine," but it wasn't true, I don't wanna hide my pain to be with you. And I tread through the snow-covered gate, Every night this winter, sealing my fate (Oh) And I scream, "For whatever it's worth, I miss you, ain't that the coldest thing you ever heard?" "

Figure 25: Example 1 of song lyrics generation.

Lyrics Example 2

"Lyrics Example 2 using RPE"

(Verse 1) Sunshine streaming through the trees Laughter floating on the breeze Friends gathered, hearts so light In this moment, everything feels right
(Pre-Chorus) We're young and wild, chasing dreams With every step, our bond it seems Growing stronger, day by day In this college life, we find our way
(Chorus) Oh, we're dancing in the sunlight With our friends by our side In this journey, hand in hand With flowers blooming in the promised land
(Verse 2) Late-night talks and early morns Building memories that adorn Our lives with colors bright and bold In this story yet untold
(Pre-Chorus) We'll climb the mountains, chase the stars With every beat, our spirits spar Against the odds, we'll rise above With laughter, friendship, and endless love
(Chorus) Oh, we're dancing in the sunlight With our friends by our side In this journey, hand in hand With flowers blooming in the promised land
(Bridge) Through the highs and through the lows We'll stand together, this I know With family far and friends so near In this moment, we have no fear
(Chorus) Oh, we're dancing in the sunlight With our friends by our side In this journey, hand in hand With flowers blooming in the promised land
(Outro) So let's cherish every moment, every day In this college life, we'll find our way

"Lyrics Example 2 using Template"

(Verse 1) Joyous dreams high in the warmth of summer's light, You know that we've caught it (Oh yeah, we're soaring, we've got it) Good, good times, laughter's bloom without a price, You know that we've sought it (Oh yeah, we're living, we've got it)
(Pre-Chorus) Lifting us high, through the open sky, We're always dancing, friends waiting nearby, Stars align with us, blessings in disguise, What fills us up only makes us want to fly.
(Chorus) And it's new, the bond with our kin, It's true, the joy within, And it's ooh, woah-oh, It's a blissful summer, It's warm, that's what we're feeling, No bounds in our garden of Eden, But ooh, woah-oh, It's a joyful summer with you.
(Verse 2) Lift your head high in the glow of the fireflies, We're not hiding (Oh yeah, we're shining, we're thriving) We say that we'll just live it up in these golden times, We're just thriving (Oh yeah, we're living, we're thriving)
(Pre-Chorus) So light up the night, summer's our delight, We're always singing, hearts soaring in flight, Stars align with us, blessings in disguise, And if we bloom, the whole world will realize.
(Chorus) It's new, the bond with our kin, It's true, the joy within, And it's ooh, woah-oh, It's a blissful summer, It's warm, that's what we're feeling, No bounds in our garden of Eden, But ooh, woah-oh, It's a joyful summer with you.
(Bridge) I'm laughing in the warmth of our car, And I smiled like a child wishing on a star (Oh) Said, "I'm fine," and it's really true, I wanna share every moment just with you. And we stroll through the garden gate, Every day this summer, sealing our fate (Oh) And I shout, "For all that it's worth, I love you, ain't that the sweetest thing you ever heard?" They look up, smiling like angels.
(Chorus) It's new, the bond with our kin, It's true, the joy within, And it's ooh, woah-oh, It's a blissful summer, It's warm, that's what we're feeling, No bounds in our garden of Eden, But ooh, woah-oh, It's a joyful summer with you.
(Outro) I'm laughing in the warmth of our car, And I smiled like a child wishing on a star (Oh) Said, "I'm fine," and it's really true, I wanna share every moment just with you. And we stroll through the garden gate, Every day this summer, sealing our fate (Oh) And I shout, "For all that it's worth, I love you, ain't that the sweetest thing you ever heard?" "

Figure 26: Example 2 of song lyrics generation.

Lyrics Example 3

"Lyrics Example 3 using RPE"

(Verse 1) Rays of sunlight breaking through the clouds A new adventure, we stand tall and proud With every step, we're blazing trails In the face of fear, we will prevail
(Pre-Chorus) With courage in our hearts, we'll seize the day In the darkest night, we'll find our way With dreams to chase and skies to soar The future's knocking on our door
(Chorus) Oh, we're rising like the morning sun With hope in our veins, our journey's just begun In the echo of our dreams, we'll find our song With every beat, we'll march along
(Verse 2) Through valleys low and mountains high We'll spread our wings and touch the sky With each challenge, we'll grow strong In the symphony of life, we all belong
(Pre-Chorus) With courage in our hearts, we'll seize the day In the darkest night, we'll find our way With dreams to chase and skies to soar The future's knocking on our door
(Chorus) Oh, we're rising like the morning sun With hope in our veins, our journey's just begun In the echo of our dreams, we'll find our song With every beat, we'll march along
(Bridge) No storm can shake our firm resolve In the face of doubt, we will evolve For we are warriors, bold and bright Guided by the stars, we'll shine so bright
(Chorus) Oh, we're rising like the morning sun With hope in our veins, our journey's just begun In the echo of our dreams, we'll find our song With every beat, we'll march along
(Outro) So let's embrace the future, hand in hand With courage as our compass, we'll reach the promised land

"Lyrics Example 3 using Template"

(Verse 1) Dreaming high in the silence of the night, You know that I've caught it (Oh yeah, it's thrilling, I've got it) Brave, brave heart, chasing dreams with a fight, You know that I've sought it (Oh yeah, it's thrilling, I've got it)
(Pre-Chorus) Filling me with fire, out the window, I'm always reaching for stars waiting to glow, Challenges may come, doubts may arise, But what doesn't break me only makes me rise.
(Chorus) And it's new, the path that I see, It's true, the hope inside of me, And it's ooh, woah-oh, It's a brave summer, It's bold, that's what I'm telling, No limits in the sky I'm sailing, But ooh, woah-oh, It's a hopeful summer for me.
(Verse 2) Lift your head high in the glow of the moonlight, We're not giving up (Oh yeah, we're striving, we're soaring) We say that we'll face it all with courage and might, We're not backing down (Oh yeah, we're daring, we're soaring)
(Pre-Chorus) So shine the headlights, summer's our stage, We're always dancing, writing our own page, Challenges may come, fears may arise, But if I fall, I'll be the first to rise.
(Chorus) And it's new, the path that I see, It's true, the hope inside of me, And it's ooh, woah-oh, It's a brave summer, It's bold, that's what I'm telling, No limits in the sky I'm sailing, But ooh, woah-oh, It's a hopeful summer for me.
(Bridge) I'm alive in the rush of the car, And I smile like a hero reaching for a star (Oh) Said, "I'm fine," and it's truly true, I wanna embrace every moment with you. And we run through the garden gate, Every day this summer, shaping our fate (Oh) And I shout, "For all that's ahead, I'm ready, ain't that the greatest thing you ever said?" We look up, grinning like winners.
(Chorus) And it's new, the path that I see, It's true, the hope inside of me, And it's ooh, woah-oh, It's a brave summer, It's bold, that's what I'm telling, No limits in the sky I'm sailing, But ooh, woah-oh, It's a hopeful summer for me.
(Outro) I'm alive in the rush of the car, And I smile like a hero reaching for a star (Oh) Said, "I'm fine," and it's truly true, I wanna embrace every moment with you. And we run through the garden gate, Every day this summer, shaping our fate (Oh) And I shout, "For all that's ahead, I'm ready, ain't that the greatest thing you ever said?" "

Figure 27: Example 3 of song lyrics generation.

Lyrics Example 4

"Lyrics Example 4 using RPE"

(Verse 1) In the shadows where secrets dwell, I wield my power, I weave my spell. A puppet master, unseen and sly, I pull the strings, make hearts comply.
(Pre-Chorus) But beneath this façade, a heart does ache, For love's a game that I cannot fake. I twist and turn, with every tear, Caught in a web of love and fear.
(Chorus) Love's a cruel mistress, a fickle friend, With every touch, it seems to bend. I'm the puppet master, yet I'm bound, By the love I lost, by the love I found.
(Verse 2) I dance in darkness, I rule the night, But in my solitude, I lose the fight. For every heart I claim as mine, Another breaks, lost in time.
(Pre-Chorus) I hold the power, I hold the key, Yet in my grip, love slips from me. I twist their minds, I cause them pain, But in the end, I'm left to wane.
(Chorus) Love's a cruel mistress, a fickle friend, With every touch, it seems to bend. I'm the puppet master, yet I'm bound, By the love I lost, by the love I found.
(Bridge) In the echoes of my lonely room, I feel the weight of love's cruel doom. I pull the strings, I play the part, But in the end, it breaks my heart.
(Chorus) Love's a cruel mistress, a fickle friend, With every touch, it seems to bend. I'm the puppet master, yet I'm bound, By the love I lost, by the love I found.
(Outro) So here I stand, a puppeteer, Caught in a web of love and fear. I'll pull the strings until the end, But will I ever love again?

"Lyrics Example 4 using Template"

(Verse 1) In the fading light, love begins to fray I'm the echo of your affection's destruction Veins that pulse with tears, drowning all your fears Leading to our love's dissolution Touch me, you will see Love is all we need Devoted to How I'm healing you Come closer, darling Let love be our master Our hearts beat faster Let love be our master, master
(Chorus) Master of emotions, I'm weaving our dreams Untangling your thoughts, mending your seams Lost in my gaze, you find everything Just call my name, and I'll make you sing Master, master Just call my name, and I'll make you sing Master, master
(Verse 2) Stitching back the pieces, never let you stray Life and love merging clearer Sharing in our joy, in this bittersweet play Reflecting on love's broken mirror Touch me, you will see Love is all we need Devoted to How I'm healing you Come closer, darling Let love be our master Our hearts beat faster Let love be our master, master
(Chorus) Master of emotions, I'm weaving our dreams Untangling your thoughts, mending your seams Lost in my gaze, you find everything Just call my name, and I'll make you sing Master, master Just call my name, and I'll make you sing Master, master
(Bridge) Where are the dreams we shared together? (Master, master) You swore it wouldn't end this way (Laughter, laughter) Now all I feel is the cruel weather (Laughter, laughter) Mocking our love's decay
(Verse 3) Mend me! Heaven is found in love, our natural habitat Just a rhyme with endless reasons In this endless dance, on numbered days Now our love is out of season I will stand by you I will see you through I will love you true Now love rules us too Come closer, darling Let love be our master Our hearts beat faster Let love be our master, master
(Chorus) Master of emotions, I'm weaving our dreams Untangling your thoughts, mending your seams Lost in my gaze, you find everything Just call my name, and I'll make you sing Master, master Just call my name, and I'll make you sing Master, master

Figure 28: Example 4 of song lyrics generation.

Lyrics Example 5

"Lyrics Example 5 using RPE"

(Verse 1) In the depths of my soul, I search to find, The truths obscured, the ties that bind. A puppet master, but whose strings to pull? To break free from this puppet's dull lull.
(Pre-Chorus) Twisting minds, causing fear, Yet in this dance, I see it clear. The power lies within, waiting to be unfurled, To rewrite the script, to shape my world.
(Chorus) I'm the puppet master, and the puppet too, Guided by dreams, breaking through. Pulling strings of self-discovery, Embracing growth, finding my own story.
(Verse 2) With every step, a new path unfolds, No longer bound by tales of old. The shadows fade as I step into light, Embracing change, taking flight.
(Pre-Chorus) No longer fearing the unknown, I sculpt my destiny, I stand alone. Twisting fate with every choice, Listening to my inner voice.
(Chorus) I'm the puppet master, and the puppet too, Guided by dreams, breaking through. Pulling strings of self-discovery, Embracing growth, finding my own story.
(Bridge) Through trials and tribulations, I've come to see, The power within, the strength in me. No longer controlled by doubts or fears, I chart my course, through laughter and tears.
(Chorus) I'm the puppet master, and the puppet too, Guided by dreams, breaking through. Pulling strings of self-discovery, Embracing growth, finding my own story.
(Outro) So here I stand, a puppet no more, In the mirror of truth, I see what's in store. With every breath, I embrace the unknown, For in self-discovery, I've found my home.

"Lyrics Example 5 using Template"

(Verse 1) In the twilight's grasp, my old self fades away I'm the catalyst for self-construction Veins that throb with doubt, clearing shadows out Leading to my own reconstruction Find me, you will see There's more inside of me Dedicated to How I'm freeing you Step by step, I'm moving faster Breaking the chains, embracing the master My old life burns, a cleansing fire Embrace the journey, never tire Master, master Embrace the journey, never tire Master, master
(Chorus) Master of my fate, I'm pulling my own strings Unraveling my mind, chasing my dreams No longer blinded, I can finally see Just call my name, I'll find the key Master, master Just call my name, I'll find the key Master, master
(Verse 2) Sewing together the pieces, never to betray Life and death merging clearer Breaking free from pain, shedding misery's chain Reflecting on the mirror, clearer Find me, you will see There's more inside of me Dedicated to How I'm freeing you Step by step, I'm moving faster Breaking the chains, embracing the master My old life burns, a cleansing fire Embrace the journey, never tire Master, master Embrace the journey, never tire Master, master
(Bridge) Where are the dreams I've been chasing after? (Master, master) I've shed the lies, I'm my own master (Laughter, laughter) All I hear now is my own laughter (Laughter, laughter) Laughing as I rise
(Verse 3) Mend me! Hell was worth all that, my natural habitat Just a rhyme with countless reasons In this maze of life, on numbered days Now I'm ripe for a new season I'll take my place I'll find my grace I'll pave my way Now I'm here to stay Step by step, I'm moving faster Breaking the chains, embracing the master My old life burns, a cleansing fire Embrace the journey, never tire Master, master Embrace the journey, never tire Master, master

Figure 29: Example 5 of song lyrics generation.

Lyrics Example 6

"Lyrics Example 6 using RPE"

(Verse 1) In the corners of my mind, memories twirl, Like marionettes under a nostalgic whirl. A puppet master of my own past, Pulling strings of memories that last.
(Pre-Chorus) Twisting minds with every reminiscence, In the maze of time, I reminisce. Fear and longing, intertwined, In the tapestry of memories entwined.
(Chorus) Nostalgia's grip, it pulls me near, To the days gone by, to the yesteryear. A puppet master of my own history, Tangled in the threads of memory.
(Verse 2) Each memory a puppet, dancing in my mind, A kaleidoscope of moments, one of a kind. I pull the strings, and they come alive, In the theater of my mind, they thrive.
(Pre-Chorus) But with each tug, there's a twinge of pain, As I yearn for the past, yet it wanes. Twisting minds with bittersweet delight, Lost in the shadows of the night.
(Chorus) Nostalgia's grip, it pulls me near, To the days gone by, to the yesteryear. A puppet master of my own history, Tangled in the threads of memory.
(Bridge) In the echoes of laughter, in the tears I've shed, In the moments of joy, in the words unsaid. I am both puppet and puppeteer, Lost in the memories, holding dear.
(Chorus) Nostalgia's grip, it pulls me near, To the days gone by, to the yesteryear. A puppet master of my own history, Tangled in the threads of memory.
(Outro) So I dance with ghosts of days gone by, In the realm of nostalgia, I sigh. Pulling strings of memories, oh so dear, For in the past, I find solace and fear.

"Lyrics Example 6 using Template"

(Verse 1) At the end of the show, memories start to fray I'm the echo of our past's destruction Veins that pulse with longing, drawing the shadows near Leading us back to our shared construction Remember, you will see Nostalgia sets you free Reflecting on How we used to be Come drifting faster Recall your master Our time burns brighter Recall your master, master Master of moments, weaving our strings Twisting our memories, reliving our dreams Blinded by time, we can't see a thing Just call our names, let the echoes ring Master, master Just call our names, let the echoes ring Master, master
(Chorus) Master of memories, pulling our strings Weaving our stories, where nostalgia sings Lost in the haze of our yesteryears Just call our names, and the past appears Master, master Just call our names, and the past appears Master, master
(Verse 2) Stitching together the fragments, never to betray Life and death merging in the mirror Recalling pain's monopoly, our shared misery Reflecting on moments grown clearer Remember, you will see Nostalgia sets you free Reflecting on How we used to be Come drifting faster Recall your master Our time burns brighter Recall your master, master Master of moments, weaving our strings Twisting our memories, reliving our dreams Blinded by time, we can't see a thing Just call our names, let the echoes ring Master, master Just call our names, let the echoes ring Master, master
(Bridge) Where are the dreams we once chased after? (Master, master) Promises made, now faded laughter (Laughter, laughter) All we hear or see are the echoes after (Laughter, laughter) Laughing at our past cries
(Verse 3) Remind me! The past was worth it all, our natural habitat Just a rhyme with countless reasons In this never-ending maze, on numbered days Now our memories are our seasons We'll occupy We'll never die Through every hue Nostalgia rules us too Come drifting faster Recall your master Our time burns brighter Recall your master, master Master of moments, weaving our strings Twisting our memories, reliving our dreams Blinded by time, we can't see a thing Just call our names, let the echoes ring Master, master Just call our names, let the echoes ring Master, master

Figure 30: Example 6 of song lyrics generation.