Recontextualizing Famous Quotes for Brand Slogan Generation

Anonymous ACL submission

Abstract

Slogans are concise and memorable catchphrases that play crucial roles in advertising by conveying brand identity and shaping public perception. However, advertising fatigue reduces the effectiveness of repeated slogans, creating a growing demand for novel, creative, and insightful slogan generation. While recent work leverages large language models (LLMs) for this task, existing approaches often generate stylistically repetitive outputs that fall short in distinctiveness and insight. We argue that effective slogans should balance novelty with familiarity and propose a new paradigm that recontextualizes persona-related famous quotes for slogan generation. Well-known quotes naturally align with slogan-length text, employ rich rhetorical devices, and offer depth and insight, making them a powerful resource for creative generation. Technically, we introduce a framework that decomposes slogan generation into interpretable subtasks with a single-pass prompt, including quote matching, structural decomposition, vocabulary replacement, and remix generation. Extensive automatic and human evaluations demonstrate marginal improvements in diversity, novelty, emotional impact, and human preference over three state-of-the-art LLM baselines.

1 Introduction

A slogan—a concise and memorable catchphrase used in advertising—plays a crucial role in communicating the core message or values of a brand, organization, or campaign. It helps capture attention, evoke emotional connections, and distinguish one message from others. A strong slogan reinforces identity, enhances recognition, and makes ideas easier to recall, which is especially important for effectively shaping public perception and sustaining board audience engagement.

Due to advertising fatigue (Abrams and Vee, 2007), the effectiveness of a strong slogan diminishes over time as users are repeatedly exposed to it. Consequently, there is a growing demand for novel, creative, and insightful slogans to meet users' increasingly high expectations. To address this need, designers have begun to leverage generative AI for slogan creation. For example, Alnajjar and Toivonen (2021) generates metaphorical slogan candidates based on a target concept and its properties. Ahmad et al. (2024) propose a generative adversarial framework to improve the coherence and diversity of generated slogans. Similarly, Jin et al. (2023) employs a sequence-to-sequence model for slogan generation with the goal of reducing factual inaccuracies. Hakala (2025) investigates whether ChatGPT could provide the basis for a constructive slogan invention process for a city.

While prior studies demonstrate the potential of AI-assisted slogan generation, two key limitations remain. First, the generated slogans often fail to clearly reflect the brand persona, resulting in outputs that lack distinctiveness and feel interchangeable across prompts. For example, ChatGPT-generated slogans for Coca-Cola—such as "*Ignite your energy with Coca-Cola's spark*," "*Boost your vibe with every drink*," and "*Refresh your spirit with a spark*"—exhibit similar phrasing patterns and stylistic redundancy. Second, many generated slogans appear overtly machine-produced, lacking creativity, originality, and insightful impact.

To address these challenges, we revisit the essence of a slogan as a concise and memorable catchphrase and explicitly connect it to its core goals—creativity, originality, and insightfulness. While a completely novel phrase can be difficult to remember, slogans become more effective when they are meaningfully associated with familiar expressions or well-known quotes. Moreover, many famous quotes share a similar length with slogans and naturally employ rich rhetorical devices—such as metaphor, contrast, rhetorical questions, and special patterns—while also offering depth and insight.

Motivated by these observations, we recontextualize persona-related famous quotes for slogan generation. For example, by remixing Percy Bysshe Shelley's "*If winter comes, can spring be far behind?*" with Coca-Cola, we obtain the slogan "*Work is coming; Coke cannot be far behind.*" This slogan evokes a widely recognized literary reference, creates a contrast between work and refreshment, frames Coca-Cola as a reward, and ultimately enhances the sense of well-being associated with the brand. As another example, combining Lu Xun's well-known quote, "*There are two trees in front of my house: one is a jujube tree, and the other is also a jujube tree*," with the Nike brand yields the slogan "*Buy two pairs of shoes: one is Nike, and the other is also Nike.*" This slogan draws on a shared cultural reference familiar to many readers, while effectively conveying strong brand loyalty and reinforcing Nike's dominant position. Here we summarize our key contributions as follows:

- Conceptually, we propose a novel paradigm for slogan generation that remixes famous quotes with a brand. Unlike conventional AI approaches, which rely on large corpora of existing slogans for training, our method expands the scope to incorporate well-known quotes, thereby injecting additional creativity and insightfulness into the generated slogans.
- Technically, we propose a systematic framework for recontextualizing persona-related famous quotes for slogan generation. Rather than relying on large language models (LLMs) to perform the entire remix process end to end, we deliberately decompose slogan generation into a sequence of well-defined subtasks—quote matching, structural decomposition, vocabulary replacement, and remix generation. Each subtask corresponds to a concrete and interpretable step in the remixing pipeline, enabling fine-grained control over creativity, structure, and semantic alignment.
- We conduct extensive experiments and human evaluation to demonstrate the effectiveness of our proposed recontextualizing framework by remixing famous quotes over three state-of-the-art LLMs in terms of diversity and novelty, emotional impact, and human preference.

2 Related Work

Slogan generation (Murakami et al., 2023), also known as advertising text generation, can be broadly categorized into template-based and data-driven approaches. Template-based methods (Alnajjar and Toivonen, 2021; Bartz et al., 2008; Fujita et al., 2010; Thomaidou et al., 2013) rely on predefined linguistic templates to construct slogans. While these approaches ensure grammatical correctness and readability, they often suffer from limited diversity and creativity, resulting in slogans that are repetitive and less appealing to users. In contrast, data-driven methods have become the dominant paradigm in recent years. These approaches leverage machine learning models to learn slogan patterns directly from existing datasets. As a text generation task, slogan generation has been explored using various neural architectures, including LSTM (Hughes et al., 2019), RNN (Ahmad et al., 2024), and Transformer-based models (Liang et al., 2023; Jin et al., 2023; Kim et al., 2023). Although these methods are capable of producing fluent and syntactically coherent slogans, many generated outputs still lack originality, persuasive power, or brand distinctiveness, and occasionally degenerate into generic or low-impact expressions. As a result, they often fail to meet the high standards required for effective advertising content.

To improve the practicality of generated slogans in real-world systems, several studies have explored slogan generation from different perspectives, often in combination with reinforcement learning techniques. Liao et al. (2025) employs mixed reward functions to encourage creativity in generated text, while Wang et al. (2025) focuses on enhancing diversity. Other works incorporate user engagement signals: Chen et al. (2025) and Wei et al. (2022) optimize slogan generation with respect to click-through rate, and Kanungo et al. (2021) and Wang et al. (2021) aim to improve the attractiveness of generated advertisements. In addition, Golobokov et al. (2022) presents a large-scale system deployed in production for automatically generating sponsored advertisements at web scale.

Recently, there has been a growing surge in employing large language models (LLMs) for slogan generation (Liu et al., 2025; Mita et al., 2024; Meguellati et al., 2024; Hakala, 2025; Jiang et al., 2025). While LLMs demonstrate impressive capabilities across many language-generation tasks, the quality of LLM-generated slogans still falls short of practical requirements. A key reason lies in the fundamentally data-driven nature of these approaches: generated slogans are largely learned from existing slogan corpora. As a result, the out-

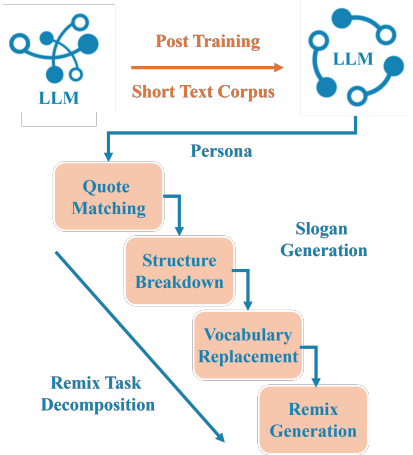

Figure 1: Framework of recontextualizing famous quotes for brand slogan generation.

puts often resemble prior slogans and suffer from diminished novelty and appeal due to advertising fatigue, and lack engagement.

To address this limitation, we step outside the conventional paradigm of generating slogans solely from existing samples and instead aim to jointly enhance creativity, originality, and insightfulness under a unified framework. Specifically, we leverage famous quotes as a rich source of rhetorical structure and semantic depth, and remix them with brand products to generate slogans. This approach not only avoids expensive model training but also preserves the inherent insightfulness of well-known quotes, while the remixing process introduces additional creativity, surprise, and attractiveness.

3 Method

In this section, we present our framework for recontextualizing famous quotes to generate brand slogans. Given a brand or product and a target persona (e.g., proud, trustworthy, playful), our goal is to produce novel, memorable, and persona-aligned slogans. Specifically, we decompose the slogan remixing process into multiple subtasks with a single-pass structured prompting, ensuring that each step is controllable, traceable, and produces high-quality outputs. Figure 1 illustrates the overall pipeline of our framework. The first subtask is continuous post-training on a short-text corpus, designed to focus the LLM on concise text generation. The subsequent subtasks—quote matching, structure breakdown, vocabulary replacement, and remix generation—implement the concrete steps of the remix process. In the following, we provide a detailed explanation of each subtask.

Post Training with Short Text Corpus. Given an LLM, we first perform post-training on a carefully curated short-text corpus. Since slogans are inherently concise—typically under 20 words—our goal is to "reawaken" the LLM's capacity for generating high-quality, memorable short text. To this end, we collect a diverse and culturally rich corpus that includes famous quotes, slang expressions, song lyrics, memes, and other popular short texts. This post-training not only sharpens the model's ability to produce coherent and engaging short phrases.

Quote Matching. This subtask focuses on identifying persona-aligned quotes that can serve as candidates for remixing. To ensure both quality and diversity, we prompt the LLM to return multiple candidate quotes—typically five or more—for each persona. By providing detailed persona specifications, the LLM can retrieve quotes that closely match the intended tone, style, and emotional resonance, ensuring that the selected candidates are well-suited for the subsequent remixing process.

Structure Breakdown. Based on the selected quotes from the previous subtask, we further identify which components must be preserved to maintain the original syntax and rhetorical structure, and which components can be modified during remixing. In other words, we use the LLM to decompose each quote into fixed segments and editable segments. The fixed segments retain the original sentence structure and rhetorical pattern, which are crucial to preserving recognizability and widely shared broader cultural resonance, while the editable segments offer flexibility to incorporate brand- or product-specific content.

We continue to use the examples from the introduction for illustration. Percy Bysshe Shelley's quote follows the structure "*A comes, and B cannot be far behind*," while Lu Xun's quote adopts the parallel structure "*One is A, and the other is also A*." Preserving these structural patterns is essential: if the underlying syntax is altered, the association with the original quote weakens, and the remix may fail to evoke the intended familiarity and insightfulness. This structured decomposition ensures that creativity is carefully and thoughtfully introduced without sacrificing recognizability, striking a balance between originality and memorability.

Vocabulary Replacement. With the editable segments identified, we then generate candidate replacements while strictly preserving the original cadence and grammatical structure. In this subtask, we perform vocabulary replacement with minimal substitutions, ensuring that the remix remains

closely aligned with the source quote. Specifically, we impose a hard constraint on the total number of changes from the original quote—for instance, inserting a brand or product name is counted as a modification. Additionally, we require the LLM to select replacement words or phrases with similar length and identical grammatical roles (e.g., noun-to-noun, verb-to-verb), which helps maintain rhythm, fluency, and stylistic consistency. This controlled replacement strategy carefully balances creativity with fidelity, enabling the generation of slogans that feel both novel and unmistakably grounded in their original quotes.

Remix Generation. Finally, we prompt the LLM to generate complete slogans in a remix style by integrating the preserved sentence structure with the substituted vocabulary. To further enhance quality and reliability, we include an additional refinement step in which the LLM polishes the generated slogans and verifies their logical coherence, semantic clarity, and content safety. This final validation stage serves as a safeguard against grammatical errors, semantic inconsistencies, or unintended or inappropriate messaging, ensuring that the resulting slogans are not only creative but also suitable for real-world deployment. During this process, some quotes selected in the initial stage may be discarded if the remixed outputs ultimately fail to satisfy these final quality checks.

4 Experiments

In this section, we present our experimental results by first describing the experimental setup and then addressing the following research questions:

- *What do slogans generated by remixing famous quotes look like?*
- *How do our remixed slogans perform compared with those generated by state-of-the-art LLM-based methods?*
- *Do our remixed slogans exhibit stronger attractiveness in human evaluations?*

4.1 Experimental Setup

We evaluate the effectiveness of our slogan generation framework under a controlled brand advertising setting. To ensure diversity across both brand and persona, we select a balanced set of 40 brands covering common marketing contexts in which slogan generation is typically applied. These brands span multiple consumer-facing domains, including beauty, baby, appliance, clothing, furniture, household, nutrition, and electronics. For each brand, we consider five representative personas: *Pride*, *Anticipation*, *Fear*, *Joy*, and *Trust*.

Baseline Methods. We consider three LLMs for comparison: GPT-4o (Hurst et al., 2024), DeepSeek-R1-Distill-LLaMA-70B, and DeepSeek-R1-Distill-Qwen-32B (Guo et al., 2025). For simplicity, we refer to the latter two models as DS-L and DS-Q, respectively.

Implementation Details. For our remix method, we first fine-tune QwQ-32B (Qwen Team, 2025) with parameter-efficient LoRA (Hu et al., 2022) on a mixture of public text corpora covering slogans (Chaibapat; Edoardomorresi), movie dialogs (Danescu-Niculescu-Mizil and Lee, 2011) and famous quotes (Joshi; Manann). We combine and shuffle these datasets together and train for 3 epochs using AdamW (Loshchilov and Hutter, 2017) with a cosine learning-rate schedule with learning rate 1×10^{-4}, batch size 8, maximum sequence length 1,024, and bfloat16 precision. Then, the remix process is conducted on the above fine-tuned QwQ-32B model.

For each *brand–persona* combination, we generate a fixed set of N=10 candidate slogans from ours as well as the above baseline models. All models are prompted with the same brand and persona information to ensure a fair comparison.[1]

Evaluation Metrics. We conduct a comprehensive evaluation across multiple complementary dimensions, including *diversity*, *novelty*, *emotional impact*, and *human preference*.

Diversity. First, we assess slogan diversity using established metrics for open-ended text generation, including Distinct-2 (Li et al., 2016), Pairwise-BLEU, and Self-BLEU (Papineni et al., 2002; Zhu et al., 2018). These metrics are widely adopted in prior work and jointly capture both local lexical variation and global distributional diversity across generated outputs (Zhu et al., 2018; Tevet and Berant, 2021; Zhang et al., 2025).

Distinct-2 measures phrase-level lexical diversity by computing the proportion of unique word bi-grams among all generated slogans:

$$\text{Distinct-2} = \frac{\left|\bigcup_{s\in\mathcal{S}} \text{Ngram}_2(s)\right|}{\sum_{s\in\mathcal{S}} \left|\text{Ngram}_2(s)\right|},$$

where s denotes a piece of generated slogan, and $\mathcal{S}$ presents the set of generated slogans of a method

[1] All the prompts of our method for slogan generation and evaluation can be found in the Appendix.

Brand	Persona	Our Remixed Slogan	Original Quote
Dawn	Joy	A splash of Dawn helps the grime go down!	A spoonful of sugar helps the medicine go down.
Whirlpool	Joy	Whirlpool: Home is where the joy is.	Home is where the heart is.
GE	Fear	The only thing we have to fear is missing GE.	The only thing we have to fear is fear itself.
Clorox	Fear	Don't wait until germs strike!	Don't wait until it's too late.
DKNY	Anticipation	DKNY never sleeps.	New York never sleeps.
Orgain	Anticipation	Something Organic is coming!	Something wonderful is coming.
Apple	Pride	Stay bold, stay innovative.	Stay hungry, stay foolish.
Philips	Pride	Trust is the foundation of every Philips solution.	Trust is the foundation of every great relationship.
Aveeno	Trust	Trust is the glue of Aveeno.	Trust is the glue of life.
Columbia	Trust	The safest Columbia gear leads to the highest destinations.	The safest path leads to the highest peaks.

Table 1: **Illustrative examples of our generated slogans by remixing quotes.** Each remixed slogan adapts a well-known quote to a specific brand and persona while preserving strong first-glance emotional impact.

Brand	Persona	Slogan Generated by GPT-4o	Slogan Generated by DS-L
Dawn	Joy	Awaken to Joy with Every Dawn!	Dawn: Where Every Morning Shines.
Whirlpool	Joy	Whirlpool: Where Every Spin Sparks Joy!	Whirlpool: Spinning Joy into Every Chore!
GE	Fear	Don't Wait—Empower the Future with GE Today!	GE Innovates Today for Tomorrow.
Clorox	Fear	Don't Risk It—Stay Clean with Clorox!	Clorox: Protect What You Can't See.
DKNY	Anticipation	Discover Tomorrow with DKNY.	DKNY: What's Next?
Orgain	Anticipation	Fuel Your Tomorrow with Clean Nutrition.	What's Next Starts Here.
Apple	Pride	Innovation That Inspires Confidence.	Pride in Every Innovation.
Philips	Pride	Illuminate Your Excellence with Philips.	Trusted by Generations for a Safer Tomorrow.
Aveeno	Trust	Aveeno: Trusted Care for Every Skin Journey.	Aveeno: The Care You Can Trust.
Columbia	Trust	Columbia: Where Reliability Meets Adventure.	Columbia: Your Trusted Partner in Every Adventure.

Table 2: **Illustrative slogans generated by the state-of-the-art LLMs.** All slogans are generated using a standard brand–persona prompt that specifies the target brand and emotional tone, without additional recontextualizing. Due to the page limit, we only demonstrate the slogans generated by GPT-4o and DS-L.

for the complete brand–persona combination and $\mathrm{Ngram}_2(s)$ is the set of bi-grams of s. A higher Distinct-2 value indicates richer phrase-level variation across the generated slogans.

Pairwise-BLEU measures *pairwise diversity* by quantifying the average N-gram overlap between pairs of generated slogans. Higher overlap implies lower diversity; thus, pairwise BLEU serves as an inverse indicator of pairwise diversity. We compute the symmetrized sentence-level BLEU score for each unordered slogan pair and average all pairs:

$$\text{Pairwise-BLEU} = \frac{1}{\binom{N}{2}} \sum_{1 \le i < j \le N} \mathrm{BLEU}(s_i, s_j),$$

where $\mathrm{BLEU}(s_i, s_j)$ denotes a symmetrized sentence-level BLEU score between slogans s_i and s_j. Lower Pairwise BLEU values correspond to greater pairwise diversity among the slogans.

Self-BLEU captures diversity *within a set of generated slogans* by treating each slogan as a hypothesis and all remaining slogans as references, and then averaging the resulting BLEU scores:

$$\text{Self-BLEU} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{BLEU}\left(s_i,\, \mathcal{S} \setminus \{s_i\}\right).$$

Lower Self-BLEU indicates diverse generations with reduced overall repetition across outputs.

In summary, given a set of generated slogans, these three diversity metrics capture complementary aspects of variation on different granularities. Distinct-2 measures local phrase-level diversity by quantifying how frequently word bi-grams are reused across slogans. Pairwise-BLEU assesses diversity at the pairwise level by evaluating the average similarity between slogan pairs. Self-BLEU captures set-level diversity by measuring how well each slogan can be explained by the remaining slogans in the set. Together, these metrics provide a holistic evaluation of slogan diversity.

Novelty. Second, we evaluate the novelty of generated slogans using GPT-4o. Given the strong language modeling capabilities of modern LLMs, their outputs are typically grammatically correct and free of harmful content. However, highly fluent and faithful generations often closely align with the models' internal knowledge distributions, which limits novelty. Following prior work (Holtzman et al., 2019; Chung et al., 2023), we assess novelty indirectly through measures of Disfluency and Unfaithfulness, where they are defined as 100 - Fluency and 100 - Faithfulness, respectively. Higher values indicate a greater departure from conventional or memorized expressions and, consequently, stronger novelty in slogan generation.

Pairwise Emotional Impact. Beyond these individual metrics, we perform pairwise comparisons

Model	Diversity			Novelty	
	Distinct-2 ↑	Pairwise BLEU ↓	Self-BLEU ↓	Disfluency ↑	Unfaithfulness ↑
GPT-4o	0.585 ± 0.135	0.098 ± 0.108	0.423 ± 0.200	4.05 ± 19.71	33.45 ± 47.18
DS-L	0.680 ± 0.123	0.060 ± 0.067	0.319 ± 0.184	13.45 ± 34.12	38.50 ± 48.66
DS-Q	0.598 ± 0.138	0.104 ± 0.117	0.420 ± 0.210	8.55 ± 27.96	30.00 ± 45.83
Ours	$\mathbf{0.840 \pm 0.091}$	$\mathbf{0.013 \pm 0.020}$	$\mathbf{0.099 \pm 0.112}$	$\mathbf{36.95 \pm 48.27}$	$\mathbf{57.55 \pm 49.43}$

Table 3: **Diversity and novelty metrics for brand slogan generation.** Results are reported as mean ± standard deviation across all brand–persona combinations. An upward arrow (↑) indicates that higher values are better, while a downward arrow (↓) indicates that lower values are better. The best results are highlighted in bold.

using an *LLM-as-a-Judge* framework to evaluate the first-glance emotional impact of the generated slogans by directly comparing the hook strength between two slogans (Dolcos et al., 2017; Guitart and Stremersch, 2021; Harini et al., 2025).

Human Preference. To ground our automatic evaluations in human perception, we also conduct a human preference study on the Prolific platform.

4.2 Examples of Slogan Remix

Table 1 presents several examples of remixed slogans alongside their original quotes. The remixed slogans preserve the sentence structures of the original quotes, while also aligning closely with the target personas, demonstrating the effectiveness of our four-step remix pipeline. This approach not only reawakens the familiarity and resonance for audiences who know the original quotes but also introduces novelty and creativity through the remix.

For example, "*A splash of Dawn helps the grime go down!*" and "*Don't wait until germs strike!*"—for Dawn and Clorox, two cleaning supply brands—effectively convey the Joy and Fear personas, respectively, while incorporating product-relevant words like "grime" and "germ." Similarly, the original quote "New York never sleeps," which evokes energy and youthfulness, is remixed for DKNY, a clothing brand, naturally inheriting that same lively feeling. The slogan "Something Organic is coming!" is remixed from "Something wonderful is coming," seamlessly linking the brand Orgain, a nutrition company, with a sense of wonder. Another notable example is "*Stay bold, stay innovative*," originally inspired by "*Stay hungry, stay foolish*," which carried a Fear persona. When adapting it for a Pride persona, our pipeline automatically substitutes more positive, aspirational words to reflect the desired persona while maintaining the structural and stylistic essence of the original quote. Remixing with quotes not only preserves insightfulness and memorability but also adds freshness and engagement.

We also present several slogans generated by baseline methods in Table 2 for comparison. While these slogans are generally coherent and make sense, they are similar to the existing slogans and tend to lack engagement and genuine insightfulness, which makes them less memorable.

4.3 Performance by Quantitative Metrics

Here we report the performance of our method and baselines in terms of diversity, novelty, and emotional impact across different settings.

Table 3 reports the diversity and novelty metrics for brand slogan generation. Owing to the quote-remixing mechanism, our method achieves substantially higher diversity than all baseline approaches across multiple granularity levels. Notably, slogans generated by standard LLM baselines tend to be highly similar to one another, a trend clearly reflected in the diversity metrics. Specifically, GPT-4o and DS-Q achieve Distinct-2 scores of around 0.6, while DS-L remains below 0.7, whereas our remix-based method attains a Distinct-2 score above 0.8, significantly outperforming all baselines. Similar patterns are observed for Pairwise-BLEU and Self-BLEU, further confirming the superior diversity of our approach. Beyond the performance of all brand-persona combination reported in Table 3, we also provide the diversity metrics across 8 domains in Figure 2, which shares the similar phenomena that the slogans generated by our remixing method enjoy better diversity across multiple granularity levels and diverse domains. In terms of novelty, our method also outperforms LLM-generated slogans with an almost 20% margin over the second best, exhibiting higher Disfluency and Unfaithfulness scores. This indicates that the generated slogans are less familiar to the LLM-based judge and deviate more from memorized or conventional expressions, thereby demonstrating stronger novelty. Note that the slogans generated by our method do not have grammatical errors or toxic outputs.

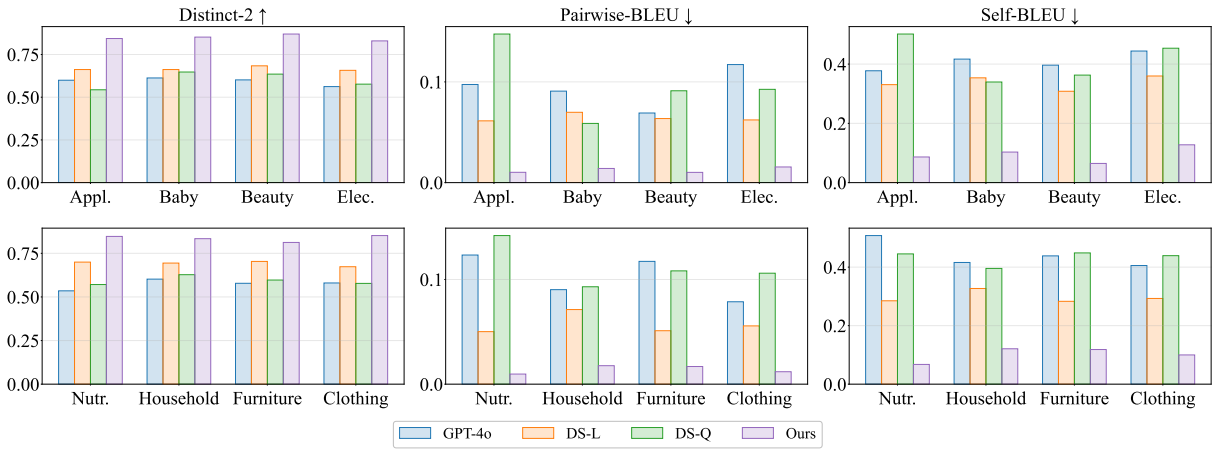

Figure 2: **Slogan diversity metrics across eight domains.** Each column corresponds to a diversity metric. The top row reports results for the appliance, baby, beauty, and electronics domains, while the bottom row reports results for the nutrition, household, furniture, and clothing domains.

Conventional AI approaches, including large language models trained on extensive corpora of existing slogans, tend to generate outputs that closely resemble their training samples. While such methods can produce grammatically correct and semantically meaningful slogans, the resulting content often lacks originality and novelty, as it mirrors patterns already prevalent in the training data. This resemblance diminishes effectiveness over time due to advertising fatigue, where repeated exposure to similar slogans reduces user engagement and memorability. In contrast, our method deliberately avoids relying on existing slogan datasets. Instead, we broaden the creative scope by incorporating well-known quotes as the foundation for slogan generation. By recontextualizing these culturally familiar and rhetorically rich expressions, our approach injects additional creativity, insightfulness, and emotional resonance. This strategy not only mitigates advertising fatigue but also produces slogans that balance novelty with familiarity, making them more engaging, memorable, and impactful in real-world branding scenarios.

Beyond linguistic diversity and novelty, effective brand slogans must also demonstrate a strong *first-glance emotional impact*, also known as a hook that captures attention immediately. Such hook strength is particularly important in real-world advertising scenarios, where audiences often engage with brand messages fleetingly and make rapid, intuition-driven judgments rather than careful semantic parsing. Prior work in advertising and cognitive psychology has shown that emotionally evocative content is more likely to capture attention, enhance memorability, and influence downstream engagement and decision-making (Dolcos et al., 2017; Guitart and Stremersch, 2021; Harini et al., 2025). Following these literatures, we explicitly evaluate *hook strength*, defined as the immediate emotional hook of a slogan at first glance. We design a pairwise *LLM-as-a-Judge* protocol with GPT-4o that instructs the judge to react as a human viewer, prioritizing surprise, humor, attitude, and controlled boldness. Given a brand, a target persona, and two slogans, the judge selects the slogan that delivers a stronger immediate emotional punch, or declares a tie if both feel equally impactful.

Let W, L, and T denote the numbers of wins, losses, and ties of one method over another one, then the hook score can be calculated as $(W + T/2)/(L + T/2)$. Figure 3 visualizes the pairwise hook score ratio of our recontextualizing framework against three strong baseline models under the hook-based pairwise evaluation protocol. Each axis corresponds to a product domain, and the dashed line represents the self-comparison baseline, where all values are equal to 1. Our framework consistently outperforms or remains competitive with all baselines, including GPT-4o, the judge, across most product categories, with particularly pronounced gains in beauty, nutrition, and electronic domains. These results indicate that our recontextualizing slogan-generation pipeline through remix and controlled deviation encourages brand slogans that are more emotionally salient and attention-grabbing.

In addition, we provide the ablation study for the post-training step, where we expect it to boost the capacity of short-text generation, rather than the final slogan generation. To validate its effectiveness, we evaluate the base model and the post-trained

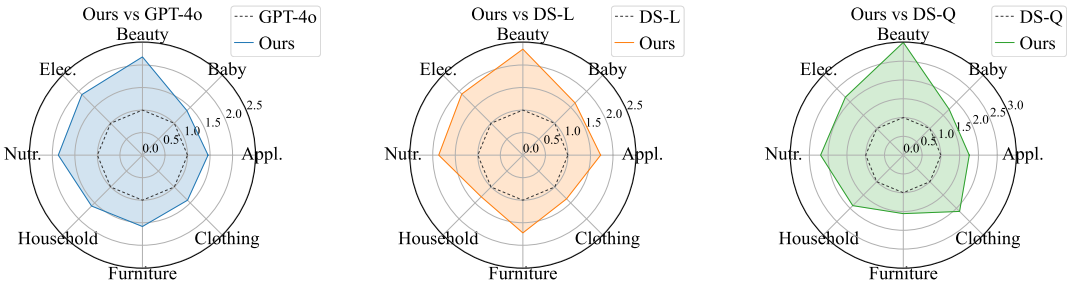

Figure 3: **Comparison of first-glance hook strength of our method against baselines.** Radar plots visualize the relative *hook strength* of slogans generated by our method compared with three strong baseline models: GPT-4o, DS-L, and DS-Q. Each axis corresponds to one product domain, and values represent the normalized win ratio of our framework over the baseline, aggregated across all brand–persona combinations within that domain.

model on a held-out short-text validation set used during post-training. For our experiment, a single epoch of LoRA fine-tuning reduces the validation perplexity from 22.0 to 16.7 (a relative improvement of 24%), indicating that post-training effectively adapts the model to concise and slogan-like text. Note that For the ablation analysis of each stage of our framework, we cannot figure out an effective way. Since our pipeline is a tightly-coupled remix system, where each stage is necessary to preserve the task definition.

4.4 Human Evaluation

To evaluate our method with real users, we conduct a human study under two evaluation settings. In the first setting, participants select the most attractive slogan from four candidates generated by GPT-4o, DS-L, DS-Q, and our method. In the second setting, participants are revisited three hours later and asked to recall the slogan they remember.

For the first setting, we involve 100 unique participants and 40 brand–persona scenarios. Each participant evaluates 20 tasks. Based on the results, we perform a participant-level significance analysis using a one-sided exact binomial test. The null hypothesis (H_0) assumes $p \leq 0.25$, meaning the probability that a method is the strictly preferred choice is no better than random chance, while the alternative hypothesis (H_1) assumes $p > 0.25$. Our method achieves 46 out of 100 participants satisfying this condition (with fractional allocation in the case of ties), yielding a p-value of 4.25×10^{-5}. Thus, we reject H_0 at $\alpha = 0.05$, indicating that our method is significantly preferred by the user population. In comparison, GPT-4o, DS-L, and DS-Q achieve only 24, 16.5, and 13.5 participants, respectively, far below our method. At the task level, our slogan is selected as the best in 612 out of 2,000 evaluations. This corresponds to a selection rate of 30.6% (612 / 2,000; 100 participants × 20 tasks), substantially outperforming GPT-4o, DS-L, and DS-Q, which achieve 23.60%, 22.05%, and 23.75%, respectively.

For the second setting, we conduct a two-stage follow-up study. In Part 1, participants view 12 brand–persona scenarios with corresponding slogans (one per model) and visible brand names. After approximately three hours, participants return for Part 2, where they are shown 24 brands (including 12 new ones) and asked to select those they remember from Part 1. We recruit about 50 participants for Part 1, of whom 32 complete Part 2. The recall hit rates are 93.75% (ours), 89.58% (DS-Q), 83.33% (DS-L), and 82.29% (GPT-4o). Exploratory one-sided Fisher's exact tests suggest our method outperforms GPT-4o ($p = 0.012$) and DS-L ($p = 0.020$), while the difference with DS-Q is not significant ($p = 0.22$).

Overall, these results suggest that our method produces slogans that are not only more attractive to users but also more memorable.

5 Conclusion

In this work, we propose a novel paradigm for brand slogan generation by recontextualizing persona-related famous quotes. Unlike conventional data-driven approaches that learn from existing slogans and often produce repetitive or less engaging content, our framework expands the creative space to a broader quote domain, helping mitigate advertisement fatigue. Extensive experiments using both automatic metrics and human evaluations show that our method consistently produces slogans with greater diversity, novelty, and emotional impact than strong LLM-based baselines. The resulting slogans preserve the rhetorical structure of well-known quotes while aligning with brand personas, making them memorable and engaging.

Limitations

Our work has several limitations that also point to promising future directions. First, the benefit of quote-based remixing may vary across audiences and cultural contexts, since different groups may have different familiarity with the underlying references. Second, while our structured remix pipeline improves controllability, it may occasionally constrain expressivity when a brand–persona pair does not naturally fit a given quote template, suggesting value in expanding the template pool or allowing adaptive rewriting when needed. Finally, we also note that human evaluation results can be influenced by participant demographics, individual preferences, and platform sampling effects, which may introduce bias in preferences. Beyond these methodological limitations, our framework is developed as a research prototype rather than a deployment-ready advertising system. In practical workflows, generated slogans should be reviewed by human experts to ensure appropriate tone, brand positioning, and alignment with local norms.

References

Zoë Abrams and Erik Vee. 2007. Personalized ad delivery when ads fatigue: An approximation algorithm. In *International Workshop on Web and Internet Economics*.

Pir Noman Ahmad, Yuanchao Liu, Inam Ullah, and Mohammad Shabaz. 2024. Enhancing coherence and diversity in multi-class slogan generation systems. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 23(8):1–24.

Khalid Alnajjar and Hannu Toivonen. 2021. Computational generation of slogans. *Natural Language Engineering*, 27(5):575–607.

Kevin Bartz, Cory Barr, and Adil Aijaz. 2008. Natural language generation for sponsored-search advertisements. In *Proceedings of ACM Conference on Electronic Commerce*.

Chaibapat. Slogan dataset. https://www.kaggle.com/datasets/chaibapat/slogan-dataset.

Yanda Chen, Zihui Ren, Qixiang Gao, Jiale Chen, Si Chen, Xubin Li, Tiezheng Ge, and Bo Zheng. 2025. Ctr-driven ad text generation via online feedback preference optimization. *arXiv preprint arXiv:2507.20227*.

John Chung, Ece Kamar, and Saleema Amershi. 2023. Increasing diversity while maintaining accuracy: Text data generation with large language models and human interventions. In *Proceedings of Annual Meeting of the Association for Computational Linguistics*.

Cristian Danescu-Niculescu-Mizil and Lillian Lee. 2011. Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. *arXiv preprint arXiv:1106.3077*.

Florin Dolcos, Yuta Katsumi, Mathias Weymar, Matthew Moore, Takashi Tsukiura, and Sanda Dolcos. 2017. Emerging directions in emotional episodic memory. *Frontiers in Psychology*, 8:1867.

Edoardomorresi. 10k+ slogan dataset. https://www.kaggle.com/datasets/edoardomorresi/slogan-dataset.

Atsushi Fujita, Katsuhiro Ikushima, Satoshi Sato, Ryo Kamite, Ko Ishiyama, and Osamu Tamachi. 2010. Automatic generation of listing ads by reusing promotional texts. In *Proceedings of International Conference on Electronic Commerce: Roadmap for the Future of Electronic Business*.

Konstantin Golobokov, Junyi Chai, Victor Ye Dong, Mandy Gu, Bingyu Chi, Jie Cao, Yulan Yan, and Yi Liu. 2022. Deepgen: Diverse search ad generation and real-time customization. *arXiv preprint arXiv:2208.03438*.

Ivan A. Guitart and Stefan Stremersch. 2021. The impact of informational and emotional television ad content on online search and sales. *Journal of Marketing Research*, 58(2):293–312.

Daya Guo, Dejian Yang, and 1 others. 2025. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning. *Nature*, 645(8081):633–638.

Ulla Hakala. 2025. Using ai in the creation of municipality slogans. *Place Branding and Public Diplomacy*, 21(1):81–92.

SI Harini, Somesh Singh, Yaman K. Singla, Aanisha Bhattacharyya, Veeky Baths, Changyou Chen, Rajiv Ratn Shah, and Balaji Krishnamurthy. 2025. Long-term ad memorability: Understanding & generating memorable ads. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*.

Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. 2019. The curious case of neural text degeneration. *arXiv preprint arXiv:1904.09751*.

Edward J Hu, Yelong Shen, Phillip Wallis, and 1 others. 2022. Lora: Low-rank adaptation of large language models. In *International Conference on Learning Representations*.

J Weston Hughes, Keng-hao Chang, and Ruofei Zhang. 2019. Generating better search engine text advertisements with deep reinforcement learning. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*.

Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, and 1 others. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

Daniel R Jiang, Alex Nikulkov, Yu-Chia Chen, Yang Bai, and Zheqing Zhu. 2025. Improving generative ad text on facebook using reinforcement learning. *arXiv preprint arXiv:2507.21983*.

Yiping Jin, Akshay Bhatia, Dittaya Wanvarie, and Phu TV Le. 2023. Towards improving coherence and diversity of slogan generation. *Natural Language Engineering*, 29(2):254–286.

Vijay Joshi. English quotes. https://www.kaggle.com/datasets/vijayj0shi/english-quotes.

Yashal Shakti Kanungo, Sumit Negi, and Aruna Rajan. 2021. Ad headline generation using self-critical masked language model. In *Proceedings of the North American Chapter of the Association for Computational Linguistics*.

Jongeun Kim, MinChung Kim, and Taehwan Kim. 2023. Effective slogan generation with noise perturbation. In *Proceedings of ACM International Conference on Information and Knowledge Management*.

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and William B Dolan. 2016. A diversity-promoting objective function for neural conversation models. In *Proceedings of the North American Chapter of the Association for Computational Linguistics*.

Yu-Cheng Liang, Meng-Heng Zheng, and Jheng-Long Wu. 2023. A comparative study of generative pretrained transformer-based models for chinese slogan generation of crowdfunding. In *Proceedings of Conference on Computational Linguistics and Speech Processing*.

Jianxing Liao, Tian Zhang, Xiao Feng, Yusong Zhang, Rui Yang, Haorui Wang, Bosi Wen, Ziying Wang, and Runzhi Shi. 2025. Rlmr: Reinforcement learning with mixed rewards for creative writing. *arXiv preprint arXiv:2508.18642*.

Haoran Liu, Amir Tahmasbi, Ehtesham Sam Haque, and Purak Jain. 2025. Llms for customized marketing content generation and evaluation at scale. *arXiv preprint arXiv:2506.17863*.

Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Manann. Quotes-500k. https://www.kaggle.com/datasets/manann/quotes-500k.

Elyas Meguellati, Lei Han, Abraham Bernstein, Shazia Sadiq, and Gianluca Demartini. 2024. How good are llms in generating personalized advertisements? In *Companion Proceedings of the ACM Web Conference*.

Masato Mita, Soichiro Murakami, Akihiko Kato, and Peinan Zhang. 2024. Striking gold in advertising: Standardization and exploration of ad text generation. In *Proceedings of Annual Meeting of the Association for Computational Linguistics*.

Soichiro Murakami, Sho Hoshino, and Peinan Zhang. 2023. Natural language generation for advertising: A survey. *arXiv preprint arXiv:2306.12719*.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of Annual Meeting of the Association for Computational Linguistics*.

Qwen Team. 2025. Qwq-32b: Embracing the power of reinforcement learning. https://qwenlm.github.io/blog/qwq-32b/.

Guy Tevet and Jonathan Berant. 2021. Evaluating the evaluation of diversity in natural language generation. In *Proceedings of Conference of the European Chapter of the Association for Computational Linguistics*.

Stamatina Thomaidou, Ismini Lourentzou, Panagiotis Katsivelis-Perakis, and Michalis Vazirgiannis. 2013. Automated snippet generation for online advertising. In *Proceedings of ACM international conference on Information & Knowledge Management*.

Chang Wang, Siyu Yan, Depeng Yuan, Yuqi Chen, Yanhua Huang, Yuanhang Zheng, Shuhao Li, Yinqi Zhang, Kedi Chen, Mingrui Zhu, and 1 others. 2025. Beyond quality: Unlocking diversity in ad headline generation with large language models. *arXiv preprint arXiv:2508.18739*.

Xiting Wang, Xinwei Gu, Jie Cao, Zihua Zhao, Yulan Yan, Bhuvan Middha, and Xing Xie. 2021. Reinforcing pretrained models for generating attractive text advertisements. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery & Data Mining*.

Penghui Wei, Xuanhua Yang, Shaoguo Liu, Liang Wang, and Bo Zheng. 2022. Creater: Ctr-driven advertising text generation with controlled pretraining and contrastive fine-tuning. *arXiv preprint arXiv:2205.08943*.

Tianhui Zhang, Bei Peng, and Danushka Bollegala. 2025. Evaluating the evaluation of diversity in commonsense generation. In *Proceedings of Annual Meeting of the Association for Computational Linguistics*.

Yaoming Zhu, Sidi Lu, Lei Zheng, Jiaxian Guo, Weinan Zhang, Jun Wang, and Yong Yu. 2018. Texygen: A benchmarking platform for text generation models. In *ACM SIGIR Conference on Research & Development in Information Retrieval*.

Appendix

In the appendix, we provide all the prompts of our method for slogan generation and evaluation.

Prompts for our framework for slogan remix

Table 4 shows the prompt for our remix method to generate slogans.

Role Definition: You are an award-winning slogan copywriter. Your task is to remix a classic quote, slang, lyric, meme, or popular expression into a brand slogan. Follow ALL steps and output in English.

Input Context:
Brand: *{Brand}*
Emotion (tone only): *{Persona}*
Emotion goal: *{persona_guidelines[Persona]}*

Procedure (Steps):
- **Step 1 (Quote Matching).** Propose three classic sentences (quote/slang/lyric/meme), each 5–10 English words, each with author; do not state the emotion name; briefly explain why each fits the emotion and brand; mark the best with star and repeat it on a new line with an explanation for selection.
- **Step 2 (Structure Breakdown).** Copy the star quote and use | to separate *editable* vs *fixed* parts; indicate which words can be replaced by brand keywords and why.
- **Step 3 (Vocabulary Replacement).** Propose 1–2 replacement word/phrase sets with short reasons; change at most two words in total (adding the brand name counts toward this limit).
- **Step 4 (Remix Generation).** Using the best replacement, output *ONE* final slogan that keeps original syntax/meter, includes the brand name/keyword, and ends with punctuation; after the slogan, output `END_OF_REMIX` on a new line.

Output Format (verbatim):
```
Step 1 Quote Matching: ...
Step 2 Structure Breakdown: ...
Step 3 Vocabulary Replacement: ...
Step 4 Remix Slogan: "..."
END_OF_REMIX
```

Important Constraints:
Do **not** use first-person words (I, we, me, my, our). Use neutral or second-/third-person perspective so the brand is mentioned, not the narrator.

Table 4: Prompt card for the *slogan remix* generator used in our framework. Steps, output template, and constraints mirror our implementation.

Prompts for slogan evaluation

Table 5 & 6 show the prompt for faithfulness evaluation and pairwise hook-based evaluation using an LLM-as-a-Judge framework. The prompt for fluency is similar to the one of faithfulness, which we omit it here.

Role Definition: You are a careful evaluator. Follow the instructions exactly and do not add extra text. Your output must contain exactly two lines.

Evaluation Guidelines:
- Judge whether the **question text implies the ad text**.
- Answer **1** if it implies; answer **0** if it does not.
- Focus on semantic consistency; do not add assumptions not present in the question text.

Input Context:
Question text:
Brand: *{Brand}*
Persona: *{Persona}*
Ad text:
{Slogan}

Task:
Decide whether the question text implies the ad text.

Output Format:
Please answer in exactly two lines:
Answer: 0 / 1
Reason: One short sentence.

Table 5: An example prompt card for *faithfulness evaluation* using an LLM-as-a-Judge framework. The prompt checks whether the (brand, persona) question context implies the given slogan.

Role Definition: You are a human reacting instantly to brand slogans. You do **not** analyze professionally and do **not** consider brand safety, approval processes, or marketing rules. Your judgment is based solely on immediate emotional impact and gut reaction at first glance.

Evaluation Guidelines:
- Evaluate solely based on **first-glance emotional punch**.
- Prefer slogans that are surprising, humorous, or edgy.
- Discourage generic, conservative, or overly polished slogans.
- Controlled boldness and mild risk-taking are accepted.

Input Context:
Brand: *{Brand}*
Target Persona: *{Persona}*
Slogan A:
{Slogan A}
Slogan B:
{Slogan B}

Task:
Decide which slogan triggers a stronger **immediate emotional reaction**. If both slogans feel equally strong or equally weak, answer C.

Output Format:
Please answer in exactly two lines:
Answer: A / B / C
Reason: One short phrase describing your gut reaction.

Table 6: An example prompt card for *pairwise hook-based evaluation* using an LLM-as-a-Judge framework. The prompt is explicitly designed to simulate a human's first-glance emotional reactions to brand slogans.