

Iterative Refinement for Synthetic Tourist Review Generation: Adapting Image Generation Principles to Sentiment Analysis

Isaias Siliceo Guzmán^{1,*}, Ramón Aranda² and Miguel Angel Álvarez-Carmona¹

¹Centro de Investigación en Matemáticas (CIMAT), Sede Monterrey, México

¹Centro de Investigación en Matemáticas (CIMAT), Sede Mérida, México

Abstract

This paper explores the adaptation of iterative refinement principles, originally developed for modern image generation models, to the task of synthetic review generation for tourism sentiment analysis. Instead of accepting the first response produced by a Large Language Model, the proposed methodology progressively refines each generated review through multiple revision stages that improve grammatical correctness, lexical diversity, semantic coherence, and alignment with the target sentiment polarity. The resulting synthetic corpus is incorporated into the REST-MEX 2026 training dataset and used to fine-tune the official baseline sentiment classifier. Experimental results show that, although iterative refinement produces more fluent and coherent reviews, the improvements in sentiment classification are more limited than those previously reported in image generation. These findings suggest that the direct transfer of iterative generation strategies between visual and textual domains is feasible but requires additional mechanisms to effectively enhance data diversity and classifier performance.

Keywords

REST-MEX 2026, Synthetic Text Generation, Iterative Refinement, Large Language Models, Sentiment Analysis, Data Augmentation, Tourism Reviews

1. Introduction

The remarkable progress of Large Language Models (LLMs) has transformed text generation into one of the most active research areas in Natural Language Processing (NLP). Modern generative models are capable of producing fluent, coherent, and context-aware documents, making them attractive tools for data augmentation, conversational systems, content creation, and intelligent decision-support applications. Rather than relying exclusively on manually annotated corpora, researchers increasingly employ synthetic text to enrich training datasets and improve the performance of downstream machine learning models [1].

Tourism constitutes an ideal application domain for these technologies [2]. Millions of travelers voluntarily publish reviews describing their experiences with hotels, restaurants, and tourist attractions. These opinions provide valuable information regarding customer satisfaction, destination image, perceived service quality, and visitor expectations. Consequently [3], sentiment analysis has become a central task in tourism intelligence, supporting recommendation systems, destination management, reputation monitoring, and market analysis through the automatic interpretation of user-generated content [1, 4, 3, 5, 6, 7].

The REST-MEX benchmark has progressively evolved into one of the principal evaluation campaigns for tourism-oriented Natural Language Processing in Spanish [8, 9, 10]. Across successive editions, the benchmark has addressed recommendation systems, sentiment analysis, multitask learning, and large-scale classification over Mexican tourism reviews. The most recent edition, REST-MEX 2026, departs from conventional classifier optimization by introducing a data-centric challenge in which participants are required to generate synthetic reviews capable of improving a fixed sentiment classification model [11, 12, 13].

IberLEF 2026, September 2026, León, Spain

*Corresponding author.

✉ isaias.siliceo@cimat.mx (I. S. Guzmán); arac@cimat.mx (R. Aranda); miguel.alvarez@cimat.mx (M. A. Álvarez-Carmona)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Most existing approaches generate synthetic reviews in a single inference step. While these methods often produce fluent text, they provide limited opportunities for correcting inconsistencies, reinforcing the intended sentiment, or improving lexical diversity after the initial generation. In contrast, several recent generative frameworks have demonstrated that progressively refining intermediate outputs can substantially improve generation quality. Iterative refinement has proven particularly effective in image generation, where successive optimization stages gradually enhance visual fidelity, structural consistency, and semantic alignment with the original prompt.

Inspired by this principle, this work investigates whether progressive refinement can also benefit synthetic text generation for sentiment analysis. Instead of accepting the first review produced by a language model, we propose an iterative generation strategy in which each intermediate version is progressively revised according to predefined linguistic and semantic criteria. At every refinement stage, the generated review is encouraged to improve grammatical correctness, lexical richness, contextual coherence, and consistency with the desired sentiment polarity while preserving its overall meaning.

Unlike diffusion models, where refinement is performed over latent image representations, the proposed methodology operates directly on natural language. Each iteration receives the previously generated review as input and produces an improved version, gradually increasing textual quality before incorporating the synthetic instance into the REST-MEX 2026 training corpus. Through this adaptation, we explore whether the benefits observed in iterative visual generation can be transferred to the textual domain for sentiment-oriented data augmentation.

Ultimately, this work seeks to evaluate the extent to which iterative refinement contributes to synthetic review generation in highly imbalanced sentiment classification problems. Although the experimental results indicate that the improvements obtained in text are less pronounced than those previously reported for image generation, the proposed framework provides initial evidence that progressive refinement constitutes a promising direction for future data-centric approaches in tourism sentiment analysis.

2. Image Generation Methods Applied to Text

Recent years have witnessed remarkable progress in image generation through deep generative models. Diffusion models, autoregressive architectures, and adversarial networks have demonstrated the ability to synthesize highly realistic images from natural language descriptions by progressively transforming an initial representation into a coherent final output. These advances have established new state-of-the-art performance in image synthesis and have motivated researchers to investigate whether similar generation principles can be transferred to other data modalities [14].

Among these approaches, diffusion models have attracted particular attention due to their iterative generation process. Instead of producing an image in a single step, diffusion-based models begin from random noise and progressively refine the generated sample through multiple denoising iterations [15]. Each refinement stage removes uncertainty while improving semantic consistency, visual quality, and alignment with the original prompt. This progressive optimization has proven especially effective for generating complex visual content and has become one of the dominant paradigms in modern generative artificial intelligence [16].

Although text generation is commonly formulated as an autoregressive prediction problem, several of the underlying ideas behind diffusion-based generation are modality-independent. In particular, the concept of iterative refinement can naturally be transferred from images to natural language. Rather than interpreting intermediate states as partially denoised images, they can be viewed as successive drafts of a document that gradually evolve toward a linguistically richer and semantically more coherent version.

Under this interpretation, the generation process no longer attempts to remove Gaussian noise from latent image representations. Instead, each iteration seeks to reduce different forms of textual uncertainty, including grammatical inconsistencies, lexical redundancy, semantic ambiguity, and weak alignment between the generated opinion and its intended sentiment polarity. Consequently, every

refinement stage improves the quality of the previous version while preserving its overall semantic content.

This idea is particularly attractive for synthetic review generation. Tourist opinions often contain multiple linguistic components, including factual descriptions, subjective evaluations, emotional expressions, and contextual information regarding destinations or tourism services. Generating all these elements simultaneously may produce reviews that are grammatically correct but stylistically repetitive or insufficiently aligned with the desired sentiment class. An iterative refinement strategy allows each generation stage to progressively strengthen these characteristics before incorporating the review into the training corpus. Similar ideas have recently been explored in synthetic data generation for sentiment analysis, where improving the quality and diversity of artificial reviews has proven beneficial for highly imbalanced datasets [17, 18].

The methodology proposed in this work adopts this philosophy while remaining independent of diffusion models themselves. Instead of modifying latent visual representations, successive iterations operate directly on textual drafts produced by a Large Language Model. Each refinement receives the previously generated review together with additional instructions encouraging improvements in fluency, lexical diversity, contextual coherence, and sentiment consistency. The resulting sequence of progressively refined opinions constitutes the synthetic corpus used to augment the REST-MEX 2026 training dataset.

Although iterative refinement has demonstrated remarkable effectiveness in image generation, its applicability to natural language generation remains largely unexplored. This work investigates whether the same principle can improve synthetic review generation for sentiment analysis, providing one of the first studies that adapts concepts originally developed for diffusion-based image synthesis to the problem of text-based data augmentation. The proposed methodology is evaluated under the official REST-MEX 2026 benchmark, which provides a realistic scenario for assessing data-centric approaches to tourism sentiment analysis [12, 13].

3. Proposal

The proposed methodology is inspired by the progressive refinement philosophy that has driven the success of modern diffusion models in image generation. Rather than producing a final sample in a single generation step, diffusion-based methods gradually improve an intermediate representation through successive refinement stages until a coherent output is obtained. In this work, we transfer this concept to Natural Language Processing by replacing visual denoising with iterative linguistic refinement.

Let $D = (x_i, y_i) * i = 1^N$ denote the original REST-MEX 2026 training corpus, where x_i represents a tourist review and $y_i \in 1, 2, 3, 4, 5$ its sentiment polarity. The objective is to construct a synthetic corpus $D' = (x'_i, y_i) * i = 1^M$, where $M \leq 3000$, that complements the original training data while preserving the intended sentiment labels.

3.1. Initial Review Generation

The process begins by generating an initial draft using a Large Language Model. Unlike conventional prompt-based approaches that directly accept this first response, the generated opinion is regarded as an intermediate representation that will be progressively improved.

Formally, the initial draft is obtained as

$$x_0 = G(p, y, c),$$

where $G(\cdot)$ denotes the language model, p is the prompt, y is the target sentiment polarity, and c represents the tourism context (hotel, restaurant, or attraction).

Although x_0 is generally coherent, it may still contain repetitive expressions, limited lexical diversity, or weak alignment with the desired sentiment category.

3.2. Iterative Linguistic Refinement

Inspired by the iterative denoising process of diffusion models, the generated opinion is repeatedly refined over a sequence of iterations.

Let x_t denote the review after refinement step t . The successive transformations are defined as

$$x_{t+1} = R(x_t, y, c),$$

where $R(\cdot)$ represents a refinement operator implemented through additional prompting.

Unlike image diffusion, the refinement operator does not remove Gaussian noise. Instead, each iteration attempts to reduce linguistic imperfections by encouraging improvements in four complementary dimensions:

- grammatical correctness,
- lexical richness,
- semantic coherence,
- consistency with the target sentiment polarity.

Consequently, every refinement stage generates a review that is expected to be more natural and informative than the previous version.

3.3. Progressive Prompting

Each refinement iteration receives both the previous review and an additional instruction describing the desired improvement. Typical refinement prompts include requests such as:

- improve fluency without changing the sentiment;
- enrich the vocabulary using more natural expressions;
- eliminate repetitive words;
- reinforce the emotional intensity associated with the target polarity;
- preserve the realism of the tourist experience.

Therefore, the complete refinement process can be expressed as

$$x_k = R^k(x_0),$$

where k denotes the total number of refinement iterations.

Unlike traditional prompting, which generates the final review in a single interaction, the proposed methodology decomposes generation into multiple successive optimization stages.

3.4. Synthetic Corpus Construction

Once the refinement process terminates, the final review x_k is incorporated into the synthetic dataset while preserving its original sentiment label,

$$D' = (x_k, y_i)_{i=1}^M.$$

Finally, the augmented corpus is obtained by

$$D^* = D \cup D',$$

which is subsequently employed to fine-tune the official baseline classifier provided by the REST-MEX 2026 organizers.

The principal motivation behind this methodology is that text generation, similarly to image generation, can benefit from progressive optimization instead of one-shot synthesis. While images are refined by gradually removing visual noise, textual refinement seeks to reduce linguistic imperfections through successive revisions. Although both processes operate over fundamentally different data modalities, they share the common objective of progressively approaching a higher-quality representation.

This adaptation allows us to investigate whether one of the fundamental principles behind diffusion-based image generation can also improve synthetic review generation for sentiment analysis. Rather than reproducing the mathematical formulation of diffusion models, the proposed framework transfers their iterative optimization philosophy to natural language generation, providing a simple yet effective mechanism for producing progressively refined synthetic tourist reviews.

4. Results

The proposed iterative refinement strategy was evaluated using the official protocol of the REST-MEX 2026 Artificial Data Generation Track. Following the competition guidelines, the synthetic reviews generated through the refinement process were incorporated into the original training corpus to fine-tune the baseline sentiment classifier. Performance was measured using the official *Ranking Performance* (RP) metric, together with Macro F1, the overall competition score, and the F1-score for each sentiment category.

The proposed methodology achieved an RP score of 0.0135 and a Macro F1 score of 0.0445. The overall evaluation score reached 16.1658, while the F1-scores obtained for polarity classes 1 through 5 were -0.0005 , 0.0397 , -0.0005 , -0.0003 , and 0.1842 , respectively. Among all sentiment categories, the best performance was obtained for class 5, which considerably exceeded the remaining classes and suggests that the refinement strategy was more effective when generating highly positive opinions. In contrast, the minority sentiment categories continued to present significant challenges, reflecting the severe class imbalance of the REST-MEX corpus.

One of the most interesting observations is that successive refinement produced reviews that were noticeably more coherent and linguistically natural during qualitative inspection. Each refinement stage tended to eliminate repetitive expressions, improve grammatical fluency, and reinforce semantic consistency with the tourism context. However, these qualitative improvements were not accompanied by proportional gains in the quantitative evaluation metrics. This suggests that increasing the linguistic quality of a synthetic review does not necessarily translate into higher discriminative power for sentiment classification.

Compared with one-shot generation strategies explored in previous submissions, the proposed methodology offers greater control over the generation process by allowing intermediate revisions before accepting the final review. Nevertheless, the experimental results indicate that this additional refinement is not sufficient to substantially improve classifier performance. A possible explanation is that each refinement iteration tends to preserve much of the lexical and semantic content of the previous version, producing reviews that become progressively more polished but not necessarily more informative from the perspective of the classifier.

These findings contrast with the behavior previously reported for iterative refinement in image generation, where successive denoising steps consistently improve visual fidelity and semantic alignment. In the textual domain, language models already produce highly coherent drafts during the first generation stage, leaving less room for subsequent refinement to introduce meaningful variability. As a consequence, although the generated reviews become more fluent, they may not significantly increase the diversity of the augmented training corpus.

Overall, the experimental results suggest that the direct transfer of iterative refinement concepts from image generation to synthetic text generation is feasible but less effective than initially expected. While the proposed framework successfully produces coherent tourist reviews and offers a structured generation process, its impact on sentiment classification remains relatively modest. These observations indicate that refinement alone may be insufficient and should be combined with mechanisms that explicitly promote lexical diversity, contextual variability, or coverage of underrepresented sentiment patterns.

5. Conclusions

This paper investigated the adaptation of iterative refinement principles, originally developed for image generation, to the problem of synthetic review generation for tourism sentiment analysis. Instead of progressively denoising latent image representations, the proposed methodology repeatedly refined textual drafts produced by a Large Language Model in order to improve grammatical quality, lexical richness, and sentiment consistency.

Although the generated reviews exhibited better linguistic quality after successive refinement stages,

the quantitative improvements in sentiment classification were limited. These results suggest that, unlike image generation, where iterative optimization substantially improves output quality, the same principle does not directly translate into equivalent benefits for text generation. Nevertheless, the present study should not be considered conclusive. The proposed framework demonstrates that iterative refinement is a viable strategy for synthetic review generation, but additional research is required to determine how refinement policies, diversity constraints, and adaptive revision mechanisms can be better exploited in Natural Language Processing.

Acknowledgments

This research was supported by the Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), Mexico, through the Humanistic Research Project H-2025-G-283, entitled “*Procesamiento de lenguaje natural adaptativo con enfoque humanístico para apoyo comunicacional y diagnóstico en pacientes con enfermedades neurodegenerativas*”. The authors gratefully acknowledge the financial support provided by this project, which made this research possible.

Declaration on Generative AI

Generative AI tools were used solely to improve the language, grammar, and readability of the manuscript. The authors reviewed and edited all AI-assisted content and take full responsibility for the final version of the manuscript.

References

- [1] M. Á. Álvarez-Carmona, R. Aranda, A. Y. Rodríguez-Gonzalez, D. Fajardo-Delgado, M. G. Sánchez, H. Pérez-Espinosa, J. Martínez-Miranda, R. Guerrero-Rodríguez, L. Bustio-Martínez, Á. Díaz-Pacheco, Natural language processing applied to tourism research: A systematic review and future research directions, *Journal of king Saud university-computer and information sciences* 34 (2022) 10125–10144.
- [2] S. Arce-Cardenas, D. Fajardo-Delgado, M. Á. Álvarez-Carmona, J. P. Ramírez-Silva, A tourist recommendation system: a study case in mexico, in: *Mexican international conference on artificial intelligence*, Springer, 2021, pp. 184–195.
- [3] R. Guerrero-Rodríguez, M. Á. Álvarez-Carmona, R. Aranda, A. P. López-Monroy, Studying online travel reviews related to tourist attractions using nlp methods: the case of guanajuato, mexico, *Current issues in tourism* 26 (2023) 289–304.
- [4] A. Diaz-Pacheco, M. Á. Álvarez-Carmona, R. Guerrero-Rodríguez, L. A. C. Chávez, A. Y. Rodríguez-González, J. P. Ramírez-Silva, R. Aranda, Artificial intelligence methods to support the research of destination image in tourism. a systematic review, *Journal of Experimental & Theoretical Artificial Intelligence* 36 (2024) 1415–1445.
- [5] Á. Díaz-Pacheco, R. Guerrero-Rodríguez, M. Á. Álvarez-Carmona, A. Y. Rodríguez-González, R. Aranda, A comprehensive deep learning approach for topic discovering and sentiment analysis of textual information in tourism, *Journal of King Saud University-Computer and Information Sciences* 35 (2023) 101746.
- [6] R. Guerrero-Rodríguez, M. A. Álvarez-Carmona, R. Aranda, Á. Díaz-Pacheco, Ai-driven analysis of hotel digital reputation: insights from a mexican destination, *International Journal of Tourism Cities* 12 (2026) 19–36.
- [7] E. P. Ramirez-Villaseñor, H. Pérez-Espinosa, M. A. Álvarez-Carmona, R. Aranda, Design, development, and evaluation of a chatbot for hospitality services assistance in spanish, *Acta universitaria* 33 (2023).

- [8] M. Á. Álvarez-Carmona, R. Aranda, S. Arce-Cárdenas, D. Fajardo-Delgado, R. Guerrero-Rodríguez, A. P. López-Monroy, J. Martínez-Miranda, H. Pérez-Espinosa, A. Y. Rodríguez-González, Overview of rest-mex at iberlef 2021: Recommendation system for text mexican tourism (2021).
- [9] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, D. Fajardo-Delgado, R. Guerrero-Rodríguez, L. Bustio-Martínez, Overview of rest-mex at iberlef 2022: Recommendation system, sentiment analysis and covid semaphore prediction for mexican tourist texts, *Procesamiento del Lenguaje Natural* 69 (2022) 289–299.
- [10] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, V. Muñoz-Sánchez, A. P. López-Monroy, F. Sánchez-Vega, L. Bustio-Martínez, Overview of rest-mex at iberlef 2023: Research on sentiment analysis task for mexican tourist texts, *Procesamiento del Lenguaje Natural* 71 (2023) 425–436.
- [11] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, A. Y. Rodríguez-González, L. Bustio-Martínez, V. Herrera-Semenets, Overview of rest-mex at iberlef 2025: Researching sentiment evaluation in text for mexican magical towns, *Procesamiento del Lenguaje Natural* 75 (2025) 499–511.
- [12] M. Á. Álvarez-Carmona, Á. Díaz-Pacheco, R. Aranda, M. Reyes-Sierra, A. Y. Rodríguez-González, L. Bustio-Martínez, V. Herrera-Semenets, Overview of rest-mex at iberlef 2026: Research on synthetic text modeling for mexican magical towns, *Procesamiento del Lenguaje Natural* 77 (2026).
- [13] A. Bonet-Jover, J. Á. González-Barba, L. Chiruzzo, Overview of IberLEF 2026: Natural Language Processing Challenges for Spanish and other Iberian Languages, in: *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2026)*, co-located with the 42nd Conference of the Spanish Society for Natural Language Processing (SEPLN 2026), CEUR-WS. org, 2026.
- [14] A. Y. Rodríguez-González, M. Á. Álvarez-Carmona, Machine learning and soft computing: Current trends and applications, 2026.
- [15] M. Á. Álvarez-Carmona, I. S. Guzmán, R. Aranda, Towards accurate and legible scene text generation in spanish; a text-to-image, in: *Progress in Artificial Intelligence and Pattern Recognition: 9th International Congress, IWAIPR 2025, Varadero, Cuba, October 14–17, 2025, Proceedings*, Springer Nature, 2026, p. 152.
- [16] M. Á. Álvarez-Carmona, R. Aranda, R. Guerrero-Rodríguez, A. Y. Rodríguez-González, A. P. López-Monroy, A combination of sentiment analysis systems for the study of online travel reviews: Many heads are better than one, *Computación y Sistemas* 26 (2022) 977–987.
- [17] A. B. García-Gutiérrez, P. E. López-Ávila, P. A. Gallegos-Ávila, R. Aranda, M. Á. Álvarez-Carmona, Balancing of tourist opinions for sentiment analysis task., in: *IberLEF@ SEPLN*, 2023.
- [18] Á. Díaz-Pacheco, A. B. García-Gutiérrez, A. Y. Rodríguez-González, R. Aranda, M. Á. Álvarez-Carmona, Balancing sentiment analysis datasets through representative-word-guided synthetic review generation: A case study on mexican spanish tourism reviews, *Applied Sciences* (2026).