CARMA 2026

8th Int. Conf. on Advanced Research Methods and Analytics
1-3 July 2026, Valencia, Spain

Manipulating customer perception via context-aware fake review injection

María Olmedilla Fernández[1], Rocío Martínez-Torres[2], José Carlos Romero[3], Sergio Toral[4]

[1]SKEMA Business School, Université Côte d'Azur, France, [2]Facultad de Ciencias Económicas y Empresariales, University of Seville, Spain, [3]Department of Marketing, Audencia Business School, 8 Route de la Jonelière, 44312 Nantes, France, [4]E. T. S. Ingeniería, University of Seville, Spain

How to cite: Olmedilla Fernández, M.; Martínez-Torres, R.; Romero, J. C.; Toral, S. (2026). Manipulating customer perception via context-aware fake review injection. In: 8th International Conference on Advanced Research Methods and Analytics (CARMA 2026). Valencia, 1-3 July 2026. https://doi.org/10.4995/CARMA2026.2026.21881

Abstract

Large language models (LLMs) can generate reviews that closely match authentic user language, raising the risk of coordinated fake-review injection on electronic word-of-mouth (eWOM) platforms. This study operationalizes a review-injection attack by fine-tuning an open LLM (Llama 3, 8B) with QLoRA on four Amazon product categories and generating polarity-controlled fake reviews via prompting. Fake reviews are injected sequentially into genuine review sets; after each step we summarize the evolving corpus with LongT5 and track (1) sentiment of the summary, (2) semantic similarity (SBERT MiniLM), and (2) lexical similarity (ROUGE-L). Across four products, only 5.3–10.6% injected negative reviews were sufficient to flip an initially positive perception, while 12.5–27.2% injected positive reviews were needed to overturn initially negative perception. Similarity metrics changed gradually, indicating that injected reviews can shift sentiment while remaining stylistically coherent.

Keywords: *Fake review injection; customer manipulation; large language models; sentiment shift; semantic similarity; information manipulation theory*

Article in Press

This work is licensed under a Creative Commons License CC BY-NC-SA 4.0
Editorial Universitat Politècnica de València