

Generative AI and Content Homogenization: The Case of Digital Marketing

Chaoran Liu

Peking University (UK Campus)

Cite as:

Liu Chaoran (2026), Generative AI and Content Homogenization: The Case of Digital Marketing.
Proceedings of the European Marketing Academy, 55th, (133815)

Paper from the 55th Annual EMAC Conference, Bath, United Kingdom, June 2-5, 2026



Generative AI and Content Homogenization: The Case of Digital Marketing

Abstract

Entrepreneurs and small businesses struggle to create high-quality marketing content due to limited resources. Generative AI (Gen AI) offers a powerful solution, but widespread adoption may inadvertently homogenize content across competitors. We examine this tension in the restaurant industry, leveraging Italy's ChatGPT ban in April 2023 as a natural experiment. Comparing Milan restaurants (treated) to restaurants in four other European cities (control), we find the ban decreased content similarity by 15% (lexical), 12% (syntactic), 2% (semantic), and 3% (language style), demonstrating that ChatGPT access drives substantial homogenization. Critically, reduced homogenization during the ban correlated with 3.5% higher consumer engagement, despite negative effects on posting frequency and content length. Our findings reveal a fundamental trade-off: while Gen AI democratizes content creation, over-reliance risks weakening differentiation and engagement. We provide actionable strategies for businesses to leverage Gen AI efficiency while preserving distinctive marketing voice.

Keywords: Generative AI, Content Homogenization, Consumer Engagement

Track: Methods, Modelling & Marketing Analytics

1 Introduction

Constrained by limited budgets and lack of specialized expertise, entrepreneurs and small business owners have historically faced challenges in accessing high-quality marketing services. Generative artificial intelligence (Gen AI) technologies are poised as a transformative solution to these limitations. By automating tasks that traditionally require human expertise, Gen AI enables efficient and affordable content creation across various contexts, including artistic production (Epstein et al. 2023), content titles (Zhang et al. 2024), advertisement development (Chen and Chan 2024), and beyond (Yilmaz et al. 2023, Burtch et al. 2023). A 2024 McKinsey study (McKinsey 2024) identifies marketing and sales as the most frequently adopted functions for Gen AI, with content support for marketing strategies ranking as the most reported use case within this domain. As such, Gen AI offers entrepreneurs and small business owners a powerful tool for creating marketing content, enabling them to maintain an active online presence and foster consumer engagement (Acar and Gvirtz 2024, Forbes 2024).

Despite this potential, wide adoption of Gen AI in content creation could also pose challenges. One concern that has been raised is the possibility of *content homogenization*. Experimental research demonstrates that adopting Gen AI for creative tasks significantly reduces semantic diversity (Anderson et al. 2024) and leads to increasingly homogenized content (Doshi and Hauser 2024), a trend that persists over time (Liu et al. 2024). For advertisement copywriting, Chen and Chan (2024) experimentally show that employing Gen AI as “ghostwriters” can result in diminished creativity and reduced semantic divergence. Similar phenomenon are documented in other settings. For example, in the context of academic writing, studies by Kobak et al. (2024), Gray (2024), and Liang et al. (2024) find that AI-generated material often lacks linguistic diversity, primarily due to an over-reliance on specific vocabulary.

In the realm of marketing, content homogenization presents a critical challenge. Businesses producing similar marketing content may dilute their brand differentiation and positioning (Pulig et al. 2006), leading to consumer satiation. For example, repeated exposure to similar content diminishes enjoyment and engagement (Ratner et al. 1999, Redden 2008). Furthermore, similarity in content may prompt negative perceptions of businesses viewed as imitators (Van Horen and Pieters 2012). Combined, content homogenization may result in reduced consumer engagement, hampering the effectiveness of marketing activities.

Against this backdrop, this paper addresses two key research questions. First, we aim to conduct a rigorous causal analysis of the impact of Gen AI access on marketing content homogenization. Second, and more critically for marketers, we seek to explore the impact of Gen AI access on consumer engagement, possibly through the channel of content homogenization.

Our empirical context is the hospitality sector, particularly restaurants. This sector is highly fragmented, with 70% of establishments owned and operated by independent entrepreneurs (Kaufman et al. 2020). Despite a wide range of marketing needs – from launching seasonal offers to local promotion – many restaurants lack access to professional marketing services

and rely on in-house resources. Consequently, they are increasingly turning to Gen AI tools to create marketing content and manage their digital presence, enabling them to produce engaging marketing materials with minimal cost and expertise (SevenRooms 2025).

Further, due to the limited marketing budgets, small restaurant owners often resort to low-cost channels such as social media to distribute their marketing content. As such, our study focuses on text content created on Instagram, which boasts a massive base of 1.47 billion monthly active users (FT 2024) and is recognized as a major social media platform in the hospitality sector (Forbes 2023a).

Estimating the causal impact of Gen AI on marketing content presents significant challenges due to endogeneity concerns, such as adoption bias. To address this, we leverage the temporary ban imposed by the Italian government on ChatGPT in 2023 as a natural experiment to identify the effects of Gen AI on marketing outcomes. The ban was announced on March 31, 2023, by the Italian Data Protection Authority and took immediate effect (Authority 2023). During this period, access to ChatGPT servers was blocked for all IP addresses originating from Italy. The ban was subsequently lifted on April 28, 2023, after regulatory bodies reauthorized OpenAI.

With the ChatGPT ban in Italy, we adopt difference-in-differences (DiD) approaches, with businesses in Milan, Italy, as the treated group, and four other cities outside Italy, including Paris, Lyon, Madrid, and Valencia, as the control group. We construct the dataset of text content on Instagram created by the restaurant businesses in these five cities from late February (week 8) to late April (week 17) in 2023.

To answer our first research question, we utilize Natural Language Processing (NLP) models to measure content similarity among Instagram posts across businesses. We assess content similarities from four perspectives: lexical (i.e., frequency of individual words, using Bag-of-words), syntactical (i.e., arrangement and order of words within sentences, using N-gram), semantic (i.e., meanings or concepts in the text, using S-BERT), and language style (i.e., the use of function words, using LIWC).

Using Generalized Synthetic Control Method (Generalized SCM) and Synthetic Difference-in-Differences (Synthetic DiD) as our estimation strategies, we find that marketing content in the treatment group exhibits decreased similarity during the ban period, highlighting the role of Gen AI in driving content homogenization. Based on Generalized SCM estimates, the treatment group experienced relative decreases of 15%, 12%, 2%, and 3% in lexical, syntactic, semantic, and language style similarity, respectively, compared to the control group during the ban. As a robustness check, we use OpenAI's text embeddings (text-embedding-3-small model), which are designed to measure the semantic relatedness of text strings, and find a similar decrease of approximately 5% in content similarity during the ban period.

Our heterogeneous treatment effect analyses further support this conclusion. First, we split the sample based on cuisine origin, assuming that restaurants with non-local cuisine are more likely to be run by non-native entrepreneurs who are more reliant on ChatGPT for marketing content creation, and would thus be more likely to be affected by the ban of ChatGPT.

Confirming this intuition, our results show a stronger homogenization effect among non-native restaurants — reinforcing the idea that increased AI tool usage contributes to greater content homogenization. Second, we split the sample based on the median drop rate in posting frequency between the ban period and the pre-ban period, with the assumption that restaurants that dropped more steeply in posting frequency during the ban were more reliant on ChatGPT. Again, we find a stronger homogenization effect among the restaurants that show a higher drop in posting frequency during the ban period.

To answer our second research question, we find that consumer engagement, proxied by the average like count, is higher in the ban period. Specifically, using a logarithm specification, we find that at the business week level, the average like count increases by approximately 3.5% during the ban period. A correlation analysis suggests a non-linear relationship between consumer engagement and content similarity. And the decrease in consumer engagement observed in our sample is likely to be driven by the increase in content similarity.

Finally, using the same estimation strategy, we find that both the frequency of posting and the average word count are significantly lower during the ban period, indicating that access to ChatGPT may also influence content volume and length. However, the increase in consumer engagement during the ban persists after controlling for these two factors, making the association between higher engagement and decreased similarity more salient.

This study presents the first comprehensive empirical analysis of how Gen AI adoption influences marketing content homogenization, as well as the resulting impact on consumer engagement. Our findings serve as a cautionary tale for entrepreneurs and small business owners, warning that over-reliance on Gen AI tools can produce homogenized content at the market level that may weaken consumer engagement and jeopardize marketing differentiation. Additionally, our results provide insights for marketers and technology developers, underscoring the need to recognize and mitigate the unintended consequences of Gen AI on marketing.

We contribute by: (1) providing the first empirical examination of Gen AI-driven homogenization using real-world marketing data; and (2) offering insights into how homogenization affects engagement.

2 Empirical Setting and Data

2.1 Italy’s ChatGPT Ban

ChatGPT dominated the Gen AI market with 60% share throughout 2023 (Forbes 2023b).¹ On March 31, 2023, Italy’s Data Protection Authority banned ChatGPT due to privacy concerns, blocking access for all Italian IP addresses until April 28, 2023 (Authority 2023). This ban provides an ideal natural experiment as it was: (1) exogenous to marketing outcomes; and (2) sudden, creating clear interruption in access.

¹Other Gen AI tools (Copilot, Gemini, MetaAI) became available in the EU only after November 2023.

2.2 Data Collection

We designate Milan as treatment and Paris, Lyon, Madrid, and Valencia as controls. We obtained restaurant information from Google Maps, identified their Instagram accounts through their websites, and subsequently extracted text content from Instagram posts. We included businesses that: (1) posted before January 1, 2023; (2) remained active after April 28, 2023; and (3) are not major chains. This yielded 486 Milan and 1,720 control businesses (Table 1).

Table 1: Sample Construction

City	Google Map	Instagram Link	Valid Business
Milan	2,695	752	486
Paris	4,401	699	429
Lyon	1,811	660	432
Madrid	3,347	801	544
Valencia	1,770	468	315

2.3 Content Similarity Measures

We measure similarity from four perspectives: **Lexical Similarity** using bag-of-words (word frequency overlap); **Syntactical Similarity** using N-grams (sentence structure); **Semantic Similarity** using Sentence-BERT (conceptual meaning); and **Language Style Similarity** using LIWC (function word usage) (Pennebaker and Chung 2007). For each business-week, we compute average pairwise similarity to all other businesses’ posts.

2.4 Estimation Strategy

Standard DiD requires parallel pre-treatment trends, which our analysis shows does not hold. We employ: **Generalized SCM** (Xu 2017) and **Synthetic DiD** (Arkhangelsky et al. 2021), constructing weighted control combinations matching treated units pre-treatment.

3 Results

3.1 Content Homogenization

Table 2 shows the ban significantly reduced similarity. Under Generalized SCM, the ban decreased lexical similarity by 0.003 (15% relative decline), syntactical by 0.003 (12%), semantic by 0.008 (2%), and language style by 0.013 (3%). Synthetic DiD confirms these findings. Homogenization is most pronounced in lexical and syntactical dimensions. As a robustness check, we employ OpenAI’s text embeddings (text-embedding-3-small model) to measure semantic relatedness and find a coefficient of -0.022, representing a 5% decline in content similarity during the ban period. Further, to ensure our results are not driven by any holiday effect during the

observation window, including Easter holiday and Italian Liberation Day, we conduct several additional robustness checks, which find consistent results.

Table 2: Impact of ChatGPT Ban on Content Similarity

	Lexical	Syntactical	Semantic	Language Style
Generalized SCM	-0.003*** (0.001)	-0.003*** (0.001)	-0.008*** (0.002)	-0.013*** (0.004)
Observations	12,004	12,004	12,004	10,354
Synthetic DiD	-0.001*** (0.000)	-0.003*** (0.001)	-0.009*** (0.001)	-0.009 (0.007)
Observations	5,780	5,780	5,780	4,420

***, **, * denote 1%, 5%, 10% significance. SE clustered at city (bootstrapped).

3.2 Heterogeneous Treatment Effects

To explore whether the AI ban’s impact varies by ChatGPT reliance, we conduct two heterogeneity analyses using proxies for usage intensity. First, we classify restaurants by cuisine origin (local vs. non-local) based on Google Maps categories. We hypothesize that non-local cuisine restaurants—more likely operated by non-native owners facing language barriers—relied more heavily on Gen AI for local marketing content. Results confirm this: non-local restaurants exhibit significantly larger decreases in content similarity across all four dimensions under both estimation methods, while local restaurants show weaker or insignificant effects. This pattern supports our argument that businesses with greater AI dependence experienced stronger treatment effects.

Second, we stratify restaurants by their posting frequency decline during the ban period, using the median drop in Italy as the cutoff. Restaurants experiencing larger posting drops (above median) likely had greater ChatGPT reliance and should therefore show stronger similarity reductions. Results validate this hypothesis: high-drop restaurants demonstrate consistently negative and significant treatment effects across all metrics, while low-drop restaurants exhibit smaller, often insignificant effects. Together, these analyses provide convergent evidence that the ban’s impact on content homogenization was concentrated among businesses most dependent on generative AI tools.

3.3 Consumer Engagement

Table 3 shows the ban increased engagement by 3.5% ($p < 0.001$). Winsorized outcomes yield similar results. Effects remain significant when controlling for word count and posting frequency, suggesting reduced homogenization enhances engagement.

Table 3: Impact of ChatGPT Ban on Consumer Engagement

	Log Avg Likes	Avg Likes (5% Wins.)
Generalized SCM	0.035*** (0.009)	8.277*** (3.636)
Observations	11,191	11,191
Synthetic DiD	0.067*** (0.020)	8.047 (6.915)
Observations	5,070	5,070

Note: Significance and SE as in Table 2.

3.4 Additional Outcomes

The ban decreased posting frequency by 0.061 posts/week ($p < 0.001$) and word count by 1.2 words/post ($p < 0.001$), demonstrating Gen AI enhances content production but increases homogenization.

4 Discussion and Conclusion

This paper provides comprehensive empirical evidence on Gen AI’s impact on marketing content. Our natural experiment exploiting Italy’s ChatGPT ban establishes that Gen AI access increases content similarity. Further, we find that reduced homogenization correlates with increased engagement (3.5%). In our ongoing experimental work, we validate that content homogenization persists with later versions of large language models such as GPT-4o. Future studies will aim to establish causality between content homogenization and consumer engagement through experimental manipulation of similarity levels."

Contributions: We provide first empirical evidence of Gen AI-driven homogenization in real-world marketing contexts (Anderson et al. 2024, Doshi and Hauser 2024), validated through both natural and controlled experiments. The convergent findings across methodologies strengthen causal inference regarding Gen AI’s role in content convergence.

Implications: While Gen AI democratizes content creation, over-reliance risks homogenization that weakens engagement and competitive differentiation. Businesses should: (1) use Gen AI strategically rather than as complete replacement; (2) employ prompt engineering emphasizing unique marketing voice; (3) maintain human oversight to inject distinctiveness; and (4) monitor similarity versus competitors. Platform developers should design features that actively preserve content diversity rather than optimize solely for linguistic fluency.

Limitations: Our empirical study is based on restaurant data from 2023, an early stage of Gen AI adoption, and continuous AI evolution necessitates ongoing monitoring of these patterns. Future research should explore other industries, platforms, long-term adaptation patterns, optimal human-AI collaboration ratios, and homogenization in multimodal content (images, videos). Additionally, our study was unable to exogenously vary similarity levels and causally estimate their impact on consumer engagement. We are currently conducting experimental

studies to address these limitations.

In conclusion, Gen AI adoption involves fundamental efficiency-homogenization trade-offs. As Gen AI integrates into marketing workflows, understanding and mitigating homogenization is essential for maintaining engagement and competitive differentiation. The persistence of these effects across real-world and experimental settings with the latest models underscores the urgency of developing strategies that preserve content diversity while leveraging AI’s efficiency gains.

References

- Acar, O. A. and Gvartz, A. (2024). GenAI Can Help Small Companies Level the Playing Field. *Harvard Business Review*.
- Anderson, B. R., Shah, J. H., and Kreminski, M. (2024). Homogenization effects of large language models on human creative ideation. In *Proceedings of the 16th Conference on Creativity & Cognition*, pages 413–425.
- Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118.
- Authority, T. I. D. P. (2023). Artificial intelligence: stop to ChatGPT by the Italian SA. Personal data is collected unlawfully, no age verification system is in place for children - 30 march 2023. Accessed: 2024-12-12.
- Burtch, G., Lee, D., and Chen, Z. (2023). The consequences of generative AI for UGC and online community engagement. *Available at SSRN 4521754*.
- Chen, Z. and Chan, J. (2024). Large language model in creative work: The role of collaboration modality and user expertise. *Management Science*.
- Doshi, A. R. and Hauser, O. P. (2024). Generative ai enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, 10(28):eadn5290.
- Epstein, Z., Hertzmann, A., of Human Creativity, I., Akten, M., Farid, H., Fjeld, J., Frank, M. R., Groh, M., Herman, L., Leach, N., et al. (2023). Art and the science of generative AI. *Science*, 380(6650):1110–1111.
- Forbes (2023a). How are food influencers changing the restaurant world? <https://www.forbes.com/sites/lizzysaxe/2018/12/07/are-food-influencers-the-restaurant-world/>. Accessed: 2024-12-12.
- Forbes (2023b). New Research Shows ChatGPT Reigns Supreme In AI Tool Sector. Accessed: 2024-12-12.
- Forbes (2024). How small businesses are using ai. <https://www.forbes.com/sites/allbusiness/2024/09/19/how-small-businesses-are-using-ai/>. Accessed: 2024-12-12.
- FT (2024). Instagram overtakes tiktok in app downloads in race for new users. Accessed: 2024-12-12.
- Gray, A. (2024). ChatGPT “contamination”: estimating the prevalence of LLMs in the scholarly literature. *arXiv preprint arXiv:2403.16887*.

- Kaufman, M. S., Goldberg, L. G., and Avery, J. (2020). Restaurant revolution: How the industry is fighting to stay alive.
- Kobak, D., González-Márquez, R., Horvát, E.-Á., and Lause, J. (2024). Delving into ChatGPT usage in academic writing through excess vocabulary. *arXiv preprint arXiv:2406.07016*.
- Liang, W., Izzo, Z., Zhang, Y., Lepp, H., Cao, H., Zhao, X., Chen, L., Ye, H., Liu, S., Huang, Z., et al. (2024). Monitoring ai-modified content at scale: A case study on the impact of chatgpt on AI conference peer reviews. *arXiv preprint arXiv:2403.07183*.
- Liu, Q., Zhou, Y., Huang, J., and Li, G. (2024). When ChatGPT is gone: Creativity reverts and homogeneity persists. *arXiv preprint arXiv:2401.06816*.
- McKinsey (2024). The state of ai in early 2024: Gen ai adoption spikes and starts to generate value. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai>. Accessed: 2024-12-12.
- Pennebaker, J. W. and Chung, C. K. (2007). Expressive writing, emotional upheavals, and health. *Foundations of health psychology*, pages 263–284.
- Pullig, C., Simmons, C. J., and Netemeyer, R. G. (2006). Brand dilution: when do new brands hurt existing brands? *Journal of Marketing*, 70(2):52–66.
- Ratner, R. K., Kahn, B. E., and Kahneman, D. (1999). Choosing less-preferred experiences for the sake of variety. *Journal of consumer research*, 26(1):1–15.
- Redden, J. P. (2008). Reducing satiation: The role of categorization level. *Journal of Consumer Research*, 34(5):624–634.
- SevenRooms (2025). 9 genius ways restaurants are using ai (and how you can, too). <https://www.forbes.com/sites/allbusiness/2024/09/19/how-small-businesses-are-using-ai/>. Accessed: 2024-12-12.
- Van Horen, F. and Pieters, R. (2012). When high-similarity copycats lose and moderate-similarity copycats gain: The impact of comparative evaluation. *Journal of Marketing Research*, 49(1):83–91.
- Xu, Y. (2017). Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis*, 25(1):57–76.
- Yilmaz, E. D., Naumovska, I., and Aggarwal, V. A. (2023). AI-driven labor substitution: Evidence from google translate and ChatGPT. *INSEAD Working Paper*.
- Zhang, X., Sun, C., Zhang, R., and Goh, K.-Y. (2024). The value of ai-generated metadata for ugc platforms: Evidence from a large-scale field experiment. *ArXiv*, abs/2412.18337.