Generative AI and Content Homogenization: The Case of Digital Marketing

Chaoran Liu
London Business School, chaoranl@london.edu

Tong Wang
London Business School, tongw@london.edu

S. Alex Yang
London Business School, sayang@london.edu

Traditionally, entrepreneurs and small businesses, constrained by limited resources, struggle to create high-quality marketing content, which limits consumer engagement and hampers growth. Generative artificial intelligence (Gen AI) presents a powerful, yet affordable tool to overcome this challenge. However, the widespread adoption of Gen AI may have unintended consequences. Specifically, content generated by popular large language models may appear more similar to each other compared to human created content, known as content homogenization. In the context of marketing, such homogenization may dampen consumer engagement and dilute brand uniqueness over time. To examine the extent of content homogenization and its potential implications, we study the restaurant industry – a highly fragmented sector where 70% of the establishments are owned by independent entrepreneurs. We leverage the country-wide ChatGPT ban in Italy during April 2023 as a natural experiment to examine how access to Gen AI affects content homogenization, consumer engagement, and volume of social media marketing content created by small business owners. Our study shows that restaurants in the treatment group—located in Milan, Italy—experienced relative decreases of 15%, 12%, 2%, and 3% in lexical, syntactic, semantic, and language style similarity in their Instagram marketing content during the ban, suggesting that access to ChatGPT leads to content homogenization. We also find the ban has a positive effect on consumer engagement – an approximate 3.5% increase in average like counts – despite negative effects on posting frequency and post length. By better underscoring an unintended consequence of Gen AI adoption, this research aims to provide valuable insights to entrepreneurs and small businesses on how to better harness the capability of Gen AI while mitigating potential drawbacks.

Key words: Artificial Intelligence, Generative AI, Content creation, Similarity, Homogenization, Consumer engagement

History: First draft 26 July 2025

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 (Pullig 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 and within 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 during the ban period exhibits decreased similarity, highlighting the role of Gen AI in driving content homogenization. Based on Generalized SCM estimates, during the ban, 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. As a robustness check, we also identify increased content similarity using OpenAI's embedding, which captures similarity measures along different dimensions.

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.

Furthermore, 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 that a non-linear relationship between consumer engagement and content similarity. Based on our estimates, an increase in similarity to the magnitude in our previous analysis is likely to attribute to a decrease in consumer engagement.

Finally, using the same estimation strategy, we find that both the frequency of posting and the average word count are significantly lower in the ban period, indicating that access to ChatGPT may also influence content volume and length. Combined, our analysis suggests that the impact of Gen AI on customer engagement is likely to be through multiple channels.

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 brand 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 market dynamics.

We contribute to the existing literature from two perspectives. First, recent research has shown that Gen AI can lead to homogenization across various domains (Epstein et al. 2023). In academic writing, for instance, research has documented a marked and disproportionate increase in the use of specific keywords following the widespread adoption of large language models (LLMs) (Kobak et al. 2024, Gray 2024, Liang et al. 2024). Such homogenization effects have been observed in creative domains, where Gen AI has been shown to reduce idea diversity(Anderson et al. 2024, Doshi and Hauser 2024, Chen and Chan 2024, Meincke et al. 2025). Extending beyond textual content, AlDahoul et al. (2025) find that text-to-image Gen AI models can produce racially homogeneous representations when generating images of people. Our study joins this discussion and extends by first empirically examining this phenomenon using real-world data, specifically focusing on homogenization in the context of marketing content creation on social media.

Second, we aim to provide deeper insights into the impact of AI-driven content homogenization on consumer engagement. On one hand, prior research in consumer behavior suggests that repetitive or overly similar content can lead to satiation, thereby reducing consumer interest over time (Rolls et al. 1984). Thus, for any brand, differentiation is a cornerstone of the effective marketing strategy (Farquhar 1989), often achieved through pricing (d'Aspremont et al. 1979) and brand image (Kuksov 2007). On the other hand, recent findings indicate that the relationship between similarity and consumer response is not strictly linear. For example, Lasaleta and Redden (2018) show that perceived similarity can, under certain conditions, reduce satiation by creating a sense of cohesive variety. Similarly, in the domain of media consumption, Woolley and Sharif (2022) document the "rabbit hole effect," where repeated exposure to thematically similar content increases the likelihood of continued engagement. This dual effect aligns with Optimal Distinctiveness Theory (Brewer 1991), which posits that content that is too distinct or too similar may fail to sustain engagement (Keller 2020). While these insights are well established in related domains, little is known about how AI-driven content homogenization affects consumer engagement in practice. Our study helps fill this gap by offering empirical evidence on the relationship between Gen AI, content similarity, and engagement outcomes in the context of social media marketing.

2. Empirical Setting and Strategy

Our analysis centers on the highly fragmented hospitality industry, with a particular focus on the restaurant sector. Due to limited marketing budgets, independent restaurateurs often rely on in-house resources and resort to low-cost channels such as social media, for various marketing needs—from launching seasonal offers to local promotions. According to MenuTiger (2024), 82% of restaurants in the United States use social media as part of their marketing efforts. Among various social media sites, Instagram, with its massive base of 1.47 billion monthly active users (FT 2024), has established itself as a crucial consumer engagement platform for the hospitality sector (Forbes 2023a). Therefore, we focus on text-based marketing content created by businesses on Instagram. With the success of ChatGPT, the adoption of Gen AI on social media has surged (Sun et al. 2024). An industry report by SevenRooms, a leading CRM software for hospitality operators, finds that 46% of the restaurant owners surveyed use AI for marketing purposes. This makes the restaurant sector an ideal candidate to study the effect of Gen AI on marketing content creation.

2.1. Italy's ChatGPT Ban in 2023

In 2023, ChatGPT, developed by OpenAI, solidified its position as a leading Gen AI tool. By January 2023, it became the fastest-growing consumer application, accumulating over 100 million users within two months of its launch. Despite the emergence of competitors, ChatGPT maintained a dominant market share of 60% throughout 2023 (Forbes 2023b).

Despite its growing success, ChatGPT faced regulatory challenges in Europe, particularly in Italy. On March 31, 2023, the Italian Data Protection Authority imposed a blanket ban on ChatGPT, immediately blocking access to its servers for all IP addresses originating from Italy (Authority 2023). The ban remained in effect until April 28, 2023, when regulatory authorities reauthorized OpenAI, allowing ChatGPT to resume operations in the country. The regulatory challenges in Europe were faced not solely by OpenAI. Other Gen AI tools including Copilot, Gemini, and MetaAI, became available in EU only after November 2023[1], June 2024[2] and March 2025[3] respectively. As a result, it was difficult for users in Italy to access other Gen AI tools during the ban period, unless they used a VPN to access ChatGPT. We provide a comment in a later section that this is not a key concern for our study.

We leverage this regulatory event, which provides an ideal setting for our research, for two key reasons. First, the ban was imposed due to privacy concerns unrelated to marketing content, ensuring the event is exogenous to our outcome of interest. Second, the sudden and unexpected nature of the ban created a clear interruption in ChatGPT access, enabling a comparison of marketing outcomes before and during the ban. This unique context allows us to isolate the impact of Gen AI on marketing content and other related outcomes.

Although the ban is relatively short-lived in Italy, existing research has shown that knowledge workers who relied on ChatGPT for tasks such as trading decisions and coding experienced reduced efficiency (Bertomeu et al. 2023, Kreitmeir and Raschky 2023). This highlights how quickly ChatGPT, as the dominant Gen AI tool, has become integral to business operations in Italy, even in early 2023.

2.2. Data Collection and Sample Construction

To establish a treatment and control group, we designate Milan as the treatment city and select four other European cities—Paris, Lyon, Madrid, and Valencia—as the control group. Milan, as one of Italy's major urban centers, is representative of national dynamics while being less affected by

[1] https://www.reworked.co/digital-workplace/microsoft-confirms-copilot-general-release-date-rolls-out-copilot-in-windows-11/

[2] https://www.independent.co.uk/tech/google-gemini-uk-launch-b2556443.html

[3] https://about.fb.com/news/2025/03/europe-meet-your-newest-assistant-meta-ai/

Table 1 Number of Business Collected by Steps

City	Step 1 Google Map	Step 2 Instagram Link	Step 3 Valid Business
Milan	2695	752	486
Paris	4401	699	429
Lyon	1811	660	432
Madrid	3347	801	544
Valencia	1770	468	315

Easter-related religious events than cities like Rome. The control cities were chosen from neighboring countries to ensure cultural and regional similarity. Moreover, we include both large metropolitan areas (Paris and Madrid) and smaller cities (Lyon and Valencia) to capture heterogeneity across urban contexts.

To capture the restaurants in these cities, we use Apify (https://console.apify.com/) to scrape public information from Google Maps. This information includes the restaurant's website (if available), business type, and overall Google review score. As this process may capture some non-restaurant businesses, we rely on Google Maps' category labels to filter out entries that are not classified as restaurants, such as cafes and grocery stores. Then, based on the restaurant's website, we develop a scraper to extract its associated Instagram account (if available). Lastly, we utilize Apify to scrape the restaurant's Instagram activity[4] since 2022, including post content and engagement metrics such as the number of likes.[5]

To ensure the robustness and relevance of our modeling sample, we apply a set of inclusion criteria based on business activity and focus. We include only those businesses that posted on Instagram before January 1, 2023, to capture entities with an established social media presence. Likewise, to prevent our findings from being skewed by natural attrition (e.g., restaurants going out of business), we exclude businesses that became inactive on Instagram after April 28, 2023, when the ban was lifted. Third, we remove major international food and hotel chains, such as McDonald's and Burger King, to focus on smaller, independent businesses. The number of businesses retained at each filtering step is summarized in Table 1.

We summarize the Instagram data of 2023 in Table 2. For each city, we report the number of businesses, along with key statistics for social media activity, including the number of posts per business, the like count, and the word count of posts. For example, among the 486 restaurants in

[4] We note that Apify Scraper only scrapes public information. See reference https://apify.com/apify/instagram-post-scraper.

[5] Instagram allows users to hide like counts to reduce social media pressure and foster a healthier online experience (BBC 2021). Consequently, our data includes only posts from businesses that choose to display this information. However, the decision to disclose likes is likely independent of Gen AI adoption.

Milan, businesses have an average of 70.6 posts in 2023, with a standard deviation of 78, and a median of 53.5 posts. On a weekly basis, a business in Milan posts an average of 2.4 times, while the most active business posts as many as 35 times in a single week. The like count is computed based on the number of likes a post has accumulated up to the date of data collection. The average like count is 735.3, but the maximum reaches 425,869, suggesting a large variation.[6] The word count per post averages at 46.1, suggesting the short format of Instagram posts.

Table 2 Summary of Social Media Data by City in 2023

City	Variable	Mean	Std	Min	Median	Max
Milan	Number of Business = 486					
	Total Number of Post	70.6	78.0	1.0	53.5	855.0
	Posting Frequency (Weekly)	2.4	2.0	1.0	2.0	35.0
	Like Count (Weekly)	735.3	6625.0	0.0	44.0	425869.0
	Post Word Count (Weekly)	46.1	40.1	0.0	37.0	347.0
Paris	Number of Business = 429					
	Total Number of Post	65.0	66.5	1.0	45.0	377.0
	Posting Frequency (Weekly)	2.4	1.6	1.0	2.0	23.0
	Like Count (Weekly)	603.9	3202.7	0.0	66.0	151749.0
	Post Word Count (Weekly)	49.8	42.1	0.0	41.0	425.0
Lyon	Number of Business = 432					
	Total Number of Post	60.2	64.4	1.0	44.5	637.0
	Posting Frequency (Weekly)	2.2	1.8	1.0	2.0	38.0
	Like Count (Weekly)	235.5	993.7	0.0	39.0	87064.0
	Post Word Count (Weekly)	53.0	43.2	0.0	43.0	399.0
Madrid	Number of Business = 544					
	Total Number of Post	85.4	76.2	1.0	73.5	717.0
	Posting Frequency (Weekly)	2.6	1.8	1.0	2.0	45.0
	Like Count (Weekly)	416.6	15347.1	0.0	38.0	3068908.0
	Post Word Count (Weekly)	47.8	39.5	0.0	40.0	396.0
Valencia	Number of Business = 315					
	Total Number of Post	83.0	89.1	1.0	65.0	990.0
	Posting Frequency (Weekly)	2.6	2.1	1.0	2.0	26.0
	Like Count (Weekly)	294.9	3400.1	0.0	39.0	416840.0
	Post Word Count (Weekly)	50.2	40.2	0.0	43.0	374.0

2.3. Variable Construction: Content Similarities

To examine Gen AI's effect on content homogenization, we first measure content similarities. Specifically, we calculate pairwise text similarities among all Instagram posts generated by different businesses within a given week in a city. To achieve a comprehensive assessment, we employ a series

[6] Since this data is a snapshot, older posts might have higher like counts simply because they've been available longer to gather engagement. However, we believe that the bulk of engagement occurs soon after a post is made public, which helps to mitigate potential measurement error due to post age.

of advanced models to compute text similarities from four distinct perspectives: *lexical*, *syntactical*, *semantic*, and *language style*.

1. *Lexical Similarity*: This measure focuses on the frequency of individual words. Therefore, we use a parsimonious bag-of-words (BoW) count model, which transforms each text into fixed-length vectors by counting word occurrences. Stop-words and URLs are removed from the text to ensure cleaner input. This method does not consider the order or meaning of the words but solely measures how often each word appears in the text. Consequently, if two Instagram posts share many common words, the lexical similarity score will be higher.
2. *Syntactical Similarity*: This measure examines the arrangement and order of words within sentences. Here, we apply an N-Gram model (up to Trigrams). The transformed vector captures the frequency of both individual words and short word sequences. Therefore, when two texts share many common n-gram sequences, their similarity score will be higher.
3. *Semantic Similarity*: This measure goes beyond word matching to assess the similarity of meanings or concepts in the text, even if different words or structures are used. For this, we use pre-trained Sentence-BERT (S-BERT), which transforms each text into a high-dimensional embedding that represents its semantic meaning. Therefore, when two texts convey similar meanings or ideas, even with varied wording or structure, their semantic similarity score will be higher.

   For the three measurements above, we compute pairwise similarities of Instagram posts across different businesses on a weekly basis using the cosine similarity metric. This metric measures the angle between two vectors in a multi-dimensional space. Cosine similarities are normalized to a range of $[-1, 1]$, where 1 represents perfect similarity, 0 represents no similarity, and -1 represents perfect opposition (Negative values are applicable only under the S-BERT approach. However, in our dataset, all semantic similarity scores are non-negative, with one exception, which we removed.).
4. *Language Style Similarity*: The previous three measurements mainly evaluate language content, which conveys what a business is writing about. Language style, on the other hand, focuses on how the message is conveyed, particularly through the use of function words. This concept has been widely explored in marketing research (e.g., Ludwig et al. 2013, Beichert et al. 2024). According to Ireland and Pennebaker (2010), language style matching (LSM) is determined by nine functional dimensions.[7] To measure these, we use the Linguistic Inquiry and Word Count (LIWC) program (Pennebaker and Chung 2007), a dictionary-based text

[7] Given the predominance of European languages in our context, we use the LIWC dictionary specific to each respective language. However, not all language dictionaries align exactly with the English categorization. To address this, we manually select the most relevant function word categories to the best of our ability.

analysis tool that classifies words into distinct linguistic categories in two steps. First, we calculate LSM for each post pair by measuring similarity across each function word category. For example, for prepositions, we calculate the LSM between post i and post j as

$$\mathrm{LSM}_{ij,\mathrm{prep}} = 1 - \frac{\left|\mathrm{prep}_{\mathrm{Post}\ i} - \mathrm{prep}_{\mathrm{Post}\ j}\right|}{\mathrm{prep}_{\mathrm{Post}\ i} + \mathrm{prep}_{\mathrm{Post}\ j} + 0.0001}. \tag{1}$$

In the second step, we average across the nine function word categories to create a composite measure of LSM for each pair:

$$\begin{aligned}\mathrm{LSM}_{ij} =& \mathrm{avg}(\mathrm{LSM}_{ij,\text{Personal pronoun}} + \mathrm{LSM}_{ij,\text{Impersonal pronoun}} + \mathrm{LSM}_{ij,\text{Articles}} + \mathrm{LSM}_{ij,\text{Auxiliary verbs}} \\ & + \mathrm{LSM}_{ij,\text{Adverbs}} + \mathrm{LSM}_{ij,\text{Prepositions}} + \mathrm{LSM}_{ij,\text{Conjunctions}} + \mathrm{LSM}_{ij,\text{Negations}} \\ & + \mathrm{LSM}_{ij,\text{Quantifiers}}).\end{aligned} \tag{2}$$

We note that, given the diverse cultural backgrounds in our studied cities, businesses may post on social media in multiple languages within the same city. To ensure a consistent framework for measuring writing styles, we restrict our sample to posts written in the official language of each country when calculating language style similarity scores. The other three methods are not subject to this limitation as they are capable of handling multi-language contexts.

After calculating the pairwise similarity scores across all Instagram posts within a week in a city, we compute, for each business, the average pairwise similarity score between its posts and those of every other business during the same week in the same city. This process yields a panel data structure where each observation corresponds to a business's similarity metric in a week when it posted. We summarize the descriptives of the across-business similarity metrics in Table 3.

2.4. Model-Free Evidence

We first present model-free evidence suggesting that access to ChatGPT can lead to content homogenization. To do this, we compute pairwise similarity scores at the week level, following the methods described earlier, where each post by a business in a given week is compared to all other posts from that same week in a city, excluding those from the same business. In Figure 1, we plot the weekly similarity score, averaged across businesses in Milan, over week 8 to 17, our modeling period. The red vertical line is week 13, when the ban took effect. We can observe a notable decline in syntactical, semantic, and language style similarities during the ban.

2.5. Estimation Strategy

The standard estimation approach under our identification strategy is difference-in-differences (DiD), which compares outcomes in treated cities (Milan) and non-treated cities (e.g., Paris) between the pre-ban and ban periods. While this approach is widely adopted for causal inference

Table 3 Summary of Similarity Metrics (Across-Business) by City in 2023

City	Metric	Mean	Std	Min	Median	Max
Milan	Lexical Similarity	0.020	0.02	0.00	0.02	0.07
	Syntactical Similarity	0.025	0.01	0.00	0.03	0.08
	Semantic Similarity	0.414	0.07	0.09	0.42	0.56
	Language Style Similarity	0.396	0.11	0.00	0.43	0.57
Paris	Lexical Similarity	0.024	0.02	0.00	0.02	0.10
	Syntactical Similarity	0.031	0.02	0.00	0.03	0.11
	Semantic Similarity	0.424	0.07	0.12	0.44	0.59
	Language Style Similarity	0.390	0.09	0.00	0.42	0.52
Lyon	Lexical Similarity	0.029	0.02	0.00	0.03	0.09
	Syntactical Similarity	0.047	0.02	0.00	0.05	0.13
	Semantic Similarity	0.435	0.07	0.03	0.45	0.58
	Language Style Similarity	0.416	0.10	0.00	0.45	0.57
Valencia	Lexical Similarity	0.021	0.01	0.00	0.02	0.07
	Syntactical Similarity	0.053	0.03	0.00	0.05	0.16
	Semantic Similarity	0.409	0.07	0.09	0.42	0.57
	Language Style Similarity	0.380	0.09	0.00	0.41	0.51
Madrid	Lexical Similarity	0.021	0.01	0.00	0.02	0.08
	Syntactical Similarity	0.053	0.03	0.00	0.05	0.15
	Semantic Similarity	0.423	0.07	0.10	0.44	0.57
	Language Style Similarity	0.377	0.09	0.00	0.40	0.50

in panel settings, it relies on the strong assumption that treated and control units would have followed parallel trends in the absence of treatment. To assess the validity of this assumption in our context, we conduct an event-study analysis, presented in Appendix A1. The results indicate that the parallel trend assumption does not hold for all the similarity metrics, undermining the validity of causal estimates derived from the standard DiD model.

As a result, we adopt two recently developed methods that relax the parallel trends requirement—the Generalized Synthetic Control Method (Xu 2017) and the Synthetic Difference-in-Differences approach (Arkhangelsky et al. 2021)—as our preferred specifications.

2.5.1. Generalized Synthetic Control Method (Generalized SCM) This method extends the conventional synthetic control method to accommodate multiple treated units, in our case, the multiple businesses in Milan.

The key innovation in Generalized SCM lies in its use of an Interactive Fixed Effects (IFE) model, which accounts for time-varying unobserved confounders through a latent factor structure (Xu 2017). Thus, it imputes counterfactual outcomes for all treated units in one run, rather than matching control units to each treated unit individually, which is computationally heavy. Specifically, the outcome variable is modeled as:

$$Y_{iw} = \theta D_{iw} + \boldsymbol{\mu}_i' \mathbf{f}_w + \varepsilon_{iw} \tag{3}$$

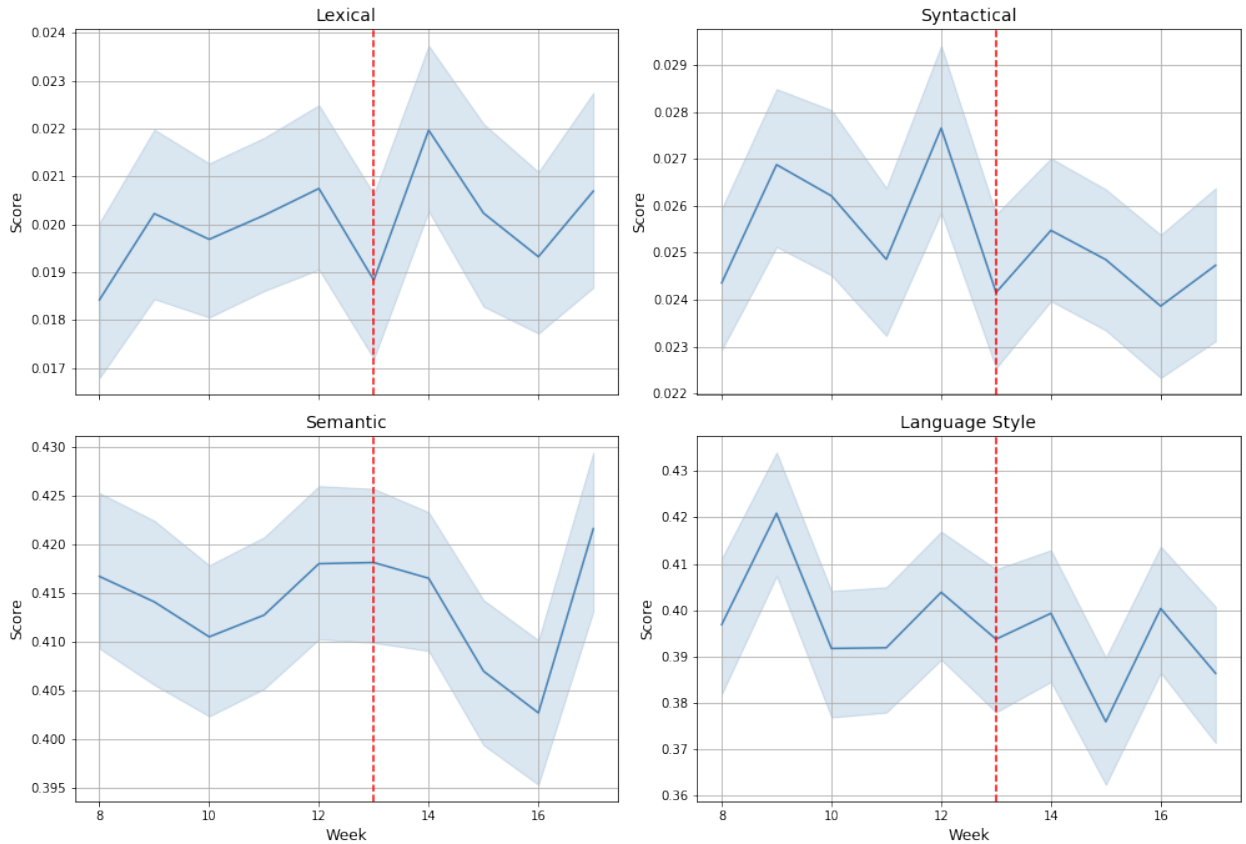


Figure 1 Similarity Scores in Milan (Treatment) before and after the ChatGPT Ban

where Y_{iw} denotes the outcome for business i during week t. We include business (α_i) and week (β_w) fixed effects to control for unobserved time-invariant characteristics of the business and time-specific trends. The treatment indicator, D_{iw}, equals one for businesses in Milan during the ban period, and θ represents the average treatment effect. Standard errors are clustered at the city level to account for the city-level shock and correlation.

Different from a standard DiD, the term $\boldsymbol{\mu}_i$ is a vector of unit-specific factor loadings, and $\mathbf{f}_w$ is a vector of unobserved common factors that evolve over time. The interactive term $\boldsymbol{\mu}_i'\mathbf{f}_w$ captures complex, unit-specific responses to unobserved temporal shocks that cannot be addressed by traditional fixed effects alone.

This structure enables Generalized SCM to flexibly model non-parallel pre-treatment trends and heterogeneous treatment effects, which are likely in real-world panel data settings. In our application, this is particularly important given the varying baseline text similarity of businesses and the possibility of region-specific communication trends. The model parameters are estimated using matrix completion techniques based on pre-treatment observations, and inference is conducted via non-parametric bootstrapping. This approach provides a robust framework for estimating the effect of the ChatGPT ban while accounting for latent confounding dynamics that could bias simpler estimators.

2.5.2. Synthetic Difference-in-Differences (Synthetic DiD) This method combines elements of traditional difference-in-differences (DiD) and the synthetic control method (SCM), offering a robust approach in situations where the parallel trends assumption is unlikely to hold strictly.

The core idea of Synthetic DiD is to construct a weighted combination of control units that closely match the treated unit(s) in the pre-treatment period, while still using a difference-in-differences framework for post-treatment comparison. Formally, the Synthetic DiD estimator is obtained by solving the following weighted least squares problem:

$$\left(\hat{\tau}^{\text{sdid}}, \hat{\mu}, \hat{\alpha}, \hat{\beta}\right) = \arg \min_{\alpha,\beta,\tau,\mu} \left\{ \sum_{i=1}^{N} \sum_{w=8}^{T} \left(Y_{iw} - \mu - \alpha_i - \beta_w - \tau D_{iw}\right)^2 \hat{\omega}_i^{\text{sdid}} \hat{\lambda}_w^{\text{sdid}} \right\} \tag{4}$$

where Y_{iw} denotes the outcome for business i at week w, μ is a global intercept, α_i and β_w are unit and time fixed effects (which are normalized to 0), and D_{iw} is the treatment indicator. The weight for unit i $\hat{\omega}_i^{\text{sdid}}$ and weight to week w $\hat{\lambda}_w^{\text{sdid}}$ are selected to balance treated and control units in the pre-treatment period. The result is an estimate of the average treatment effect $\hat{\tau}^{\text{sdid}}$ that is more robust to violations of the parallel trends assumption than standard DiD. We highlight that Synthetic DiD relies on a balanced sample, which we construct accordingly.

For robustness, we plot out the pre-treatment fit from Generalized SCM and Synthetic DiD in Appendix A1. The plots indicate a good quality of match and support the credibility of the estimated treatment effects.

3. Impacts of Gen AI on Content Homogenization

We leverage Italy's temporary ban on ChatGPT in 2023 as a natural experiment to identify the effects of Gen AI on marketing content homogenization as described in Section 2.1. This exogenous shock randomizes access to ChatGPT at the city level, enabling us to compare the marketing content in the treatment city with that in the control cities. Our modeling period extends from week 8 to week 17 of 2023, when the ban started on week 13 (week commencing 27 March 2023) and ended on week 17. We estimate the effect of Italy's ChatGPT ban using the Generalized Synthetic Control Method and Synthetic Difference-in-Differences, as outlined in the preceding sections. The results, reported in Table 4, indicate that the estimated treatment effects are negative across all similarity measures.

Specifically, under the Generalized SCM estimator, the ban is associated with a reduction of 0.003 in lexical and syntactical similarity, 0.008 in semantic similarity, and 0.013 in language style similarity. To contextualize the magnitude of these effects, we refer to the descriptive statistics in

Table 4 Impacts on Marketing Content Similarity from ChatGPT Ban

	Lexical (1)	Syntactical (2)	Semantic (3)	Language Style (4)
Generalized SCM	-0.003***	-0.003***	-0.008***	-0.013***
	(0.001)	(0.001)	(0.002)	(0.004)
Observations	12,004	12,004	12,004	10,354
Synthetic DID	-0.001***	-0.003***	-0.009***	-0.009
	(0.000)	(0.001)	(0.001)	(0.007)
Observations	5,780	5,780	5,780	4,420

Note: ***, **, and * represent 1%, 5%, and 10% significance levels, respectively. Standard errors are clustered at the city level. Bootstrapped Standard errors are used in Generalized SCM and Synthetic DID methods.

Table 3. These coefficients correspond to relative declines of approximately 15%, 12%, 2%, and 3% in lexical, syntactical, semantic, and language style similarity, respectively.[8] This suggests that the content homogenization is particularly pronounced through the use of similar words and sentence structure.

As a robustness check, we also construct a fifth similarity metric using OpenAI embedding and find consistent results. See details in Appendix A2.1. 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. We refer the readers to Appendix A2.2 and A2.3 for details. While we could not rule out that some users in Milan might access ChatGPT via VPN (Virtual Private Network), which masks IP addresses by routing through another country, this possibility implies that the negative impact identified in our result is a conservative estimation.

3.1. Heterogeneous Treatment Effects

Our identification strategy exploits the country-level exogenous shock introduced by the ChatGPT ban. However, the underlying hypothesis assumes that the ban would disproportionately affect businesses that rely more heavily on ChatGPT for content creation. Since we do not observe direct usage data, we turn to two proxies to explore heterogeneity in ChatGPT usage: 1) cuisine origin-based heterogeneity, and 2) posts frequency drop-based heterogeneity.

We conduct two heterogeneity analyses to explore whether the treatment effect of the AI ban varies systematically across subgroups.

[8] These are computed by dividing the coefficient estimates by the mean of the similarity score.

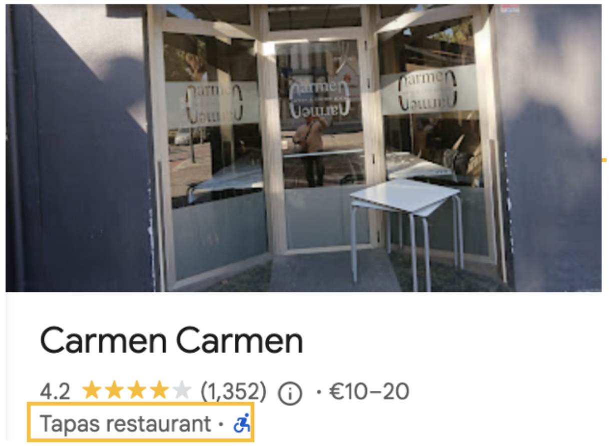


Figure 2 **Restaurant Cuisine Information on Google Map**

3.1.1. Cuisine Origin-Based Heterogeneity We classify restaurants based on the origin of the cuisine they serve, distinguishing between local and non-local cuisines. The underlying hypothesis is that non-local cuisines are more likely to be operated by non-native owners, who may face greater language or cultural barriers in local marketing. These owners may have been more reliant on Gen AI tools such as ChatGPT for generating marketing content in the local language. As such, we hypothesize that non-local cuisine restaurants exhibit a larger negative response to the AI ban.

We rely on business category information provided by Google Maps to classify the origin of each restaurant's cuisine. As a guiding principle, if the category contains explicit geographic indicators – such as Italian Restaurant, Greek Restaurant, or Japanese Restaurant – we map the cuisine to one of the following groups: Italian, Other European (e.g., German, Greek), or Non-European (e.g., Chinese, Japanese, Indian, Mexican). Next, based on the country where the business operates, we label the restaurant's cuisine as local or non-local. For example, in Figure 2, Google Map categorizes the restaurant "Carmen Carmen" as a tapas restaurant, which offers Spanish cuisine. Since this restaurant is in Valencia, we thus label this restaurant as local for the market where the restaurant operates.

If the category is vague or lacks a clear geographic reference—such as Salad Bar, Bistro, or Fast Food Restaurant—and no additional geographic information is available, we exclude the restaurant from the cuisine-based heterogeneity analysis. This ensures that our classification captures meaningful differences in likely owner background and potential reliance on Gen AI for creating content in local languages.

We then define the indicator variable d_{local} to equal 1 if the restaurant's cuisine matches the country in which the restaurant operates, and 0 otherwise. Using this classification, we stratify the

Table 5 Heterogeneous Treatment Effects - Cuisine Origin

Generalized SCM	(1) Lexical	(2) Syntactical	(3) Semantic	(4) Language Style
Local Restaurants	-0.001	-0.003**	-0.008*	-0.002
	(0.001)	(0.001)	(0.004)	(0.006)
Observations	1,792	1,792	1,792	1,667
Non-Local Restaurants	-0.003***	-0.003**	-0.013***	-0.022 ***
	(0.001)	(0.001)	(0.002)	(0.007)
Observations	3,596	3,596	3,596	2,978
Synthetic DID	Lexical	Syntactical	Semantic	Language Style
Local Restaurants	-0.002	0.000	-0.006	0.017**
	(0.002)	(0.002)	(0.005)	(0.006)
Observations	800	800	800	710
Non-Local Restaurants	-0.002***	-0.002	-0.015***	-0.017**
	(0.000)	(0.002)	(0.002)	(0.007)
Observations	1,840	1,840	1,840	1,340

Note: ***, **, and * represent 1%, 5%, and 10% significance levels, respectively. Standard errors are clustered at the city level. Bootstrapped Standard errors are used in Generalized SCM and Synthetic DID methods.

sample and conduct the analysis separately for local and non-local cuisine subgroups. The results are presented in Table 5. Consistent with our hypothesis, we find that the negative effect, the decrease in content similarity across businesses during the ChatGPT ban, is more pronounced for non-local restaurants. This strengthens our argument that the businesses that are more likely to rely on Gen AI tools would see a stronger effect from the ban.

3.1.2. Posting Frequency Drop-Based Heterogeneity Our second heterogeneity analysis leverages on the drops in posting frequency following the AI ban. Intuitively, if a restaurant reduces their posting frequency during the ChatGPT ban period more than their peers, then it is likely that that the restaurant also had a higher reliance on Gen AI for content creation. Therefore, we hypothesize that restaurants with the larger drops in posting activity are those most affected by the ban and thus demonstrate a stronger treatment effect.

To operationalize this, we classify restaurants based on the percentage drop in their average posting frequency between the pre-ban and ban periods. The drop is calculated as the relative change in mean posting frequency across the two periods at the business level. We use the median percentage drop among restaurants in Italy as the cutoff point. Restaurants with an absolute drop

Table 6 Heterogeneous Treatment Effects - Posting Frequency Drop

Generalized SCM	(1) Lexical	(2) Syntactical	(3) Semantic	(4) Language Style
High Drop Restaurants	-0.002*	-0.004***	-0.009***	-0.022***
	(0.001)	(0.001)	(0.001)	(0.007)
Observations	5,368	5,368	5,368	4,536
Low Drop Restaurants	-0.003*	-0.002	-0.006***	-0.004
	(0.001)	(0.001)	(0.002)	(0.010)
Observations	6,636	6,636	6,636	5,818
Synthetic DID	Lexical	Syntactical	Semantic	Language Style
High Drop Restaurants	-0.001***	-0.003***	-0.010***	-0.020**
	(0.000)	(0.001)	(0.002)	(0.008)
Observations	2,360	2,360	2,360	1,760
Low Drop Restaurants	-0.000	-0.002	-0.008***	-0.001
	(0.000)	(0.002)	(0.002)	(0.005)
Observations	3,420	3,420	3,420	2,660

Note: ***, **, and * represent 1%, 5%, and 10% significance levels, respectively. Standard errors are clustered at the city level. Bootstrapped Standard errors are used in Generalized SCM and Synthetic DID methods.

greater than the median are classified as high-drop, while those with a drop less than or equal to the median are classified as low-drop. This binary classification allows us to stratify the sample and estimate treatment effects separately across the two groups.

The results are reported in Table 6. As shown, the high-drop subsample exhibits consistently negative treatment effects, whereas the low-drop subsample shows smaller effect sizes and only sporadically significant results. This implies that for restaurants that are more likely to have relied on Gen AI tools, the ban has a stronger effect on content similarity, consistent with our hypothesis.

4. Impacts of Gen AI on Consumer Engagement

The previous section establishes the causal impact of Gen AI access on marketing content homogenization. Next, we investigate our second research question: Does access to Gen AI tools affect consumer engagement?

To capture consumer engagement, we focus on the average number of likes a business receives per week. The outcome variable, denoted as AverageLikes_{iw}, represents the average number of likes on Instagram posts by business i during week w.[9] We note that the distribution of AverageLikes_{iw}

[9] Our data is based on a snapshot taken at the end of 2024, meaning that the recorded like counts reflect accumulated engagement over time.

is highly skewed, with a median of 46.25, a maximum of 131,018, and a standard deviation of $1,571.25$. To address this skewness, we use a log-transformed version of the outcome variable, implemented as $\log(\text{AverageLikes} + 1)$.

Following the estimation strategies outlined in Section 3, we apply both the Generalized Synthetic Control Method (Generalized SCM) and Synthetic Difference-in-Differences (Synthetic DiD) to derive causal estimates. The results are presented in Table 7. In Column (1), both methods yield positive and statistically significant effects, suggesting an increase in average likes during the ChatGPT ban period. In Column (2), we also estimate models using a winsorized version of the outcome variable, in which extreme values are capped at the 5% level. The results remain consistent in both direction and significance, confirming that the observed effect is not driven by a small number of outliers.

To further account for potential confounding factors affecting consumer engagement, we introduce additional control variables, including the average word count per post and the weekly posting frequency. The positive effect of the ban on consumer engagement persists even with these controls. Detailed results from these extended models are reported in Appendix A3.

Due to data limitations, we are unable to rule out the effect of other activities conducted by restaurants on Instagram, such as paid sponsorship. However, we argue that such activities should not be systematically dependent on the ChatGPT ban. Another possibility is that the higher engagement is due to Instagram's algorithm preference for non-AI content. However, we note that our observation window is early 2023. As Meta officially announces its AI detection tools in February 2024, it is unlikely that the observed effect is driven by AI-motivated platform policies.[10]

In conclusion, the evidence suggests that during the ChatGPT ban period, average likes per post increased, pointing to a potential rise in consumer engagement.

4.1. Correlation between Marketing Content Similarity and Consumer Engagement

The preceding analysis confirms that the access to Gen AI tool could affect consumer engagement with the marketing content. In this section, we aim to offer an explanation to this observed pattern in relation with content homogenization.

Our hypothesis is grounded in Optimal Distinctiveness Theory (Brewer 1991), which posits that individuals are motivated by two opposing needs: the need for assimilation and inclusion – a desire to belong within social groups – and the need for differentiation and distinctiveness, which reflects the desire to assert one's uniqueness. In marketing contexts, this duality is particularly relevant to brand positioning, as consumers seek to engage with brands that both reflect social belonging and affirm their personal distinctiveness (Keller 2020).

[10] https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads/

Table 7 Impacts on Consumer Engagement from ChatGPT Ban

	Log Average Like Count (1) All	Average Like Count (2) 5% Winsorized
Generalized SCM	0.035***	8.277***
	(0.009)	(3.636)
Observations	11,191	11,191
Synthetic DiD	0.067***	8.047
	(0.0198)	(6.915)
Observations	5,070	5,070

Note: ***, **, and * represent 1%, 5%, and 10% significance levels, respectively. Standard errors are clustered at the city level. Bootstrapped Standard errors are used in Generalized SCM and Synthetic DID methods.

Based on this framework, we expect an inverted U-shaped relationship between content similarity and consumer engagement. That is, consumer response may increase with similarity up to a certain point, where the content is familiar, but decline when content becomes either too homogeneous or too distinct. To test this relationship empirically, we specify the following model:

$$LogAverageLikeCount_{iw} = \alpha_i + \beta_w + \theta_1 Similarity_{iw} + \theta_2 {Similarity_{iw}}^2 + \theta_3 P_{iw} + \epsilon_{iw} \tag{5}$$

where we employ a quadratic specification to capture the hypothesized U-shape relationship between similarity and consumer engagement. P_{iw} controls for the average word count of the posts by business i in week w. This is to control for post characterists that could affect consumer engagement. The other variables are the same as previously described. We report the results in Table 8, where each column corresponds to a different similarity measure. Across all specifications, we consistently find a positive coefficient on the similarity variable and a negative coefficient on its squared term, supporting the hypothesized inverted U-shaped relationship. However, this quadratic relationship reaches statistical significance only in the cases of Syntactical and Language Style Similarities. We also identify a positive relationship between post length and consumer engagement, suggesting that consumers may respond more actively to posts that offer more detailed or informative content.

A simple calculation of the turning points yields meaningful thresholds: 0.04 for Lexical Similarity, 0.05 for Syntactical Similarity, 0.50 for Semantic Similarity and 0.32 for Language Style

Similarity.[11] This suggests that beyond these thresholds, further increases in similarity begin to negatively affect consumer engagement, likely due to a loss of perceived distinctiveness.

Taken together, these results suggest that moderate similarity can boost engagement, but excessive similarity becomes counterproductive. This implies that reduced content similarity can have varying effects on engagement depending on whether content was previously above or below the optimal threshold. If the market was saturated with highly similar content, reducing homogenization could enhance consumer engagement. Conversely, if similarity was already low, further differentiation might reduce engagement by making the content appear overly distinct or fragmented. Whereas in this case, based on the descriptive summary of across business similarity metrics in Table 2, it is possible that the overall similarities have reached beyond the critical point in some period. As a result, the ChatGPT ban, which reduced content similarities across businesses, can have a positive impact on consumer engagement.

Combined with our earlier findings that 1) the ChatGPT ban reduced content homogenization while increasing consumer engagement; and 2) the evidence supporting an inverted U-shaped relationship between content similarity and engagement, we find a consistent narrative: Gen AI tools, while helpful in content creation, may inadvertently lower consumer engagement by promoting excessive content similarity. This highlights the importance of balancing efficiency gains from Gen AI with content differentiation to maintain consumer engagement. Taken together, the results suggest that moderate content similarity enhances consumer engagement, while excessive similarity can be counterproductive. Thus, the effect of reduced similarity depends on the starting point: in oversaturated markets, less homogenization may boost engagement; when similarity is already low, further differentiation could make content feel disjointed or inconsistent.

In this context, the descriptive statistics in Table 3 indicate that similarity across businesses may have exceeded the optimal threshold during certain periods. Accordingly, the ChatGPT ban – which reduced content similarity – may have contributed to increased engagement.

5. Impacts of Gen AI on Other Marketing Outcomes

We also analyze changes in other marketing-related outcomes that are likely to be affected by access to Gen AI tools. These include posting frequency and length of posts.

5.1. Posting Frequency

We examine whether access to ChatGPT influences the frequency of posting. To model the frequency of posting, we build the outcome variable Frequency_{iw} which represents the number of posts by business i in week w. In our modeling sample, on average, business posts 1.34 times a

[11] The critical point is calculated based on $-\frac{\theta_1}{2\theta_2}$ from Equation 5.

Table 8 Correlation: Content Similarity and Consumer Engagement

	Log Average Like Count			
OLS Model	(1)	(2)	(3)	(4)
Lexical	2.826			
	(1.549)			
Lexical square	-33.122			
	(38.157)			
Syntactical		2.874^{*}		
		(1.317)		
Syntactical square		-28.431^{**}		
		(9.295)		
Semantic			2.296^{*}	
			(1.070)	
Semantic square			-2.280	
			(1.380)	
Language Style				0.868^{**}
				(0.269)
Language Style square				-1.365^{**}
				(0.489)
Average Word Count	0.004^{***}	0.004^{***}	0.004^{***}	0.003^{***}
	(0.001)	(0.001)	(0.001)	(0.000)
Week	Yes	Yes	Yes	Yes
Business	Yes	Yes	Yes	Yes
Observations	11,191	11,191	11,191	9,668
R^2	0.87	0.87	0.87	0.87

Note: ***, **, and * represent 1%, 5%, and 10% significance levels, respectively. Standard errors are clustered at city level.

week, with the maximum reaching 28 posts in a single week. Weeks with no posts are recorded as a value of 0.

We use the same estimation strategies as in Section 3. Table 9 reports the results. Column (1) shows a negative effect from the ban on the frequency of posting. This means that the ban has led to a decrease in posting frequency. While readers might challenge that this decreased posting frequency can lead to selection bias, our main results in Section 3 are robust with the Synthetic DiD method, which relies on a balanced sample with restaurants that have posted every week during our observation window. This rules out the possibility that the reduced similarity during the ban period is due to businesses' reduction in posting.

5.2. Length of Post

Next, we evaluate the impact of ChatGPT on post length, measured by word count. The underlying intuition is that Gen AI tools can enhance the comprehensiveness of text, potentially leading to

Table 9 Impacts on Other Outcomes from ChatGPT Ban

	Frequency	Average Word Count
	(1)	(2)
Generalized SCM	-0.061*** (0.024)	-1.210*** (0.283)
Observations	21,530	12,004
Synthetic DiD	-0.056*** (0.019)	-0.637*** (0.220)
Observations	21,530	5,780

Note: ***, **, and * represent 1%, 5%, and 10% significance levels, respectively. Standard errors are clustered at city level. Bootstrapped Standard errors are used in Generalized SCM and Synthetic DID methods.

longer and more detailed posts. The outcome is the $AverageWordCount_{iw}$, the average word count of business i during week w, with results reported in Column (2) of Table 9. The number of observations differs from Column (1) because if a business does not post in a given week, no observation is recorded for that period.

Our finding indicates a negative effect on post length, with an estimated reduction of approximately one word following the ChatGPT ban. We acknowledge that this effect is relatively small in absolute terms. But given the short format of social media posts in our context, even a few words' change could affect the performance. For example, Zhang et al. (2024) find that even the change of titles, facilitated by AI technology on a UGC platform, can significantly boost content consumption.

6. Concluding Remarks

As one of the first attempts to assess the impact of Gen AI on marketing content homogenization using observational data, this paper finds a consistent and significant reduction in content similarity across businesses during the ban period, when ChatGPT became suddenly unavailable. This suggests that Gen AI tools like ChatGPT contribute to greater content homogenization in the market, likely by enabling businesses to produce content that is similar in vocabulary, structure, topic, and style at scale. The effect is more pronounced among businesses that are more likely to have relied on the tool, such as non-local restaurant owners. Further, we find that consumer engagement increased during the ban period. A correlation analysis suggests that this improvement might be attributed to decreased content similarity. Lastly, we find that the ChatGPT ban also led to a decline in both posting frequency and average word count, suggesting a reduction in content production capacity without access to Gen AI tools.

Our findings carry important implications for businesses leveraging Gen AI in their marketing strategies. While tools like ChatGPT can enhance efficiency in content creation (increased frequency and longer posts), they may also lead to content homogenization, possibly dampening consumer engagement. Therefore, businesses should strike a balance: leveraging Gen AI to support content creation, but doing so thoughtfully to preserve differentiation and topical variation, for example, through prompt engineering or model fine-tuning.

Our research could also be extended along different dimensions. First, while we find strong correlation between access to Gen AI and consumer engagement, our data does not allow us to causally identify the relationship between content similarity and engagement. Alternative methods, such as lab or field experiments, may help establish concrete causal evidence. Second, as Gen AI tools continue to expand beyond text to include image, video, and multi-modal content, it will be important to examine whether similar dynamics of homogenization and engagement emerge in these richer content formats. These newer forms of content may have an even greater influence on marketing strategies.

References

Acar OA, Gvirtz A (2024) GenAI Can Help Small Companies Level the Playing Field. *Harvard Business Review* .

AlDahoul N, Rahwan T, Zaki Y (2025) Ai-generated faces influence gender stereotypes and racial homogenization. *Scientific reports* 15(1):14449.

Anderson BR, Shah JH, Kreminski M (2024) Homogenization effects of large language models on human creative ideation. *Proceedings of the 16th Conference on Creativity & Cognition*, 413–425.

Arkhangelsky D, Athey S, Hirshberg DA, Imbens GW, Wager S (2021) Synthetic difference-in-differences. *American Economic Review* 111(12):4088–4118.

Authority TIDP (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. URL `https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9870847`, accessed: 2024-12-12.

BBC (2021) Instagram lets users hide likes to reduce social media pressure. URL `https://www.bbc.co.uk/news/technology-57254488`, accessed: 2024-12-12.

Beichert M, Bayerl A, Goldenberg J, Lanz A (2024) Revenue generation through influencer marketing. *Journal of Marketing* 00222429231217471.

Bertomeu J, Lin Y, Liu Y, Ni Z (2023) Capital market consequences of generative AI: Early evidence from the ban of ChatGPT in Italy. *Available at SSRN* 4452670.

Brewer MB (1991) The social self: On being the same and different at the same time. *Personality and social psychology bulletin* 17(5):475–482.

Burtch G, Lee D, Chen Z (2023) The consequences of generative AI for UGC and online community engagement. *Available at SSRN 4521754* .

Chen Z, Chan J (2024) Large language model in creative work: The role of collaboration modality and user expertise. *Management Science* .

d'Aspremont C, Gabszewicz JJ, Thisse JF (1979) On Hotelling's" Stability in competition". *Econometrica: Journal of the Econometric Society* 1145–1150.

Doshi AR, Hauser OP (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 MR, Groh M, Herman L, Leach N, et al. (2023) Art and the science of generative AI. *Science* 380(6650):1110–1111.

Farquhar PH (1989) Managing brand equity. *Marketing research* 1(3).

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. URL `https://www.forbes.com/sites/chriswestfall/2023/11/16/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. URL `https://www.ft.com/content/7db1c1b3-5a61-4dee-a922-ade8b9c77522`, accessed: 2024-12-12.

Gray A (2024) ChatGPT "contamination": estimating the prevalence of LLMs in the scholarly literature. *arXiv preprint arXiv:2403.16887* .

Ireland ME, Pennebaker JW (2010) Language style matching in writing: synchrony in essays, correspondence, and poetry. *Journal of personality and social psychology* 99(3):549.

Kaufman MS, Goldberg LG, Avery J (2020) Restaurant revolution: How the industry is fighting to stay alive. URL `https://hbswk.hbs.edu/item/restaurant-revolution-how-the-industry-isfighting-to-stay-alive`.

Keller KL (2020) Consumer research insights on brands and branding: a jcr curation. *Journal of Consumer research* 46(5):995–1001.

Kobak D, González-Márquez R, Horvát EÁ, Lause J (2024) Delving into ChatGPT usage in academic writing through excess vocabulary. *arXiv preprint arXiv:2406.07016* .

Kreitmeir DH, Raschky PA (2023) The Unintended Consequences of Censoring Digital Technology–Evidence from Italy's ChatGPT Ban. *arXiv preprint arXiv:2304.09339* .

Kuksov D (2007) Brand value in social interaction. *Management Science* 53(10):1634–1644.

Lasaleta JD, Redden JP (2018) When promoting similarity slows satiation: The relationship of variety, categorization, similarity, and satiation. *Journal of Marketing Research* 55(3):446–457.

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, Li G (2024) When ChatGPT is gone: Creativity reverts and homogeneity persists. *arXiv preprint arXiv:2401.06816* .

Ludwig S, De Ruyter K, Friedman M, Brüggen EC, Wetzels M, Pfann G (2013) More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of marketing* 77(1):87–103.

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.

Meincke L, Nave G, Terwiesch C (2025) Chatgpt decreases idea diversity in brainstorming. *Nature human behaviour* 1–3.

MenuTiger (2024) Boost restaurant social media strategy with a QR code menu software. URL `https://www.menutiger.com/blog/restaurant-social-media-statistics`, accessed: 2024-12-12.

Pennebaker JW, Chung CK (2007) Expressive writing, emotional upheavals, and health. *Foundations of health psychology* 263–284.

Pullig C, Simmons CJ, Netemeyer RG (2006) Brand dilution: when do new brands hurt existing brands? *Journal of Marketing* 70(2):52–66.

Ratner RK, Kahn BE, Kahneman D (1999) Choosing less-preferred experiences for the sake of variety. *Journal of consumer research* 26(1):1–15.

Redden JP (2008) Reducing satiation: The role of categorization level. *Journal of Consumer Research* 34(5):624–634.

Rolls BJ, Van Duijvenvoorde P, Rolls ET (1984) Pleasantness changes and food intake in a varied four-course meal. *Appetite* 5(4):337–348.

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.

Sun Z, Zhang Z, Shen X, Zhang Z, Liu Y, Backes M, Zhang Y, He X (2024) Are we in the ai-generated text world already? quantifying and monitoring aigt on social media. *arXiv preprint arXiv:2412.18148* .

Van Horen F, 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.

Woolley K, Sharif MA (2022) Down a rabbit hole: how prior media consumption shapes subsequent media consumption. *Journal of Marketing Research* 59(3):453–471.

Xu Y (2017) Generalized synthetic control method: Causal inference with interactive fixed effects models. *Political Analysis* 25(1):57–76.

Yilmaz ED, Naumovska I, Aggarwal VA (2023) AI-driven labor substitution: Evidence from google translate and ChatGPT. *INSEAD Working Paper* .

Zhang X, Sun C, Zhang R, Goh KY (2024) The value of ai-generated metadata for ugc platforms: Evidence from a large-scale field experiment. *ArXiv* abs/2412.18337, URL `https://api.semanticscholar.org/CorpusID:274992114`.

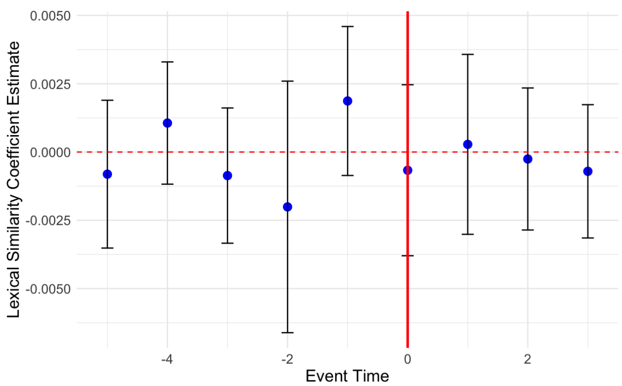
Appendix A1: Parallel Trend Tests

A1.1. Standard DiD

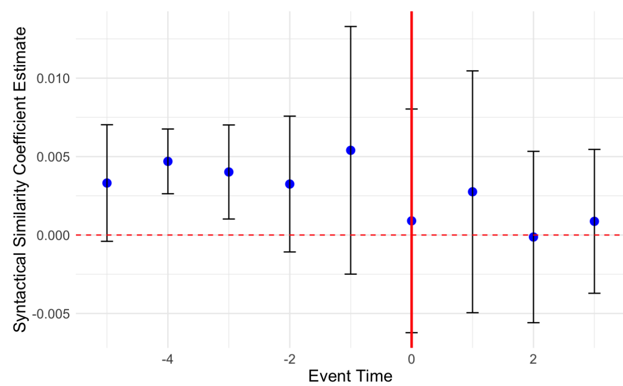
Implementing standard DiD requires a strong assumption of parallel trends. If this assumption fails, the control group will not be a good counterfactual for Milan leading to potential bias in causal estimates. To assess the validity of this assumption, we use the following model specification:

$$Y_{iw} = \alpha_i + \beta_w + \sum_{w=8}^{w=17} \omega_w D_{iw} \epsilon_{iw}, \tag{6}$$

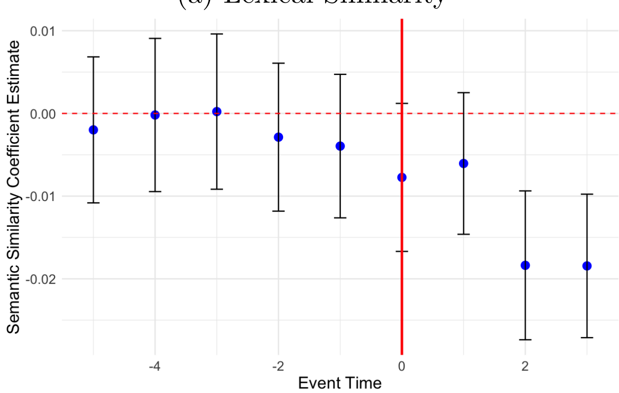
We plot the estimated coefficients ω along with their confidence intervals in Figure A1, comparing overall lexical, syntactical, semantic and language style similarities of marketing contents in Milan with control cities. The figures show that prior to the ban of ChatGPT (indicated by the red vertical lines on the figures), the coefficients are not all close to zero especially for syntactical similarity, in violation of the parallel trend assumption for standard DiD approach.



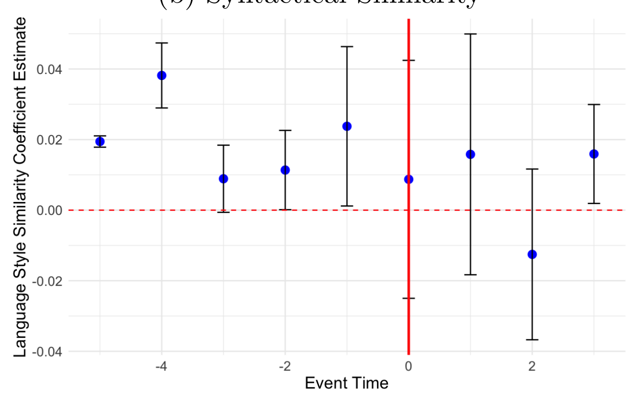
(a) Lexical Similarity



(b) Syntactical Similarity



(c) Semantic Similarity

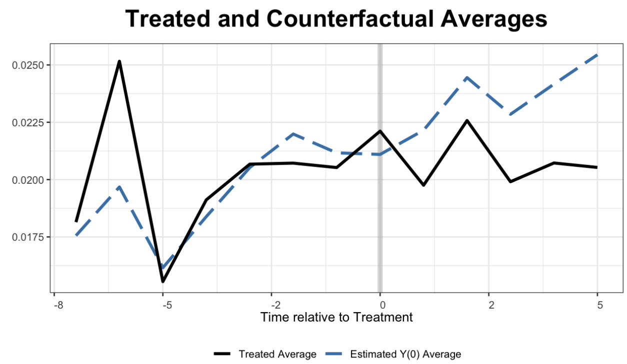


(d) Language Style Similarity

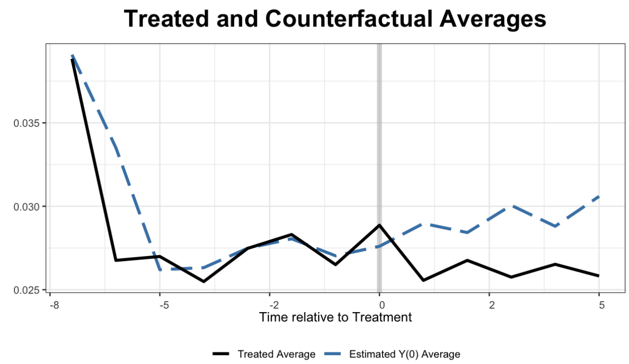
Figure A1 Changes in Content Similarities

A1.2. Generalized SCM

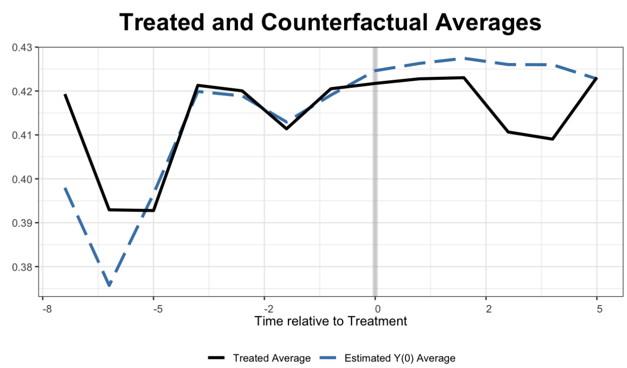
Figure A2 presents the average outcomes for the treated units and their corresponding synthetic controls generated by the Generalized SCM method over the 10 weeks (from week 8 to week 17). The solid line represents the treated, while the dotted line denotes the synthetic. Although the two trajectories are not perfectly aligned, the method achieves a strong pre-treatment fit especially closer to the treatment time, indicating that the synthetic control effectively captures the treated unitsâ trends prior to the treatment.



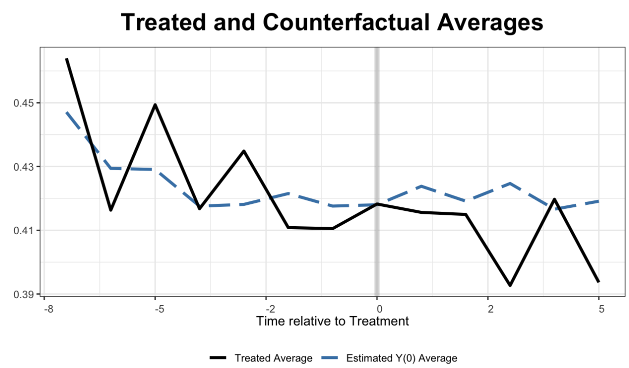
(a) Lexical Similarity



(b) Syntactical Similarity



(c) Semantic Similarity



(d) Language Style Similarity

Figure A2 Generalized SCM Pre-treatment Fit and Estimated Effect of Content Similarities

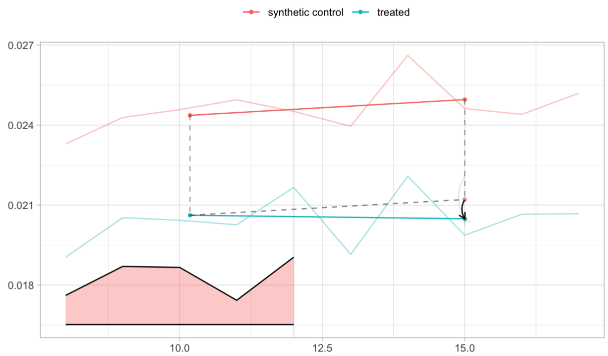
A1.3. Synthetic DiD

Figure A3 presents the average outcomes for the treated units and their corresponding synthetic controls generated by the Synthetic DiD method over the 10 weeks (from week 8 to week 17). The blue line represents the treated units, while the red line denotes the synthetic controls. Overall, the figure demonstrates that the Synthetic DiD estimator achieves a strong pre-treatment fit, indicating that the constructed synthetic control closely replicates the treated units' trends prior to the intervention (before week 13).

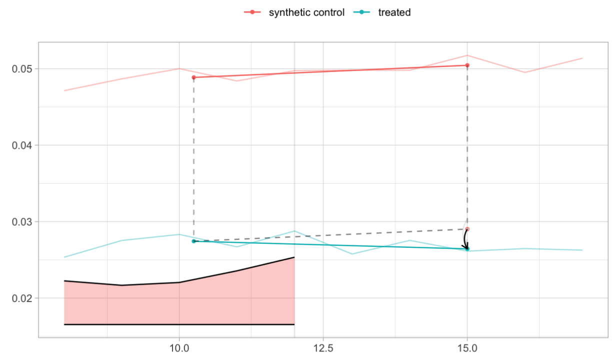
Appendix A2: Robustness Checks and Placebo Test: Content Similarity

A2.1. Robustness Check: Similarity measured by OpenAI Embedding

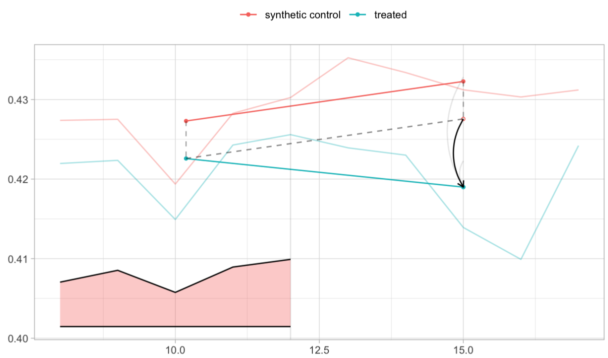
Our main analysis uses methods that are acknowledged in the past literature to capture the four distinct but related aspects of text. As a robustness check, we complement this with an alternative approach using OpenAIâs text embeddings, which are designed to measure the semantic relatedness of text strings. These



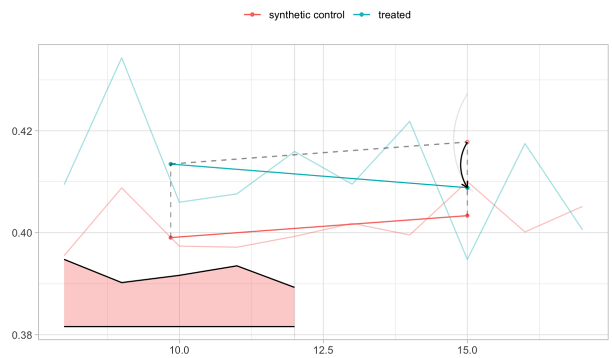
(a) Lexical Similarity



(b) Syntactical Similarity



(c) Semantic Similarity



(d) Language Style Similarity

Figure A3 Synthetic DiD Pre-treatment Fit and Estimated Effect of Content Similarities

embeddings have been widely applied in tasks such as text clustering and diversity measurement, among others, according to OpenAI's documentation.[12]

Specifically, we use the text-embedding-3-small model to generate embeddingsâvectors of floating-point numbersâfor each Instagram post. We then compute the cosine similarity between every pair of posts. Following the same aggregation strategy as in our primary analysis, we average the similarity scores at the businessâweekâcity level, resulting in a panel-like dataset.

We repeat our main regression analysis using this alternative similarity metric and report the results in Table A1. Consistent with our original findings,

A2.2. Easter Holiday

In 2023, Easter fell on 9th April, with Good Friday on 7th April and Easter Monday on 10th April. This period falls within our ban period, which began in week 13, starting 27th March 2023. Easter is a major public holiday celebrated across most European countries, often accompanied by extended weekends, school holidays, and increased travel activity. Many people use this time to go on vacation, both within their own countries and abroad. As a result, one potential explanation for the observed reduction in similarity could be that businesses shift their marketing strategies to target a more transient, tourist-oriented customer base.

[12] `https://platform.openai.com/docs/guides/embeddings`

Table A1 Impacts on Content Similarity Measured by OpenAI Embedding

Method	Similarity based on OpenAI Embedding
Generalized SCM	-0.022*** (0.001)
Observations	12,004
Synthetic DID	-0.011*** (0.001)
Observations	5,780

Note: ***, **, and * represent 1%, 5%, and 10% significance levels, respectively. Standard errors are clustered at the city level and are calculated using bootstrapping.

This is particularly relevant in a city like Milan, which typically experiences a surge in international visitors during festive periods such as Easter.

To address this alternative explanation, we note first that our control group includes other major tourist destinations, such as Madrid and Paris, which are similarly affected by Easter-related travel. Second, we conduct a robustness check by excluding the Easter periodâspecifically, weeks 14 and 15âto ensure that our results are not driven by holiday-specific effects. We report the results in Table A2. The results are consistent. This suggests that our results are not driven by the special dynamics due to Easter holiday.

Table A2 Robustness Check: Excluding Easter Periods

	Lexical (1)	Syntactical (2)	Semantic (3)	Language Style (4)
Generalized SCM	-0.004*** (0.001)	-0.003*** (0.001)	-0.007*** (0.001)	-0.009** (0.004)
Observations	9,653	9,653	9,653	8,336
Synthetic DID	-0.001* (0.000)	-0.003*** (0.001)	-0.009*** (0.001)	-0.009 (0.006)
Observations	5,136	5,136	5,136	3,936

Note: ***, **, and * represent 1%, 5%, and 10% significance levels, respectively. Standard errors are clustered at city level. Bootstrapped Standard errors are used in Generalized SCM and Synthetic DID methods.

A2.3. Italian National Holidays

Week 17, commencing 24 April 2023 in our data, includes two significant days in Italy. 25 April is Liberation Day (Festa della Liberazione), a national holiday commemorating the end of Nazi occupation during World War II. While 24 April is not an official public holiday, it is commonly taken as a "bridge day" (ponte), allowing many Italians to enjoy a long weekend. To ensure our results are not biased by irregular patterns in consumer or business activity associated with this extended holiday period, we exclude week 17 as part of our robustness checks. We find that the results remain largely consistent in direction, indicating a decrease in content homogenization. However, the previously significant decline in language style becomes statistically insignificant. This suggests that the inclusion of Week 17 may have amplified the observed effect due to atypical behavior during the national holiday. Nonetheless, as the results remain consistently significant across the other three measures, we maintain that our overall conclusion remains robust and convincing.

Table A3 Robustness Check: Excluding Italian National Holidays

	Lexical (1)	Syntactical (2)	Semantic (3)	Language Style (4)
Generalized SCM	-0.002***	-0.003**	-0.010***	-0.010
	(0.002)	(0.001)	(0.003)	(0.008)
Observations	10,792	10,792	10,792	9,301
Synthetic DID	-0.001***	-0.003**	-0.010***	-0.008
	(0.000)	(0.001)	(0.001)	(0.008)
Observations	5,346	5,346	5,346	4,158

Note: ***, **, and * represent 1%, 5%, and 10% significance levels, respectively. Standard errors are clustered at city level. Bootstrapped Standard errors are used in Generalized SCM and Synthetic DID methods.

Appendix A3: Robustness Check: Impacts on Consumer Engagement

To account for other factors that may influence consumer engagement, we re-estimate the model using the Generalized Synthetic Control Method (Generalized SCM) by including a business's average word count and posting frequency in a given week as additional covariates in the specification. The results are presented in Table A4. As shown, the estimates of the treatment interaction term remain consistently positive.

Appendix A4: Impacts on Consumer Engagement: Number of Comments

To measure consumer engagement, we primarily use the average number of likes received by each business. We replicate the analysis using the average number of comments as an alternative engagement metric and report the results in Table A5. Although the estimated coefficients are positive, the results are not statistically significant. This may be due in part to the nature of commenting on Instagram, which tends to be a less direct and less frequent form of consumer interactionâas evidenced by the low median number of comments, which is just one per post.

Table A4 Impacts on Consumer Engagement from ChatGPT Ban: with Covariates

	Average Like Count		Log Average Like Count
	(1)	(2)	(3)
	1% Winsorized	5% Winsorized	All
Generalized SCM	4.684	9.393***	0.039***
	(4.640)	(3.318)	(0.009)
Observations	11,191	11,191	11,191

Note: ***, **, and * represent 1%, 5%, and 10% significance levels, respectively. Standard errors are clustered at city level. Bootstrapped Standard errors are used in Generalized SCM.

Table A5 Impacts on Consumer Engagement (Comment) from ChatGPT Ban

	Log Average Comment Count	Average Comment Count
	(1)	(2)
	All	5% Winsorized
Generalized SCM	0.010	-0.042
	(0.018)	(0.078)
Observations	12,004	12,004
Synthetic DiD	0.027	0.011
	(0.032)	(0.127)
Observations	5,780	5,780

Note: ***, **, and * represent 1%, 5%, and 10% significance levels, respectively. Standard errors are clustered at the city level. Bootstrapped standard errors are used in Generalized SCM and Synthetic DiD methods.