Food Neologisms and Word Formation Trends Identified on Social Media Posts Using LLMs for Hashtag Collection.[1]

Malamatenia Panagiotou[1,†], Konstantinos Gkatzionis[1,†] and Efstathios Kaloudis[2,*,†]

[1] *Laboratory of Consumer and Sensory Perception of Food & Drinks, Department of Food Science and Nutrition, University of the Aegean, Metropolite Ioakeim 2, Myrina, 81400, Lemnos, Greece*

[2] *Computer Simulation, Genomics and Data Analysis Laboratory, Department of Food Science and Nutrition, University of the Aegean, Metropolite Ioakeim 2, Myrina, 81400, Lemnos, Greece*

Abstract

Social media platforms can provide researchers with massive authentic data. In consumer studies, social media platforms have been recently used for linguistic and cultural data collection to gain insight into consumers' attitudes, responses, and expectations. The use of modern computational tools, such as Large Language Models and Generative Artificial Intelligence (GenAI) can automate the processes of data analysis, thus saving time and effort. The present study aims at investigating into traditional and local food consumption (case studies: snail dishes, and cheeses of the North-Aegean Sea), and at identifying key concepts and the relevant specific words used by Greek consumers. Instagram posts on the cases under study were collected, and using GenAI applications, sentiment analysis was performed on the posts to identify positively related concepts. The hashtags collected revealed patterns in word formation on social media. GenAI applications were used in an attempt to automate analysis tasks. Hashtag coappearances in posts also revealed key concepts, food trends, as well as conceptual networks. The methodology can be transferred to a) specialized lexicography (e.g. to compile domain-specific word lists), b) linguistic and cultural studies (e.g. to study word formation patterns and conceptual networks in linguistic/cultural studies and in comparative studies), and c) terminology (e.g. to identify neologisms and check term usage over time).

Keywords

social media, large language models, generative artificial intelligence, food, neologisms

1. Introduction

Despite high consumer liking scores in laboratory and in-house tests, product failures in the marketplace are common. Consequently, sensory and consumer studies are increasingly employing alternative methods to gather data from consumers in real-world settings. One such method involves leveraging social media platforms. Online social media networks, content communities, reviews, forums, and blogs offer a vast and rich source of qualitative data, which can be quantitatively analyzed [1]. Current social media platforms utilized for research in language and food studies include Facebook, Instagram, Twitter, and Reddit.

Manual handling is impractical for managing large datasets and cannot serve as a measure of accuracy. However, it can offer insights into the precision of current Natural Language Processing (NLP) tools [1]. Data collection can be efficiently performed using web scraping tools, which

[1] *1st International Workshop on Terminological Neologism Management (NeoTerm 2025), June 18, 2025, Thessaloniki, Greece.*

* Corresponding author.

† These authors contributed equally.

teniapanag@aegean.gr (M.Panagiotou); kgkatzionis@aegean.gr (K.Gkatzionis); stathiskaloudis@aegean.gr (E.Kaloudis)

0000-0001-8179-1163 (M.Panagiotou); 0000-0002-8904-6448 (K.Gkatzionis); 0000-0001-7602-3282 (E.Kaloudis)

rapidly and automatically gather online data from social media websites, extracting it into a well-structured format that is human-readable, machine-readable, easily accessible, and lightweight for storage [2]. Numerous programming languages, including Python, support the reading and processing of collected data.

Sentiment Analysis, a method initially used in politics and marketing, is also used in food and consumer studies. It is the computational study of people's opinions, emotions, and attitudes towards entities, topics etc. [3]. Machine Learning algorithms, which learn how to identify the valence of each word, i.e., the dimensional aspect of emotional experience varying from pleasant to unpleasant [4] within a specific context, are commonly used in sentiment analysis tasks. When every word of the post has been assigned a score, the sum of scores is computed, thus determining whether the post is positive, negative, or neutral (and how much so) [5]. Posts can thus be classified as either positive or negative, or -in some cases- as neutral. Food producers and sellers, and marketing companies want to know how consumers feel about their products and brand. Sentiment analysis can also provide insight into how words "behave" in context and how they correlate with other words.

The objective of this study is to explore Greek consumers' conceptualizations of traditional and local food consumption through a novel methodology incorporating NLP models. The primary goals are: a) to discern how consumers think, feel, and express themselves on social media about traditional and local foods compared to new products with the same basic ingredient, b) to examine how specific aspects of food consumption —such as sensory attributes, geographical characteristics, nutritional value, and environmental concerns— influence consumer choices, and c) to identify pertinent concepts. The developed methodology was tested in case studies involving: a) snails, a traditional Mediterranean food and a sustainable meat alternative, and b) local cheeses from the North-Aegean Sea islands.

2. Methodology

Instagram was the social media platform chosen for this study, because Instagram users interact with companies more often than on other platforms in Greece, and cooking comes second (together with health/ fitness) among the most common interests of Greek Instagram users [6].

For data collection, Apify [7], a web scraping tool was employed to perform automatic multi-word searches on publicly available posts. Posts related to snails and the cheeses under investigation were identified using hashtags (i.e., words or phrases preceded by the symbol # used to classify the accompanying text) in Greek, English, and Greeklish (a non-standardized form of Greek transliterated using the Latin alphabet based on pronunciation or spelling). Social media users frequently use English hashtags to broaden their audience reach, even if English is not their native language. The presence of both English and Greek hashtags suggested that the account was of Greek origin, and thus, these posts were included in the study. A preliminary search was conducted using specialized software to identify Greeklish and misspelled forms of the relevant hashtags, as misspellings are common on social media.

Posts from April 2012 (the release of the Instagram platform) to March 2023 were collected. The posts were consolidated into a single file, and duplicates were removed. The data collected for each post included: post ID, type of post (image, sidecar, video), shortCode (shortened URL of the post), caption, hashtags, number of comments, number of likes, timestamp, and whether the post appeared on a professional account. The data were further refined by removing hashtags in languages other than Greek, Greeklish, or English, nonsensical words, parts of speech not relevant to the study (such as pronouns, articles, and prepositions), names of businesses (e.g., snail farming and selling companies, and restaurants), and the original hashtags used for data collection. In all, 1773 posts on snails containing 1844 hashtags, and 18,026 posts on the cheeses under study containing 54,780 hashtags were collected for further analysis.

ChatGPT [8], an innovative Artificial Intelligence (AI) application (version 3.5), was utilized to automatically apply the food-relatedness criterion, ensuring that only food-related posts were

retained for further analysis, and perform sentiment analysis of posts, considering captions, emojis/emoticons, and hashtags. ChatGPT was selected for its capabilities: a) processing data in multiple languages, b) accessibility and integration into Python applications through an API, and c) being cost-effective. Subsequently, the research team's linguist conducted a manual verification of food-relatedness and sentiment analysis to assess the agreement between human and machine responses and to explore the potential of the machine in replacing manual data management, thereby saving time and effort.
The posts were finally categorized by sentiment (positive, neutral, negative) and by hashtag. The hashtags collected for both case studies were then merged into one file, put into alphabetical order, and duplicates were removed.

3. Results and Discussion

3.1. Food trends identified

Within the hashtags repeated key concepts that pertain to food consumption were identified. These concepts were grouped using English monolingual dictionaries and a name was given to each group as follows:

- Preparation: homemade, handmade, easy, simple
- Origin: traditional food, local food, slow food, Protected Designation of Origin-PDO, agriculture, fusion cuisine, Mediterranean cuisine
- Quality: healthy food/ diet, fresh, bio/ organic, quality, natural/ real food
- Content/ Diet: Mediterranean diet, clean eating, low-carb, (high-) protein, vegetarian, vegan, gluten-free
- Presentation: comfort food, gourmet
- Sensory: tasty food, pleasure first

These key concepts were looked up on Google Trends [9] and Wikipedia page search statistics [10] to collect data on user interest over time (since January 2004 for Google Trends, since July 2015 for Wikipedia), interest by country (Figure 1), related topics, and related queries. Various comparisons can be made using this data to gain insight into food consumer-related phenomena (Figure 2).

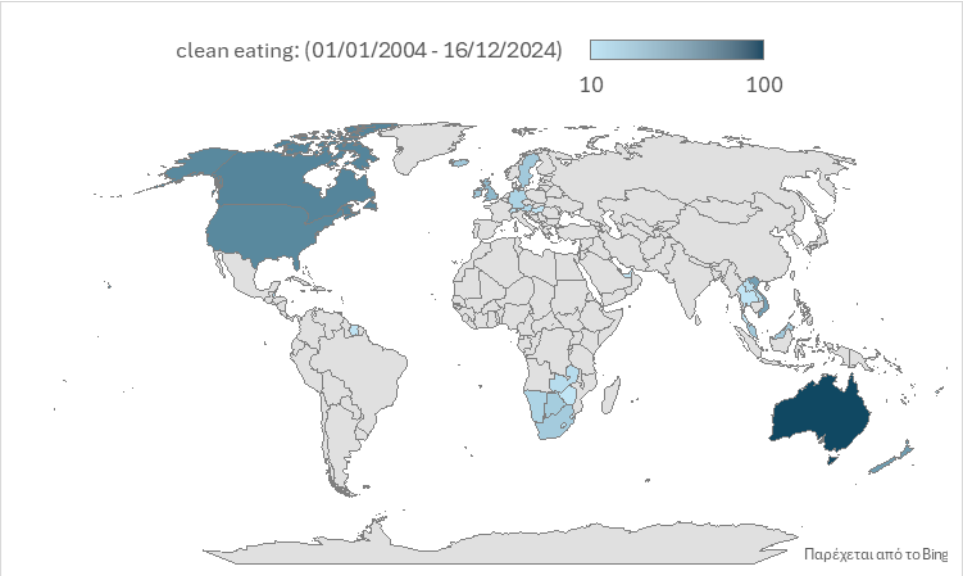

Figure 1: Visualization of interest by country in “slow food” as a consumption-related trend, identified in our case studies, based on Google Trends.

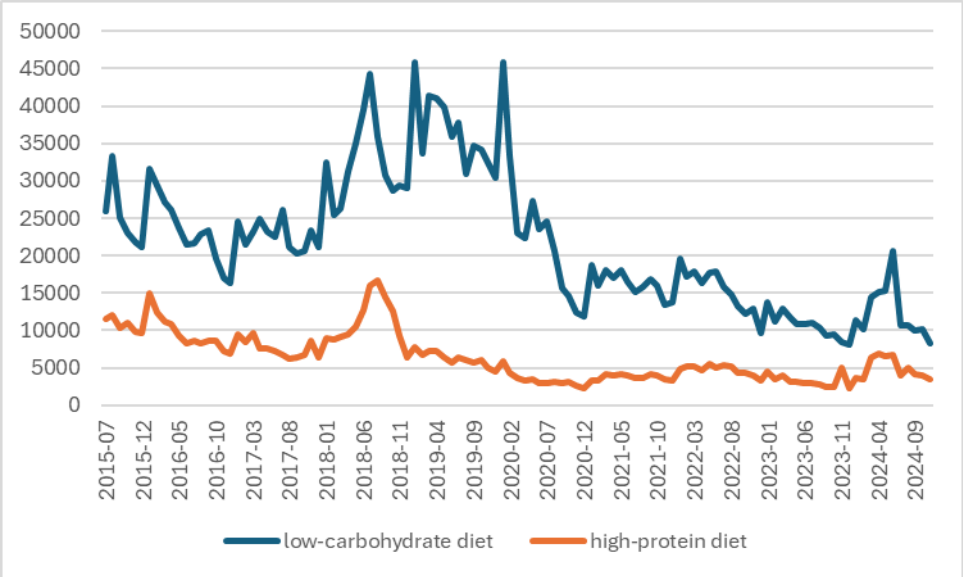

Figure 2: Comparison of two food-related trends (low-carbohydrate and high-protein diets), identified as key concepts in our case studies, based on Wikipedia page search statistics. The graph depicts searches (user interest) over time.

3.2. Word formation trends

Specific prefixes and suffixes involved in word formation were identified, and several word formation patterns were presented in Table 1. These prefixes and suffixes were further investigated using Google Trends to analyze user interest over time (Figure 3), interest by country, related topics, and related queries.

Table 1
Prefixes and suffixes that participate in word formation as identified in food-related hashtags collected on social media.

Prefix	Suffix	Hashtags identified
insta-		instafood, instafoodgram, instagood, instagreece, instafoodie, instamood
	-insta	foodinsta
	-stagram	cheesestagram, greecestagram, foodstagram, cakestagram, breakfastagram, eatstagram
	-gram	foodgram, cheesegram, foodiegram, travelgram, cookingram, pastagram, gastronogram
	-ofinstagram	foodiesofinstagram, chefsofinstagram, foodofinstagram, bakersofinstagram
	-feed	buzzfeed, feedfeed
foodie-		foodie, foodiegram, foodielife, foodiesofinstagram
	-foodie	instafoodie, greekfoodie, athensfoodie, localfoodie
food-		foodaddict, foodstyle,

		foodstylist, foodinstagram, foodofinstagram
	-porn	foodporn, cheeseporn, pastaporn, burgerporn
	-gasm	foodgasm, dessertgasm, beergasm, cretegasm
buzz-		buzzfeed, buzzfeedfood(s), buzzfeast
	-spotting	tastespotting
	-yearsstrong	50yearsstrong
	-yearsyoung	50yearsyoung

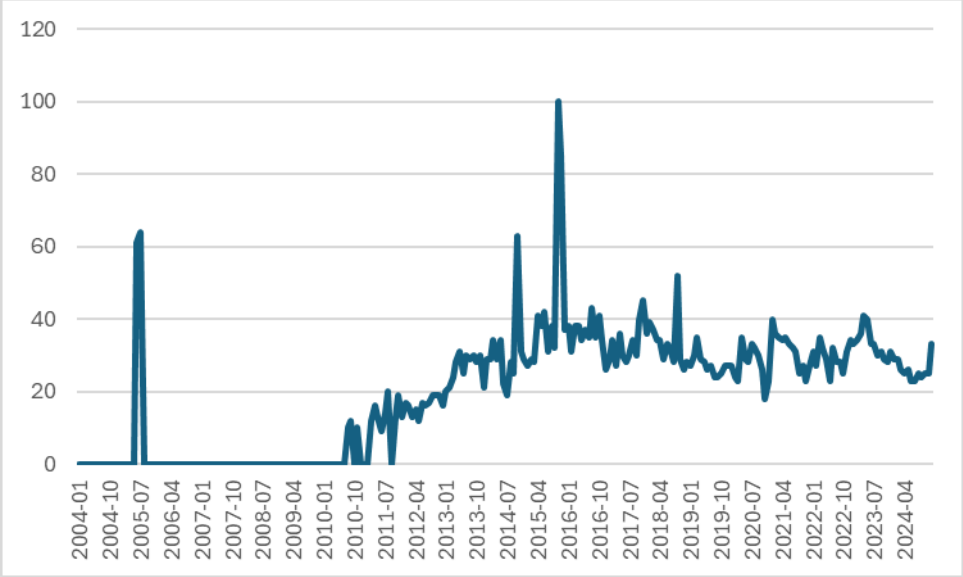

Figure 3: Google searches of the word "foodgasm", identified as a trending hashtag in our case studies, based on worldwide user interest over time data from Google Trends.

ChatGPT was consulted to provide historical context, including their origins, trends and movements leading to their emergence, changes in meaning over time, and their representation in dictionaries. Each case pertains to a social media/ internet language trend. For example, regarding the word "buzz" ChatGPT provided the following response (on December 17th, 2024):

"The word **"buzz"** has an interesting history in both language and culture. While primarily recognized as an onomatopoeic word describing the sound of bees or other buzzing noises, it has evolved over time to take on broader meanings, including excitement, chatter, or activity. Its use as a **prefix** and **suffix** in word formation is relatively modern and tied to its metaphorical extensions. Here's an exploration of its history and usage:

1. Origins of "buzz"

Onomatopoeic Root: The word **"buzz"** originated in Middle English (around the 15th century) as an imitation of the sound made by bees, insects, or other buzzing noises.

Early figurative uses: By the 16th century, **"buzz"** began to acquire metaphorical meanings related to *noise*, *chatter*, or *excitement*. For example, it could refer to a rumor or a bustling environment.

2. "buzz" as a Standalone Word

19th–20th Century Expansions: "buzz" became widely associated with excitement, attention, or energy, often describing a sense of rapid movement or liveliness.

Example: "The room was buzzing with excitement."

In advertising and popular culture, "buzz" was used to describe public interest or excitement around a person, product, or event (e.g., "media buzz" or "generate buzz").

3. "buzz" as a Prefix

Modern Usage: The use of **"buzz-"** as a prefix is a relatively modern development (20th century onwards), and it reflects the word's association with chatter, popularity, and energy.

Common Examples:

Buzzword (first recorded in the 1940s): Refers to a trendy or popular term often used in specific fields or contexts, reflecting the "noise" around a concept.

Buzzkill (popularized in the 1980s–1990s): Describes a person or event that dampens enthusiasm or excitement, implying a disruption of the "buzz."

Buzzfeed (founded in 2006): Combines "buzz" with "feed" to suggest a platform generating and delivering shareable, attention-grabbing content.

Characteristics of Buzz as a Prefix:

Often implies excitement, trendiness, or disruption.

Frequently used in contexts involving social or cultural dynamics (e.g., language, media, or technology).

4. "buzz" as a Suffix

Emergence of "-buzz": The use of **-buzz** as a suffix is less common than its prefixal form, but it is still productive, especially in informal or playful contexts.

Examples:

Earbuzz: Used to describe a persistent sound or auditory phenomenon.

Buzzworthy (late 20th century): Denoting something that generates excitement or public interest.

Characteristics of Buzz as a Suffix:

Suggests association with excitement, noise, or attention.

Often conveys a sense of liveliness or trending appeal.

5. Cultural and Linguistic Influence

Buzz in Marketing and Technology: The metaphorical extension of "buzz" has been embraced by marketing and tech industries, emphasizing excitement and viral appeal (e.g., "buzz marketing" or "Google Buzz," a short-lived social platform).

Buzz in Social Media: In the age of digital communication, "buzz" is often used to describe online chatter, virality, or hype (e.g., "create buzz for a campaign").

6. Summary

The word **"buzz"** originated as an onomatopoeic term describing a sound, but it evolved to capture ideas of excitement, chatter, and activity. As a **prefix**, "buzz-" is commonly used to denote trendiness, popularity, or disruption (e.g., "buzzword," "buzzkill"). As a **suffix**, "-buzz" is less frequent but still conveys associations with attention, liveliness, or auditory phenomena (e.g., "buzzworthy"). Its rise in modern contexts, particularly in media and technology, reflects its ability to evoke the dynamic and fast-moving nature of trends and communication."

ChatGPT's response was cross-checked as valid using other sources (e.g., dictionaries, social media posts). The same procedure was performed for other food-related words from the list.

Copilot [11], another popular AI application at which we have an institutional account, was asked to analyze the list of hashtags for compounding innovations, to test whether AI can be used for morphological analysis of hashtags. The response provided was the following:

"Here are some examples of compound innovations identified:

1. **2018worlds50bestrestaurants**: Combines the year, a global ranking, and a specific category (restaurants).
2. **Caesars_sauce**: Combines a proper noun (Caesar) with a food item (sauce).
3. **handmadephylloDont**: Combines a descriptor (handmade) with a food item (phyllo) and an additional word (Dont).
4. **newyearseveLets**: Combines a holiday (New Year's Eve) with an action (Lets).
5. **OldSpaghettiFactorys**: Combines a descriptor (Old), a food item (Spaghetti), and a place (Factory).
6. **spaghettifactorys**: Similar to the previous example, combining a food item (spaghetti) with a place (factory).
7. **threecheeses**: Combines a number (three) with a food item (cheeses).
8. **todaysrecipes**: Combines a time reference (today's) with a food item (recipes).

These compound innovations often blend multiple words to create new, meaningful hashtags that convey specific ideas or themes."

Copilot's response was quite satisfactory albeit not thorough as it only provided examples and not an exhaustive list. The same task was assigned to ChatGPT which was able to identify the following word formation patterns: a) blends (it identified hashtags that consisted of more than one word as blends, e.g. *mummysrecipe, heresmyfood, aegeancuisine*), b) compounds (it also identified hashtags that consisted of more than one word as compounds, e.g. *greekcheese, healthyeating, instafood, loveforfood*), c) code-switching [hashtags that contained words in different languages, e.g. *μyκουζινα* (English (my) + Greek (kitchen)], d) innovative suffixation or prefixation (hashtags that ended in -lover, -addict, -holic, and-ista, e.g. *cheeselover, cheeseaddict, cheesaholic, cheeseanista*), and

e) use of Greeklish. Further exploring of how to prompt and train AI to successfully identify innovative word formation is required.

3.3. New words identified

Attempts were made to process the list of hashtags using ChatGPT and Copilot, giving the prompt: "I will give you a list of hashtags. I need you to group them into 3 groups: certainly not neologisms, certainly neologisms, and not sure whether neologisms or not. Remember that hashtags are not always one word but can contain two or more words written as one. Some hashtags are in English, some in Greek, and others in Greeklish (meaning Greek transcribed in Latin alphabet based on how they are written or pronounced)". These attempts were not successful. The AI applications used were unable to successfully process hashtags in Greeklish and hashtags that contained more than one word. As a result, the list of hashtags had to be manually checked for neologisms. In the future, this task needs to become automated.

The methodology for analyzing the hashtag data involves constructing a co-occurrence network based on hashtags extracted from posts. The hashtag frequencies are calculated to identify the most used hashtags. A graph is then constructed using NetworkX [12], a Python package for the study of the structure, dynamics, and functions of complex networks, where nodes represent hashtags, and edges signify co-occurrence relationships, weighted by the frequency of co-appearance. Community detection is performed using the Louvain method [13] to identify clusters of related hashtags. Finally, the graph is visualized using Pyvis [14], an interactive network visualizations Python library, enabling exploration of the network structure, centrality measures, and community groupings (Figure 4). This approach provides insights into the relationships and thematic groupings among hashtags. If the file is populated further with hashtags from other food case studies, we will be able to determine if food consumption revolves around certain key concepts and to identify these concepts according to food category, noting any variations.

The sentiment analysis of posts reveals whether the words under study are predominantly used in positive, negative, or neutral contexts within a specific domain. For example, the hashtags containing *buzz-, -porn,* and *-gasm* exclusively appeared in positive posts. This information is valuable when identifying and studying neologisms.

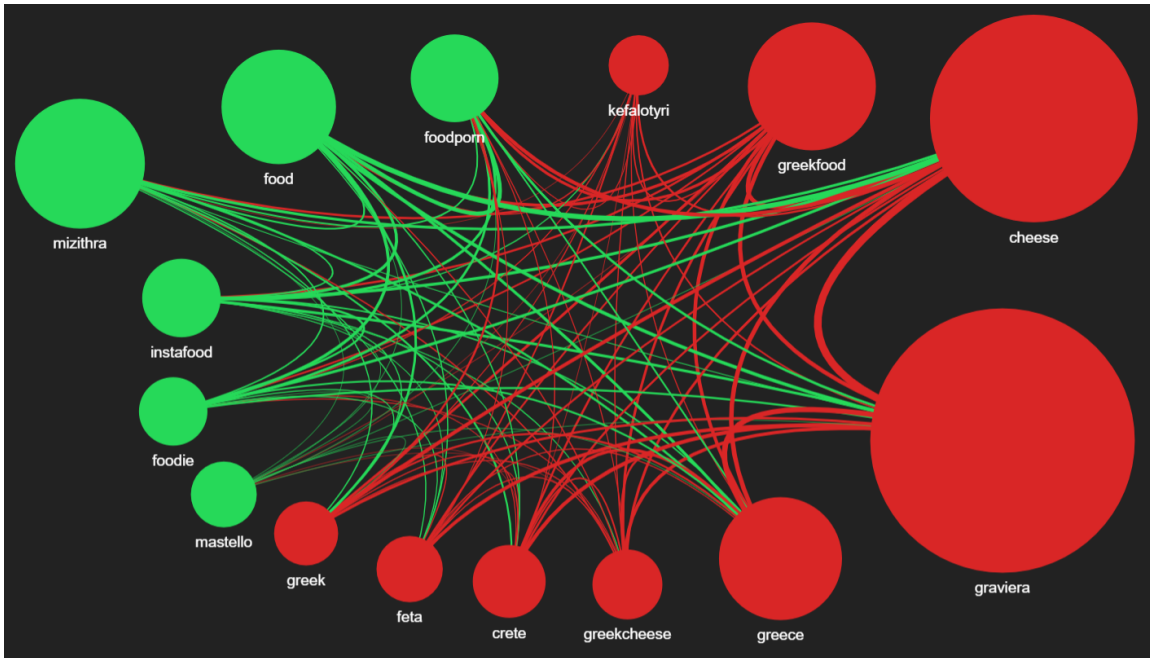

Figure 4: Network analysis of cheese-related hashtags identified in posts collected from Instagram. Nodes represent hashtags, and edges signify co-occurrence relationships, weighted by the frequency of co-appearance.

4. Conclusions

This study highlights the importance of leveraging advanced methodologies, such as NLP models and sentiment analysis, to gain deeper insights into food consumption trends and language use on social media. By analyzing hashtag co-appearances and sentiment, we can identify key concepts and understand the context in which food-related terms are used. Although current AI applications faced challenges in processing certain hashtags, the manual verification process provided valuable data. Future efforts should focus on automating these tasks to enhance efficiency and accuracy. For this purpose, an open access manually annotated dataset for LLM validation is being prepared. Overall, the integration of linguistic analysis and AI tools offers a promising approach to studying consumer behavior and language trends in the digital age.

Declaration on Generative AI

During the preparation of this work, the authors used Chat-GPT-4 in order to: Grammar and spelling check. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] S. C. Hutchings, Y. Dixit, M. Al-Sarayreh, D. D. Torrico, C. E. Realini, S. R. Jaeger, M. M. Reis, A critical review of social media research in sensory-consumer science, Food Res. Int. (2023) 112494. doi:10.1016/j.foodres.2023.112494.

[2] H. Nigam, P. Biswas, Web Scraping: From Tools to Related Legislation and Implementation Using Python, in: Innovative Data Communication Technologies and Application, Springer Singapore, Singapore (2021)149–164. doi:10.1007/978-981-15-9651-3_13.

[3] C. C. Aggarwal, C. Zhai, A Survey of Text Classification Algorithms, in: Mining Text Data, Springer US, Boston, MA (2012) 163–222. doi:10.1007/978-1-4614-3223-4_6.

[4] L. F. Barrett, Solving the Emotion Paradox: Categorization and the Experience of Emotion, Personal. Soc. Psychol. Rev. 10.1 (2006) 20–46. doi:10.1207/s15327957pspr1001_2.

[5] D. Tao, P. Yang, H. Feng, Utilization of text mining as a big data analysis tool for food science and nutrition, Compr. Rev. Food Sci. Food Saf. 19.2 (2020) 875–894. doi:10.1111/1541-4337.12540.

[6] J. Gewiese, S. Rau, Instagram Users in Greece, Statista, 2023. URL: https://www.statista.com/study/141743/instagram-users-in-greece/.

[7] Apify. Full-stack web scraping and data extraction platform. URL: https://apify.com/store

[8] OpenAI. (2024). ChatGPT (Version 4.0) [Large Language Model]. URL : https://chat.openai.com/

[9] Google Trends. (2024) URL: https://trends.google.com

[10] Wikimedia Foundation (2024) Wikipedia pageviews analysis. URL: https://pageviews.wmcloud.org/

[11] Microsoft (2024) *Copilot* [Large Language Model]. Microsoft. URL: https://www.microsoft.com/en-us/copilot

[11] A.A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using NetworkX", in Proceedings of the 7th Python in Science Conference (SciPy2008), G. Varoquaux, T. Vaught, and J. Millman (Eds), (Pasadena, CA USA), pp. 11–15, Aug 2008.

[12] V.D. Blondel, J. L. Guillaume, R. Lambiotte, E. Lefebvre. Fast unfolding of communities in large networks. J. Stat. Mech 10008, 1-12(2008). doi:10.1088/1742-5468/2008/10/P10008

[13] https://pyvis.readthedocs.io