Methods in Marketing: Leveraging Artificial Intelligence for Academic Information Dissemination

Maximilian Witte
University of Hamburg
Keno Tetzlaff
University of Hamburg
Martin Reisenbichler
University of Hamburg
Mark Heitmann
University of Hamburg

Cite as:
Witte Maximilian, Tetzlaff Keno, Reisenbichler Martin, Heitmann Mark (2024), Methods in Marketing: Leveraging Artificial Intelligence for Academic Information Dissemination. *Proceedings of the European Marketing Academy*, 52nd, (119835)

Paper from the 53rd Annual EMAC Conference, Bucharest, Romania, May 28-31, 2024

Methods in Marketing: Leveraging Artificial Intelligence for Academic Information Dissemination

Abstract

In the dynamically evolving field of marketing research, it is imperative that scholars follow the substantive and methodological developments. However, in the context of thousands of newly published articles each year, increasing diversity in methods, and further acceleration through the advent of artificial intelligence (AI), keeping track becomes a daunting task for marketing researchers. Some classical vehicles for information dissemination exist (e.g., editorials, benchmarks), but are typically not designed to cover developments in the entire marketing area and rely on error-prone manual work. We propose an AI-based approach for comprehensive and structured analysis of the entire marketing literature from 2010 onward. We demonstrate the increasing complexity in marketing (i.e., increase in both articles and methods per article), identify recent developments in the field, and strive to provide more structured access to article information.

Keywords: *text analysis, research methods, marketing domains*
Track: *Methods, Modelling & Marketing Analytics*

1 Introduction

In a landscape inundated with the publication of thousands of marketing articles per year (Bornmann and Mutz 2015), and with rapid transformations driven by emerging trends such as generative AI, big data, and machine learning that are gaining momentum (Kumar 2018), it becomes challenging for researchers to keep track of their dynamic field. However, a comprehensive understanding of their research domain is essential for scholars. Profound familiarity with previous research empowers academics to choose topics that matter and articulate their thoughts with precision, facilitating their active participation in academic discourse.

That is why the extraction and retrieval of information from previous research is a key to academic success, as exemplified by the many review articles that attempt to provide an overview and guidance on issues of importance within the academic marketing sphere (e.g., Grewal et al. 2020; Donthu et al. 2021; Shankar and Parsana 2022). However, while such means are helpful, extracting and providing information to researchers remains a challenging task for several reasons.

For example, the increased availability of novel data sources requires researchers to adapt new methods (Grewal et al. 2021). Identifying a suitable starting point and relevant resources might not be easy for some researchers (Berger et al. 2020) and, as Moorman et al. (2019) note, internal forces limit marketing researchers in the adoption of novel methods and stick to established topics, methods, and like-minded people instead.

Prior literature coined this exploration-exploitation trade-off (Sudhir 2016): exploration of novel topics and methods that oppose the exploitation and elaboration of known themes. While exploitation is a valid way to go, exploration is a prime route to follow when aiming to generate novel insights. However, current research often relies on repeating established and well-known methods and procedures (Inman et al. 2018), sometimes rendering the method obsolete by the time the article is published (Toubia 2022). For example, we observe these tendencies in the rather novel domain of text analysis. Transformers are objectively superior for sentiment classification for most applications. However, many recent marketing articles rely on dictionaries such as LIWC for sentiment classification (e.g., Woolley and Sharif 2021; Becker et al. 2022; Peng et al. 2022) and reference usage in a few older hallmark articles such as Berger and Milkman (2012) or Ludwig et al. (2013) that are up to 10 years old. Similarly, a latent moderation analysis Pieters et al. (2022) revealed that 95% of articles use the same method, although it is not optimal in most situations.

Collectively, this underscores the essential requirement for scholars to have a profound understanding of their research domain, remain attuned to the dynamic trends in market-

ing methodologies, and use algorithms skillfully to access pertinent knowledge. Within the context of this article, we present an innovative AI-based methodology capable of comprehensively analyzing the entirety of the marketing literature. This approach enables the extraction and identification of methodological and topical trends embedded within the literature. Our research holds tangible benefits for both researchers and practitioners. It facilitates easy access to novel methods and applications in marketing, thereby streamlining the process of staying abreast of advances in the field. Furthermore, our methodology contributes to strengthening the case for exploration over exploitation (Sudhir 2016), highlighting the importance of embracing new avenues and possibilities in marketing research and practice.

In the subsequent sections, we survey tools scholars currently rely on when turning to understanding their field including (i) journal editorials, (ii) method tutorial papers, and (iii) empirical method benchmarks. Then, we introduce our novel AI-based method to automatically extract knowledge from the marketing literature published from 2010 onward. Finally, we illustrate how themes and methods in marketing research develop over time, and how the marketing field generally evolves in terms of substantive and method contributions, exemplified on articles entailing LIWC and Topic Modeling applications.

2 Vehicles of Academic Information Dissemination

To provide the reader with some background on how scholars currently access information, we now turn to illustrate typical vehicles of scholarly information dissemination: editorials and commentaries (e.g., Yadav 2010; Moorman et al. 2019; Grewal et al. 2020), tutorial articles (e.g., Podsakoff et al. 2012; Hulland et al. 2018; Berger et al. 2020), and method benchmarks (e.g., Pieters et al. 2022; Shankar and Parsana 2022; Tetzlaff et al. 2023). Although each of these vehicles provides substantial value to a specific case, they are not designed to provide structured information for scholars on developments in the entire marketing field. That is why we propose a novel AI-based approach that allows for large-scale, automated analysis and grants researchers innovative angles on information dissemination.

First, editorials and commentaries analyze the development of the marketing field (e.g., Grewal et al. 2020, for JMR)). These articles are generic, cover broader developments, and attempt to distill future research directions by reflecting on the past (e.g., Moorman et al. 2019, for JM and marketing research at large).

Second, tutorial articles provide researchers with more specific information compared to editorials. For example, Berger et al. (2020) summarize applications of natural language processing (NLP) methods in marketing. Similarly, Hulland et al. (2018) collect best prac-

tices for survey research, while Podsakoff et al. (2012) inform about method bias in social sciences.

Third, with the advent of big data and AI in marketing, method benchmarks are an increasingly important tool for informing researchers about methods. Focused on specific tasks such as latent moderation analysis (Pieters et al. 2022), text analysis (Shankar and Parsana 2022), or image classification (Tetzlaff et al. 2023), they aim to provide guidance for the application of task-specific methods. For example, Pieters et al. (2022) compare six methods for Latent Moderation Analysis and Shankar and Parsana (2022) provide an empirical comparison of the NLP methods commonly used in marketing.

Each of the three tools mentioned has inherent design limitations. Editorials, for example, may lack comprehensive coverage due to topical interests or journal-specific focuses. Tutorials and benchmarks might exhibit bias by concentrating on specific subsections of marketing research, relying on keyword-based searches in publication databases, or unintentionally relying on citation networks (including like-minded articles, while excluding others).In summary, while these approaches remain viable, their inherent limitations prevent them from being optimal for large-scale, objective dissemination of research information.

3 Using AI to Inform Scholars on their Field

As emphasized in the preceding section, conventional avenues for accessing academic information on ongoing research and methodological trends remain valid, although not entirely ideal due to inherent shortcomings. Thus, we propose an innovative AI-based approach towards structured information dissemination from article full-texts. In this section, we delineate the framework of this approach that aims to rectify the limitations inherent in existing information channels.

3.1 Proposing an Automated Information Extraction Framework

Following previous literature on large-scale scientometric analysis (e.g., Grewal et al. 2020; Donthu et al. 2021), we extract rich information from a set of 20,000 marketing articles on the type of contribution, data sources and types, analytical methods, investigated constructs, and key findings. This structured information helps us to inform our analysis and thus enables marketing scholars to understand and explore developments in the field. To do so, we build an automatic extraction pipeline that analyzes the full-text manuscripts of all articles. This pipeline follows a five-step process: (1) define relevant information properties, (2) create datasets for the classification of relevant sentences in the article, (3) fine-tune a

sentence classification model to identify relevant text, (4) extract information using a Large Language Model (LLM), and (5) validate the extracted information using keyword search and manual inspection.

3.2 Extractive AI Model Tuning

Based on a small set of manually identified sentences, we create training data sets and fine-tune a few-shot text classification model (Tunstall et al. 2022) to detect potentially relevant sentences in the full text of the articles. This reduces noise from irrelevant context (e.g., literature review), which increases extraction precision and reduces computational cost (Liu et al. 2023). We use OpenAI's ChatGPT API to extract all information entities from the filtered sentences of the publications. Lastly, to minimize possible hallucinations of the language model, we validate all extracted categorical information through a keyword search in the full text of the publication and manual inspection.

4 Results

Next, we put our proposed AI approach to use and extract exemplary methodological information from thousands of marketing publications to inform scholars on the development of their field. More precisely, our sample contains close to 20,000 publications from journals in marketing research published between 2010 and 2023, published in all journals related to marketing, including all major journal rankings (FT50, UT Dallas, ABCD, Cnrs, EJL, ABS, Den, and Scimago) with a minimum rank of B.

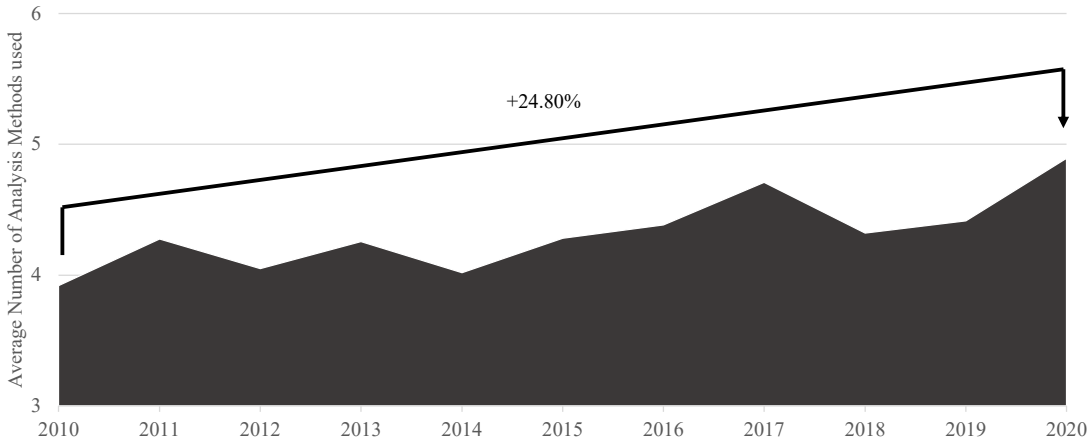

Figure 1: The average number of analysis methods used per publication between 2010 and 2020. The mean number of methods used increased by 24.80% between 2010 and 2020.

Publications were accessed through Web of Science and their full-text manuscripts were manually downloaded by research assistants. We exclude articles without any empirical dataset from our analyses (i.e., editorials, tutorials, and commentaries). Due to computational limits, our preliminary sample at the time of submission includes 6,208 publications between the years 2010 and 2020.

Figure 1 shows the average number of analysis methods used per publication between 2010 and 2020. This number grew from 3.92 in 2010 to 4.89 in 2020, an increase of 24.80% that indicates a greater methodological complexity within individual articles. Probably, this is driven in part by the introduction of novel analysis methods (e.g., text classification) that feed into econometric models and are seldom used standalone. In addition to the increasing number of analysis methods used per publication, the increasing complexity for marketing researchers to navigate the field is highlighted by an increasing number of methods that are used each year in marketing publications (average growth rate of 9.25%).

Table 1 illustrates the developments in the use of data sources in marketing research. Although the use of survey and experimental data remains roughly equal between 2010 and 2020, the percentage of publications using observational data increased by 11.72% to 1,213 in 2020. This reflects the transformation in marketing research that Kumar (2018) commented on in his recent editorial, where the growing popularity of digital platforms allows easier access to observational data and simplifies conducting field experiments in online environments.

Data Source (Year)	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	Pct. Change
Uses experimental data (%)	34.81	33.65	37.94	35.41	32.05	35.26	34.76	34.39	36.94	35.40	34.22	-01.69
Uses observation data (%)	**53.85**	**53.69**	**56.03**	**56.14**	**57.88**	**59.54**	**63.59**	**60.14**	**59.50**	**61.07**	**60.16**	**+11.72**
Uses simulation data (%)	03.85	04.35	02.33	03.02	02.75	02.31	03.62	02.65	03.55	03.52	02.14	-44.42
Uses survey data (%)	66.15	65.78	63.62	68.61	66.30	65.70	67.87	69.84	67.67	67.45	67.51	+02.06

Table 1: The percentage of publications between the year of 2010 and 2020 that use the specific source of data. While the use of most data sources remain roughly stable, the percentage of publications that use any form of observation data increased by 11.72% (bold text).

Lastly, Figure 2 investigates how the use of specific analytical methods in marketing is related to the impact of an article. Traditional methods of analysis (e.g., regression models and analysis of variance) are still the most popular forms of analysis in marketing research.

Not surprisingly, machine learning (ML) methods in general and specific analytical methods from the wider machine learning domain, such as Linguistic Inquiry and Word Count, are increasingly used in marketing research publications (see left panel of Figure 2). The right panel of Figure 2 compares the impact of traditionally popular methods related to ML on an article's number of citations.

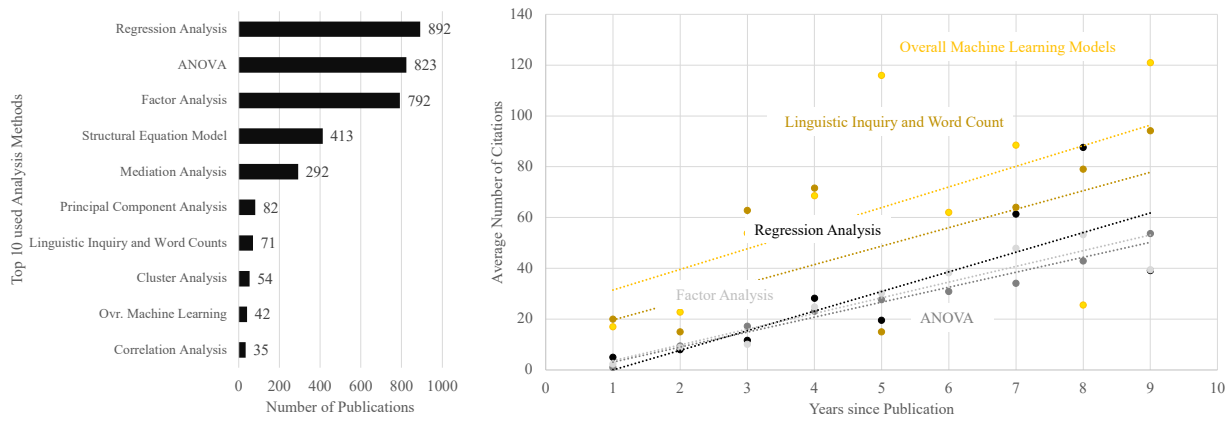

Figure 2: The left panel displays the top 10 most used methods of analysis in marketing research. The graph highlights that only a limited number of research methods are used in most publications. The right panel showcases the impact of method use on citation count. Especially, emerging methods driven by machine learning influence a fast growth in citations.

Publications that exploit traditional and frequently used methods of analysis (i.e., regression analysis, analysis of variance, and factor analysis), accumulate an average of 2.78 citations during the first year of publication. Superior to that, articles that explore novel ML-based methods achieve an average of 18.50 citations during the first year. In addition, a publication will accumulate an average of 44.10 citations during the first 10 years of publication, whereas a publication that leverages ML-based methods will be cited 107.60 times on average (+143.97% compared to traditional analysis methods).

In addition to our analysis, we will extend our analysis with a supplementary, free-of-cost web application (URL BLINDED FOR PEER-REVIEW), to support marketing researchers and practitioners in navigating the vast amount of literature published each year, and provide article search based on the structured information we extracted.

5 Discussion

Informing scholars on the development of themes and methods in their field of research is key to academic success. Although there are classic vehicles for information dissemination, such as editorials, tutorials, and method benchmarks, none of these are optimal in design when it comes to exploring novel directions, methods, and themes in a structured and systematic way.

To the best of our knowledge, we are the first paper to present a novel AI-based approach for the automatic extraction of information from thousands of marketing articles from 2010 onward. With the help of state-of-the-art, few-shot classification, and conversational AI, we

extract six relevant information properties from the full-text of marketing research publications. Analyzing a sample of 6,208 marketing publications between 2010 and 2020, our preliminary results indicate an increase in the diversity of analysis methods that are applied in marketing research over time. Moreover, we find that observational data is increasingly being used. Together, this suggests that marketing scholars put greater emphasis on the use of non-reactive (online) observations and quantitative methods, as they benefit from the realm and availability of big data through digitization and online platforms. Importantly, our results indicate a positive link between the explorative use of emerging methods for analysis (i.e., machine learning) and the impact of an article (i.e., citation count) that surpasses the impact of popular traditional methods (e.g., Structural Equation Modeling).

Our results highlight the applicability of AI for scholarly information extraction and insight generation at scale.These are of great value to researchers who want to gain a more profound understanding of fast-paced developments and trends in the marketing field. Through structured information and easier access to relevant articles, our results help scholars identify interesting topical areas alongside relevant methods for these. We hope this gives rise to a better alignment of the exploration-exploitation trade-off in research, where scholars can leverage past knowledge to explore and apply new themes and methods. With our innovative approach towards academic information dissemination, we resolve many shortcomings of existing attempts (e.g., editorials and benchmarks) and common dangers for the academic business such as falling into citation circles, constantly re-using and citing old methods, themes, and insights.

This analysis includes some limitations that call for future research. First, our data sample includes only publications published after 2010. Although 2010 seemed like a reasonable cut-off point for our data collection process, due to the recent shift of research driven by digitalization and machine learning-based research methods, the use of analytical methods in marketing research is influenced by the field's historical development. Future research could extend the analysis time frame to gather additional information. Lastly, our analysis is limited to only a few information properties extracted from the publications. Many other points of view, despite the focus on methods and data sources used in our analysis, could help researchers navigate the transforming field of marketing research.

References

Becker, Laura, Kristof Coussement, Marion Büttgen and Ellen Weber (2022), "Leadership in innovation communities: The impact of transformational leadership language on member participation," *Journal of Product Innovation Management*, 39 (3), 371–393.

Berger, Jonah, Ashlee Humphreys, Stephan Ludwig, Wendy W. Moe, Oded Netzer and David A. Schweidel (2020), "Uniting the Tribes: Using Text for Marketing Insight," *Journal of Marketing*, 84 (1), 1–25.

Berger, Jonah and Katherine L. Milkman (2012), "What Makes Online Content Viral?" *Journal of Marketing Research*, 49 (2), 192–205.

Bornmann, Lutz and Rüdiger Mutz (2015), "Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references," *Journal of the Association for Information Science and Technology*, 66 (11), 2215–2222.

Donthu, Naveen, Werner Reinartz, Satish Kumar and Debidutta Pattnaik (2021), "A retrospective review of the first 35 years of the International Journal of Research in Marketing," *International Journal of Research in Marketing*, 38 (1), 232–269.

Grewal, Rajdeep, Sachin Gupta and Rebecca Hamilton (2020), "The Journal of Marketing Research Today: Spanning the Domains of Marketing Scholarship," *Journal of Marketing Research*, 57 (6), 985–998.

Grewal, Rajdeep, Sachin Gupta and Rebecca Hamilton (2021), "Marketing Insights from Multimedia Data: Text, Image, Audio, and Video," *Journal of Marketing Research*, 58 (6), 1025–1033.

Hulland, John, Hans Baumgartner and Keith Marion Smith (2018), "Marketing survey research best practices: evidence and recommendations from a review of JAMS articles," *Journal of the Academy of Marketing Science*, 46 (1), 92–108.

Inman, J Jeffrey, Margaret C Campbell, Amna Kirmani and Linda L Price (2018), "Our Vision for the Journal of Consumer Research: It's All about the Consumer," *Journal of Consumer Research*, 44 (5), 955–959.

Kumar, V. (2018), "Transformative Marketing: The Next 20 Years," *Journal of Marketing*, 82 (4), 1–12.

Liu, Nelson F., Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni and Percy Liang (2023), "Lost in the Middle: How Language Models Use Long Contexts," arXiv:2307.03172 [cs].

Ludwig, Stephan, Ko de Ruyter, Mike Friedman, Elisabeth C. Brüggen, Martin Wetzels and Gerard Pfann (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.

Moorman, Christine, Harald J. van Heerde, C. Page Moreau and Robert W. Palmatier (2019), "Challenging the Boundaries of Marketing," *Journal of Marketing*, 83 (5), 1–4.

Peng, Ling, Geng Cui, Ziru Bao and Shuman Liu (2022), "Speaking the same language: the power of words in crowdfunding success and failure," *Marketing Letters*, 33 (2), 311–323.

Pieters, Constant, Rik Pieters and Aurélie Lemmens (2022), "Six Methods for Latent Moderation Analysis in Marketing Research: A Comparison and Guidelines," *Journal of Marketing Research*, 59 (5), 941–962.

Podsakoff, Philip M., Scott B. MacKenzie and Nathan P. Podsakoff (2012), "Sources of Method Bias in Social Science Research and Recommendations on How to Control It," *Annual Review of Psychology*, 63 (1), 539–569.

Shankar, Venkatesh and Sohil Parsana (2022), "An overview and empirical comparison of natural language processing (NLP) models and an introduction to and empirical application of autoencoder models in marketing," *Journal of the Academy of Marketing Science*, 50 (6), 1324–1350.

Sudhir, K. (2016), "Editorial—The Exploration-Exploitation Tradeoff and Efficiency in Knowledge Production," *Marketing Science*, 35 (1), 1–9.

Tetzlaff, Keno, Jochen Hartmann and Mark Heitmann (2023), "The Bigger Picture: A Comprehensive Review and Recommendations for Automated Image Classification in Marketing," *SSRN Electronic Journal*, .

Toubia, Olivier (2022), "Editorial: A New Chapter or a New Page for Marketing Science?" *Marketing Science*, 41 (1), 1–6.

Tunstall, Lewis, Nils Reimers, Unso Eun Seo Jo, Luke Bates, Daniel Korat, Moshe Wasserblat and Oren Pereg (2022), "Efficient Few-Shot Learning Without Prompts," arXiv:2209.11055 [cs].

Woolley, Kaitlin and Marissa A. Sharif (2021), "Incentives Increase Relative Positivity of Review Content and Enjoyment of Review Writing," *Journal of Marketing Research*, 58 (3), 539–558.

Yadav, Manjit S. (2010), "The Decline of Conceptual Articles and Implications for Knowledge Development," *Journal of Marketing*, 74 (1), 1–19.