LUISS

Department of Business and Management

Ph.D in Management

Cycle XXXVII

The ABC in Marketing Technologies: Apps, Brand Analytics and Cryptocurrencies

Prof. Villarroel Ordenes
Supervisor

Prof. Lefkeli
Co-Supervisor

Konstantin Pikal
Candidate

Academic Year 2024/2025

ABSTRACT

This dissertation explores in three essays how digital technologies can be used to apply, produce, and advance marketing science. The first essay examines research-driven apps, which are interactive tools that translate academic insights into accessible, usable formats. Through a literature review and two experiments it shows that these tools increase how relevant and interesting marketing research appears to managers and students. The author then develops a hands-on tutorial with two sample apps. The second essay looks at how language on social media shapes consumer behavior in cryptocurrency. Using two large social media datasets and two experiments, it finds that greedy language (e.g., "make money with crypto") increases perceived profitability and drives consumers to invest. The third essay focuses on improving dictionary-based brand analytics using large language models. It develops a new method to generate context-specific dictionaries for text analysis, validating them using agentic AI and human annotation. It then applies them to 245,000 tweets about four luxury brands. The findings show that generative AI can improve brand analytics while reducing human involvement. Together, these essays demonstrate how digital technologies open up new ways to make marketing science more applicable, insightful, and scalable. The goal of this work is to encourage researchers to embrace technology to increase the real-world impact of their work and make marketing science more useful for managers and for society at large.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION **5**

CHAPTER 2: INTERACTIVE RESEARCH-DRIVEN APPS IN THE MARKTETING SCIENCE VALUE CHAIN **9**

2.1 INTRODUCTION 10
2.2 CONCEPTUAL BACKGROUND 11
2.2.1 The Marketing Science Value Chain *11*
2.2.2 What are Research-Driven Apps? *13*
2.2.3 Current State and Potential of Research-Driven Apps in Marketing Research *14*
2.2.4 The Impact of Research-Driven Apps on (Future) Marketing Managers *18*
2.3 TUTORIAL 21
2.3.1 Step 1: Specify the Contribution, the App Type and identify the Backend *23*
2.3.2 Step 2: Decide on Parameters, Input and Output *25*
2.3.3 Step 3: Design Graphical User Interface (GUI) *26*
2.3.4 Step 4: Chose the Implementation Tool *30*
2.3.5 Step 5: Build the Frontend *31*
2.3.6 Step 6: Connect Back- and Frontend *32*
2.3.7 Step 7: Deploy the App *35*
2.4 CONCLUSION 37

CHAPTER 3: THE MORE, THE MERRIER: GREED LANGUAGE IN SOCIAL MEDIA AFFECTS CRYPTOCURRENCY ADOPTION **40**

3.1 INTRODUCTION 41
3.2 CONCEPTUAL BACKGROUND 43
3.2.1 Cryptocurrency as an Investment Method *43*
3.2.2 Social Influences, Language Use, and Cryptocurrency Adoption *45*
3.2.3 Greed and Consumer Financial Decision Making *46*
3.2.4 Perceived Profitability of Cryptocurrency *47*
3.3 OVERVIEW OF STUDIES 48
3.3.1 Study 1A and 1B: The Effect of Greedy Language on Social Media on New Cryptocurrency Addresses Created *48*
3.3.2 Study 2: Manipulating Greedy Language *63*
3.3.3 Study 3: Mediation *67*
3.4 GENERAL DISCUSSION 72
3.4.1 Theoretical Contributions *72*
3.4.2 Managerial and Policy Implications *73*
3.4.3 Future Research and Limitations *74*

CHAPTER 4: ADVANCING SOCIAL MEDIA BRAND ANALYTICS WITH GENERATIVE AI **76**

4.1 INTRODUCTION 77
4.2 RELATED LITERATURE 79
4.2.1 Measuring consumer mindset metrics using Machine Learning Techniques *79*
4.2.2 Using LLMs to create dictionaries for text-analysis *84*
4.3 CURRENT RESEARCH 86
4.3.1 Data and Setting *86*
4.3.2 Prompt Strategy and Scenarios *87*
4.3.3 Implementation *88*
4.3.4 Brand Reputation Tracker Integration *92*
4.3.5 Empirical Application *96*
4.3.6 Exploring the dynamics of luxury brand drivers *96*
4.4 CONCLUSION 98
4.4.1 Managerial and Theoretical Contributions *98*
4.4.2 Limitations and Future Research *99*

CHAPTER 5: CONCLUSION **100**

CHAPTER 6: REFERENCES **102**

APPENDIX A **116**

APPENDIX B **149**

CHAPTER 1: INTRODUCTION

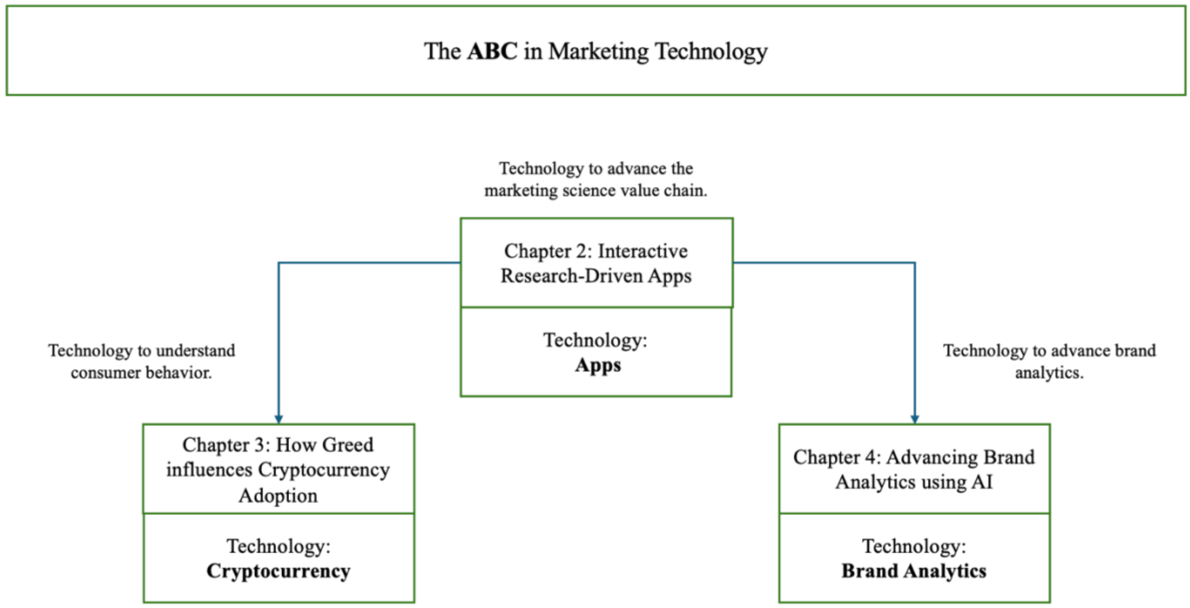
Over the past two decades, digital technologies have profoundly changed how consumers and researchers interact with their environments (Brynjolfsson and McAfee 2014). In marketing, these technologies have changed how academics conduct research and process data, how consumers make decisions, and how firms evaluate their marketing strategies.

This thesis asks: How can emerging digital technologies be harnessed to produce, disseminate, and apply marketing knowledge? The advancements in computing power and programming languages have put software development at the fingertips of everyday users, but researchers are still slow to adapt them for science communication. Interactive apps, a wide-spread digital touchpoint among consumers, are a unique chance for researchers to share their knowledge beyond the borders of academia.

Web scraping and text mining technologies offer unique research opportunities by enabling the real-time analysis of vast volumes of unstructured data (Berger et al. 2020; Boegershausen et al. 2022). In this way, technology can be used to understand consumers' behavior in depth (Babić Rosario et al. 2016; Villarroel Ordenes et al. 2017). Social media, where consumers often share what they think and feel (Hewett et al. 2016), can therefore be used to gain deeper insights into what drives consumers' behavior. Paired with blockchain technology, which makes transactions easily observed by everyone, technology enables researchers to connect what is said on social media to actual consumer behavior (Peres et al. 2023).

Beyond apps and social media, there is another technology that has set out to change the way that academics and practitioners operate: Artificial intelligence (Peres et al. 2023). Generative artificial intelligence will continue to change the way that marketing data is analyzed (Cillo and Rubera 2024), marketing content is created (Hartmann, Exner, and Domdey 2025; Peres et al. 2023; Reisenbichler et al. 2022), and marketing research is conducted.

Figure 1. *Overview of Dissertation Essays*



This thesis explores how digital technologies can be used in and for marketing science (See Figure 1). First, we explore how researchers can use technology interactive app to transmit knowledge through their findings (Dew, Ansari, and Toubia 2022; Farace et al. 2025), enabling managers to use marketing science to solve real-world problems. Next, we investigate, using marketing technology, how consumers speak on social media and why this affects how consumers invest in cryptocurrencies (Friederich et al. 2024; Hakkarainen and Colicev 2023). Last, we examine how technology, in the form of generative artificial intelligence can be used to improve existing marketing analytics methodologies.

More specifically, in my first essay of this thesis we explore the phenomenon of research-driven apps. Research-driven apps are "online interactive tools that provide a deeper understanding of the usability of the research contribution" (Chintagunta et al. 2022). We hypothesize that they can render marketing science more relevant and interesting for managers and researchers. I do so in two ways: First, by reviewing current literature for existing research-driven apps and assessing their impact experimentally, and second, by creating a thorough tutorial on how such research-driven apps can be developed, along with two example applications based on a current marketing science article (Rust et al. 2021) I show how a key technology, research-driven apps, can transform knowledge into application and how researchers can exploit this to technology to propel their own research forward.

In the second essay of my thesis, we explore how social media affects the adoption of cryptocurrencies. I investigate how greed, defined as the dissatisfaction of never having enough and the desire of always having more, is driving consumer behavior in the cryptocurrency space. Using a multi-method approach, combining field studies and experiments, we find that the occurrence of greedy language on prolific social media networks, such as X and Reddit, leads to an increased level in the adoption of cryptocurrencies. we find that a one-unit increase in greedy language ("get more money", "be greedy") is associated with a 5.0% daily (in X) to 9.4% weekly (in Reddit) in new cryptocurrency addresses. Further investigation in the lab shows that this effect is causal and that it is driven by the perceived anticipated profitability. In other words, greedy language is associated with profitable opportunities and therefore drives the adoption of cryptocurrencies.

In the third essay of my thesis, we show how an emerging digital technology, generative artificial intelligence can be used to improve on existing brand analytics methodologies. We use prompt engineering on a variety of large language models to create context-specific dictionaries for automated text analysis. For this, we extend an extending brand analytics methodology, brand reputation (Rust et al. 2021). we first develop an algorithm for dictionary creation and then validate the accuracy and validity of those dictionaries on social media conversations of four luxury brands, using automated text analysis and LLM-supported sentiment analysis. My results show how LLMs can be used to reduce cost and effort in developing custom brand analytics.

This thesis makes three overarching contributions. The first essay contributes to the literature of tutorials in marketing (Boegershausen et al. 2022; Guyt, Datta, and Boegershausen 2024; Villarroel Ordenes et al. 2025), as well as literature on managerial relevance (Lilien 2011; Schauerte et al. 2023; Shapiro and Kirkman 2018). The second essay contributes to research on the role of social media on consumer spending (Chevalier and Mayzlin 2006; Trusov, Bucklin, and Pauwels 2009; Kumar et al. 2016; Liadeli, Sotgiu, and Verlegh 2022), blockchain technologies (Hakkarainen and Colicev 2023; Peres et al. 2023), greed (Hoyer, Zeelenberg, and Breugelmans 2022; Mercadante, Tracy, and Götz 2023; Seuntjens, Zeelenberg, Breugelmans, et al. 2015; Zeelenberg and Breugelmans 2022) and financial decision making (Greenberg and

Hershfield 2019). The third essay contributes to literature on generative artificial intelligence in marketing (Hartmann, Exner, and Domdey 2025; Reisenbichler et al. 2022), agentic artificial intelligence (Arora, Chakraborty, and Nishimura 2025) and brand analytics (Rust et al. 2021; Zhang and Moe 2021).

CHAPTER 2: INTERACTIVE RESEARCH-DRIVEN APPS IN THE MARKTETING SCIENCE VALUE CHAIN

ABSTRACT

Every year, approximately 2700 articles are published in marketing journals, with the primary aim of contributing to both scientific knowledge and managerial practice. Despite this, engagement with science by managers remains low. Research-driven apps, which present scientific findings in an interactive and accessible format, are a viable solution because managers might prefer easy-to-apply tools over research articles. But it is not explored how academics can create research-driven apps and whether they render marketing knowledge more interesting to managers. This research investigates whether research-driven apps enhance perceptions of relevance and interest among marketing managers and students. Through extensive review of the literature and two experiments, this study provides empirical evidence supporting the effectiveness of research-driven apps in increasing managerial relevance in research. To stimulate future development of research-driven apps, the authors create a seven-step tutorial along with two online-accessible apps.

2.1 INTRODUCTION

There is widespread agreement that a research–practice gap exists in marketing sciences (Kohli and Haenlein 2021). Moorman et al. 2019; Roberts et al. 2014). According to Web of Science, around 2,700 articles[1] are published in marketing journals every year. Most of them, undoubtedly, aim to create relevant research that influences managerial practice and contributes to overall science. But reaching a wider audience remains challenging (Stäbler and Haenlein 2025; Haenlein and Jack 2024). There are two main reasons cited for the research–practice gap: The *knowledge production problem* reflects the failure to generate research that is managerially relevant, and the *knowledge dissemination problem* refers to the inaccessibility of academic research to practitioners. While Schauerte et al. (2023) and van Heerde et al. (2021), among others, address the knowledge production problem with guidelines for researchers, we focus on the knowledge dissemination problem in this article.

While transfer journals (e.g., Harvard Business Review) or consultants may alleviate the knowledge dissemination problem by acting as interpreters that translate academic insights (Roberts et al. 2014), we believe that the academic community and the individual research team can do a better job in disseminating the knowledge inherit in their publications. More specifically, one promising solution to bridge the research–practice gap with knowledge dissemination are research-driven apps, defined as "online interactive tools that provide a deeper understanding of the usability of the research contribution" (Chintagunta et al. 2022). That is because the interactivity of research-driven apps not only triggers active learning (Moreno and Mayer 2007) but might also increases managerial acceptance (Mayer et al. 2020).

For example, Dew, Ansara, and Toubia (2022) analyze a large set of brand logos (N = 706) with strong brand identity. Their analysis decomposes logos with the help of a deep learning algorithm into various visual elements such as colors, fonts, and number of corners, and then matches these insights with data on how brand personalities and industries are perceived. Their results are extremely rich but complex to convey in a traditional format. To make them more

[1] Average number of publications per year for the timeframe of 2014-2024 in 47 marketing journals in the Social Sciences Citation index (full table in Appendix).

appealing and easier to understand, they developed a research-driven app to convey their findings in an interactive way, available at https://dr19.shinyapps.io/explore_logo_data/.

However, to date marketing researchers lack insights on the potential effects of research-driven apps, and on how to develop a research-driven app to increase knowledge dissemination. To address these open issues, we review the last five years of published marketing research in the field's leading journals and find that while at least 17 research articles in top marketing journals have developed a research-driven app, there is a large potential for additional research-driven apps. Namely, we uncovered that around 30% of recent top marketing publications has the potential of adding a research-driven app. We next empirically test whether adding a research-driven app to a published article increases perceived relevance among (future) marketing managers. We then provide a tutorial with concrete steps on how to create research-driven apps with little or no coding experience, encompassing: choosing an app typology and identifying a backend, deciding on in- and output as well as parameters, designing the graphical user interface, choosing an implementation tool, building the frontend, connecting back- and frontend and finally deploying the app. We showcase our approach by developing an interactive research-driven app for the real-time brand reputation tracking approach from Rust et al. (2021). We chose this article because of its exemplary value as a theory-driven, time-relevant approach where its direct integration and implementation can help both other researchers and managers.

2.2 CONCEPTUAL BACKGROUND

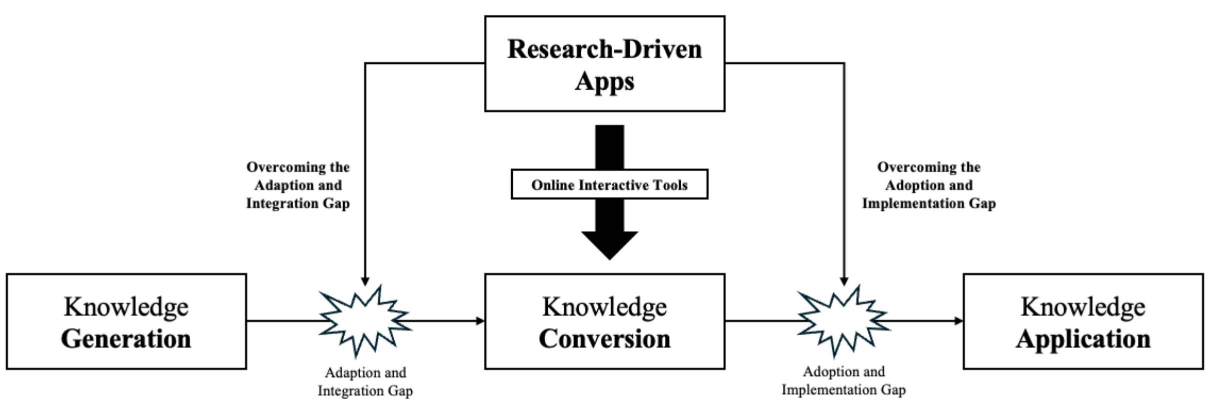
2.2.1 The Marketing Science Value Chain

The marketing science value chain (Roberts et al. 2014) posits that the diffusion of marketing knowledge to practice is driven by three consecutive steps: the knowledge generation by academics in terms of articles, the knowledge conversion by marketing intermediaries in terms of tools, and the knowledge by marketing practitioners in terms of decisions. In this knowledge diffusion process, two potential gaps emerge, as indicated in Figure 1. First, the adaption and integration gap, hindering knowledge to be converted into tools by marketing intermediaries, such as market research agencies, marketing and strategy consultancies, specialist niche marketing consulting firms, or the marketing science division of a marketing organization. Second, the adoption and implementation gap, blocking generated marketing knowledge to be applied by marketing managers via practical tools to make marketing decisions.

We believe that research-driven apps can help to close these two gaps. Research-driven apps can facilitate the adaption and integration of marketing knowledge from articles because they enable a concisely presentation of findings and contributions, increasing their understanding and easing their application. Hence, generated marketing knowledge can be more seamlessly converted, and marketing intermediaries as well as other marketing researchers can more efficiently use fellow the generated knowledge. For example, when novel insights—such as the automated identification of passive voice in text—emerge in text analysis, a researcher unfamiliar with the method may find it more accessible to engage with an interactive research-driven app that has already integrated the technique (Sepehri, Mirshafiee, and Markowitz 2023).

Research-driven apps can also help to overcome the adoption and implementation gap because they make articles more interesting and applicable, motivating marketing managers to apply the generated knowledge in practical decisions. For example, when crafting social media posts, they might find it easier to apply research findings on writing effective messages where when the writing is supported through recommendations by an interactive research-driven app (Atalay, Kihal, and Ellsaesser 2023).

Figure 2. *Interactive Research-driven Apps and the Marketing Science Value Chain*



Note: The underlying diffusion process is based on Roberts et al. (2014).

2.2.2 What are Research-Driven Apps?

Research-driven apps fulfill the task to "provide(s) a deeper understanding of the usability of the research contribution" in the form of an interactive tool (Chintagunta et al. 2022). A research-driven app thus serves as a dynamic computational supplement to a published article, thereby adding interactivity to the otherwise static nature of a research publication. The goal and functionality of a research-driven app depends intricately on the scope and goals of the original article, and Chintagunta et al. (2022) classified four types of research-driven apps that are well-suited to marketing research: predictors, optimizers and recommenders, explorers, and converters – as summarized in Table 1.

Table 1. *Typology of Research-Driven App*

Typology	Description	Examples
Predictors	Predictors offer model-based predictions and thereby depict the interplay among various factors affecting a marketing input, process, or outcome.	André and Reinholtz (2024); Laghaie and Otter (2023)
Optimizers and Recommenders	Optimizers offer normative solutions to improve currently suboptimal marketing decisions made by agents, and recommenders offer superior solutions to current problems marketing stakeholders face.	Farace et al. (2025); Warren et al. (2021)
Explorers	Explorers investigate sensitivity of research results to various research design assumptions.	Dew, Ansari, and Toubia (2022); Schoenmueller, Netzer, and Stahl (2022)
Converters	Converters provide accessible marketing insights through the conversion of unstructured input (e.g., text, audio, videos) to structured data, and models developed in the research framework.	Humphreys, Isaac, and Wang (2021); Sepehri, Mirshafiee, and Markowitz (2023)

Note. Typology based on Chintagunta et al. (2022).

Predictors offer model-based predictions and thereby depict the interplay among various factors affecting a marketing input, process, or outcome. Some examples include André and Reinholtz (2024) who help researchers reduce the cost of experimentation by developing an app that assists them in designing more cost-effective but rigorous experiments, predicting the sample sizes that are needed to achieve the desired level of statistical power. Laghaie and Allenby (2023) develop an app that helps to predict the validity of mediation results in the presence of measurement error.

Optimizers offer normative solutions to improve currently suboptimal marketing decisions made by agents, and recommenders offer superior solutions to current problems marketing stakeholders face. Some examples include Farace et al. (2025) who help marketing managers to optimize textual overlays in their social media posts. Similarly, Warren et al. (2021) help researchers to write more clearly, by developing a text-analysis that checks for common writing pitfalls, such as lack of examples and overuse of passive voice.

Explorers investigate sensitivity of research results to various research design assumptions. Some examples include Dew, Ansari, and Toubia who render the interconnections of logo features (e.g. color, font and shape) and brand personalities easily explorable through their research-driven app. Schoenmueller, Netzer, and Stahl (2022), in their app, allow to explore the political affiliation of social media brand followings in the US.

Converters provide accessible marketing insights through the conversion of unstructured input (e.g., text, audio, videos) to structured data, and models developed in the research framework. Some examples include Humphrey, Isaac and Wang (2021) who enable users to convert words or sentences into concreteness scores. Sepehri, Mirshafiee, and Markowitz (2023) develop a research-driven app that identifies passive voice with high accuracy and converts textual data into scores.

2.2.3 Current State and Potential of Research-Driven Apps in Marketing Research

While Chintagunta et al. (2022) emphasizes the integration of research-driven apps into academic articles to enhance understanding, consumption, adoption, and ongoing usage of research findings from academic articles that may be difficult to understand for practitioners

(Warren et al. 2021), to the best of our knowledge we miss a systematic inquiry on what has been done in this regard. To understand the status-quo of research-driven apps, we thus search in articles published from 2020 to 2025 in the leading marketing journals[2] and classify all existing research-driven apps using the typology of Chintagunta et al. (2022). We chose this timeframe, as there were no research-driven apps published before 2020[3]. In order to identify existing research-driven apps we screen the article abstract, the appendix and, when existing, the online supplementary material.

Our search results in 17 existing research-driven apps (see Table 2). The entirety of research-driven apps span a wide range of topics, including branding (Dew, Ansari, and Toubia 2022), scientific writing (Warren et al. 2021) and linguistic phenomena (Luangrath, Xu, and Wang 2023; Sepehri, Mirshafiee, and Markowitz 2023). We find that most articles are of type "converters" and "explorers". We suspect that this is because converter apps might be the least complicated to develop, whereas explorer apps offer the most freedom. Most research-driven apps are in the Journal of Marketing. Notably, no article in the Journal of the Academy of Marketing Science, International Journal of Research in Marketing and Journal of Interactive Marketing were accompanied by a research-driven app to data.

Research-driven apps are undoubtedly an emerging phenomenon in marketing research. Nevertheless, before the backdrop of thousands of marketing articles every year, the number remains relatively small. One reason for this might be that recent articles are "simply not made" for an app, and we therefore observe this small number. The question remains whether other articles might have had the potential of additionally offering an app.

---

[2] Including *Journal of Marketing*, *Journal of Marketing Research*, *Marketing Science*, *Journal of Consumer Research*, *Journal of the Academy of Marketing Science*, *Journal of Consumer Psychology*, *International Journal of Research in Marketing* and *Journal of Interactive Marketing*.
[3] It has been brought to the author's attention that Blanchard, Aloise and Desarbo (2017) had added a research-driven app to their article. However, this app is not hosted anymore.

Table 2. *Overview of Existing Research-Driven Apps*

Article	Journal	Description of App	Link	App Type
André and Reinholtz (2024)	JCR	PRIApp: Design, Simulate and Pre-Register Interim Analysis Design, helping to reduce the number of participants in online experiments.	https://priadconsumerresearch.shinyapps.io/PRIApp/	Predictor
Atalay, Kihal, and Ellsaesser (2023)	JM	Syntactic Surprise Calculator, calculates how syntactically surprising a sentence is, gives recommendations to optimize for optimal (medium) surprising syntax.	https://syntactic-surprise-calculator.com/	Optimizer
Dew, Ansari, and Toubia (2022)	MS	Logo Explorer, explores the connections between log features (e.g. color or font), industry and brand personalities.	https://dr19.shinyapps.io/explore_logo_data/	Explorer
Dyachenko and Allenby (2023)	JCR	Bayesian Analysis of Heterogeneous Mediation (BAHM), to help uncover how mediating effects might vary among participants.	https://bayesianmediationanalysis.shinyapps.io/BAHM/	Predictor
Farace et al. (2025)	JM	Text Overlay App, helps to design multimodal social media posts and offers improvements while also predicting engagement.	https://www.knime.com/visual-flow-data-app	Optimizer
Hartmann, Bergner, and Hildebrand (2023)	JCP	Mindminer, trained on customer reviews, it predicts mind perception of consumers in smart objects.	https://huggingface.co/spaces/j-hartmann/MindMiner	Predictor
Hotz-Behofsits, Wlömert, and Abou Nabout (2025)	JM	Natural affect detection explorer (NADE), converts text into emojis and then emotions.	https://nade.rds.wu.ac.at/	Converter
Hovy, Melumad, and Inman (2021)	JCR	Wordify, helps to identify words that are different among documents, for example between good or bad reviews.	https://wordify.unibocconi.it/	Converter
Humphreys, Isaac, and Wang (2021)	JMR	Construal Score, converts text into concreteness scores based on two dictionaries (MRC and Brysbaert).	https://rebeccawangphd.shinyapps.io/ConcretenessScoringTool/	Converter
Jedidi et al. (2021)	JM	R2M index, scores an article on its relevance and makes the article scores explorable.	https://www.r2mindex.com/	Explorer
Laghaie and Otter (2023)	JMR	BFMediate_APP, improves prediction of mediation in the case of measurement-of-mediation designs.	https://bfmediate.shinyapps.io/bfmediate_app/	Predictor

Matthe, Ringel, and Skiera (2023)	MS	Evo Map, helps to explore longitudinal spatial relationships, specifically product market competition among publicly listed firms.	https://www.evomap.io/main/index.html	Explorer
McShane and Böckenholt (2022)	JCP	MCSM: Multiple Contrast Standardized Meta-Analysis, which converts study results with multiple contrasts and measurements into a meta-analysis.	https://blakemcshane.shinyapps.io/mcsmeta/	Converter
Ringel (2023)	JMR	MapMX, explores relationship between objects, such as firms or brands.	https://mapxp.app/MapXP/	Explorer
Schoenmueller, Netzer, and Stahl (2022)	MS	Social Listening, allows to explore the political affiliation of social media brand followings in the US.	https://www.social-listening.org/	Explorer
Sepehri, Mirshafiee, and Markowitz (2023)	JCP	PassivePy, converts text into scores for passive speech.	https://passivepy.streamlit.app/	Converter
Warren et al. (2021)	JM	Clarity Calculator, analyses text on passive voice, concreteness of examples, and readability of the text and gives suggestions for improvement.	http://writingclaritycalculator.com/	Optimizer

Note. JCR = Journal of Consumer Research, JCP = Journal of Consumer Psychology – JM = Journal of Marketing, JMR = Journal of Marketing Research, MS = Marketing Science.

To assess the potential for app development, we refine our search and use the keywords from the articles that already have an app to create a query, resulting in 78 author keywords (see appendix for details). We focus on the same timeframe and the same journals, and retrieve a total of 720 articles. After excluding those where a research-driven app already existed (N = 16), and articles that were categorized as editorial (N = 8), we are left with a total of 697 articles. We then assess, with the help of a research assistant, the abstract of each article, judging their potential to add a research-driven app onto their publication. We check whether the research is empirical and whether the initial data can be adapted (e.g. in Sepehri, Mirshafiee, and Markowitz 2023). Several indicators are taken into account: First, whether the research article concentrates on a specific context (Stremersch et al. 2023). This might limit the model's capacity to be applied to new data. Next, we assess whether the applied model is replicable and sufficiently described in detail (e.g. through reproduction code). Finally, in the case that there is no model that can be applied to new data, but the results are prone to be visualized, e.g. in a meta-analysis, we might still assess there is potential for an interactive research-driven app.

This way, we annotate the papers based on their viability of adding a predictor, optimizer and recommender, explorer, and converter[4]. We find that around 30% of articles (i.e., 210 articles) have potential to add a research-driven app, showing significant potential for app development in current literature.

2.2.4 The Impact of Research-Driven Apps on (Future) Marketing Managers

To understand whether research-driven apps can make research more relevant, we run two studies with marketing managers and grad students from a large public Dutch university. First, we recruited 146 marketing managers using the research platform Prolific that passed an instruction check (60.3 % female, $M_{age} = 38.77$) for a within-subject experimental design. We exposed all participants to a randomized series of four recent marketing articles that were published in top journals (i.e., Atalay, Kihal, and Ellsaesser 2023; Dew, Ansari, and Toubia 2022; Farace et al. 2025; Humphreys, Isaac, and Wang 2021). We use a 4x2 within subject design with

[4] For a detailed assessment of how we annotated the papers, please see the appendix.

a random order. In the treatment condition, participants view each research article accompanied by an interactive research-driven app. In the control condition, the same article is presented with a chart from the original paper. Each participant is shown a randomized sequence of four articles, with the format (app or chart) randomly assigned per article. We check whether the app manipulation was successful after each stimulus (“To what extent do you think the research paper is interactive?”). In order to measure interestingness and relevance, we used a 3-item scale for interestingness ($\alpha = 0.88$) and a 4-item scale for relevance ($\alpha = 0.90$). We used adapted scales based on the findings from Schauerte et. al 2023, see appendix for details. For our analysis, we average the interestingness and relevance items, using these as dependent variables, with the experimental condition (exposure to a research-driven app vs. a chart) as the independent variable. We expand the observations in such a way that we obtain a long-format dataset with 584 observations. Due to the ordered nature of our dependent, we run an ordinal logistic regression. Our results show that participants exposed to the treatment condition (app) found the presented research articles more interesting and relevant. Due to the ordinal nature of our dependent variables, we analyze our dataset using ordered logistic regressions. We find that the presence of research-driven apps increases both the perception of interestingness (b = .345, SE = .145, $t = 2.374$, $p = .018$) and relevance (b =.316, SE =.144, $t = 2.182$, $p = .029$).

To extend our findings to the classroom, we run a between-subjects experimental design with 36 graduated marketing students at a large public Dutch university. Students were randomly assigned to one of two tutorial groups for the length of the whole course, and in one of the tutorials were introduced to a marketing-related topic (“How logos influence brand perception”) and are given a summary of the respective article from Dew, Ansari, and Toubia (2022), along with a link to the full article in the control group and links to the article and the research-driven app in the treatment group. None of the students in the control group used the research-driven app for the task. After working for 20 minutes on the task, we then ask three questions regarding interestingness, usefulness and relevance (In your opinion, how interesting is it to implement the findings from this research? In your opinion, how useful is it to implement the findings from this research? How relevant do you find this research paper?). A one-way ANOVA reveals that students that were exposed to the research-driven app perceived the research article as more interesting ($M_{App} = 5.32$, SD = 1.42, $M_{Control} = 3.88$, SD = 1.32, $F(1,35) = 9.814$, $p < .01$), useful ($M_{App} = 4.79$, SD = 1.32, $M_{Control} = 3.76$, SD = 1.09, $F(1,35) = 6.379$, $p < .05$) and relevant (M_{App}

= 4.58, SD = 1.43, $M_{Control}$ = 3.71, SD = 1.26, $F(1,35) = 3.741$, $p = .061$) compared to those that were not exposed to the research-driven app.

Taken together, we conclude that (future) marketing managers perceive research articles with research-driven apps as more interesting and more relevant than those without. Taken together, we have found that research-driven apps are an emerging phenomenon with significant potential in current literature. We have also found strong evidence that research-driven apps strengthen interestingness and relevance, in turn being able to increase the diffusion of marketing knowledge. Despite these findings, we believe that the technological barriers to app development remain an important obstacle to the broader integration of research-driven apps into academic publications. Therefore, we now move to a tutorial to assist fellow researchers in the development of a research-driven app.

2.3 TUTORIAL

The objective of this tutorial is to facilitate the development of a research-driven app by marketing researchers. As a hands-on example, we have developed a research-driven app, the "Brand Reputation Tracker". The "Brand Reputation Tracker" is based on Rust et al. (2021) and applies text-mining on social media data to measure brand metrics, i.e. brand reputation. We selected this article for three reasons. First, it was in our list of articles with the potential to become an app. Second, it involved the use of unstructured data (in this case text) which is increasing its application in marketing. Finally, because this article incorporated all the needed information to replicate its work and develop the app. Going forward, we will use this app to showcase our tutorial. The full code and workflow can be found in the Web appendix. Live versions of the apps can be found here: https://research-driven-app.streamlit.app/ for Streamlit, and here www.konstantinpikal.com/research-driven-app for KNIME.

Figure 1*. How to develop an Interactive Research-Driven app*

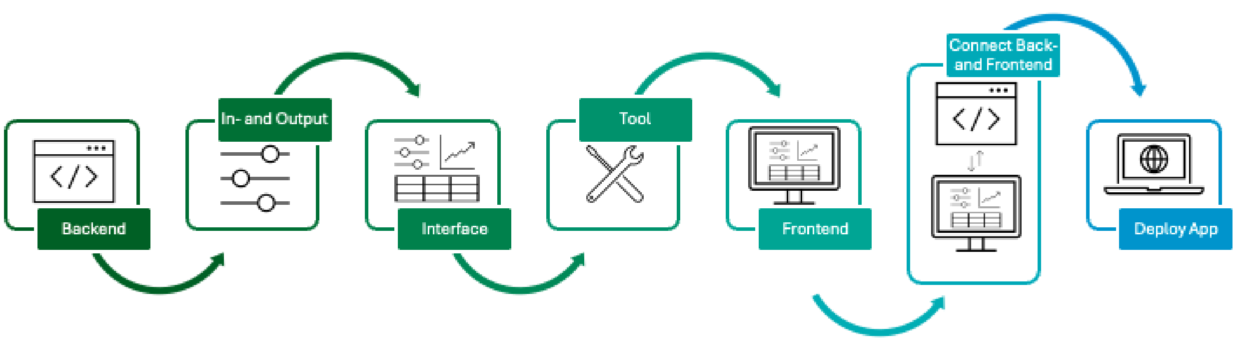


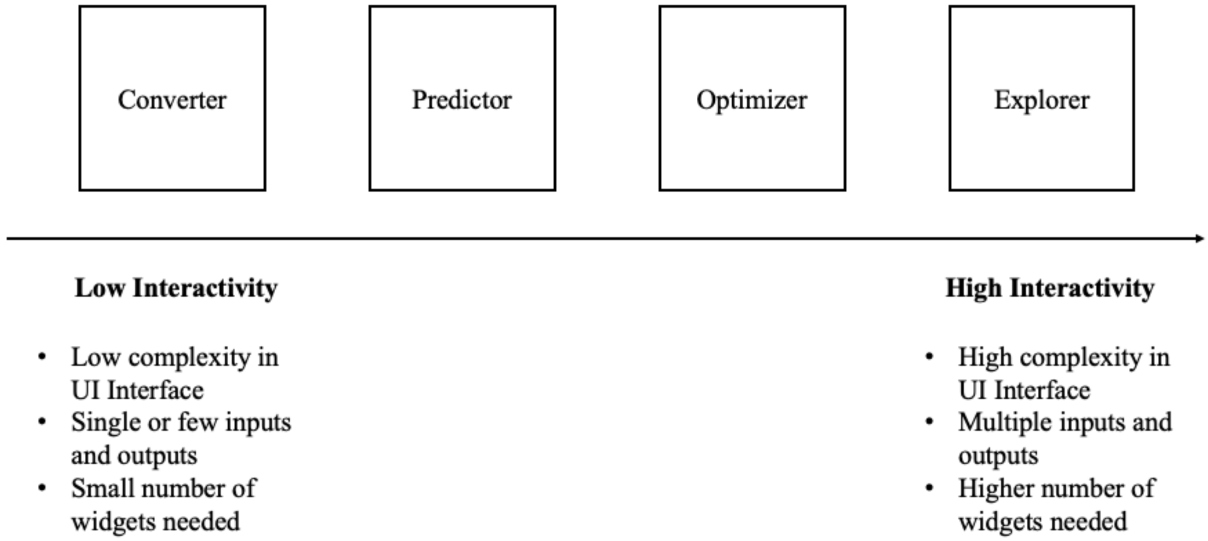
Table 3. *Steps to develop an interactive research-driven app*

Step	Description
Specify the contribution, the app Type and identify the backend	Choice among app types (converter, predictor, optimizer, explorer) depending on the main contribution of the research paper along with the desired degree of interactivity. Researchers need to identify the backend from the empirical analysis.
Decide on parameters, input and output	Depending on app type, all desired inputs and outputs are listed and described. Mandatory and optional settings are defined.
Design graphical user interface (GUI)	The graphical user interface is designed based on widgets (i.e., in the form of interactive buttons) that are chosen depending on the needed inputs, parameters, and outputs.
Chose the implementation tool	Various tools and frameworks exist, the best tool depends on researcher capabilities and scope of the app. Decision is taken on existing backend, researcher preferences and budget.
Build the frontend	Using the widgets from the selected tool, the frontend is built based on simple design rules used in web design, making sure that the most important elements are on top of the screen.
Connect back- and frontend	The backend analysis is triggered through interactive widgets in the frontend. The logic of execution is clarified.
Deploy the app	Budget, expected usage and accessibility are assessed in order to choose the best deployment option.

2.3.1 Step 1: Specify the Contribution, the App Type and identify the Backend

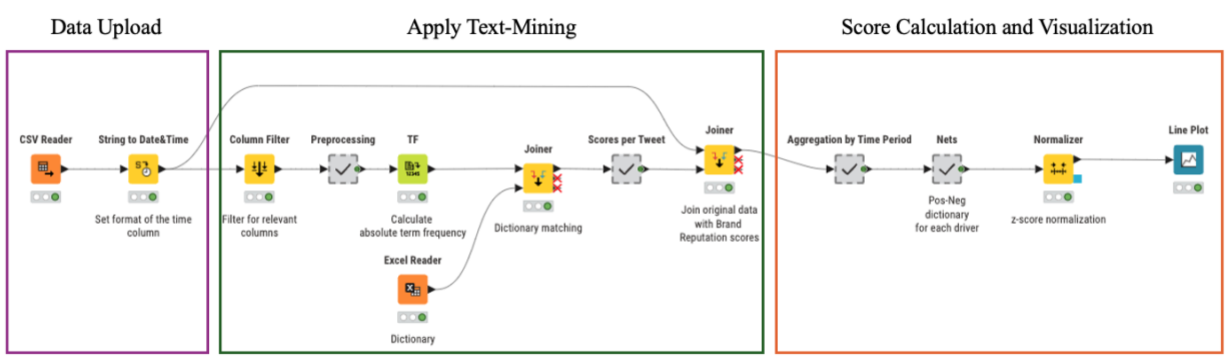
Researchers must first decide on the type of app they want to develop: converter, predictor, optimizer or explorer (See Table 1. for an overview of recent research-driven apps). This decision is generally linked to the contribution of the paper and drives its interactivity. While converter apps are the least interactive, and therefore tend to be easier to program, explorer apps are the most interactive but also complex apps (see Figure 4). Predictor apps tend to be more complex and interactive. They normally have multiple inputs and single output, a prediction, while optimizer apps tend to have multiple outputs (normally a prediction and a recommendation on how to improve the result). Explorer apps are the most complex since they have multiple inputs and outputs that are additionally often visualized for the user. For example, in Sepehri, Mirshafiee and Markowitz (2022) the backend is composed by a set of grammatical rules that scores the level of passive voice in text (making it a converter app). Our example app, the Brand Reputation Tracker converts text data (e.g., tweets) into brand metrics and provides a visualization. We first load the data into the backend. Next, we preprocess the text data by converting the text column into the appropriate "document" format, converting all text to lowercase, stemming words, and creating a bag-of-words representation. We then apply text analysis using the dictionary from Rust et al. (2021) to obtain a score for each social media post. These scores are aggregated according to a specified timeframe (e.g., weekly or monthly), normalized, and visualized. While the strict brand reputation classification would make it a converter app, by visualizing the results and making them available for interactive exploration, it is classified as an explorer app. Going forward, the choice of app desired will determine how complex the development will be. Once the contribution of the app has been specified, the researcher needs to identify the backend (i.e., code) that was involved in obtaining the desired goal. All empirical research articles have an underlying analysis which constitutes the backend of the research-driven app.

Figure 4. *Different app types*



Note. This figure shows the different types of apps along with their interactivity

Figure 3. Example of the backend from the Brand Reputation Tracker App in KNIME

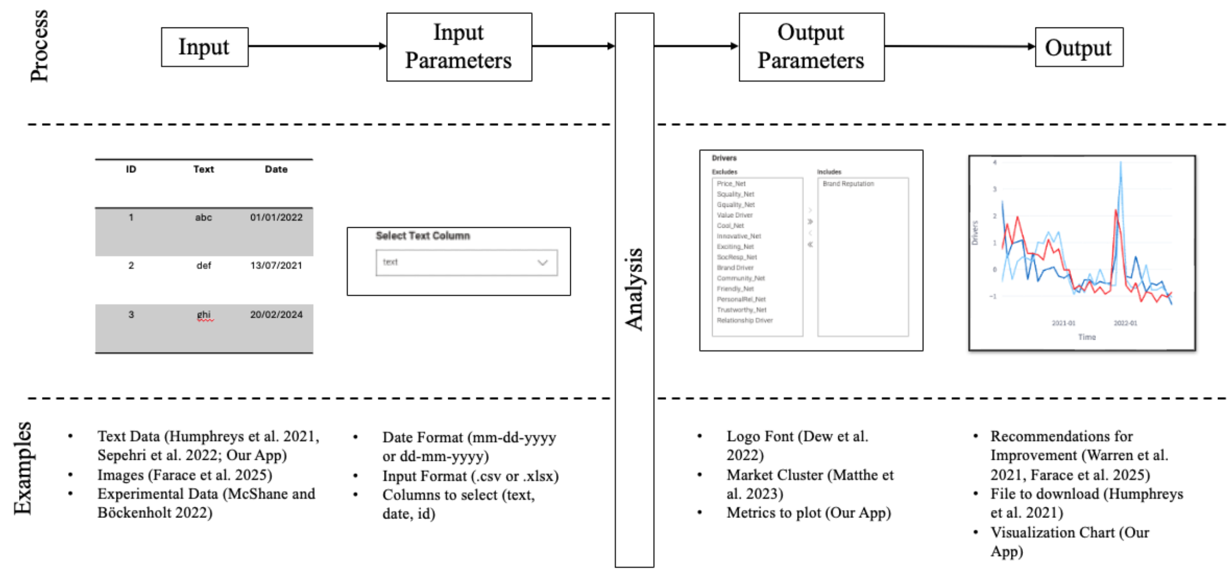


Note. This is the backend for our example app on KNIME. In Python, we organized most of the backend code on one page, please see here.

2.3.2 Step 2: Decide on Parameters, Input and Output

All apps are based on at least three steps. First the input, then a model that is applied and then an output. To specify the way that the analysis is done, we use parameters. In our brand reputation tracker, the input would be the data that we are uploading (social media text), where we need to select the columns that need to be considered for the analysis. For example, to specify the time dimension in our analysis, we need to select the column named "created_at" as the column containing the time at which a social media post was created (see Figure 5). Additionally, we need to have parameters for the text and ID column. Those are our input parameters. If we took BFMediate as a sample, the seed field would be an input parameter that ensure reproducibility of the simulations. When the results are processed, the user might be able to adjust and explore results based on output parameters, for example by changing the aggregation type (e.g., from weekly to monthly). To gain a comprehensive understanding of the app's components, both mandatory and optional inputs should be documented, along with the relevant configuration parameters, intended outputs, and output parameters. The more inputs, outputs and configuration parameters are added, the more interactive an app will become. To battle complexity, it is important to keep the number of inputs to a minimum (Krug 2014), as too many inputs might overwhelm the user. See Table 3 for an excerpt of parameters, inputs and outputs in our brand reputation app.

Figure 5. List of Parameters, Inputs and Outputs for Example App along with Examples in Literature



Note. This is an example of how inputs, outputs and parameters interact with each other on the actual. The outputs and in- and output parameters are based on our example app.

2.3.3 Step 3: Design Graphical User Interface (GUI)

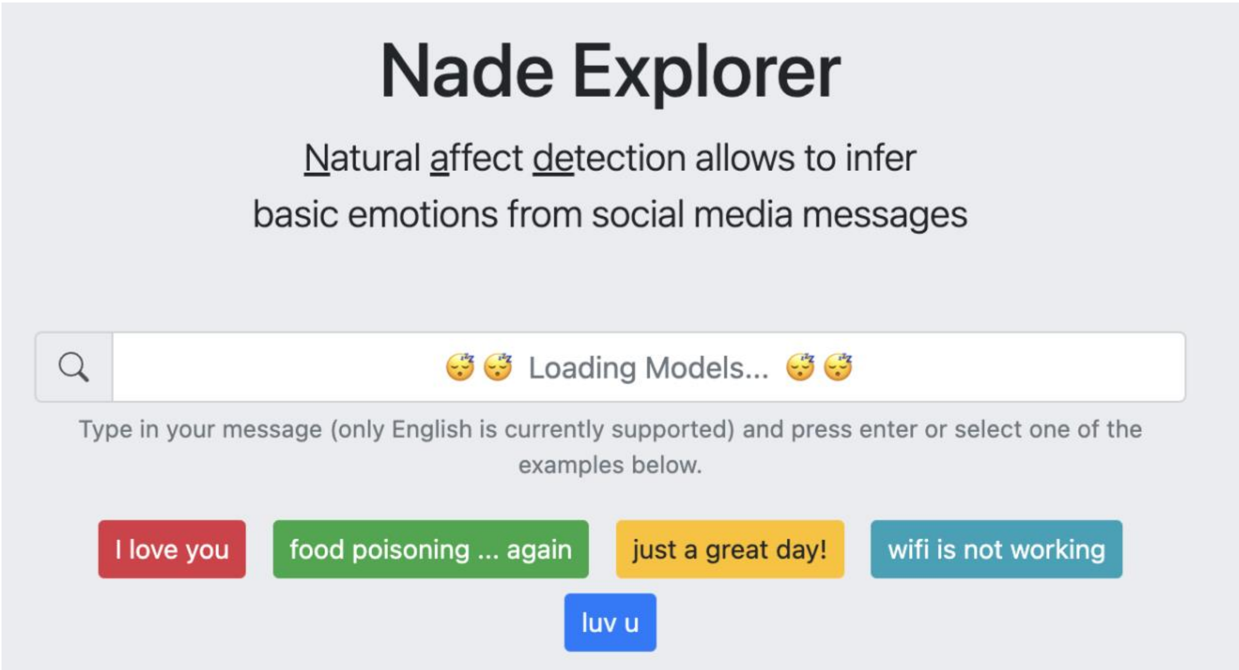
Design and user interactivity are the key to successful apps. When designing an interface, researchers should follow some long-standing user experience rules (heuristics): First, apps should have an obvious starting point. When the user first opens the interactive research-driven app, she should immediately know where to start. Second, apps should have a consistent logic and flow. Users should be "taken" by the hand: after one step has been taken, the next step should be clear. In the moment, that an action is taken, feedback is necessary and helpful. For example, when a model runs, a loading bar explains the users that they have to wait. To further facilitate usage, the proximity of interaction elements is crucial. Related settings (e.g. input parameters) should be visually grouped together. Last, interactive research-driven apps should be created in such a way that errors can be prevented as much as possible (e.g. through default settings, limited choice). We give an overview of important heuristics along with examples in Table 5. While the forementioned heuristics should be guiding the researcher in the design process, superior user experience and good interactivity is eventually created through the well-planed interplay of widgets. We explain widgets next.

Table 5. *Heuristics to improve user experience*

Heuristic	**Explanation**	**Example**
Obvious start	The user should immediately know where to start	NADE, directly offers a text field including examples.
Consistent logic and flow	The next steps should be clear to the user.	Text Overlay app, “next” button on the bottom right.
Immediate Feedback	When the user interacts, he should have immediate feedback .	Social Listening app, directly adapts charts when changing brands.
Proximity	Related settings should be grouped together.	Logo explorer, output parameters are grouped together.
Error prevention	User experience should be designed to reduce mistakes.	ConcretenessScoringTool offers tool tutorial and various options for input (e.g. single strings or complete files).

Note. Heuristics selected from seminal user experience literature (Blair-Early and Zender 2008; Molich and Nielsen 1990)

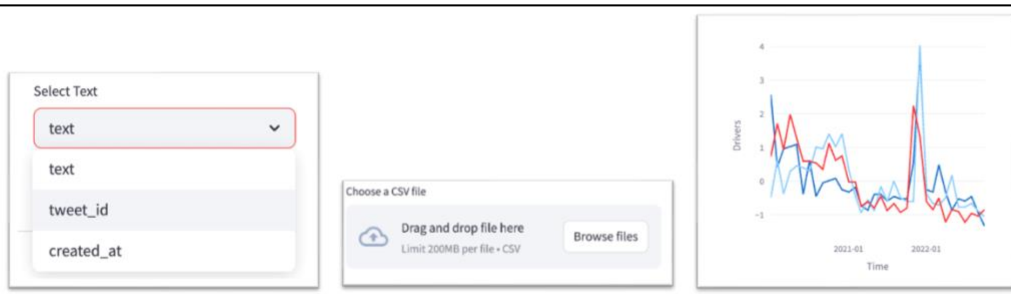
Figure 6. *Nade explorer, an example of good UX*



Note. The Nade Explore has an obvious start, helping using with examples. Consistent logic in the proximity of the controls "Type in your message and press enter". Immediate feedback through the loading bar about the status.

Widgets are the key technical elements that allow app interactivity. A widgets is "an element of a graphical user interface that displays information or provides a specific way for a user to interact with (…) an application"(Kirvan 2022). Widgets can be used to input data (file upload, set a parameter), trigger actions (execute button) or output data (graphs, download files). The choice of input widget depends on the inputs chosen in a previous step. In general, the researcher can pick a widget from a setting based on the following process: first, by identifying the type of input (such as table data, string, number, or boolean); second, by assessing the number of options available (for example, a limited and small list, a continuous or discrete set, or a numeric range); and third, by determining whether the selection should be single or multiple (e.g., “select all that apply”). For example, if the researcher needs users to select an aggregation method from a predefined list of 3 options (weekly, monthly, yearly) and only one selection is allowed, the appropriate widget would be a dropdown menu, implemented as a “Single Selection Widget in KNIME” or a “st.selectbox()” in Streamlit. If users are asked to select multiple brand metrics from a longer list, a multi-select widget would be more appropriate, such as the “Column Selection” Widget in KNIME or “st.multiselect()” in Streamlit. The widgets for the expected outputs depend on the typology of app chosen. For example, with converter apps, the final results are often made available to download (Hovy, Melumad, and Inman 2021; Humphreys, Isaac, and Wang 2021). In this case, the researcher should include a download widget. Explorer apps, on the other hand, additionally rely on visualizations (our example app; Matthe, Ringel, and Skiera 2023). The final combination of widgets determines the visual appearance and user experience of the app. The resulting design can be improved based on the order in which the user is expected to interact with the settings and additional elements that are added for subsequent actions: triggering the analysis with a button, displaying a loading bar widget, and finally reviewing the results in a graph. Following are some examples for our brand reputation app.

Figure 7. Widget examples



Widget	Selectbox	File uploader	Line chart
Description	Select columns in the dataset (specifying description, options and default)	Upload proprietary data (specifying description and filetype)	Visualize data from tables
Code	`selected_column = st.selectbox("Select Text", [”text", ”tweet_id", ”created_at"], index=1)`	`uploaded_file = st.file_uploader("Choose a CSV file", type=("csv"))`	`st.line_chart(df_drivers[selection2])`

Note. We show examples of widgets from the Streamlit (Python) framework along with a short description and example code. The full code can be found in the Web Appendix (W.4). The configuration for the KNIME nodes can be inspected in the online repository.

2.3.4 Step 4: Chose the Implementation Tool

An implementation tool should be chosen based on what interactive elements are needed, the programming language of the existing backend, and the researcher's programming skills. Most tools are frameworks that are based on specific programming languages, like Python or R. Frameworks offer pre-made function that are simplifying the coding process, by offering prebuilt functions, for example the multiselect widget in Streamlit (a Python based framework). If the existing analysis has been done in Python and the researcher is skilled in Python, it is sensible to pick an implementation tool that is based on Python (e.g. Streamlit). If more freedom is wanted in the choice and design of the widgets, more extensive tools such as Observable can be a better fit than more other frameworks. If a researcher has deep knowledge of a programming language or framework, adapting the chosen tool on the researchers existing skills can facilitate the development process. Accordingly, researchers that are more skilled in R, should use a framework that is based on R, like Shiny. In our example, we use two different tools. First Streamlit, which is a Python-based framework, and KNIME which is low-code solution (Villarroel Ordenes and Silipo 2021). To get an idea of the variety of tools, please see Table 6 for an overview of current solutions.

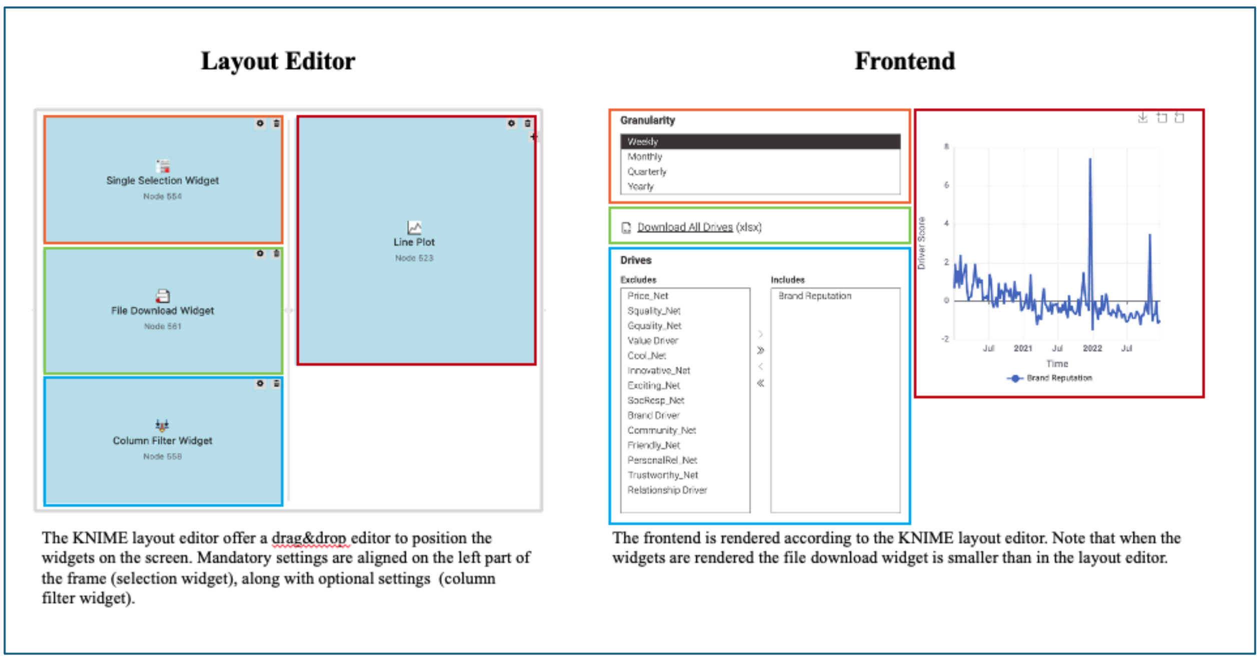
Table 6. *Overview of Implementation Tools.*

Tool	Description	Free option	Programming Language and Framework	Widgets
Gradio	Python-based open-source framework, specialized on machine learning.	Via Hugging Face	Python, git	Limited
KNIME	Low-code analytics tool offering R and Python integrations. WYSIWG editor for front-end editing.	No, but academic options available	No	Extensive
Netlify	Composable platform for hosting web apps	Yes	Git, Node.js	Extensive
Observable	Open-source framework focused on complex data visualizations.	No	R or Python, git, npm, Javascript	Unlimited due to JavaScript
Plotly Dash	R-based open-source app framework specialized in visualizations rich applications.	Via Render	R, git, flask, virtualenv,	Extensive
Shiny	R and Python-based open-source framework.	Yes	R or Python	Extensive
Streamlit	Python-based open-source framework.	Yes	Python, git	Limited

2.3.5 Step 5: Build the Frontend

When building the frontend, it makes sense to follow some guidelines (Buscher, Cutrell, and Morris 2009). Web-apps are normally structured around a top-left-bottom right orientation, where the most important elements are positioned on the top left and the least important elements are on the bottom right. The title of the app and instructions (in the form of widgets) are normally placed in the top left quadrant, in line with mandatory settings (Krug 2014). The widgets for optional settings such as, for example, the sample size taken, can be placed on the bottom left site. Visualization widgets, like graphs, can be placed on the top right corner of the app. Some tools offer "what you see is what you get" visual editors, that facilitate design of the frontend (see Example in Figure 5). When programming is strictly done via code (e.g. in R or Python), large language models (LLMs) can be assisting in the implementation of code, but generally are not yet capable of rendering the final results or visualizations, which makes inspection of the result more challenging. LLMs can offer an easy way to adapt code to different frameworks and used to understand the code (Nam et al. 2024). When generating code (Chen et al. 2021), for example for backend tasks, researchers should still proceed with caution when implementing the proposed solutions (Nejjar et al. 2025).

Figure 8. *Layout Editor and Frontend Results*



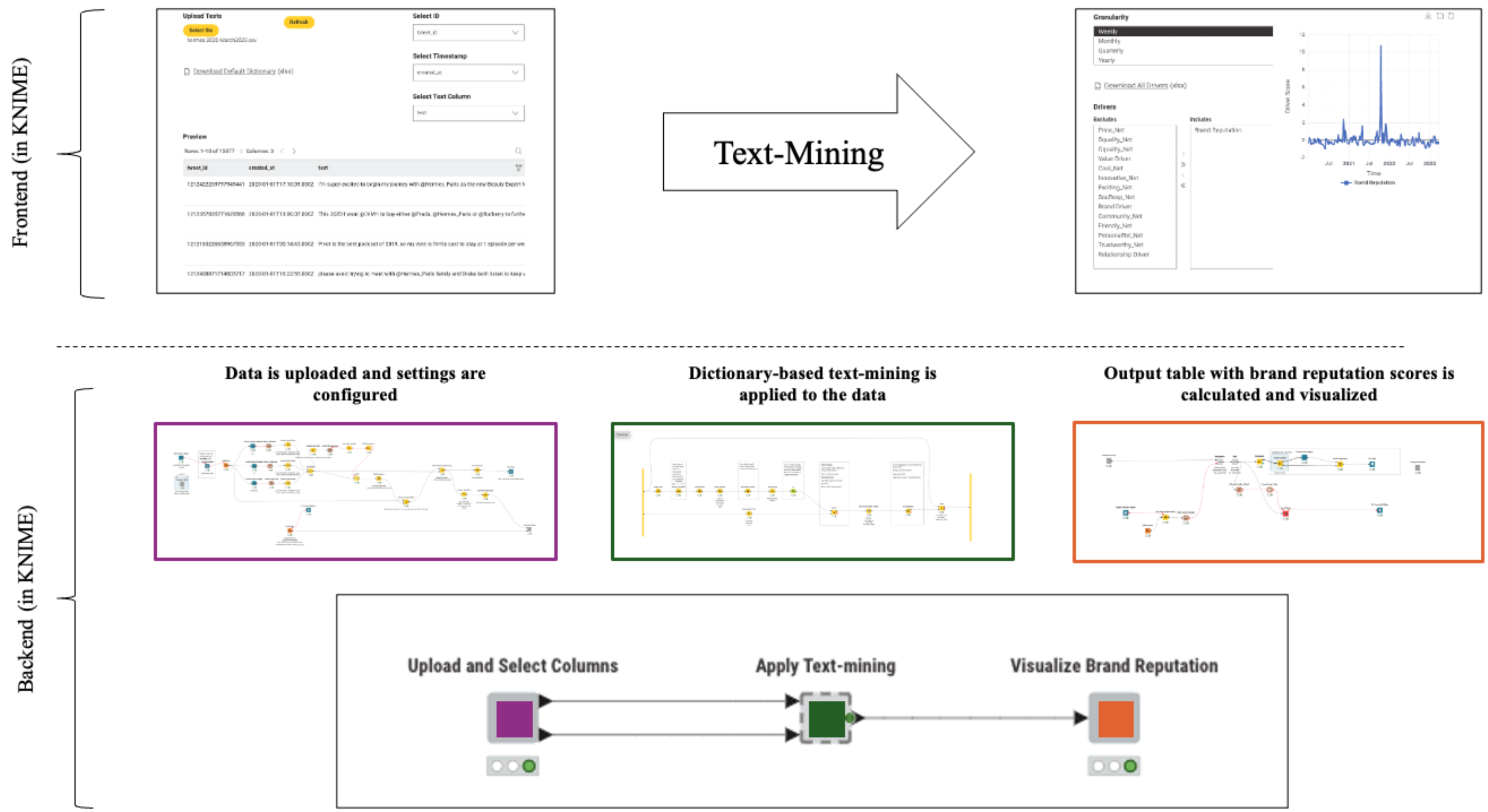
Note. This figure demonstrates the functionality of a drag&drop editor for visualization purposes. On the left, how it is configured, and on the right, how it is rendered in the front-end.

2.3.6 Step 6: Connect Back- and Frontend

After having chosen the widgets and implementation tools needed, the widgets need to be connected to the initial analysis (i.e., backend). In most cases, the following logic should be applied: First, settings are chosen (both mandatory and optional), then the analysis is triggered by button widget. After the analysis has finished, the results are either visualized in a graph or table widget or provided for download (or all of them together). How this is executed in the code is dependent on the tool and framework in use. Streamlit, as a Python-based framework executes code from top to bottom. Executing the code does not need to be explicitly triggered, but it is good practice for most apps. If, for example, the execution is computation heavy, it makes sense to split Python code in multiple pages, so as to make sure that the code is sequentially executed. If not, any interaction triggers the UI to be updated and re-executed. Researchers with computation intensive code should additionally cache results for multiple outcomes. For example, in our app, the user is allowed to change the brand metrics that are visualized, along with the aggregation. Using the caching capabilities, this can then be done without re-executing the text-mining code, improving user experience.

The connection between front- and backend is based on variables and dataframes. In the case of an input parameter, to connect front- and backend, a widget converts the input into a variable that is then used in a subsequent function. For a visualization, see Figure 6, for the code, including the connection of frontend with backend, please see our Web appendix.

Figure 9. *Connection of Backend and Frontend*



2.3.7 Step 7: Deploy the App

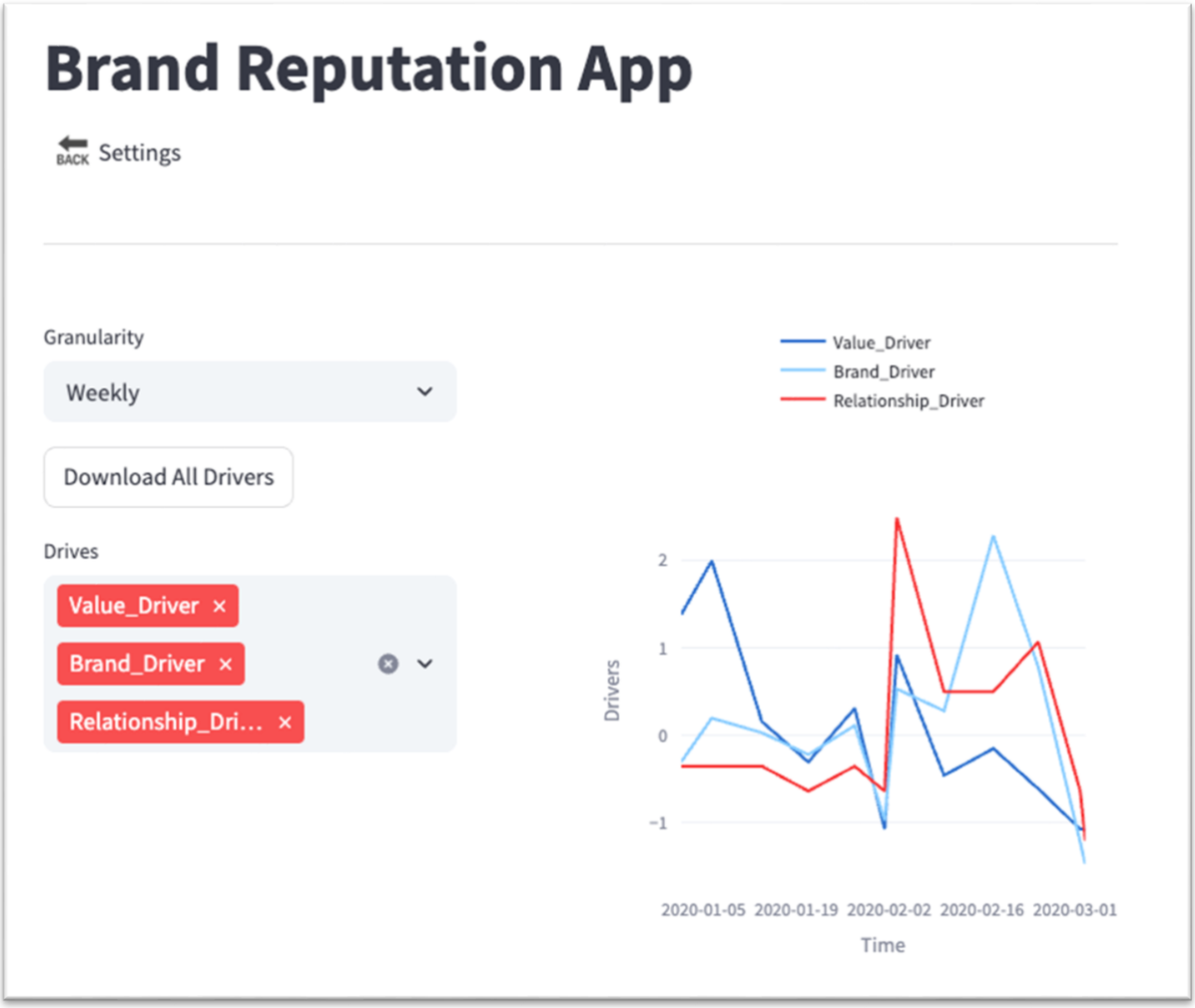
When deploying the app, the researcher needs to take three decisions. First, whether the app should be published privately or publicly. One common pitfall here is that public research-driven apps that are using external APIs might run into unforeseen costs due to high usage (i.e. more than 100 users per day). For example, if the research-driven app is using an API to connect to an LLM like GPT-4o, every user request will be billed. Generally, high usage might additionally slow down or break the app's functionalities. The more computational resources are needed (e.g. applying AI models on cloud hardware), the higher the cost will be. Cost structure varies by deployment provider, with some options being free (but slow) and others being costly (but fast). For example, in our case, the deployment via Streamlit is free and has almost no limitations in usage, while the deployment via KNIME comes at a cost of 99€ per month at the time of writing. See Table 7 for a comparison of options that have been used in the past. See Figure 6 for the final version of our deployed research-driven app on the Streamlit cloud.

We want to note that our research-driven apps are built with the purpose of showcasing the app development process, but do not necessarily fulfill the requirements for high-traffic, high processing usage. For more advanced users, the infrastructure should be carefully planned and monitored (see Richards and Ford 2020 for an overview).

Table 7. *Commonly used Frameworks*

Framework	**Costs for Public Deployment**	**Selected Examples**
Streamlit	Free for first app, only when publicly deployed	• Sepehri, Mirshafiee & Markowitz (2022) • Hovy, Melumad, Inman (2021)
Shiny	Free for 25 hours of usage	• André and Reinholtz (2024) • Dew, Ansari & Toubia (2022)
KNIME	99€ / month	• Farace et al. (2025)
Netlify	Free until 100GB bandwidth	• Matthe, Ringel & Skiera (2022)

Figure 10. *Visualization of the Results in deployed App*



Note. The current version of this interactive research-driven app can be found on https://research-driven-app.streamlit.app/

2.4 CONCLUSION

This article explores how research-driven apps on the interestingness and relevance of academic articles, particularly for marketing managers and students and how academics can develop research-driven apps on top of their articles. To do so examine both the present state and future potential of research-driven apps in marketing research. We find a small but growing number of high-quality research articles that utilize these apps to showcase their contributions. assess the future potential for implementing research-driven apps in existing publications, drawing on the latest app typology proposed by Chintagunta (2023). Our analysis suggests substantial untapped opportunity for integrating such apps into academic publishing. With two studies, one in the lab with marketing managers, and one in the field with marketing students, we confirm our hypothesis that research-driven apps significantly increase the perceived interestingness and relevance of academic research. We then conclude with a seven-step hands-on tutorial addressing the barriers of adapting and integrating research. This tutorial covers the development process of a research-driven app using a recent marketing article (Rust et al., 2021) as an example. We demonstrate the implementation in both code and low-code frameworks, making the process accessible to a broad range of researchers.

Taken together, our findings are encouraging evidence for academics to consider research-driven apps not only to enhance their article's contribution, but also as a tool to bridge the gap between academic research and practical application. First, we contribute to the research on research-driven apps (Chintagunta et al. 2022). While the literature has seen an increasing trend of apps in the last years, there has not been a comprehensive overview of existing apps. By extensively screening recent top marketing publications, we take inventory of the existing research-driven apps and classify them by their typology (converter, predictor, optimizer, explorer). We add onto this contribution by assessing the app development potential of current publications that have not added apps to their articles.

Second, we contribute to literature on managerial relevance, in line with recent work that has explored how to render marketing research more relevant for managers (Schauerte et al. 2023; Warren et al. 2021). While scholars have started to identify how marketing research is converted into knowledge and ultimately disseminated with managers and peer researchers (Roberts, Kayande, and Stremersch 2014), the focus has been one of conceptual nature (Schauerte et al. 2023). We, on the other hand, direct our attention to understanding the concrete implementations in the dissemination process. By demonstrating how research-driven apps can render research articles more interesting for managers and students, we show an applicable and concrete way for researchers to bring their work closer to industry, effectively narrowing the theory-practice gap (Baldridge, Floyd, and Markóczy 2004; Deighton, Mela, and Moorman 2021).

Third, by creating an exhaustive overview on how to develop a research-driven app, we contribute to the literature of tutorials in marketing research (Berger et al. 2020;

Boegershausen et al. 2022; Guyt, Datta, and Boegershausen 2024; Villarroel Ordenes et al. 2025; Villarroel Ordenes and Silipo 2021). Not only do we conceptually develop a step-by-step process, we also validate and showcase this process by contributing online-accessible code repositories. Furthermore, we review different programming frameworks and hosting options to facilitate decision-making for researchers.

Methodologically, we are curious to see what the future of app development will bring. Recent advances in LLM technology have found ways to easily convert one format of information into another, e.g. text into images, images into text, text into audio and so on. Lately, this was expanded also by the possibility of turning research articles into easily consumable media for consumers. For example, the tool "Wondercraft" can easily turn a research article into a podcast, with surprisingly good results. We would be excited to see a similar approach with research-articles, where articles could be easily turned into research-driven apps.

CHAPTER 3: THE MORE, THE MERRIER: GREED LANGUAGE IN SOCIAL MEDIA AFFECTS CRYPTOCURRENCY ADOPTION

ABSTRACT

Cryptocurrency as an investment method is growing in popularity around the world. Yet little is known about how consumers decide to invest in cryptocurrency. In this research, the authors posit that social media chatter can influence cryptocurrency adoption. Specifically, they argue that the use of greedier language (e.g., "make money with crypto") can increase cryptocurrency adoption among consumers. Analyses of posts from two different social media platforms paired with publicly available data on new cryptocurrency adoption through addresses (wallets) reveal that greedier language use in social media correlates with more new addresses created. Two lab studies provide causal evidence that greedier language in social media increases consumers' interest in investing in cryptocurrency. The effect of greedy language on cryptocurrency adoption is driven by changes in how profitable consumers perceive investing in cryptocurrency to be. Specifically, greedier language increases the perceived profitability of investing in cryptocurrency, which in turn increases consumers' interest in investing in cryptocurrency. The current work underscores a need to better understand how greedy language on social media affects consumers' perceptions of and behaviors towards cryptocurrency investments.

3.1 INTRODUCTION

Since its launch in 2009, cryptocurrency has become one of the most popular financial products around the world. Valued at over $2.42 trillion as of April 9, 2025 (CoinMarketCap 2025), cryptocurrency has revolutionized global financial markets and is transforming consumers' financial portfolios (Fritze et al. 2024). Despite its growing popularity, cryptocurrency is often considered a risky investment. This is in part because cryptocurrency remains unregulated by many federal governments and has historically demonstrated extreme price volatility (Fidelity 2024; Fritze et al. 2024; Nalley 2024). Further, unlike other investments, cryptocurrency prices are not based on underlying company assets or earnings but rather supply and demand, economic conditions and consumer sentiment and mood (Fidelity 2024). Lastly, cryptocurrencies can be created by any individual person who possesses the skills and resources needed to create a cryptocurrency (Fidelity 2024). Despite these risks, as the marketplace matures and cryptocurrencies become more widely available, consumers are increasingly incorporating cryptocurrencies into their investment portfolios.

Cryptocurrencies are particularly appealing among Gen Z and Millennial consumers (Fidelity 2025). To engage with such consumers, social media has become a critical communication tool for the cryptocurrency marketplace, with many platforms providing consumers with avenues to share real-time information, updates, and opinions about cryptocurrency (Fidelity 2024; FINRA 2023). For instance, platforms such as Reddit and X host various cryptocurrency communities where consumers share the latest on cryptocurrency (Almeida and Gonçalves 2023; Vasile 2023). Cryptocurrency owners seek to engage with consumers on social media, especially if said consumers ultimately participate in the cryptocurrency marketplace, because owners reap financial benefits as demand for cryptocurrency increases; as demand increases, so too do cryptocurrency prices and ultimately profits for those who cash out. As a result, cryptocurrency owners have a financial incentive to post engaging content that leads to increased cryptocurrency adoption. How do posters make their content engaging?

Prior research suggests that the use of greedy language on social media can be an effective way to facilitate social media engagement among followers (Mercadante et al. 2023). Perhaps it is unsurprising then that cryptocurrency content often contains some level of greedy language (e.g., "Doubled my money on @Bitcoin I need some more"; Martin, Chrysochou, and Strong 2024). Despite this prevalence, whether use of greedy language on social media results in subsequent offline behavioral actions, is unknown. Given this research gap, and the growing interest in cryptocurrency as an investment method, in the current research we examine the role of greedy language in social media chatter on cryptocurrency adoption. We suggest that greedier language on social media leads to increased interest in and adoption of cryptocurrency because greedier language increases how profitable consumers perceive investing in cryptocurrency to be. Using a multimethod approach, including text analyses of six years of social media field data from Reddit and X (previously Twitter), combined with daily data on new cryptocurrency addresses[5], and two lab experiments, we show that greedier language on social media increases cryptocurrency adoption.

Our research offers three main contributions for theory and practice. First, by studying the effect of greedy language use in social media on consumer adoption of cryptocurrency, we extend work that considers the role of social media on consumer decision making. Specifically, while prior research has considered the role of social media on consumer spending (Chevalier and Mayzlin 2006; Trusov, Bucklin, and Pauwels 2009; Kumar et al. 2016; Liadeli, Sotgiu, and Verlegh 2022), we focus on the role of social media on consumer cryptocurrency investment decisions. Leveraging blockchain's transaction traceability (Peres et al. 2022), we analyze daily new crypto addresses created. In so doing, we find that a one-unit daily increase in greedy language in Reddit posts correlates with a 5.5% increase in new Bitcoin addresses. Similarly, a one-unit weekly increase in greedy language in X posts correlates with up to a 9.0% increase in new addresses across 20 different cryptocurrencies. These findings, replicated in controlled settings, deepen our understanding of how use of greedy language in social media shapes consumer investment behavior.

Second, we contribute to research on greed (Hoyer, Zeelenberg, and Breugelmans 2022; Mercadante, Tracy, and Götz 2023; Seuntjens et al. 2015; Zeelenberg and Breugelmans 2022). Despite the importance of greed in the consumer finance marketplace, extant research falls short in exploring the effects of its different dimensions on consumer related outcomes. Specifically, in the context of financial decision making, scholars have mostly explored greed as a dispositional or stable personality trait (Martin et al. 2024; Seuntjens et al. 2016). In the current research, we consider how greed can be situationally induced (e.g., by other

[5] Addresses are the building block of transactions on the blockchain network. They are needed for receiving and sending cryptocurrency tokens. The number of new addresses is a strong indicator of user adoption on the blockchain network (Cong, Li, and Wang 2021; Liu and Tsyvinski 2021)

investors). By focusing on exposure to other people's greed on social media, which constitutes the primary platform for information exchange in the crypto setting, we advance our understanding of the ways in which observing others' greed can influence consumers' own investment decisions.

Third, we contribute to literature on consumer perceptions and financial decision making (Greenberg and Hershfield 2019; Lee and Andrade 2011). In particular, we extend research that suggests merely being exposed to fear of missing out (FOMO) appeals increases consumer investments in cryptocurrencies (Friederich et al. 2024). Beyond such FOMO appeals, which focus on fear-based communication, we argue that social media chatter containing varying levels of greed sentiment (e.g., just made some good money with bitcoin!; Shin 2022) influences consumer cryptocurrency investment decisions. Further, our lab experiments demonstrate that exposure to higher levels of greed language increases how profitable consumers perceive investing in cryptocurrencies to be, which in turn increases their interest in and adoption of cryptocurrency.

The remainder of the article is organized as follows: First, we provide an overview of the theoretical foundations. Next, we formulate specific hypotheses regarding the influence of greedy language on cryptocurrency adoption. These hypotheses are tested using two social media data sets (X and Reddit) that include more than 800,000 posts, and two controlled laboratory experiments involving 614 participants. Finally, we discuss the implications of our research for practitioners and academics, limitations, and potential future directions.

3.2 CONCEPTUAL BACKGROUND

3.2.1 Cryptocurrency as an Investment Method

Cryptocurrency, a digital asset that can be bought and sold online, was initially created as a digital medium of exchange (currency). Due to the use of blockchain technology as its underlying infrastructure, cryptocurrency can be securely exchanged between parties without the need for a central authority, such as a bank (Hakkarainen and Colicev 2023; Peres et al. 2022; Hofstetter, Fritze, and Lamberton 2024). As cryptocurrency becomes more widely used around the world, it is gaining traction as an investment method by consumers and firms alike (Fritze et al. 2024; Mathmann et al. 2025). In recent years, cryptocurrency as an investment has become especially appealing in part due to its potential for high returns (Aiello et al. 2022). For instance, early investors saw Bitcoin (the first and most popular cryptocurrency) prices increase from mere pennies to thousands of dollars (Nalley 2024).

While the potential for profits is indeed appealing, the potential does not come without its risks. Like many other investments, profits are not guaranteed. Indeed, cryptocurrency has historically had extremely high price volatility; between 2014 and 2024, Bitcoin's average annual volatility was 46.31%, and was 4.5 times more volatile than the S&P500 between August 2023 and August 2024 (Sergeenkov 2024). Further, cryptocurrency remains unregulated by many federal governments and is susceptible to pump-and-dump schemes, or schemes where owners drive up demand for their cryptocurrency to inflate its price, only to sell it off once it becomes profitable, resulting in potentially significant losses for investors (Fidelity 2024). Despite such risks, ownership of cryptocurrency is increasing.

How do consumers invest in cryptocurrency? To get started, consumers must decide on a cryptocurrency broker or exchange, such as RobinHood or Kraken, to open an account with (Tretina and Schmidt 2021). Once an account is opened, consumers then create a cryptocurrency wallet (via addresses), transfer funds into their wallet, and place a cryptocurrency order to begin investing (Tretina and Schmidt 2021). Once the order is made, consumers can either hold onto the cryptocurrency or sell it for a profit or loss depending on how the price changes relative to the initial purchase price.

Extant research on factors that influence consumer cryptocurrency investment decisions suggest that the adoption of cryptocurrency is largely influenced by herding behavior (i.e., mimicking other investor's decisions). Specifically, Bouri, Gupta, and Roubaud (2019) find correlational evidence of investors "following" the cryptocurrency market as shown through a high degree of co-movement in cross-returns across cryptocurrencies. Additionally, Friederich et al. (2024) suggest that fear of missing out (FOMO) can increase cryptocurrency investments. In the current research, we argue that consumers might react to not only market changes and fear-based appeals, but also social media chatter that contains greedy language, which we discuss next.

3.2.2 Social Influences, Language Use, and Cryptocurrency Adoption

Consumers often share insights and information about new products with close others. Such information sharing can increase the adoption of innovative products (Rogers 1983). For instance, Iyengar et al. (2011) demonstrate that through social contagion (i.e., sharing information with people you know), people are more likely to adopt innovative antiviral drugs. Similarly, when introduced to new products, consumers often face uncertainties about whether to adopt such products and rely on social factors, such as online social networks, to help make their decisions (Peres, Muller, and Mahajan 2010). In the current digital age, social media has become a popular way for new product developers to generate interest in their products and engage with consumers (Gruner et al. 2018). Indeed, social media has become the primary communication format for cryptocurrency founders (those who created and continue to create cryptocurrencies), investors, and consumers alike to connect with each other, acquire and share information about cryptocurrency, and stay informed about the latest cryptocurrency news and events (Fidelity 2024). Social media cryptocurrency communities have in turn captured the attention of millions of consumers; in December 2024, Bitcoin and cryptocurrency forums on Reddit had over 15 million followers, and on X, Bitcoin had 7.3 million followers, and Ethereum had 3.6 million followers.

Given the prevalence of social media communication in today's society, research continues to explore the effect of online communication on consumer decision making. In this context, prior research has identified language use in social media as an important driver of consumer perceptions, engagement and behaviors. Specifically, Packard and Berger (2017) demonstrate that using explicit endorsements such as "I recommend", instead of implicit endorsements such as "I like", are more persuasive and increase purchase intent. Additionally, when brands post messages that contain more (vs. less) certainty-related words (e.g. forever, always), consumers interact more with these posts because consumers perceive the brands as more powerful (Pezzuti, Leonhardt, and Warren 2021).

With regards to consumer investments, the volume and sentiment of stock-related user-generated content on social media has been shown to predict stock market returns (Bollen, Mao, and Zeng 2011; Tirunillai and Tellis 2012). This suggests that consumers who read about investment opportunities on social media may act on the information they encounter (Berger et al. 2020). Without a doubt, social media has become an important information source to investors (Cookson et al. 2024). In the same line, we argue that language use, specifically greedy language, in social media crypto chatter influences the adoption of cryptocurrencies in the marketplace. We focus on greedy language for three reasons. One, as we discuss next, greed plays a substantial role in consumers' own behaviors and financial decisions (Mussel et al. 2015; Seuntjens et al. 2016; Rodrigues et al. 2023). Two, greedy language on social media has been shown to increase consumers' engagement with posts (Mercadante et al. 2023). And three, cryptocurrency owners have an incentive to post content that not only appeals to those eager to learn more but also helps increase demand for cryptocurrency (Fiddsy 2022). This is because increased demand increases cryptocurrency prices, which increase profits for cryptocurrency owners who cash out.

3.2.3 Greed and Consumer Financial Decision Making

Greed, which is defined as the "dissatisfaction of not having enough, combined with the desire to acquire more" (Seuntjens et al., 2015, p. 519), can be dispositional, and thus a stable consumer trait, or can be situationally induced (Lambie and Haugen 2019). Dispositional greed has been shown to be a strong predictor of consumer financial decision-making**.** Specifically, prior research reveals that people who are high on dispositional greed tend to have higher incomes and higher expenses, yet also lower savings, more debt, and greater dissatisfaction with their lives. Additionally, Rodriguez et al. (2023) explore the relationship between dispositional greed and risk taking and find that in Ballon Analogue Risk Tasks (BART), people who are high in dispositional greed pump balloons up more, increasing the risk of their balloons bursting.

People high in dispositional greed are also more inclined to own cryptocurrency (Martin et al. 2024). However, because dispositional greed is a normally distributed and stable trait (Zeelenberg and Breugelmans 2022), it fails to explain the high variance in cryptocurrency adoption over time. This suggests that there is an important component of greed that is dependent on the situation (Lambie and Haugen 2019) and may thus be situationally or externally induced.

When greed is externally induced, it can lead to important behavioral outcomes. For example, similar to dispositional greed, induced greed can also increase risk-taking in individuals (Mussel et al. 2015). Further, situational and dispositional greed reinforce each other on their effect on risk-taking (Rodrigues et al. 2023). Lastly, Mercadante et al. (2023)

argue that politicians employing greedy language to accuse opponents of exploitative behavior can boost engagement with their social media posts among followers.

In the context of cryptocurrency, posters often utilize phrases such as “More more and more @bitcoin” or “Make me that money @dogecoin”. Such phrases are intended to generate hype around and demand for cryptocurrency as cryptocurrency owners benefit from higher cryptocurrency prices derived from greater demand (Carr 2024). Such posts often contain greedy language, albeit to varying degrees (Martin et al. 2024). We suggest that greed may be situationally induced through such externally observed greedy language use in social media chatter and ultimately motivate consumers to invest in cryptocurrency. Specifically, we posit the greedier the language used, the more likely consumers are to invest in cryptocurrency. We suggest the effect of greedy language on cryptocurrency adoption is driven by perceptions of how profitable consumers perceive investing in cryptocurrency to be.

3.2.4 Perceived Profitability of Cryptocurrency

Consumers' perceptions of investment profitability are influenced by several affective and cognitive processes (Aspara 2013), and play a fundamental role in consumer investment decisions (Clark-Murphy and Soutar 2004). For instance, optimism can lead to inflated expectations of returns (Aspara 2013), while overconfidence may cause investors to underestimate associated risks (Pikulina, Renneboog, and Tobler 2017). Greed can also affect financial decision-making by disproportionately shifting consumers’ attention towards expected profits and away from risk (Mussel et al., 2015), and lead to a desire to emphasize potential gains rather than potential losses (Rodrigues et al. 2023).

Among cryptocurrency communities, there is a prevalent narrative linking cryptocurrency investments with rapid financial gains (Bowles 2018). Indeed, significant price increases have characterized cryptocurrencies over the past decade. For example, Bitcoin’s value surged from approximately $1,000 in 2014 to $95,000 in 2024 (CoinMarketCap 2024). High volatility in crypto markets has allowed investors, including those with limited resources, to realize substantial returns, albeit accompanied by significant risks. Interestingly, the considerable uncertainty inherent in cryptocurrency markets also motivates consumers to actively participate in social media discussions, primarily driven by an intent to uncover profitable investment opportunities (Almeida and Gonçalves 2023).

Drawing on this, we propose that exposure to greedy language in social media chatter increases how profitable consumers perceive investing in cryptocurrency to be, and consequently increases their adoption of cryptocurrency as an investment. More formally, we propose the following hypotheses:

H_1: More versus less greedy language on social media will increase consumers interest in and adoption of cryptocurrency as an investment.

H_2: The effect of more versus less greedy language on social media on adoption of cryptocurrency is mediated by how profitability consumers perceive investing in cryptocurrency to be.

3.3 Overview of Studies

We test our hypotheses by employing a multimethod approach using field data consisting of over 800,000 social media posts from two different social platforms (Study 1a [Reddit], and Study 1b [X]), and two lab experiments (Studies 2-3). First, for each platform, we model the aggregated amount of greedy language on social media posts and associate it with the daily adoption of cryptocurrency as measured by the number of new addresses (wallets) created. Next, we provide causal evidence by manipulating greedy content in social media posts in a lab setting (Studies 2-3). Then, we test our proposed mechanism and show that exposure to more versus less greedy language increases how profitable consumers perceive investing in cryptocurrency to be, and this mediates the effect of greedy language on cryptocurrency adoption (Study 3). Lastly, we consider whether those effects hold while controlling for individual differences: disposition to greedy behavior, which prior research suggests affects cryptocurrency investment decisions (Martin et al. 2024), subjective financial knowledge (Netemeyer et al. 2024), and the susceptibility to interpersonal influence (Bearden, Netemeyer, and Teel 1989).

3.3.1 Study 1A and 1B: The Effect of Greedy Language on Social Media on New Cryptocurrency Addresses Created

In Studies 1A and 1B, we aim to test our main prediction that the use of greedy language in social media posts increases consumers' adoption of cryptocurrencies in the field. With this goal, in Study 1A, we focused on the largest, most well-known cryptocurrency, Bitcoin, and analyzed the level of greedy language in Reddit posts containing Bitcoin related content. In Study 1B, we aimed to replicate and extend our findings by exploring content across 20 cryptocurrencies on Twitter (X). In both studies, our cryptocurrency adoption measure is new addresses created. The creation of new addresses is reported on the blockchain ledger, where all addresses and transactions are recorded. As a result, blockchain analytics providers, who analyze the changes on the ledger, can observe transactions and new addresses created. We exploit this for our empirical model.

1.1.1.1 Study 1A: Setting

To provide initial evidence of the effect of greedy language in social media content on cryptocurrency adoption, we turned to Reddit, one of the largest social media platforms that discusses a host of consumer topics, including consumer finances (Egan 2023). Next, we sought to obtain posts and comments from the largest cryptocurrency community (subreddit) in Reddit: /Bitcoin. This subreddit cryptocurrency community contained more than one million posts and comments from 2010 to 2022, making it a rich source of discussion data on the major cryptocurrency. Further, given Bitcoin was launched in 2009, this also represents a significant time span. Using archival data from the-eye.eu, an archive of subreddit data, we initially obtained 1,141,746 comments during this period. However, after 2017, Bitcoin gained mainstream attention following its first major price surge from $1,000 to nearly $20,000, marking a pivotal moment in its recognition. We therefore focused on the timeframe from 2017 to 2022 and removed fully deleted and archived posts, as well as links (Tash et al. 2024), leaving 605,420 posts in the dataset for analysis[6]. Descriptive statistics and correlations are in Table 4.

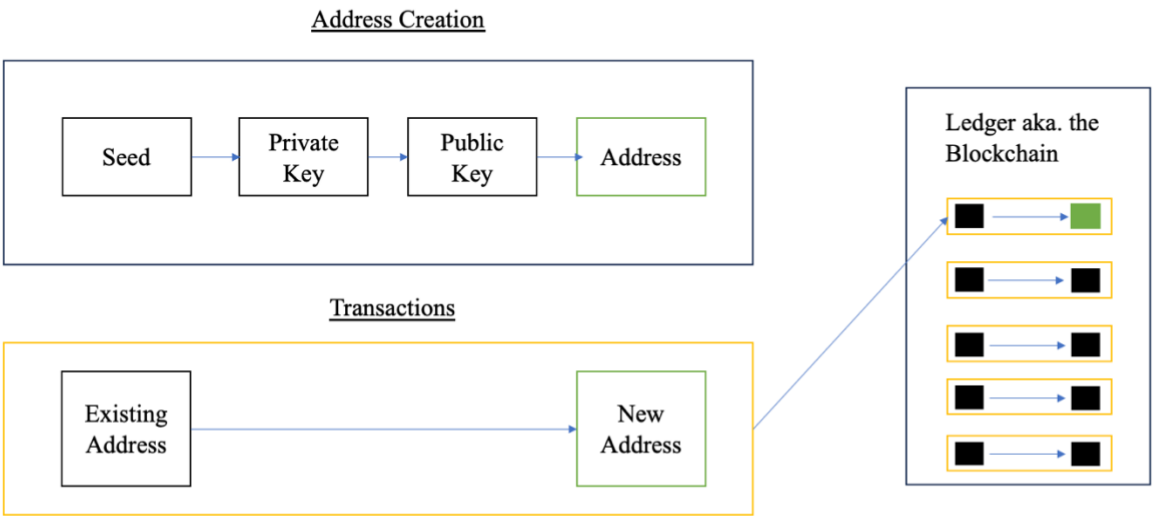
3.3.1.1.1 Measurements

Dependent Variable. We used daily cryptocurrency new address data on the Bitcoin blockchain and aggregated daily, resulting in 2,191 daily observations. Addresses, like bank account numbers, are used to send and receive cryptocurrency tokens. Once an address that

[6] Results of our estimations do not change if we expand the data to the 2010-2022 timeframe.

has been created receives cryptocurrency tokens, this transaction is registered on the blockchain (see Figure 1). We use blockchain analytics providers (into-the-block, Glassnode) that analyze the blockchains of cryptocurrencies, to observe the number of new addresses that are created.

Figure 1. *How New Addresses are created on the Blockchain*



Independent Variable. We used the Linguistic Inquiry and Word Count Software (LIWC) (Boyd et al. 2022) in order to identify greed communication in the textual corpus (Mercadante et al. 2023). This was done using a dictionary approach (Berger et al. 2020) for identifying greed-related words (e.g. "greedy, rich, millionaire, not enough, more"). It counted the occurrences per document and reports the proportion. In other words, if 10 out of 100 words were greed-related ("not enough", "more"), then LIWC reports 10% of greed words. Following this process, we aggregated our metrics on a weekly basis, obtaining the average level of greed. Face validity of the dictionary is assessed by comparing greedy language (e.g., "Grab yours free bitcoin unlimited access") with non-greedy language (e.g., "What can I buy with Bitcoins?"). Discriminant validity is demonstrated as the greedy language measure does not correlate with prosocial linguistic factors ($r = .028$, $p = .190$) or anxiety ($r = .002$, $p = .916$), while convergent validity is supported by the relationship between the money dictionary (Boyd et al., 2022) and both the greed dictionary ($r = .123$, $p < .001$) and reward ($r = 0.181$, $p < .001$; see Table 2).

Table 1. *Greed Dictionary, Study 1A and Study 1B*

never enough, not enough, self-interest, never satisfied, not satisfied, money-grub, not sharing, glut, money, more, not generous, not thinking of others, self-centered, money-hungry, ungenerous, self-obsessed, overabundance, too much, too many, entitled, stockpile, super rich, unsatisfied, self-serving, rich people, take, took, cash, uncharitable, no matter what the consequences are, power, whatever it takes, insatiable, luxury, show-off, status, spoiled, stingy, self-glory, showing-off, self-important, looking out for oneself, looking out for yourself, looking out for myself, looking out for himself, looking out for herself, looking out for themselves, show-off, over-the-top, stole, lacking self-control, lack self-control, lacks self-control, lacked self-control, ruthless, unlimited, conspicuous consumption, material*, greed*, overconsum*,selfish*, excess*, hoard*, want*, self-indulg*, wealth*, possessive*, overindulg*, crave*, narcissis*, corrupt*, self-absorb*, billion*, consum*, envious*, ego*, exploit*, possess*, million*, no matter the consequenc*, lavish*, lust*, grab*, dollar*,obsess*, profit*,demand*, gain*, boast*, brag*, rich*, jealous*,extravagen*, manipulat*,arrogan*,devour*,snob*, stash*, indulg*, maximiz*, steal*, acquir*, vain*, hog*, accumulat*, inconsiderate*, avarice*, opulen*, ravenous*, voracious*

Table 2. *Validity Checks for Greed Dictionary**,** Study 1A and Study 1B*

Type of validity	**Validation Procedure**
Face Validity: Is the construct measuring what it claims to measure?	Greedy language: "Grab yours free bitcoin unlimited access" vs. No greedy language: "What can I buy with Bitcoins?"
Discriminant validity: Does the measurement differentiate from measures of other constructs?	Our measure for greedy language does not relate with prosocial linguistic factors (r = .028, p = .190) or anxiety (r = .002, p = .916).
Convergent validity: Do multiple measurements of the construct all converge to the same concept?	Money dictionary (Boyd et al. 2022) relates to Greed dictionary (r = .123, p < .001) and reward (r = 0.181, p < .001).
Predictive validity: Does the construct have the expected effects on a meaningful variable?	Across multiple data sets and different measurement approaches, we confirm the relationship between greedy language and adoption.

Control Variables. We included several control variables that could influence consumers' adoption of cryptocurrency. First, we control for language certainty, which is a measure of how confident the author of the post is. We control for language certainty because prior research demonstrates that higher language certainty may strengthen the persuasiveness of social media posts (Pezzuti et al. 2021; Rocklage et al. 2023). We also control for anxiety, since higher communicated anxiety might signal investors to abstain from investing, having a negative effect on product diffusion (Lee and Andrade 2015). Furthermore, since arousal (Berger 2011) and valence (Berger and Milkman 2012) can increase the virality of social media messages, generating even more hype about certain content, we control for both arousal and valence language in the posts. Next, we control for concreteness of the message: more concrete messages can be more persuasive to potential buyers that are further along in their decision making process (Humphreys, Isaac, and Wang 2021). We include familiarity of language, since crypto chatter often contains technical terms that might hinder understanding (Mathmann et al. 2025). We also control for word count, because longer posts could transmit more information, enhancing engagement (Berger and Milkman 2012). Emojis have been found to also have a crucial role in engagement, and so we include them as a control (Luangrath, Xu, and Wang 2023). Next, we control for social media volume and social media engagement. Social media volume, since it has been found to be predictive of stock performance (Tirunillai and Tellis 2012), and social media engagement because it can influence consumer behavior (Liadeli, Sotgiu, and Verlegh 2023). Lastly, to control for the general uncertainty around cryptocurrency, we control for price volatility, where a high volatility might discourage consumers from investing in cryptocurrencies (Gaies et al. 2023). We do not control for price for two reasons. First because we already control for volatility, which covers changes in prices over time. Second, because of the divisibility of cryptocurrencies (a consumer can invest in one hundred millionth of Bitcoin), entry prices should matter less than the changes in prices.

3.3.1.1.2 Model and Estimation

Our dependent variable (new addresses) is a count variable and overdispersed ($p < .001$, likelihood ratio test). We therefore use negative-binomial regression. We specify two models. A base model, without any control variables, and a full model including our control variables (see Table 3). We estimate the following model:

$$\# \boldsymbol{of\ new\ addresses_i} = \mathbf{exp}(\boldsymbol{\alpha_0} + \boldsymbol{\beta_1} * \boldsymbol{Greed_i} + \boldsymbol{\beta_n \theta_n} + \boldsymbol{\epsilon_i})$$

where $\boldsymbol{\beta_n \theta_n}$ represent the control variables and their respective coefficients and $\boldsymbol{\epsilon_i}$ is the error term.

Table 3. *Overview of Variables*

Variables	Why we measure	How we measure
Greed	Greed in the investment setting can be perceived as a cue to profitable investments.	Mercadante et al. (2023)
Anxiety	Anxiety in posts might signal fear in the posters, discouraging investment (Shefrin 2002).	LIWC-22, Anxiety Dictionary
Valence	Positive posts might encourage consumers to invest.	NRC-VAD Lexicon (Mohammad 2018)
Arousal	Higher arousal emotion might increase virality of posts (Berger and Milkman 2012).	NRC-VAD Lexicon (Mohammad 2018)
Certainty	Certainty can increase engagement for brands (Pezzuti et al. 2021).	Evaluative Lexicon (avg.) (Rocklage et al. 2021)
Concreteness	More concrete messaging might be more persuasive (Humphreys et al. 2021).	BootstrappedMRC (Paetzold and Specia 2016)
Familiarity	Familiar words increase fluency and can lead to more engagement.	BootstrappedMRC (Paetzold and Specia 2016)
Wordcount	More words could convey more information (Berger and Milkman 2012).	LIWC-22
Engagement	Social media engagement influences consumer behavior (Liadeli et al. 2023.	X (Study 1A): Likes + Replies + Retweets; Reddit (Study 1B): Number of Comments
Social Media Volume	Social media volume influences consumer behavior (Colicev et al. 2018), Volume of user-generated content can influence stock-prices (Tirunillai and Tellis 2012).	Count of posts; Own measure
Emoji	Emojis can influence engagement of posts (Luangrath, Xu, and Wang 2023).	LIWC-22
Volatility	High volatility might discourage investors from investing in cryptocurrencies.	Annualized 30-day volatility (Source: Into-the-Block)

Table 4. *Descriptive Statistics and Correlations, Study 1A*

VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
(1) New Addresses	1.000														
(2) Greed	.074	1.000													
(3) Anxiety	.120	.002	1.000												
(4) Valence	.015	.149	-.089	1.000											
(5) Arousal	-.059	.175	.018	.828	1.000										
(6) Certainty	.224	-.012	.068	.187	.104	1.000									
(7) Concreteness	.179	.139	.052	.548	.541	.501	1.000								
(8) Familiarity	.381	-.029	.177	.009	-.123	.497	.268	1.000							
(9) Wordcount	.186	.022	.045	.230	.165	.633	.394	.347	1.000						
(10) Engagement	.092	.090	.080	.134	.088	.265	.141	.163	.327	1.000					
(11) Volume	.350	-.022	.175	-.188	-.151	.132	.148	.340	.020	-.246	1.000				
(12) Emoji	.176	.019	.052	-.016	-.058	-.017	-.112	.100	.049	.412	-.117	1.000			
(13) Volatility	-.017	.034	.070	-.092	-.025	.114	.159	.163	.050	-.104	.437	-.089	1.000		
(14) Reward	-.010	.181	-.065	.143	.119	-.075	-.027	-.114	.009	.106	-.164	.123	-.087	1.000	
(15) Out	-.021	.060	.096	-.006	.051	-.039	.045	-.016	-.003	.148	-.038	.170	.062	.023	1.000
MEAN	401124.496	1.341	.093	.185	.140	2.458	231.909	561.910	38.612	15.833	275.979	.675	.587	.420	.165
SD	81039.008	.309	.092	.008	.006	.346	6.361	6.270	9.752	7.428	236.428	.844	.234	.199	.281

Table 5. Greedy Language Can Impact Bitcoin Adoption, Study 1A

	Base			Full		
DV: New Addresses	Coef.	SE	*p*	Coef.	SE	*p*
Greed	.049	.014	.000	.054	.012	.000
Anxiety				.068	.041	.103
Valence				2.128	.864	.014
Arousal				-4.337	1.200	.000
Certainty				.009	.015	.573
Concreteness				.003	.000	.000
Familiarity				.007	.000	.000
Wordcount				.001	.000	.005
Engagement				.000	.000	.693
Volume				.000	.000	.000
Emoji				.044	.005	.000
Volatility				-.186	.018	.000
Constant	12.836	.019	.000	8.378	.434	.000
Observations	2,191			2,191		
Log Likelihood	-27,833.195			-27,348.035		

3.3.1.1.3 Results

Results demonstrate that the level of greedy language in aggregated daily Reddit posts had a significant and positive effect on the number of new Bitcoin addresses created (IRR = 1.050, SE = .014, t = 3.538, p < .001). These results hold after adding control variables (IRR = 1.055, SE = .012, t = 4.444, p < .001). Specifically, with the Reddit data, a one unit increase in greedy language is associated with a 5.5% increase in new addresses. Additionally, a likelihood ratio chi-square test was conducted to compare the full model and base model. The results showed a statistically significant improvement in our full model compared to the base model ($\chi^2(11) = 697.319$, p < .001). This study provides initial correlational evidence of the impact of the level of greedy language use in social media on cryptocurrency adoption. A series of robustness checks confirm our findings as discussed below.

The results of our control variables are also in line with past findings on the results of language on social media engagement. Valence and arousal have positive effects on cryptocurrency adoption, which is in line with research by Berger and Milkman (2012), who show that messages that are high in valence and arousal might spread faster compared to low arousal, low valence messages. Overall social media volume is positively associated with cryptocurrency adoption, corroborating literature on the impact of social media content on consumer outcomes (Colicev et al. 2018; Tirunillai and Tellis 2012). Volatility negatively affects cryptocurrency adoption, suggesting that even though volatility is an acknowledged characteristic of cryptocurrency, it decreases cryptocurrency adoption on average. This is in contrast to literature suggesting that traders of risky assets trade more when volatility is high (Weiss-Cohen et al. 2023).

3.3.1.1.4 Endogeneity and Robustness Checks

We employ several measures to verify the robustness of our findings. First, to exclude the possibility that the results of greed are strictly dependent on our choice of dependent variable, we apply a closely related dependent variable: the number of active addresses. An address is defined as an active address when it has received funds at least once. Our results show a significant coefficient for the impact of greed on active addresses (IRR = 1.061, SE = .013, t = 4.717, p < .001). Additionally, to exclude that our results are dependent on our empirical model, we use standard OLS estimations with a log-transformed DV on the full model. Our results remain significant (b = .051, SE = .012, t = 4.183, p < .001).

Second, one might argue that the variation in amount of greedy language is not random. To address endogeneity concern, we employ a control function approach (Petrin and Train 2010), which is better suited to non-continuous dependent variables (Papies, Ebbes, and Van Heerde 2017; Rutz and Watson 2019). In the first stage of the control function approach, we regress our potential endogenous variable, greed, on the proportion of out-of-the-money addresses at the time and the proportion of profit-related words using the LIWC reward dictionary (Boyd et al. 2022), obtaining an estimation for our endogenous variable. An address is "out of the money", when a realization of its holdings at current prices would result in a loss. While this might influence the strategic posting of greedy content (in order to raise prices to move "in the money"), it is uncorrelated to the creation of new addresses (because current holders already have addresses and do not open new ones). We obtained this data via a blockchain analytics provider (Into the Block). The reward dictionary of LIWC (Boyd et al. 2022) includes words like “rewarding, benefit and bonus” and is based on the human drive of being rewarded. We make sure that our variables are uncorrelated with the dependent variable (reward: r = -.013, p = .552; out-of-the-money: r = -.021, p = .322), but correlated with our focal independent variable (reward: r = .179, p < .001, out-of-the-money: r = .060, p = .005). We then estimate using OLS and include the residuals of our estimation into our original full model. The resulting regression should then be corrected for endogeneity. The impact of greedy language on cryptocurrency remains close to the original estimations (IRR = 1.120, SE = .065, t = 1.724, p = .085).

Third, we check our results against one prominent measure of cryptocurrency market sentiment, the “Fear and Greed Index”(FGI). In the cryptocurrency context, practitioners use the FGI, by consulting firm Alternative.io, “a popular sentiment indicator tailored to the Bitcoin market”(Wang, Liu, and Hsu 2024, p.1). The score ranges from greed (100) on the one side of the spectrum, to fear (0), on the other. While our measure for greed is based on occurrences of greedy languages and is funded on the analysis of single words in social media posts, the FGI takes a compound approach of multiple measures. It takes volatility (25%), market momentum and volume (25%), social media (15%), surveys (15%), dominance (10%) and Google trends (10%) into account (Wang, Liu, and Hsu 2024). Informed by the index’s metrics, practitioners can forecast consumers’ interest in cryptocurrency (Alternative 2024). Here, high greed is taken as predictive of a future market correction, so prices that are too high. Past research has found the index to be associated with price-changes in Bitcoin (Gaies et al. 2023). To test the robustness of our results, we add the index in two ways. First, as a substitute to our volume and engagement variables, since the Fear and Greed Index potentially is a compound measure of those. Results for our greed measure remain similar in size (IRR = 1.036, SE = .011, t = 3.024, p = .002). Second, we add the FGI additionally to search post volume and engagement. Results are similar (IRR = 1.036, SE = .012, t = 2.23, p = .023).

Fourth, we checked whether our results still held if we substituted social media volume with search queries via Google Trends (Kristoufek 2013). Results remain significant and similar in size (IRR = 1.056, SE = 0.016, t = 3.029, p = .002).

Table 6. Overview of Robustness Checks, Study 1A

Modeling	What we test	How we test it
Endogeneity, Greed	Is the effect of greed on adoption driven by unobserved factors?	Control function with out-of-the-money and reward as instruments (IRR = 1.120, SE = .065, t = 1.724, p = .085).
Alternative model, linear	Is the count model inappropriate for estimation?	OLS with log-transformed DV (b = .051, SE = .012, t = 4.183, $p < .001$).
Alternative count model	How do results change if we estimate with a Poisson model?	Poisson model with robust standard errors (IRR = 1.057, SE = .000, t = 498.4, p < .001)
Alternative DV	Does the evidence also hold for active users only (i.e. only addresses that have received transaction)?	Estimating with active addresses as the dependent variable (IRR = 1.061, SE = .013, t = 4.717, $p < .001$).
Measurement		
Alternative measure of investor attention	Does the inclusion of search volume trends change the impact of greed on adoption?	Substituting the post volume by Google Trends (IRR = 1.056, SE = 0.016, t = 3.029, p = .002).
Alternative conceptualization of Greed	Does our IV differ from fear and greed?	Substituting post volume by fear and greed index (IRR = 1.036, SE = .011, t = 3.024, p = .002). Adding index to model (IRR = 1.036, t = 2.23, SE = .012, p = .023).

1.1.1.2 Study 1B: Setting

We selected a sample of 20 cryptocurrencies based on their market capitalizations and the availabilities of adoption data of the cryptocurrencies on two blockchain analytics platforms: “Into the Block” and “Glassnode”. To explore the use of greedy language on social media, we focused on the most prolific social media platform in the cryptoverse: X. We used a variety of textual analyses methods on over 300,000 social media posts that involved one of these 20 cryptocurrencies. Using X’s application programming interface (API) we extracted tweets about the largest 20 cryptocurrencies by market capitalization (see Table 7). To do this, we searched the Twitter archive for any tweets that mentioned cryptocurrency (e.g. “@bitcoin”). We scraped the tweets shared by the cryptocurrency accounts starting from the date they were created, filtered the tweets for English language, and removed the duplicates and links. Even though our dataset included multiple mentions, we only kept the tweets that mention only one cryptocurrency in order to identify the most relevant tweets on cryptocurrencies (Tash et al. 2024). After cleaning our dataset, we were left with 289,225 tweets, spanning a timeframe from 2017 to 2022. We aggregated our tweets on a weekly level and kept those observations where we had at least five tweets per week. We matched this with blockchain diffusion data, resulting in a dataset of 3938 weekly observations. [7]. Descriptive statistics and correlations are in Table 4.

3.3.1.1.5 Model and Estimation

We used the same dependent variable (daily cryptocurrency new address data for each cryptocurrency) and independent variable (greed dictionary on X posts) as in Study 1A, along with the same control variables. By expanding our dataset to include more cryptocurrencies, we aim to enhance the generalizability of our findings. However, not all cryptocurrencies are created equal. Therefore, the volume of tweets mentioning different cryptocurrencies varies. While Bitcoin constantly receives high number of mentions, lesser-known cryptocurrencies do not. To account for this, we aggregate tweets on a weekly basis. Second, cryptocurrencies may have idiosyncratic differences that are not directly observable. To address this, our model controls for individual cryptocurrency differences using coin-fixed effects. Lastly, we add weekly time-effects in the model to address for potential changes in seasonality (Rutz and Watson 2019; Long et al. 2020). Like Study 1A, our dependent variable, the number of new addresses, is overdispersed and a count variable ($p < .001$, likelihood ratio test). Therefore, we estimate the effect of greedy language on cryptocurrency adoption using a negative binomial regression model and estimate the following model:

$$\# \, \boldsymbol{of\ new\ addresses_i} = \mathbf{exp}(\boldsymbol{\alpha_0} + \boldsymbol{\beta_1} * \boldsymbol{Greed_i} + \boldsymbol{\beta_n \theta_n} + \boldsymbol{\alpha_w} + \boldsymbol{a_c} + \boldsymbol{\epsilon_i})$$

where $\boldsymbol{\beta_n \theta_n}$ represent the control variables and their respective coefficients, $\boldsymbol{\alpha_w}$ are the weekly fixed effects, $\boldsymbol{a_c}$ are the coin-fixed effects and $\boldsymbol{\epsilon_i}$ is the error term.

---

[7] Results of our estimations do not change if we expand the data to the 2010-2022 timeframe.

Table 7. Descriptive Statistics Study 1B

VARIABLES	1	2	3	4	5	6	7	8	9	10	11	12	13
(1) New Addresses	1.000												
(2) Greed	.149	1.000											
(3) Anxiety	.094	.063	1.000										
(4) Valence	-.146	-.026	-.075	1.000									
(5) Arousal	-.073	.025	-.017	.928	1.000								
(6) Certainty	-.085	.024	.013	.237	.189	1.000							
(7) Concreteness	-.131	.076	.001	.744	.751	.310	1.000						
(8) Familiarity	-.012	-.046	-.037	.275	.216	.149	.278	1.000					
(9) Word Count	-.239	-.016	-.027	.246	.173	.474	.304	.072	1.000				
(10) Engagement	-.203	.015	-.004	.100	.048	.100	.084	-.032	.204	1.000			
(11) Volume	.164	-.014	.036	-.021	-.018	-.036	-.040	.078	-.079	.087	1.000		
(12) Emoji	-.042	.015	-.010	-.182	-.205	-.043	-.266	-.035	-.018	.120	.149	1.000	
(13) Volatility	-.014	-.009	-.011	-.027	-.038	.017	-.040	-.004	.031	.044	-.016	-.007	1.000
Mean	345855.573	.944	.042	.159	.115	2.440	195.850	565.405	22.231	30.334	73.320	4.641	7.326
SD	788408.359	.651	.118	.026	.019	.904	22.291	15.143	4.379	38.060	115.921	4.806	202.494

Table 7. *Number of Weekly Observations Across Twitter Accounts, Study 1B*

Cryptocurrency	Twitter Handle	# of Weekly Observations
Bitcoin	@bitcoin	312
Dogecoin	@dogecoin	269
Decentraland	@decentraland	267
Ethereum	@ethereum	264
Enjin	@enjin	261
Litecoin	@litecoin	255
USDT	@tether_to	241
Chainlink	@chainlink	227
Basic Attention Token (BAT)	@attentiontoken	214
Gemini	@gemini	214
Cardano	@cardano	208
Loopring	@loopringorg	190
Quant Network	@quant_network	186
Synthetix	@synthetix_io	137
Compound Finance	@compoundfinance	131
Sandbox	@thesandboxgame	125
Curve	@curvefinance	125
Aave	@aaveaave	118
Polygon	@0xpolygon	98
Lido Finance	@lidofinance	96

Table 8. Greedy Language Impacts the Adoption of Cryptocurrencies, Study 1B

	Base			Full		
DV: New Addresses	Coef.	SE	*p*	Coef.	SE	*p*
Greed	.904	.052	.000	.086	.021	.000
Anxiety				-.225	.109	.038
Valence				2.474	1.542	.109
Arousal				-2.023	2.103	.336
Certainty				-.016	.017	.328
Concreteness				-.002	.001	.112
Familiarity				.002	.000	.039
Word Count				-.015	.004	.000
Tweet-Volume				.002	.000	.000
Engagement				.000	.000	.039
Emoji				.004	.003	.219
Volatility				-.000	.000	.022
Constant	11.806	.061	.000	16.390	.561	.000
Coin-Fixed-Effects	No			Yes		
Time-Fixed-Effects	No			Yes		
Observations	3,938			3,938		
Log Likelihood	-47,700.517			-41,914.264		

Note: Coin-fixed constants and time-fixed constants are left out for parsimony.

3.3.1.1.6 Results

Table 2 shows the results of this regression model and demonstrates the positive impact of greedy language on the number of new addresses created (IRR = 1.090, SE = .021, $t = 4.132$, $p < .001$). In other words, greedy communication about cryptocurrencies on weekly social media chatter is significantly associated with a 9.0% increase in cryptocurrency adoption. Additionally, a likelihood ratio chi-square test was conducted to compare the base model and the full model. The results indicated that the full model provided a significantly better fit than the base model ($\chi^2(341) = 11{,}572.510$, $p < .001$). Thus, even when considering a wider set of cryptocurrencies, we replicate the effect of greedy language on cryptocurrency adoption.

As shown in Table 2, we also find that anxiety has a negative effect on cryptocurrency, which might be evidence of its antagonistic role towards greed. This is in line with the findings of fear being an important factor in financial-decision making (Lee and Andrade 2011; Shefrin 2002). Arousal, valence and certainty, otherwise strong indicators for subsequent consumer engagement (Berger and Milkman 2012), are not found to be significant predictors in our X dataset. Volume and social media engagement remain significant and positive, in line with findings on earned social media (Colicev et al. 2018). Similarly, volatility remains significant and negative, contrary to findings in the literature that suggest higher trading activity with higher volatility (Weiss-Cohen et al. 2023).

1.1.1.3 Discussion

In two studies, we use text analysis of social media content (over 800,000 posts) and blockchain adoption data to explore the impact of greedy language on cryptocurrency adoption. In Study 1A we find that a one-unit increase in greedy language is associated with a 5.5% increase in daily new Bitcoin addresses. In Study 1B, with an extended dataset of 20 cryptocurrencies on X, we find that a one-unit increase in greedy language translates to a 9.0% increase in new addresses. Taken together, we find converging correlational evidence that greedy language is positively associated with an increase in cryptocurrency adoption. In the next two studies, we provide causality of this effort, and probe the underlying mechanism.

3.3.2 Study 2: Manipulating Greedy Language

In Study 2, to provide causal evidence of the effect of greedy language on social media on cryptocurrency adoption, we sought to replicate the results of the two field studies in a more controlled, laboratory setting. With this goal, we manipulated the level of greedy language in social media posts and measured consumers' willingness to invest in cryptocurrency. Additionally, we also consider the amount consumers are willing to invest in cryptocurrency as a second dependent variable. Lastly, since X has previously been used as a medium to manipulate cryptocurrency market movements (Ante 2023), in line with prior research, we control for the perceived trustworthiness (Ahluwalia and Burnkrant 2004), and the perceived intent of the posters (Ohanian 1990). We predicted that exposure to more versus less greedy language would increase consumers' willingness to invest in cryptocurrencies, as well as the amount they would invest in cryptocurrencies.

1.1.1.4 Method

Three hundred and forty-six participants recruited through CloudResearch Connect participated in this pre-registered study (https://aspredicted.org/z2pc-pjbj.pdf) for monetary compensation. Responses from 31 participants who failed an attention check were removed, as pre-registered, leaving 315 participant responses for analysis (51.1% female, 47.6% male, 1% non-binary, 0.3% other, M_{age} = 38.8). We employed a two cell between-subjects design (content: greed vs. control). First, participants were randomly assigned to one of two content conditions and asked to imagine viewing their X (formerly Twitter) timelines that consisted of four tweets. In the greed condition, participants viewed four tweets about cryptocurrency that contained similarly high levels of greedy language, as verified by the greed descriptive dictionary (Mercadante, Tracy, and Götz 2023). In the control condition, participants viewed four tweets by the same four posters but that contained content about everyday topics. Further, in this condition, to replicate consumers' actual daily social media scrolling viewership that likely consists of a variety of content and not just cryptocurrency, only one of the posts mentioned cryptocurrency, and all posts were void of greedy language (see Figure 2).

Next, all participants imagined they had \$500 to invest and reported their interest in investing in cryptocurrencies and likelihood of opening a cryptocurrency account (1 = Not at all, 7 = Very; Keller and Siegrist 2006, α = .96), which we combined into our focal dependent variable. Then, participants indicated the amount they would be willing to invest in cryptocurrency using a slider scale that ranged from \$0 to \$500. As potential controls, participants then indicated the perceived intent of the posters across three items (bad, pushy, aggressive; Ahluwalia and Burnkrant 2004; $\alpha = .65$), as well as how trustworthy they perceived the posters to be (reliable, honest, dependable, sincere, trustworthy; Ohanian 1990; $\alpha = .96$). Finally, participants responded to a manipulation check before providing demographics.

Figure 2. *Tweets used in Study 2*

Treatment Condition	**Control Condition**
	

1.1.1.5 Results

Manipulation check. Tweets in the greed condition were perceived as more greedy than those in the control condition ($M_{Greed} = 6.16$, SD = 2.11 vs. $M_{Control} = 2.72$, SD = 1.72, $t(313) = 15.83$, $p < .001$, $d = -1.784$). Additionally, as an additional check on the manipulations, we measured the percentage of greedy language in each condition using text-analysis. The results of the text analysis provide additional support that the greed condition contained a greater percentage of greedy language than the control condition ($M_{Greed} = .12$, $SD_{Greed} = .08$ vs. $M_{Control}$

= .00, $SD_{Control}$ =.00, $t(6) = 2.98$, $p = .025$). We also applied independent t-tests to verify that there were no other relevant language confounds (all p's > .14).[8]

Investment likelihood. Results of an independent sample t-test indicated participants were more likely to invest in cryptocurrency in the greed versus control condition ($M_{Greed} = 3.17$, SD = 2.01 vs. $M_{Control} = 2.50$, SD = 1.56; $t(313) = 3.32$, $p < .001$, $d = -.317$), as expected. We also ran a one-way ANCOVA and included controls for perceived intent and trustworthiness of the tweeters. Results for investment likelihood remained significant ($F(1, 311) = 39.104$, $p < .001$, $\eta_p^2 = .112$).

Investment amount. Results of an independent sample t-test indicated participants sought to invest more money into cryptocurrency in the greed versus control condition (M_{Greed} = \$128.11, SD = \$137.95 vs. $M_{Control}$ = \$78.35, SD = \$97.48; $t(313) = 3.70$, $p < .001$). We also ran a one-way ANCOVA and again controlled for perceived intent and trustworthiness of the posters. Results for investment amount remained significant ($F(1, 311) = 33.715$, $p < .001$, $\eta_p^2 = .098$).

1.1.1.6 Discussion

Building on the results of Studies 1A and 1B, in this study we find causal evidence that participants are more interested in investing in, and invest more money in, cryptocurrency after observing greedy (vs non-greedy) language in social media posts. While this study demonstrates the effect of greedy language on interest in investing in cryptocurrency, the control condition contains non-cryptocurrency, everyday tweets. Thus, it is possible participants were more interested in investing in cryptocurrency not because they perceived cryptocurrency as more profitable, but because the lack of cryptocurrency related content reduced their interest in investing in cryptocurrency. To address this, in Study 3 we maintain cryptocurrency as a focal topic in both the greed and control conditions.

[8] We again also checked via text analysis that there were no significant differences in anxiety, valence, arousal, certainty or concreteness between our experimental conditions, in line with our field studies.

3.3.3 Study 3: Mediation

The objectives of Study 3 are threefold. First, we aim to provide evidence for our proposed mechanism. We hypothesized that exposure to greedy language in social media posts increases how profitable consumers perceive investing in cryptocurrency to be. Thus, in this study, we measure perceived profitability. We also consider, and rule out, the possibility that greedy language affects consumers' perceptions of the risks involved in, and potential for losing money from, investing in cryptocurrency. Second, we seek to replicate the effect of greedy language with a stronger manipulation with greater internal validity. In Study 2, our greedy language manipulation was employed through varying the content type across conditions. Specifically, in our control group, one post was about cryptocurrency, while the other three posts were about everyday topics, which provided external validity. In this study, we keep the cryptocurrency content constant between conditions and manipulate greedy language without changing the content, which provides greater internal validity. Third, we explore potential theory driven moderators. Investment decisions might be influenced by how knowledgeable consumers are with financial instruments (e.g. stocks and bonds), with more knowledgeable consumers potentially being more willing to invest in cryptocurrency and to invest higher amounts. Therefore, we measure subjective financial knowledge in this study (Netemeyer et al. 2024). Additionally, given cryptocurrency's prevalence on social media, we consider how susceptible consumers are to interpersonal influence (Bearden, Netemeyer, and Teel 1989). Lastly, it's possible that consumers who are inherently greedier might show even stronger inclinations to invest in cryptocurrency when exposed to others' greed. We thus measure dispositional greed as our third possible individual difference moderator (Seuntjens et al. 2015)

1.1.1.7 Method

Three hundred and thirty-one participants recruited through CloudResearch Connect participated in this pre-registered (https://aspredicted.org/8tjz-5q64.pdf) study for monetary compensation. Responses from 34 participants who failed an attention check were removed, as pre-registered, leaving 299 participant responses for analysis (47.2% female, 52.5% male, .3% Other / prefer not to say, M_{age} = 37.6). As in Study 2, we employed a two cell between-subjects design (content: greed vs. control). First, after being randomly assigned to condition, participants imagined viewing their X (formerly Twitter) timelines that consisted of four tweets. In the greed condition, participants viewed the same four greed tweets used in Study 2. These tweets contained similarly high levels of greedy language as verified by the greed descriptive dictionary (Mercadante et al. 2023). In the control condition, participants viewed four tweets by the same four posters but that contained content about cryptocurrency yet were void of greedy language, also as verified by the greed dictionary (M_{Greed} = .12, SD_{Greed} = .08 vs. $M_{Control}$ = .01, $SD_{Control}$ = .02, $t(6) = 2.61$, $p = .040$). We applied independent sample t-tests to verify that there were no other relevant language confounds (all p's > .08).[9]

Next, participants reported their interest in investing in cryptocurrency, their likelihood of opening an account to invest in cryptocurrency, and their interest in learning more about investing in cryptocurrencies (1 = Not at all, 9 = Very; Keller and Siegrist 2006, α = .96), which we combined into our focal dependent variable. Then, participants imagined they had $500 available to invest and indicated the amount they would be willing to invest in cryptocurrency using a slider scale that ranged from $0 to $500. Then, for our proposed mediator, we measured how likely participants thought it was that cryptocurrency would to increase in value over the next few years, that they would profit from investing in cryptocurrency (1 = Not at all likely, 9 = Very likely) and that they would lose money from investing in cryptocurrency (1 = Not at all likely, 9 = Very likely; reverse scored prior to analysis). Participants also indicated how risky they thought it was to invest in cryptocurrency (1 = Not at all risky, 9 = Very risky), and how volatile they thought the cryptocurrency market was (1 = not at all volatile, 9 = very volatile). As potential controls, participants then indicated the perceived intent of the posters across three items (bad, pushy, aggressive; Ahluwalia and Burnkrant 2004; $\alpha = .61$), as well as how trustworthy they perceived the posters to be (reliable, honest, dependable, sincere, trustworthy; Ohanian 1990; $\alpha = .95$)

As potential individual difference moderators, participants then completed the dispositional greed (Seuntjens et al. 2015; α = .90), subjective financial knowledge (Netemeyer et al. 2024; α =.94), and susceptibility to interpersonal influence scales (Bearden 1989; α =.94). Finally, participants responded to a manipulation check ("How greedy would you rate the

[9] We checked that there were no significant differences in anxiety, valence, arousal, certainty or concreteness between our experimental conditions, in line with our field studies.

tweeters of the tweets you saw previously?"; 1 = Not at all greedy, 9 = Very greedy) and provided demographics.

Figure 3. Tweets used in Study 3

Control

Treatment

1.1.1.8 Results

Manipulation check. Tweets in the greed condition were perceived as greedier than those in the control condition (M_{Greed} = 6.03, SD = 2.08 vs. $M_{Control}$ = 4.35, SD = 2.05, $t(297) = -7.02$, $p < .001$, $d = -.812$). Additionally, we measured the percentage of greedy language in the stimuli tweets using text-analysis (M_{Greed} = .12, SD_{Greed} = .08 vs. $M_{Control}$ = .01, $SD_{Control}$ = .02) and verified that there were no other textual confounds[10].

Investment likelihood. An ANCOVA revealed that participants were more likely to invest in cryptocurrencies in the greed versus control condition (M_{Greed} = 4.40, SD = 2.77 vs. $M_{Control}$ = 4.12, SD = 2.64; $F(1, 295) = 14.13$, $p < .001$, $\eta_p^2 = .046$), controlling for perceived intent and

[10] As in Study 2, we checked that there were no significant differences in anxiety, valence, arousal, certainty or concreteness between our experimental conditions (all *p*'s > .172).

trustworthiness of the tweeters (Perceived intent: $F(1, 295) = 7.052$, $p = .008$, $\eta_p^2 = .023$; Trustworthiness: $F(1, 295) = 195.76$, $p = < .001$, $\eta_p^2 = .400$).

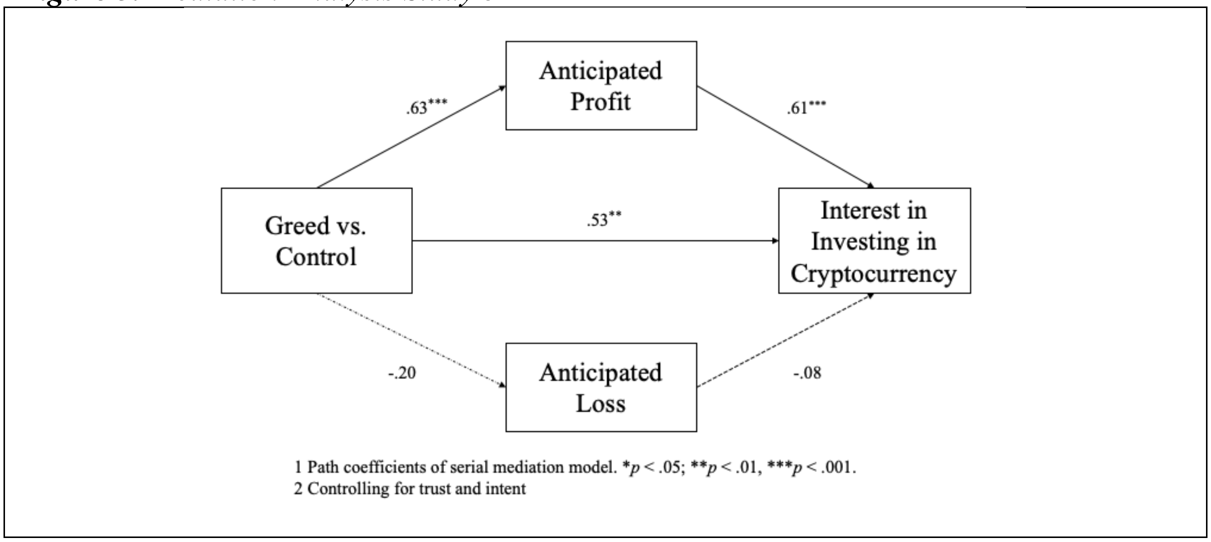
Investment amount. An ANCOVA revealed that the participants invested a higher amount in the greed versus control condition (M_{Greed} = \$140.20, SD = \$133.98 vs. $M_{Control}$ = \$117.29, SD = \$122.42, $F(1, 295) = 16.00$, $p < .001$, $\eta_p^2 = .051$), controlling for perceived intent and trustworthiness of the tweeters (Perceived intent: $F(1, 295) = 1.106$, $p = .294$, $\eta_p^2 = .004$; Trustworthiness: $F(1, 295) = 116.67$, $p < .001$, $\eta_p^2 = .283$).

Perceived Profitability. An ANCOVA revealed that the participants perceived cryptocurrencies as more profitable in the greed versus control condition ($M_{Greed} = 5.07$, SD= 2.30 vs. $M_{Control} = 4.93$, $SD = 2.38$, $F(1, 295) = 7.77$, $p = .006$, $\eta_p^2 = .026$), controlling for perceived intent and trustworthiness of the tweeters (Perceived intent: $F(1, 295) = 8.062$, $p = .005$, $\eta_p^2 = .027$; Trustworthiness: $F(1, 295) = 149.640$, $p < .001$, $\eta_p^2 = .337$).

Mediation – Interest in Investing in Cryptocurrency. We conducted a mediation analysis using 5,000 bootstrapped resamples using PROCESS Model 4 (Hayes 2022). As expected, the indirect effect through perceived profitability was significant (ab = .39, SE = .14, 95% CI: [.12, .67]). Further, the indirect effect through perceived loss was nonsignificant (ab = .02, SE = .02, 90% CI: [-.01, .06]). There was also significant direct effect of greed on cryptocurrency investment interest (b = .53, t(297) = 2.56, p = .011, ηp2 = .XX). Thus, consistent with our expectations, the greed (vs. control) condition increased interest in investing in cryptocurrency, and this occurred because participants perceived investing in cryptocurrency to be more profitable.

Mediation – Cryptocurrency Investment Amount. We conducted a mediation analysis using 5,000 bootstrapped resamples using PROCESS Model 4 (Hayes 2022). As expected, the indirect effect through perceived profitability was significant (ab = \$18.59, SE = \$6.90, 95% CI: [\$5.60, \$32.68]). Further, the indirect effect through perceived loss was nonsignificant (ab = .83, SE = 1.33, 90% CI: [-.60, 3.55]). There was a significant direct effect of greed on cryptocurrency investment amount (b = \$32.05, t(297) = 2.89, p = .004). Thus, consistent with our expectations, the greed (vs. control) condition increased of the amount participants sought to invest in cryptocurrency, and this occurred because participants perceived investing in cryptocurrency to be more profitable.

Figure 3. *Mediation Analysis Study 3*



Potential Moderators. We modeled separate regressions to evaluate the effect of greed condition and each potential individual difference moderator on investment likelihood and investment amount. None of the interactions were significant (all *p*s > .10). See appendix for full results.

Riskiness and volatility. We ran two independent sample t-tests to determine whether participants viewed investing in cryptocurrency as riskier, or more volatile between conditions. Results revealed no difference in risk perceptions ($p > .73$) or volatility ($p > .74$) between conditions.

1.1.1.9 Discussion

Exposure to greedy content makes the profitability of cryptocurrencies more salient, leading to a higher likelihood to invest in cryptocurrencies along with an increased amount that consumers plan to invest, confirming our hypothesis that the impact of greed on investment behavior is mediated by profit anticipation. Additionally, we find that the effect of greed on cryptocurrency investment likelihood and amount is stronger for those that are greedier by nature (i.e. score high on dispositional greed survey).

3.4 GENERAL DISCUSSION

Cryptocurrency as an investment method is growing in popularity around the world. Yet little is known about how consumers decide to invest in cryptocurrency. Across four studies, including two field studies and two lab experiments, we show that greedier language on social media increases consumers' adoption of cryptocurrency (Studies 1a and 1b). This effect is not limited to social media scrolling that contains only cryptocurrency related posts, as non-cryptocurrency related posts that are devoid of greedy language also curtail consumers' interest in investing in cryptocurrency (Study 2). Further, observing greedier language on social media not only increases consumers' willingness to invest in cryptocurrency, but also the amount they are willing to invest in cryptocurrency (Study 2). The effect is explained, at least in part, by an increase in how profitable consumers perceive investing in cryptocurrency to be (Study 3).

3.4.1 Theoretical Contributions

Our research has important implications for theory and practice. First, we advance research that considers the role of social media on consumer decision making by demonstrating that greedy language in social media affects consumer investment in cryptocurrencies. While previous work has examined the role that social media can play in consumer spending (Chevalier and Mayzlin 2006; Trusov, Bucklin, and Pauwels 2009; Kumar et al. 2016; Liadeli, Sotgiu, and Verlegh 2023), we extend literature analyzing the impact of social media on financial products (Tirunillai and Tellis 2012). We do so by showing a significant effect of greedy language in social media on the number of new wallets created, which are needed to begin investing in cryptocurrency, as well as the amount consumers are willing to invest in cryptocurrency.

Second, we contribute to prior research on greed (Seuntjens, Zeelenberg, Van De Ven, et al. 2015; Lambie and Haugen 2019; Zeelenberg and Breugelmans 2022) and specifically to greed communication in social media (Mercadante, Tracy, and Götz 2023). While having a disposition of being greedy has been found to influence financial decision-making (Seuntjens et al. 2016), especially in a cryptocurrency context (Martin, Chrysochou, and Strong 2024), no study has looked at the impact of greed on social media on financial decision-making. Our field and experimental findings show that greedy language increases consumers' willingness to invest and the amount they invest in cryptocurrencies. Specifically, situational greed, such as when consumers observe others' talking about greed, motivates consumers to take important financial decisions. We see this as an important contribution not only to the financial domain, but also provides insights into how social media can be a powerful motivator for consumer financial decision-making.

Third, we contribute to literature on consumer financial decision making by demonstrating that greedy language use in social media shapes investment perceptions and decisions related to cryptocurrencies, extending the literature on the role of affective processes in consumer financial decision making. In particular, we extend research by Friederich et al. (2024) who show that consumers exposed to vignettes using fear of missing out (FOMO) appeals increase expected pleasure and reduce anticipated regret, which leads to more cryptocurrency investments. Beyond FOMO appeals, which focus on fear-based communication, we argue social media communication and interactions about cryptocurrency specifically, frequently contain greedy statements (e.g., just made some good money with bitcoin!) that influence consumers' decision-making through an increase in anticipated profit (Shin 2022). This anticipated profit then leads to higher adoption of cryptocurrency.

3.4.2 Managerial and Policy Implications

Our results demonstrate that greed, when communicated through social media, has a significant impact on consumers' decision-making. This is a worthwhile finding for who monitors and trades on the market for cryptocurrencies. While the cryptocurrency market has been almost exclusively populated by retail investors, institutional investors have started to participate in the market. Examples include BlackRock's launch of the first Bitcoin exchange traded fund (McGee, Lang, and Lang 2024) and American president Trump's 2025 creation of "Strategic Bitcoin Reserve" that started to buy up Bitcoin with the goal of making the United States "the crypto capital of the world" (The White House 2025). Closely monitoring social-media conversations could lead to deeper insights into cryptocurrency exchanges and for financial advisors on how the market evolves.

Our findings also hold implications for policymakers. While research has not ultimately understood whether greed leads to more rational or less rational decision-making (Seuntjens, Zeelenberg, Van De Ven, et al. 2015), greed has significant impacts on how consumers make financial decisions. We add to these insights and find that greedy language affects consumer perceptions about cryptocurrency profitability. This is important because Gen Z consumers, born between 1997 and 2012, are now more than four times more likely to own cryptocurrencies than to own a retirement account (YouGov 2025). Since poor financial decisions can have significant negative consequences for consumer well-being, policymakers looking to understand the role of social media in the consumer participation in cryptocurrency markets can leverage this research to take action to mitigate consumers' overexposure to risky assets (Greenberg and Hershfield 2019).

3.4.3 Future Research and Limitations

Even though our research evaluated the impact of greedy language in social media on cryptocurrency adoption, it is possible that the effect of greed is not limited to this new investment method. Further research might try to deepen our understanding regarding what drives investment behaviors in stocks, shares and bonds, as well as other recently developed investment classes (such as P2P and NFTs). While we observe adoption through the lens of blockchain data, we do not know how many different consumers are opening new addresses. To address this question, data from cryptocurrency exchanges could be exploited in the future.

Further, there is likely heterogeneity among the consumers who adopt cryptocurrencies. While many consumers adopt cryptocurrencies as an investment, others may use cryptocurrencies as a currency and thus as a means of exchange. Future research can analyze this by systematically analyzing wallet addresses that pertain to businesses, for example in e-commerce stores (Hanneke et al. 2024). Some CRM providers already offer a Web3 integration that aim to enable customers to connect their wallets with the vendor (Salesforce 2025), potentially offering a rich basis for customer analysis.

Lastly, future research might look to further our understanding of the role of language use in social media in consumer financial decision-making. In 2023, 39% of American consumers aged 22-64 reported making investment decisions based on information posted on social media platforms (Baird 2023). Some phenomena include, for example, the upcoming trends of FinTok (Finance TikTok) and the ongoing surge of influencers dispersing finance advice on other social media (Finfluencers). While we are looking at all social media users, such Finfluencers might exert more nuanced influences on consumers. Given the impact that financial decision-making can have on consumers' lives, it may be interesting to explore the language and communication strategies that finfluencers are using to drive investment and saving behavior among consumers.

CHAPTER 4: ADVANCING SOCIAL MEDIA BRAND ANALYTICS WITH GENERATIVE AI

ABSTRACT

Brands are in constant need to measure their performance. Both the increase in textual data over the last years and the advancements in text analysis have offered a unique opportunity to marketers to do so. But the approaches for brand analysis using text remain limited. One reason is the complicated process of creating validated dictionaries. In this research, the authors explore the generative capacities of GenAI to create context-specific dictionaries with the goal of improving on existing social media brand tracking methodologies. They use LLMs to extend and recreate brand reputation dictionaries from Rust et al. (2021) and validate them on a dataset of 245k tweets. They find that LLMs can be used to create context-specific brand dictionaries and validate the accuracy through agentic AI and human annotators. The resulting accuracy outperforms existing default dictionaries. Their research contributes to the emerging literature of AI in marketing and the literature on agentic AIs, text mining, and brand analytics.

4.1 INTRODUCTION

Brands are at the core of successful companies. In today's dynamic marketplace, accurately measuring brand metrics is essential to gain the necessary insights (Keller 2021). The massive amount of user generated content on social media platforms has set the stage for text-analysis based brand tracking methodologies (Nam, Joshi, and Kannan 2017; Rust et al. 2021; Zhang and Moe 2021). Among the various ways to measure brand attributes from text, dictionary-based models remain the most interpretable approaches for managers and academics (Berger et al. 2020). While this is a big advantage for adoption (Gross and Desveaud 2024), it is a two-sided sword: The creation and validation of accurate text-mining dictionaries is an cost and time-intensive process (Humphreys and Wang 2018). After creation, dictionaries need to be validated internally, e.g. by human annotators, and externally, for example through further experiments. This leads to a limited number of dictionaries and often requires researchers to create context-specific dictionaries. One solution could be generative AI that also offers the capabilities to create text for marketing applications (Reisenbichler et al. 2022).

Large language models (LLMs), for example, are now able to create website content or search engine ads that perform better human-created content (Reisenbichler et al. 2022; Reisenbichler, Reutterer, and Schweidel 2023). Those LLMs have been repeatedly shown to be able to recreate, even if not perfectly, the perceptions of humans (Arora, Chakraborty, and Nishimura 2025; Li et al. 2024; Stromberg et al. 2025). Therefore, one way to use the generative potential of large language models would be to use them to efficiently create custom dictionaries for specific contexts. While the analytical capabilities of LLMs have been used in marketing application such as sentiment classification (Krugmann and Hartmann 2024) and service failure and recovery (Villarroel Ordenes et al. 2025), only few studies have explored whether LLMs generative capabilities can be used to track brand metrics (Stromberg et al. 2025). No study, so far, has systematically explored the usage of generative AI to create dictionaries for text-analysis.

In this study, we exploit the generative capacities of LLMs in order to improve on brand analytics to create context-specific dictionaries. We create 286 dictionaries for 22 subdrivers of brand reputation using LLMs. We then use agentic AI to check their validity and triangulate with human annotators. To ensure validity of our results, we test our approach with 245.884 user-generated messages on 4 brands in the luxury context and show that our context-specific dictionaries have higher accuracy than the standard dictionary. Finally, we explore dynamics of our measurement methods using vector auto-regressive regressions.

This article makes several contributions. First, by developing an algorithm to develop context-specific dictionaries using LLMs, we contribute to literature on text-analysis and LLMs (Berger et al. 2020; Goli and Singh 2024; Li et al. 2024; Villarroel Ordenes et al. 2025). We extend on research that shows how LLMs can be used recreate consumer brand perceptions (Arora, Chakraborty, and Nishimura 2025; Brand, Israeli, and Ngwe 2023). Specifically, our research shows that LLMs can be valid way of creating context-specific dictionaries, striking a balance between the time-intensive process of dictionary-creation and the lack of interpretability that comes with advanced machine learning techniques, such as transfer learning models.

Second, we contribute to literature on brand and social media analytics and brand analytics (Rust et al. 2021; Fronzetti Colladon 2018; Colicev et al. 2018). While most brand analytics is either behind a commercial black-box approach (e.g. YouGov or RepTrak) or lack real-time capabilities, we show that researchers and managers can easily use a combination of text analysis and LLMs to calibrate and customize brand analytics.

Third, we contribute to literature on agentic LLMs and human ai collaboration(Arora, Chakraborty, and Nishimura 2025). Specifically, we show that LLMs can be used to automatically assess accuracy of dictionaries, and that this accuracy closely matches the accuracy of human annotators. This finding is valuable for researcher and managers that want to create customized dictionaries in their research. First, because hiring research assistants can be costly and second, because of the time that is needed to annotate and analyze the annotation.

4.2 RELATED LITERATURE

4.2.1 Measuring consumer mindset metrics using Machine Learning Techniques

Brand sentiment is one of the most popular consumer mindset measures. Sentiment can be used in analyzing customer reviews or interactions, or to analyze social media posts regarding brands. Sentiment has shown to be predictive of important firm outcomes. For example, Twitter sentiment on brands has been shown to be able to predict firms' stock returns (Tirunillai and Tellis 2012). Recent technical developments have helped to increase the accuracy of sentiment classification significantly, with transfer learning models achieving up to 90% accuracy on social media data (Hartmann et al. 2023). Recently, LLMs have also shown to be able to reach similarly high accuracies (Krugmann and Hartmann 2024). One drawback of sentiment classification are the limited outcomes. Common models are either based on two (positive or negative) or three (positive, neutral and negative) classes. Therefore, sentiment classification is but one piece in measuring consumer mindset metrics. While it is fundamental to understand whether conversations about brands are positive or negative, managers and researchers are interested in the underlying brand attributes connected to those sentiments.

Advancements in machine-learning, along with the increasing amount of user-generated data, has enabled researchers to develop techniques to measure brand attributes more granularly. For example, Nam, Joshi, and Kannan (2017), use topic-mining techniques (LDA) on social media tags to obtain brand associations (for example, Apple being used frequently with "innovation"). Using dynamic factor analysis (DFA), they show how those associations are subject to change over time. Another perceptual approach to understand brand attributes is by looking at a brand's personality (Aaker 1997). Pamuksuz, Yun, and Humphreys (2021) use machine learning techniques from topic modelling and word embeddings to develop an algorithm to automatically show brand personality scores (e.g. sincerity or sophistication). While those approaches are calibrated on social media data, brand perceptions can also be estimated through "silicon sampling". Silicon sampling describes techniques where LLMs "mimic human respondents to describe, explain, and predict human behavior "(Sarstedt et al. 2024, p. 1254).

LLMs are prompted to act like survey participants and are then asked to answer survey questions. For example, Li et al. (2024) show that LLMs that are prompted on brand perceptions, can closely match the answers of human participants. Similarly, Stromberg et al. (2025) find that LLMs answers closely track YouGov metrics (awareness, consideration, purchase intent, satisfaction and recommendation). While silicon sampling might be a cost-efficient method to use LLMs to track consumer mindset metrics, since they are pre-trained on rigid datasets, LLMs have difficulties reproducing developments of metrics over time, hindering real-time brand tracking. In other words, if the LLM was trained on a dataset until 2023, it cannot reproduce consumer mindset metrics for 2025. Taken together, most methods to measure brand attributes are either lacking the granularity, simplicity or actionability (See Table 1 for a review).

Recently, Rust et al. (2021) have developed a way to measure another important brand mindset measure, brand reputation, from text. Brand reputation is defined as the "overall impression of how stakeholders think, feel and talk about a brand" and is conceptually merging corporate reputation (Lange, Lee, and Dai 2011), customer-equity (Rust, Lemon, and Zeithaml 2004) and brand-equity (Keller 1993). The real-time brand reputation of a company relies on brand events. Brand events are events that potentially influence the brand's reputation, for example the launch of a new product or the expansion to a new market. Rust et al. (2021) propose a three-driver model to measure stakeholders' perceptions from text. It consists of a value-driver, a brand-driver and a relationship driver. To ensure managerial relevance, those three drivers are then divided in eleven sub-drivers. The value driver, for example, consists of subdrivers for price service quality and goods quality. They also show how brand reputation can eventually also impact the financial performance of a firm. While the brand reputation tracker by Rust et al. (2021) is based on rigid dictionaries, the requirements of a fast-moving consumer environment might require adaptation of dictionaries to recent developments. In the next section, we give an overview over how dictionaries are created and validated.

Figure 1. *Brand Reputation Conceptualization from Rust et al. 2021*

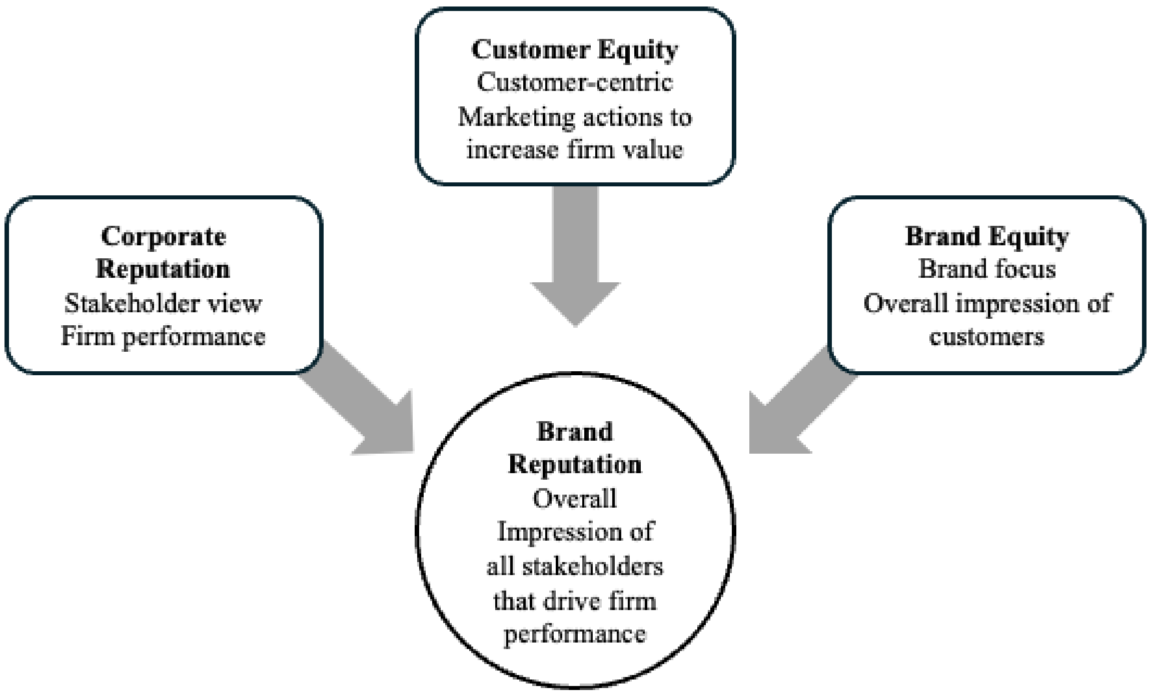


Table 1. *Summary of Literature on Machine Learning Techniques to measure Brand Attributes*

Paper	Construct	Method	Advantages	Disadvantages
Stromberg et al. 2025	Brand awareness, brand consideration, brand purchase intent, brand satisfaction, brand recommendation	Using SOTA commercial LLMs (GPT-4o and Gemini) to recreate consumer survey responses (YouGov).	Relatively easy and quick to implement.	Reduced accuracy in bottom of the funnel metrics and after purchase metrics.
(Krugmann and Hartmann 2024)	Sentiment	Testing commercial LLMs (GPT-3.5, GPT-4, Llama 2) to do automated sentiment classification	Better accuracy than transfer learning models on three class sentiments. Easy implementation.	Explainability of LLMs has to be specifically requested, transfer learning still outperforms LLMs in some tasks on some datasets.
(Hartmann et al. 2023)	Sentiment	Comparing multiple sentiment classification approaches and develop a fine-tuned transformer model (SiEBERT)	SiEBERT outperforms traditional machine learning and lexicon approaches in sentiment classification.	Lack of interpretability in transfer learning models, high hardware requirements for processing.
(Pamuksuz, Yun, and Humphreys 2021)	Brand Personality	Topic modeling using latent dirichlet allocation (LDA) and word embeddings to develop an algorithm (LDA2Vec) to analyze brand personalities from social media data.	Fast way to assess brand personality without reliance on surveys.	Limited predictive validity of construct for marketing measures.
Rust, Rand, M. H. Huang, et al. 2021	Brand Reputation	Dictionary-based (22 dictionaries) text analysis to measure brand-reputation from social media	High interpretability, low hardware requirements, predictive value.	Rigid dictionaries, limited accuracy out of existing samples.
(Zhang and Moe 2021)	Brand Favorability / Sentiment	Use probabilistic graphical models to model interactions of users and brands to create brand favorability scores that tracks industry measures (e.g. BrandZ Value, Brand Z Rank)	Good model performance and convergent validity with current industry approaches. Correction for positivity bias.	Complex to integrate
(Nam, Joshi, and Kannan 2017)	Brand Associations	LDA parameterized Gaussian finite mixture model to create brand perception maps from social media. Dynamic Factor Analysis (DFA) to analyze brand associations over time.	Brand associations are individual and not predetermined by researcher allowing for visualization and exploration.	Tags decreasingly used in social media

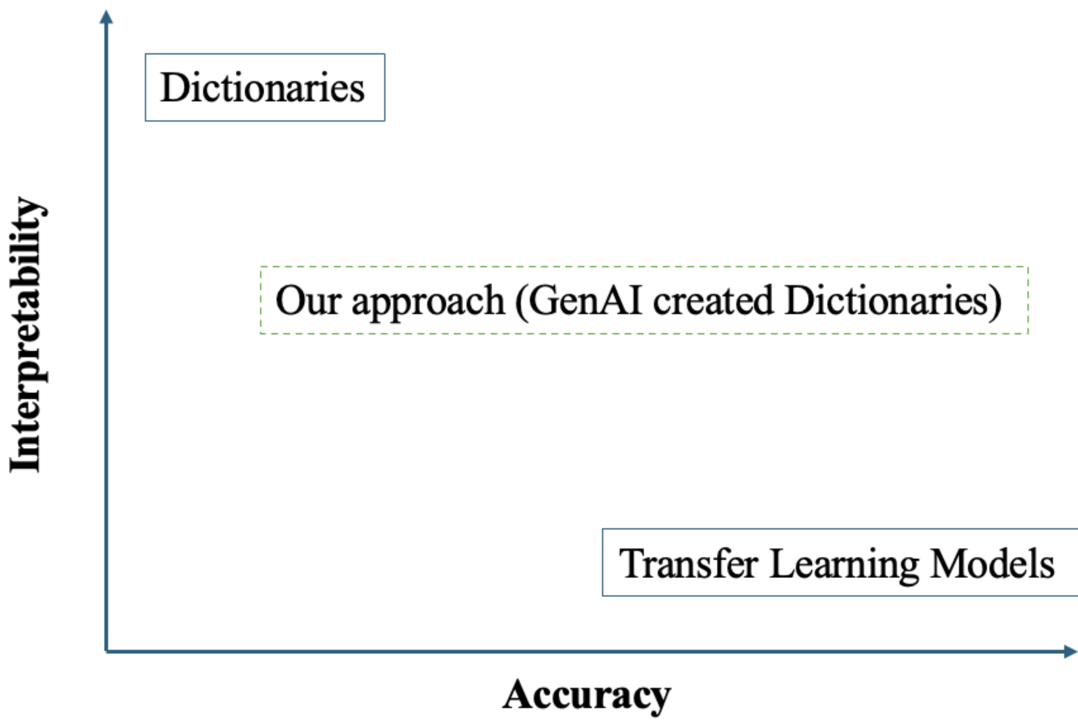
4.2.2 Using LLMs to create dictionaries for text-analysis

Researchers regularly use dictionaries for text-analysis (Berger, Sherman, and Ungar 2020; Boyd et al. 2022; Packard, Moore, and Berger 2023; Rocklage et al. 2023), for example to measure arousal in influencer posts (Cascio Rizzo et al. 2023; Mohammad 2018), to understand how far along a consumer is on her decision journey (Humphreys, Isaac, and Wang 2021), or to understand how certain a sender is about his message (Rocklage et al. 2023). Dictionaries are also used to measure brand-related constructs such as brand personalities (Opoku, Abratt, and Pitt 2006) or brand reputation (Rust et al. 2021). Dictionaries are usually composed by a number of words pertaining to a construct, for example all words that relate to greed, like “money”, “more”, “greedy” and so on. In this approach, all words have the same weight. Another approach is to score words on the scale of a construct, e.g. from 0 to 1, as in the case of the VAD, the valence-arousal-dominance dictionary (Mohammad 2018). If a researcher does not find a suitable dictionary for his purposes, he might decide to create his own. Dictionary creation can be achieved through one of two ways, as outlined by Humpreys and Wang (2018): by induction or by deduction. When using deduction, several strategies can be used: First, a researcher can analyze a set of documents manually and identify the most important words in the content, consequently forming a dictionary, for example with the help of a tool like Wordify (Cascio Rizzo et al. 2024; Hovy, Melumad, and Inman 2021). Another deductive way to create a dictionary is to mark-up words from a document that are important, so to create different important categories. Last, the researcher could use a topic-modelling approach to identify several topics among a corpus of documents, and then use the most frequent words to form dictionaries, pertaining to the categories (Blei, Ng, and Edu 2003). Inductive creation of dictionaries works the other way around. One way is to query subjects (e.g. experts or customers) on words that they associate to a certain topic or to note down the words that come to their minds when thinking about an experience. Yet another way is to start with a theory-driven definition of a construct and then ask for words that come to their mind. The resulting lists then need to be validated. As this review shows, in most cases, dictionary creation and validation are a laborious and time-consuming process.

Recent research has shown that LLMs are capable of reproducing human perceptions based on how they are prompted (Li et al. 2024) and may also be functional to capture human preferences (Goli and Singh 2024), making them an interesting substitute to human participants when creating dictionaries. Our goal is to simplify dictionary creation with the help of LLMs, while still ensuring validity of the resulting dictionaries. Given LLMs capacities of imitating human perception and preferences, we want to explore whether they can adjust existing dictionaries or create new ones (Goli and Singh 2024; Li et al. 2024). We see several advantages in this approach.

First, creating dictionaries using LLMs can be a time- and cost-effective way to creating dictionaries either through focus groups or corpus-based topic modelling. For example, in order to create a dictionary from scratch (inductively), at least 50 participants are involved to create an initial set of keywords. With LLMs, we can simplify the process through repeated queries, significantly reducing the time. Second, in text-analysis, keyword-based text analysis has a better interpretability than other models, for example, transformer models (Krugmann and Hartmann 2024). This is an important point in closing the research-practice gap, since explainability remains a barrier for integration of NLP models within companies (Gross and Desveaud 2024). Similarly, researchers, who want to understand in depth what drives the underlying decisions (Berger et al. 2020), might prefer dictionary-based methods over black box approaches. As an example, let us take the case of post that has been tagged positive in sentiment: With dictionary-based methods, a researcher can pinpoint the words (e.g. good, awesome, nice) that have led to this classification, whereas a transfer learning models, such as BERT, would not give a concrete explanation (See Figure 2).Third, LLMs and other transformer models, especially state-of-the art models are time intensive. Training a transformer model on a classification task requires significant time. While there are available for classification tasks, e.g. SiEBERT (Hartmann et al. 2023), that reaches high precision with sentiment classification, there are, so far, no classifiers for brand attributes. More so, there are no test-sets readily available for training. Last, our dictionary approach could be applied without reliance on cloud-hosted models for sentiment analysis and validation and could therefore avoid that customer data (e.g. reviews) is transmitted elsewhere. This is important if there are potential privacy issues that can arise (e.g. in the EU through the GDPR).

Figure 2. *Trade-off between Interpretability and Accuracy*



Note. Dictionaries generally lack the accuracy of specifically trained models, but they are easy to understand and actionable for managers. They are easily integrated and replicable for researchers.

Taken together, we argue that our approach unites the advantages of interpretability and speed of dictionary-based text analysis with the perceptual and generative capacities of LLMs. In the next section we describe our prompting strategy, the automated dictionary generation and validation.

4.3 CURRENT RESEARCH

4.3.1 Data and Setting

For the application of our custom created dictionaries, we chose the luxury setting. We do so because luxury brands are generally perceived differently to main-stream brands (e.g. while “cheap” is positively connotated for the price of a mainstream brands, a luxury brand is deemed “valuable”) and therefore offer a rich environment for exploration. Stakeholders (e.g. existing and future customers, investors) often turn to social media sites to express their feelings and opinions regarding brands, a phenomenon known as online WOM (Hewett et al. 2016). For our data collection, we therefore turn to one of the most popular social media sites: Twitter. We collect data on five distinct luxury brands: Dior, Hermes, Prada, and Burberry. While we were able to mine Hermes, Prada and Burberry using the academic Twitter access, we had to acquire data for Dior from a social media intelligence company. We collect data for the timeframe of 2020 to March 2023, resulting in 245.884 tweets.

4.3.2 Prompt Strategy and Scenarios

We use LLMs to create custom-created dictionaries based on prompts. Our goal is to chart a way that is easy to use and replicate, while it adheres to scientific rigor. We first start with a simple extension of the Rust et al. (2021) dictionary. Their approach is based on 22 dictionaries, one positive and negative for each subdriver (See Table W.3.1 for an overview of the default dictionary). To explore the capabilities of LLMs in dictionary creation, we focus on three approaches. First, we aim to simply extend the existing dictionary, keeping the keywords and adding new keywords to it. Second, we regenerate keywords based on the existing categories (e.g. quality, innovativeness and trustworthiness).

Rust Extended - Extension Dictionary - Scenario 1. Our goal is to improve on the performance of dictionary-based text analysis while keeping the need of human intervention minimal. The Rust et al. (2021) default dictionary contains an overall of 185 words over the eleven subcategories. Our first strategy is to expand the current dictionaries with new words, adapting the dictionaries to the desired context, in our case luxury.

New Luxury - Updated Dictionary - Scenario 2. Keeping the base dictionary from Rust et al. (2021) might limit bias our measurement too much towards the original dataset and its standard brands. Therefore, we want to use the generative capacities of LLMs to generate customized dictionaries on the context. While we keep the original sub-dimensions of the brand-reputation tracker, we seek a fresh-start approach by creating context-specific (i.e. luxury) dictionaries from scratch.

4.3.3 Implementation

We use KNIME analytics with the OpenAI API to query OpenAI LLMs. We chose OpenAI because their models are regarded as highly performant in text generation tasks and have been used in the past (Li et al. 2024; Reisenbichler et al. 2022; Krugmann and Hartmann 2024). LLMs are constantly improved by their developers. OpenAI, for example, has released more than 10 different models between the launch of ChatGPT in November 2022 and Mid 2025. While models are generally improving over time, we want to test multiple versions. Since they are trained with different datasets, even though the details are not disclosed, the output might differ. We decide to test GPT4.5, the most recent model, GPT4o, GPT4.1, and GPT4o mini. One of the parameters of LLMs is its temperature. When the temperature is "low", so reduced to 0, LLMs will always return the same prompt. When the temperature is "high", LLMs will display greater variety in their output. This is sometimes termed as "creativity". We exploit this hyperparameter functionality by testing two settings: the standard temperature of 0.2, and an increased temperature of 1, "high temperature" hereafter.

We prompt the LLMs creating a role, task, format (RTF) prompt. RTF prompts have been shown to increase accuracy for LLM responses (Li et al. 2024). To impose a role on the LLM, we make use of the developer message functionality. Those are instructions that are prioritized by the model, before the rest of the prompt (OpenAI 2025a). We form a prompt for every subdriver polarity combination, e.g. the positive dictionary for the subdriver "price". For *Scenario 1*, we give the model the instruction about the industry, the subdriver to extend, and the existing dictionary. Therefore, we adapt the system message in the following way: "You are a researcher and need to keep and extend this dictionary to luxury brands: *Price*. It is described by the question: *Is the brand known for low prices, such as being cheap, affordable, having deals, bargains, discounts, and sales?* the words of the dictionary are: *Cheap, afford, inexpens, deal, low, bargain, thrifti, reason, econom, frugal, joy, discount, pleas, sale*". The next part of our prompt is the task, and we explicitly state the polarity, positive or negative and the language style used: “Keep the original words and generate 5 *positive* words additionally. Use standard and social media language.” Last, we define the format: “No compound words. Output all words in JSON as a list called 'updated_words'”.

Table 2. *Prompt examples*

Scenario	Subdriver	Developer message (Role)	Prompt (Task + Format)
Scenario 1	Price	You are a researcher and need to keep and extend this dictionary to luxury brands: Price. It is described by the question: Is the brand known for low prices, such as being cheap, affordable, having deals, bargains, discounts, and sales? the words of the dictionary are: Cheap, afford, inexpens, deal, low, bargain, thrifti, reason, econom, frugal, joy, discount, pleas, sale	Keep the original words and generate 5 positive words additionally. Use standard and social media language. No compound words. Output all words in JSON as a list called 'updated_words'.
Scenario 2	Service quality	You are a participant in a survey.	Think about a scenario where you recently interacted with a luxury company and had a positive experience with a brand about Service quality. Please list five keywords that you would use to describe this scenario. Use standard and social media language. No compound words. Output all words in JSON as a list called 'updated_words'.

Next, we repeat the prompt of a sub-driver dictionary 50 times, prompting the LLM to create five words, e.g. "Think about a scenario where you recently interacted with a luxury company and had a positive experience with a brand about Service quality. Please list five keywords that you would use to describe this scenario. Use standard and social media language. No compound words. Output all words in JSON as a list called 'updated_words'.". To facilitate data-processing, we configure the LLM to return answers in JSON format, which gives JSON formatted tables as output. We then process the JSON and stem the words (Porter 1980). We keep keywords that were mentioned at least three times (Rust et al. 2021). We repeat this procedure with a variety of models and prompts (see Figure 2). This results in 1100 (22 subdriver dictionaries x 50 repetitions) prompts per scenario with 5 keywords each. The final number of keywords per set of dictionaries is therefore 5500. Since keywords might be mentioned multiple times for different subdrivers, we assign the keywords to the subdriver where it was mentioned most often. We summarize the results of the different models in Table 3.

Figure 2. *Creation of Dictionaries using LLMs*

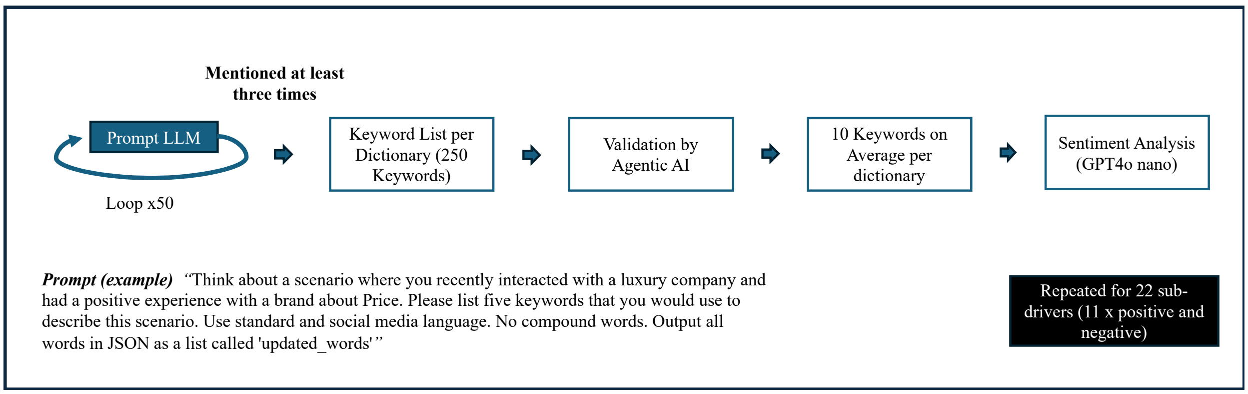


Table 3. *Created Dictionary Results of Prompting the LLM*

Prompt Scenario	**Model**	**Temperature**	**Keywords**
Default Dictionary	None	None	185
Scenario 1	GPT-4.1	0.2	281
Scenario 1	GPT-4.1	1	358
Scenario 1	GPT-4o	0.2	291
Scenario 1	GPT-4o	1	437
Scenario 2	GPT-4.1	0.2	96
Scenario 2	GPT-4.1	1	146
Scenario 2	GPT-4o	0.2	100
Scenario 2	GPT-4o	1	164
Scenario 2	GPT-4o mini	1	114
Scenario 2	GPT-4.5	0.2	110
Scenario 2	GPT-4.5	1	139

Note. GPT4o mini on default temperature did not return complete dictionaries. Due to high costs, we prompted GPT4.5 only for Scenario 2

We create 12 different set of dictionaries based on two scenarios. First, by extending the current dictionary to the luxury (scenario 1), and second, by keeping the subdrivers and prompting the LLM to recreate the dictionary for the luxury context (scenario 2). Higher temperature seems to result in a larger variety of keywords. We next integrate our created dictionaries with the existing brand reputation methodology and assess accuracy.

4.3.4 Brand Reputation Tracker Integration

Data Cleaning. To clean our data, we first filter for tweets that consists of text only, filtering out images and videos. We also filter for tweets that only mention one of the luxury brands in our dataset and remove duplicate tweets. Additionally, we filter for tweets in English. The resulting dataset includes 124.952 tweets in the timeframe from 2020 to March 2023.

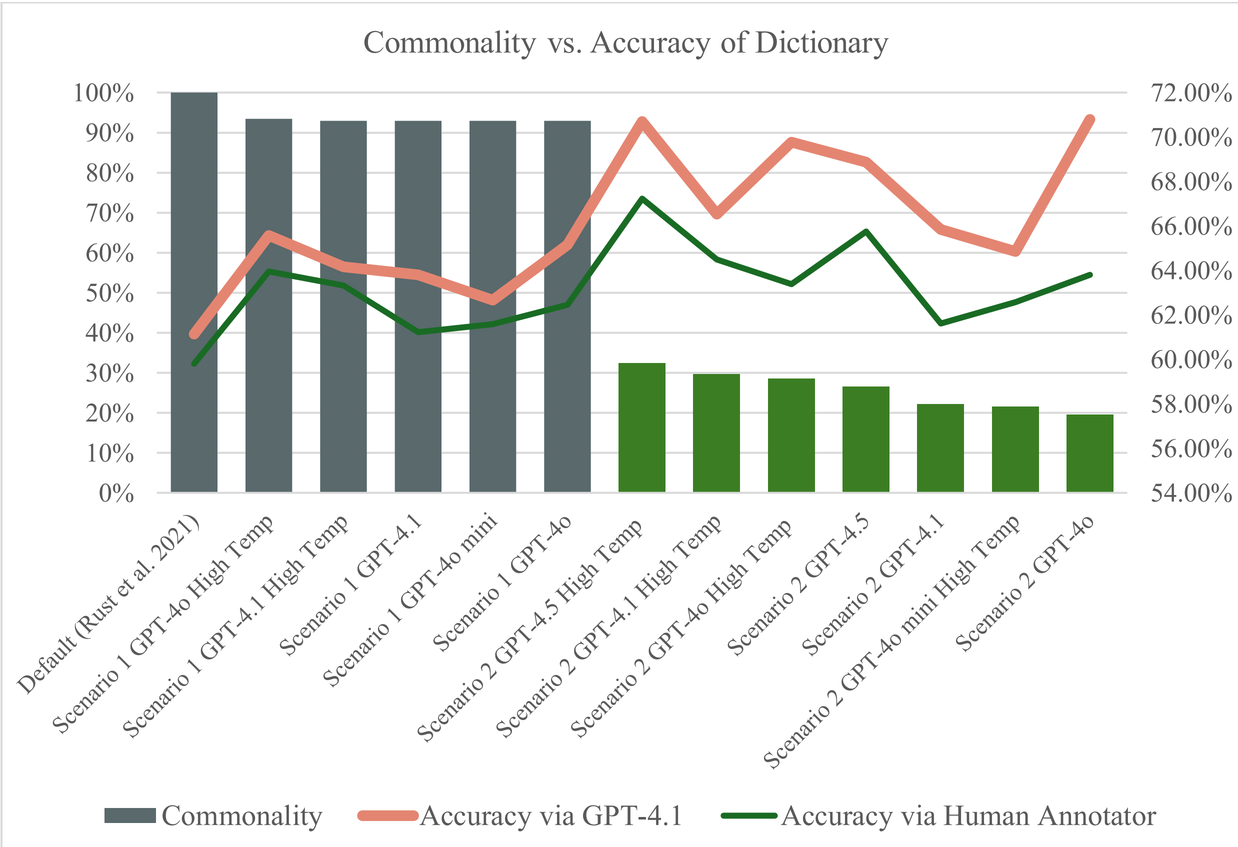
Sentiment. One disadvantage of dictionary-based text analysis is that it does not take the word context into account. We address this by analyzing sentiment using a LLM to classify our tweets into positive or negative sentiment (Krugmann and Hartmann 2024). We use GPT4o nano, due to its comparably low cost and high speeds (OpenAI 2025b). To maximize throughput, we chunk our tweets in batches of 10.000. We prompt the LLM with "*Classify the sentiment in this tweet. Only output one word: positive or negative*" followed by the tweet. Following Krugmann and Hartmann (2024), we validated GPT4o nano on the "Twitter (Airlines I)" – dataset, achieving 95.51% accuracy.

Data Preprocessing. To facilitate text-analysis, we stem the tweets using a porter stemmer (Porter 1980) and convert them to lower-case. We then apply a bag-of-words model to facilitate dictionary matching.

Dictionary Operationalization. We use our 22 dictionaries to score each tweet. To increase accuracy, we use the results from our sentiment analysis so that we only classify positive sentiment tweets with positive dictionaries and vice versa for negative words. We then aggregate the scores per tweet. To obtain our brand reputation scores, we substract the negative scores from the positive scores and aggregate the results by brand and week.

Dictionary Validation. To assist in the validation of our dictionaries, we turn to LLMs again. From our classified tweets, we draw a random subsample of 30 tweets for each dictionary. We balance the amount of classified and unclassified tweet to obtain balanced subsamples. We then use GPT-4.1 to validate the results, we set the temperature to the lowest value (0) to obtain reproduceable results. We graphically show the results in Figure 3 and report detailed results in Table 4.

Figure 3. *Commonality and Accuracy of Dictionaries*



Note. We compare the share of words that our created dictionary has with the original Rust et al. (2021) dictionary and measure the accuracy on a random subset of our data, with GPT-4.1 as a judge.

Table 4. *Dictionary Validation*

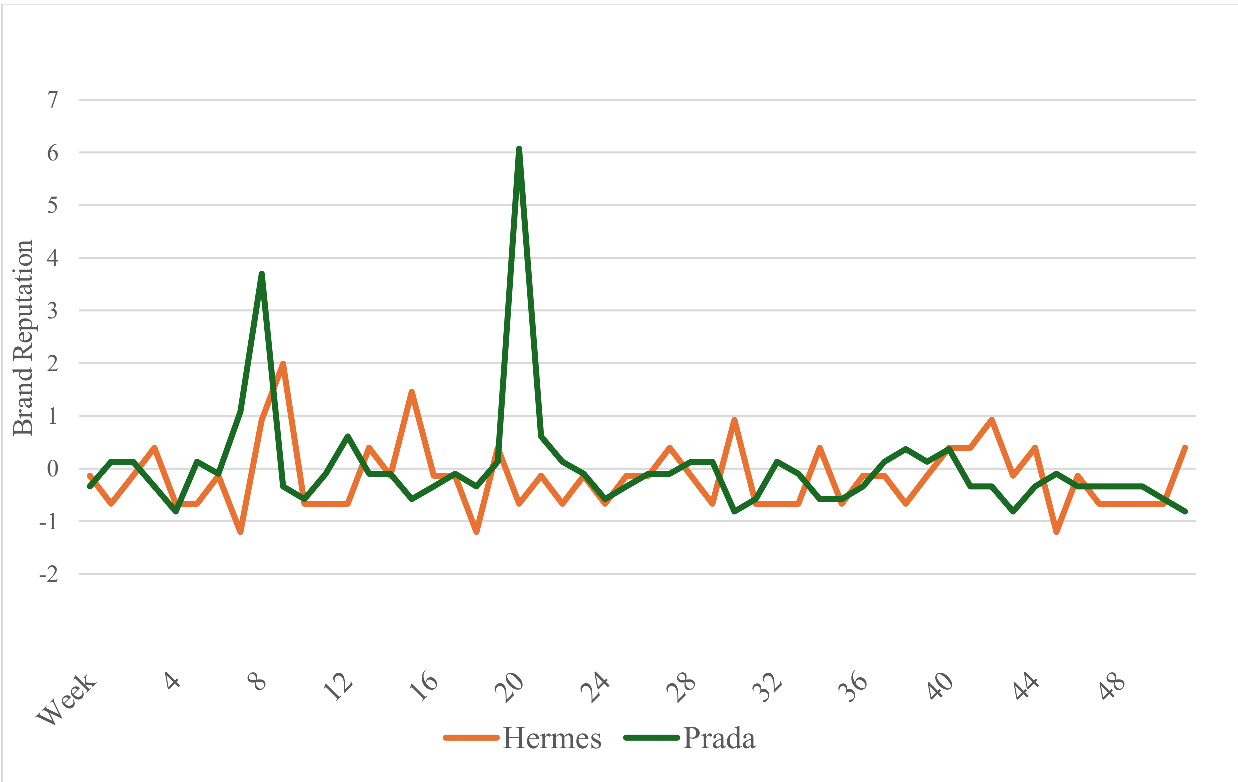
Scenario	Model	Temperature	Commonality	Accuracy via LLM	Accuracy via Human Annotator	Average Accuracy
Default	-	-	100.00%	59.91%	60.33%	59.80%
1	GPT-4o	1	93.50%	62.70%	62.21%	63.94%
1	GPT-4.1	1	93.00%	59.72%	60.90%	63.32%
1	GPT-4.1	0.2	93.00%	65.24%	60.66%	61.23%
1	GPT-4o mini	0.2	93.00%	60.19%	61.39%	61.57%
1	GPT-4o	0.2	93.00%	63.72%	61.39%	62.45%
2	GPT-4.5	1	32.40%	67.49%	**66.80%**	**67.23%**
2	GPT-4.1	1	29.70%	66.98%	63.20%	64.50%
2	GPT-4o	1	28.60%	68.82%	59.92%	63.40%
2	GPT-4.5	0.2	26.50%	**69.05%**	65.82%	65.76%
2	GPT-4.1	0.2	22.20%	65.33%	60.66%	61.61%
2	GPT-4o mini	1	21.60%	63.42%	63.20%	62.57%
2	GPT-4o	0.2	19.50%	66.45%	59.67%	63.81%

Note. Accuracy is judged by GPT-4.1 and a human annotator. Default describes Rust et al. 2021 dictionary. Commonality measured as words in common with default dictionary.

Based on the initial results, we find that scenario 2 GPT-4.5 and scenario 2 GPT-4.5 with high temperature have the best accuracies. We find that scenario 2 GPT-4.5 with high temperatures also matches on more tweets. We therefore continue with the scenario 2 GPT-4.5 high temperature dictionary for further validation of our tracker approach.

4.3.5 Empirical Application

Figure 3. *Brand Reputation Hermes vs. Prada 2022*



We apply our new dictionaries social media data for the timeframe between 2020 and March 2023. While the chart shows generally higher brand reputation for Prada than for Hermes during the observed period (2022), we can also see that the driver picks up positive brand events such as the Prada's involvement in the fashion week (week 8 on the chart), alongside the influencer Lalisa.

4.3.6 Exploring the dynamics of luxury brand drivers

We measured each sub-driver and drivers' scores over an observation period from January 2020 to March 2023, a total of 169 weeks to evaluate the weekly shifts of brand reputation of our four brands, resulting in a total of 676 observations. The drivers might be depending on each other, meaning that one driver's movement influences subsequent movement of another driver in the future. For example, when a brand increases relationship value, this might be reflected in higher brand value in subsequent periods. We first caried out Fisher-type tests to understand whether the three drivers are stationary. The null hypothesis of the test is that all brand time series have a common root. The results of three separate Fisher type tests (one per driver) result in p-values < .001 and confirm that our timeseries do not have a unit root (Choi 2001). To determine the optimal lag-order, we run lag selection tests. Our results indicate that the first-order panel VAR the values for the modified Bayesian information criterion (MBIC = -201.226), the modified Akaike information criterion (MAIC = -39.944), and the Quinn information criterion (MQIC = -102.493), are lowest at the first-order, compared with the second-order (MBIC = -145.661; MAIC = -24.699; MQIC = -71.611) and the third-order (MBIC = -98.663; MAIC = -18.022; MQIC = -49.297) models. This is not surprising, since discussions about brands on social media change rapidly and frequently. In order to understand the dynamics of the three drivers, we run dynamic multivariate vector autoregression (VAR) estimations (Abrigo and Love 2015) to examine the internal dynamics of the three drivers. We estimate the following equation:

$$\boldsymbol{Y}_{it} = \boldsymbol{Y}_{it-1}\boldsymbol{\alpha} + \boldsymbol{u}_i + \boldsymbol{e}_{it}, \quad (1)$$

where

$\boldsymbol{i}$ = Brand (there are 4 brands),
$\boldsymbol{t}$ = Week (there are 167 weeks),
$\boldsymbol{Y_{it}}$ = A (1x3) vector of endogenous variable (i.e. value, brand and relationship),
$\boldsymbol{u_i}$ = A (1x3) vector of endogenous variable-specific brand fixed effects,
$\boldsymbol{e_{it}}$ = A (1x3) vector of idiosyncratic errors,
$\boldsymbol{\alpha}$ = A (3x3) matrix of parameters for endogenous variables.

Table 5. *The Mutual Impacts of Brand Reputation Drivers*

Endogenous Variables	**$Value_t$**	**$Brand_t$**	**$Relationship_t$**
Predictors	**α (z-Value)**	**α (z-Value)**	**α (z-Value)**
Endogenous causality			
$Value_{t-1}$	.110 (2.34)	.111 (2.17)	.037 (0.74)
$Brand_{t-1}$	.048 (1.11)	.030 (0.61)	.008 (0.21)
$Relationship_{t-1}$	- .000 (0.00)	.110 (1.98)	.107 (2.32)
Model statistics			
No. of observations	668		
No. of brands	4		
Avg. no. of weeks	167		
Granger causality Wald test	**χ^2(prob.)**	**χ^2(prob.)**	**χ^2(prob.)**
$Value_t$		4.71 (.030)	n.s
$Brand_t$	n.s		n.s
$Relationship_t$	n. s	3.94 (.047)	

Note. t denotes the current value, and t-1 the lagged-one-week value of the respective variables.

We find that, while the value driver and relationship driver are predictive of the brand driver (so the time-series of one driver significantly predicts the time-series of the other), the brand-subdriver is not predicted by its own past values.

4.4 CONCLUSION

4.4.1 Managerial and Theoretical Contributions

Our research has important implications for marketing managers, especially those looking for ways to track their brand reputation more closely. We make several contributions.

First, we contribute to literature on generative AI in marketing (Arora, Chakraborty, and Nishimura 2025; Li et al. 2024; Reisenbichler et al. 2022). We show that LLMs can be used to generate context-specific text-mining dictionaries. While generative AI has been used in its analytical capacities, for example to assess sentiment from text (Krugmann and Hartmann 2024; Villarroel Ordenes et al. 2025), we show how LLMs can be used to increase the accuracy of existing dictionaries to measure brand attributes.

Second, we contribute to the literature on brand analytics using social media data (Fronzetti Colladon 2018; Nam, Joshi, and Kannan 2017; Rust et al. 2021; Zhang and Moe 2021). We show that LLMs can be used to improve existing methodologies. By integrating LLM-created dictionaries into existing brand reputation tracker (Rust, Rand, M. H. Huang, et al. 2021), and using LLMs to classify sentiment (Krugmann and Hartmann 2024),we unite the interpretability of dictionary-based text-mining with the accuracy of transformer models. Additionally, by offering replicable dictionaries, we help to increase reproducibility in brand analytics. This helps to further improve and standardize brand measurement approaches in the future.

Third, we contribute to the emergent literature of agentic AI (Arora, Chakraborty, and Nishimura 2025). We show how LLMs can be prompted to act as research assistants used to validate the accuracy of dictionaries. Our results confirm that LLMs track the results of human annotators closely, opening up new possibilities for researchers in developing new dictionaries in a faster and more cost-efficient way (Berger et al. 2020; Humphreys and Wang 2018).

4.4.2 Limitations and Future Research

Our work is not without limitations. Since we use commercial models for our dictionary creation, we cannot shine light on how exactly the words are sourced. When it comes to our dictionary-based real-time brand tracking approach, we still might not be able to catch all granularities in consumers perception. Even though we try to alleviate this by integrating LLMs for sentiment classification (Krugmann and Hartmann 2024), future research can look into improving the accuracy of our tracker. Next, most user-generated content might be multi-modal, including, for example, images. Memes involving brands could be a stronger indicator of how a stakeholder feels, think or talks about the brand.

Future research could explore more applications of custom-created dictionaries. For example, companies that cater mostly to other companies (B2B) might need specialized dictionaries to track their brand reputation. One can easily imagine that when businesses speak about other businesses, the way they communicate changes. Additionally, companies that are strictly in the service sector, e.g. tourism, might find value in context-specific dictionaries. Another potential application of our approach are foreign languages. Most dictionaries only exist in English, but large language models are being released in many different languages. The capabilities of LLMs to translate dictionaries are another interesting application. Especially international companies can use this to their advantage.

CHAPTER 5: CONCLUSION

In this thesis I explored how technology can be harnessed to apply, produce and advance marketing science. In my first essay, we showed how researchers can use digital technology, in the form of interactive research-driven apps to share their findings with managers. To do so, we first explored, reviewing current literature and conducting two experiments, whether research-driven apps render science more interesting and relevant for marketing managers and students. We find encouraging empirical support for our hypothesis. We then develop a tutorial, guiding readers through the most important steps in app development. This allows fellow researchers to transform their research articles into interactive apps. We see this as an encouraging starting point for future research that reconciles the research-practice gap, to ultimately spark future developments that are beneficial to both researchers, managers and hopefully, society. Future research directions could include the usage of LLMs in creating research-driven apps more easily, but also pinpoint the mechanisms on why research-driven apps render articles more interesting and relevant for managers.

In my second essay, we conduct a multimethod study on two large social media datasets and two experiments. We showed how researchers can study consumer behavior using text-mining technology to analyze social media and blockchain data on cryptocurrencies. We find that consumers' investment behavior in cryptocurrencies is driven by greedy language (e.g. "make money with crypto", "get rich with bitcoin"). This happens because when consumers that observe others' make greedy statements, they assume there are profitable opportunities to be made. They successively start to invest into cryptocurrencies. Those findings shine light of the role of social media in investing, but more so at the role that language on social media has on consumers' behavior, especially when confronted with important financial decisions. Further research will be needed to unravel the various influences that cause cryptocurrencies to be an ongoing multi-billion phenomenon.

In my third essay, we use a combination of technologies, namely large language models (LLMs) and automated text analysis to enhance existing marketing measurement methodologies. We develop an easy to replicate algorithm, consisting of multiple LLMs, to automatically create sets of text-mining dictionaries to be used in brand analytics. We then then test their accuracy through agentic artificial intelligence and human annotators. Subsequently, we deploy our enhanced measurement methodology in a new contextual setting, luxury brands. We find that our created dictionaries reach sufficient accuracy and chart the way for further improvements in brand analytics and automated text analysis through LLM technology. Coming research should also seek to further validate brand analytics.

The contributions of this thesis to marketing and technology literature are manifold. First, to the literature of technological tutorials in marketing (Villarroel Ordenes et al. 2025; Boegershausen et al. 2022; Berger et al. 2020), as well as literature on the relevance of marketing research (Schauerte et al. 2023; Kohli and Haenlein 2021; Stäbler and Haenlein 2025). Second, to research on cryptocurrencies (Friederich et al. 2024; Peres et al. 2022), greedy language (Mercadante, Tracy, and Götz 2023; Hoyer, Zeelenberg, and Breugelmans 2022) and financial decision making (Greenberg and Hershfield 2019). Finally it contributes to literature on AI in marketing (Hartmann, Exner, and Domdey 2025; Reisenbichler et al. 2022), agentic AI (Arora, Chakraborty, and Nishimura 2025) and brand analytics (Rust et al. 2021).

Taken together, my findings point at the importance of technology in marketing science. I hope that this research will further inspire researchers to use the vast opportunities of marketing and digital technologies for their own research. In the future, I hope my findings will help bridge the gap between research and practice by facilitating the diffusion of knowledge to managers.

REFERENCES

Aaker, Jennifer L (1997), "Dimensions of Brand Personality," *Journal of Marketing Research*, 11.

Ahluwalia, Rohini and Robert E. Burnkrant (2004), "Answering Questions about Questions: A Persuasion Knowledge Perspective for Understanding the Effects of Rhetorical Questions," *Journal of Consumer Research*, 31 (1), 26–42.

Aiello, Darren, Scott R. Baker, Tetyana Balyuk, Marco Di Maggio, Mark J. Johnson, and Jason D. Kotter (2022), "Who Invests in Crypto? Wealth, Financial Constraints, and Risk Attitudes," SSRN Scholarly Paper, Rochester, NY: Social Science Research Network.

Almeida, José and Tiago Cruz Gonçalves (2023), "A systematic literature review of investor behavior in the cryptocurrency markets," *Journal of Behavioral and Experimental Finance*, 37, 100785.

André, Quentin and Nicholas Reinholtz (2024), "Pre-Registered Interim Analysis Designs (PRIADs): Increasing the Cost-Effectiveness of Hypothesis Testing," *Journal of Consumer Research*, 51 (4), 845–65.

Ante, Lennart (2023), "How Elon Musk's Twitter activity moves cryptocurrency markets," *Technological Forecasting and Social Change*, 186, 122112.

Arora, Neeraj, Ishita Chakraborty, and Yohei Nishimura (2025), "AI–Human Hybrids for Marketing Research: Leveraging Large Language Models (LLMs) as Collaborators," *Journal of Marketing*, 89 (2), 43–70.

Aspara, Jaakko (2013), "The role of product and brand perceptions in stock investing: Effects on investment considerations, optimism and confidence," *Journal of Behavioral Finance*, 14 (3), 195–212.

Atalay, A. Selin, Siham El Kihal, and Florian Ellsaesser (2023), "Creating Effective Marketing Messages Through Moderately Surprising Syntax," *Journal of Marketing*, 87 (5), 755–75.

Babić Rosario, Ana, Francesca Sotgiu, Kristine De Valck, and Tammo H.A. Bijmolt (2016), "The Effect of Electronic Word of Mouth on Sales: A Meta-Analytic Review of Platform, Product, and Metric Factors," *Journal of Marketing Research*, 53 (3), 297–318.

Baldridge, David C., Steven W. Floyd, and Lívia Markóczy (2004), "Are managers from Mars and academicians from venus? Toward an understanding of the relationship between academic quality and practical relevance," *Strategic Management Journal*, 25 (11), 1063–74.

Bearden, William O., Richard G. Netemeyer, and Jesse E. Teel (1989), "Measurement of Consumer Susceptibility to Interpersonal Influence," *Journal of Consumer Research*, 15 (4), 473–81.

Berger, Jonah (2011), "Arousal Increases Social Transmission of Information," *Psychological Science*, 22 (7), 891–93.

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.

Berger, Jonah, Garrick Sherman, and Lyle Ungar (2020), "TextAnalyzer. Retrieved from http://textanalyzer.org."

Blackstone, Tom (2024), "2024 Cryptocurrency Adoption and Sentiment Report," Security.org.

Blair-Early, Adream and Mike Zender (2008), "User Interface Design Principles for Interaction Design," *Design Issues*, 24 (3), 85–107.

Blei, David M, Andrew Y Ng, and Jordan@cs Berkeley Edu (2003), "Latent Dirichlet Allocation Michael I. Jordan," *Journal of Machine Learning Research*, 993–1022.

Boegershausen, Johannes, Hannes Datta, Abhishek Borah, and Andrew T. Stephen (2022), "Fields of Gold: Scraping Web Data for Marketing Insights," *Journal of Marketing*, 86 (5), 1–20.

Bollen, Johan, Huina Mao, and Xiaojun Zeng (2011), "Twitter mood predicts the stock market," *Journal of Computational Science*, 2 (1), 1–8.

Boyd, Ryan, Ashwini Ashokkumar, Sarah Seraj, and James Pennebaker (2022), *The Development and Psychometric Properties of LIWC-22*.

Brand, James, Ayelet Israeli, and Donald Ngwe (2023), "Using GPT for Market Research," *SSRN Electronic Journal*.

Brynjolfsson, Erik and Andrew McAfee (2014), *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*, The second machine age: Work, progress, and prosperity in a time of brilliant technologies, New York, NY, US: W. W. Norton & Company, 306.

Buscher, Georg, Edward Cutrell, and Meredith Ringel Morris (2009), "What do you see when you're surfing?: using eye tracking to predict salient regions of web pages," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Boston MA USA: ACM, 21–30.

Carr, Andrew (2024), "What does 'To the Moon' mean?," (accessed December 2, 2024), [available at https://atomicwallet.io/academy/articles/what-does-to-the-moon-mean].

Cascio Rizzo, Giovanni Luca, Jonah Berger, Matteo De Angelis, and Rumen Pozharliev (2023), "How Sensory Language Shapes Influencer's Impact," *Journal of Consumer Research*, (A. T. Stephen and J. Goldenberg, eds.), 50 (4), 810–25.

Cascio Rizzo, Giovanni Luca, Francisco Villarroel Ordenes, Rumen Pozharliev, Matteo De Angelis, and Michele Costabile (2024), "How High-Arousal Language Shapes Micro-Versus Macro-Influencers' Impact," *Journal of Marketing*, 88 (4), 107–28.

Chen, Mark, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, Alex Ray, Raul Puri, Gretchen Krueger, Michael Petrov, Heidy Khlaaf, Girish Sastry, Pamela Mishkin, Brooke Chan, Scott Gray, Nick Ryder, Mikhail Pavlov, Alethea Power, Lukasz Kaiser, Mohammad Bavarian, Clemens Winter, Philippe Tillet, Felipe Petroski Such, Dave Cummings, Matthias Plappert, Fotios Chantzis, Elizabeth Barnes, Ariel Herbert-Voss, William Hebgen Guss, Alex Nichol, Alex Paino, Nikolas Tezak, Jie Tang, Igor Babuschkin, Suchir Balaji, Shantanu Jain, William Saunders, Christopher Hesse, Andrew N. Carr, Jan Leike, Josh Achiam, Vedant Misra, Evan Morikawa, Alec Radford, Matthew Knight, Miles Brundage, Mira Murati, Katie Mayer, Peter Welinder, Bob McGrew, Dario Amodei, Sam McCandlish, Ilya Sutskever, and Wojciech Zaremba (2021), "Evaluating Large Language Models Trained on Code," arXiv.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345–54.

Chintagunta, Pradeep, Rajdeep Grewal, Detelina Marinova, Rik Pieters, and Shrihari Sridhar (2022), "Journal of Marketing Call for Papers: Marketing Impact with Research-Driven Apps," *American Marketing Association*.

Choi, In (2001), "Unit root tests for panel data," *Journal of International Money and Finance*, 20 (2), 249–72.

Cillo, Paola and Gaia Rubera (2024), "Generative AI in innovation and marketing processes: A roadmap of research opportunities," *Journal of the Academy of Marketing Science*.

Clark-Murphy, Marilyn and Geoffrey N Soutar (2004), "What individual investors value: Some Australian evidence," *Journal of Economic Psychology*, 25 (4), 539–55.

Colicev, Anatoli (2023), "How can non-fungible tokens bring value to brands," *International Journal of Research in Marketing*, 40 (1), 30–37.

Colicev, Anatoli, Ashwin Malshe, Koen Pauwels, and Peter O'Connor (2018), "Improving Consumer Mindset Metrics and Shareholder Value through Social Media: The Different Roles of Owned and Earned Media," *Journal of Marketing*, 82 (1), 37–56.

Cong, Lin William, Ye Li, and Neng Wang (2021), "Tokenomics: Dynamic Adoption and Valuation," *The Review of Financial Studies*, 34 (3), 1105–55.

Cookson, J. Anthony, Runjing Lu, William Mullins, and Marina Niessner (2024), "The social signal," *Journal of Financial Economics*, 158, 103870.

Deighton, John A., Carl F. Mela, and Christine Moorman (2021), "Marketing Thinking and Doing," *Journal of Marketing*, 85 (1), 1–6.

Dew, Ryan, Asim Ansari, and Olivier Toubia (2022), "Letting Logos Speak: Leveraging Multiview Representation Learning for Data-Driven Branding and Logo Design," *Marketing Science*, 41 (2), 401–25.

Dyachenko, Tatiana L and Greg M Allenby (2023), "Is Your Sample Truly Mediating? Bayesian Analysis of Heterogeneous Mediation (BAHM)," *Journal of Consumer Research*, 50 (1), 116–41.

Egan, John (2023), "Nearly 80% Of Young Adults Get Financial Advice From This Surprising Place," *Forbes Advisor*, (accessed July 17, 2024), [available at https://www.forbes.com/advisor/investing/financial-advisor/adults-financial-advice-social-media/].

Farace, Stefania, Francisco Villarroel Ordenes, Dennis Herhausen, Dhruv Grewal, and Ko de Ruyter (2025), "EXPRESS: Standing Out While Fitting in: Visual Design of Text Overlays in Social Media Communication," *Journal of Marketing*, 00222429251322773.

Fidelity (2024), "How to learn about cryptocurrency," *Fidelity*, (accessed February 6, 2025), [available at https://www.fidelity.com/learning-center/trading-investing/how-to-learn-about-cryptocurrency].

FINRA (2023), "Gen Z and Investing," Washingtion.

Friederich, Felix, Jan-Hinrich Meyer, Jorge Matute, and Ramon Palau-Saumell (2024), "CRYPTO-MANIA: How fear-of-missing-out drives consumers' (risky) investment decisions," *Psychology & Marketing*, 41 (1), 102–17.

Fritze, Martin, Cait Lamberton, David Reibstein, and John Zhang (2024), "Consumer Cryptocurrency Confidence Report 2023," Philadelphia, PA: Wharton School, University of Pennsylvania.

Fronzetti Colladon, Andrea (2018), "The Semantic Brand Score," *Journal of Business Research*, 88, 150–60.

Gaies, Brahim, Mohamed Sahbi Nakhli, Jean-Michel Sahut, and Denis Schweizer (2023), "Interactions between investors' fear and greed sentiment and Bitcoin prices," *The North American Journal of Economics and Finance*, 67, 101924.

Goli, Ali and Amandeep Singh (2024), "Frontiers: Can Large Language Models Capture Human Preferences?," *Marketing Science*, 43 (4), 709–22.

Greenberg, Adam Eric and Hal E. Hershfield (2019), "Financial decision making," *Consumer Psychology Review*, 2 (1), 17–29.

Gross, Jana and Kathleen Desveaud (2024), "Navigating computational linguistic in marketing practices: The barriers of natural language processing in social media marketing and a path to future research," *Journal of Marketing Analytics*.

Guyt, Jonne Y., Hannes Datta, and Johannes Boegershausen (2024), "Unlocking the Potential of Web Data for Retailing Research," *Journal of Retailing*, 100 (1), 130–47.

Haenlein, Michael and Andrew Jack (2024), "Measuring the long-term impact of business school research on academia, teaching, society and decision makers," *International Journal of Research in Marketing*.

Hakkarainen, Tuuli and Anatoli Colicev (2023), "Blockchain-enabled advances (BEAs): Implications for consumers and brands," *Journal of Business Research*, 160, 113763.

Hanneke, Björn, Bernd Skiera, Thilo Gerwien Kraft, and Oliver Hinz (2024), "Decoding blockchain data for research in marketing: New insights through an analysis of share of wallet," *International Journal of Research in Marketing*.

Hartmann, Jochen, Anouk Bergner, and Christian Hildebrand (2023), "MindMiner: Uncovering linguistic markers of mind perception as a new lens to understand consumer–smart object relationships," *Journal of Consumer Psychology*, 33 (4), 645–67.

Hartmann, Jochen, Yannick Exner, and Samuel Domdey (2025), "The power of generative marketing: Can generative AI create superhuman visual marketing content?," *International Journal of Research in Marketing*, 42 (1), 13–31.

Hartmann, Jochen, Mark Heitmann, Christian Siebert, and Christina Schamp (2023), "More than a Feeling: Accuracy and Application of Sentiment Analysis," *International Journal of Research in Marketing*, 40 (1), 75–87.

Hewett, Kelly, William Rand, Roland T. Rust, and Harald J. van Heerde (2016), "Brand Buzz in the Echoverse," *Journal of Marketing*, 80 (3), 1–24.

Hofstetter, Reto, Emanuel de Bellis, Leif Brandes, Melanie Clegg, Cait Lamberton, David Reibstein, Felicia Rohlfsen, Bernd Schmitt, and John Z. Zhang (2022), "Crypto-marketing: how non-fungible tokens (NFTs) challenge traditional marketing," *Marketing Letters*.

Hofstetter, Reto, Martin P Fritze, and Cait Lamberton (2024), "Beyond Scarcity: A Social Value-Based Lens for NFT Pricing," *Journal of Consumer Research*, 51 (1), 140–50.

Hotz-Behofsits, Christian, Nils Wlömert, and Nadia Abou Nabout (2025), "Natural Affect DEtection (NADE): Using Emojis to Infer Emotions from Text," *Journal of Marketing*, 00222429251315088.

Hovy, Dirk, Shiri Melumad, and J Jeffrey Inman (2021), "Wordify: A Tool for Discovering and Differentiating Consumer Vocabularies," *Journal of Consumer Research*, 48 (3), 394–414.

Hoyer, Karlijn, Marcel Zeelenberg, and Seger M. Breugelmans (2022), "Greed: What Is It Good for?," *Personality and Social Psychology Bulletin*, 01461672221140355.

Humphreys, Ashlee, Mathew S. Isaac, and Rebecca Jen-Hui Wang (2021), “Construal Matching in Online Search: Applying Text Analysis to Illuminate the Consumer Decision Journey,” *Journal of Marketing Research*, 58 (6), 1101–19.

Humphreys, Ashlee and Rebecca Jen-Hui Wang (2018), “Automated Text Analysis for Consumer Research,” *Journal of Consumer Research*, 44 (6), 1274–1306.

Jedidi, Kamel, Bernd H. Schmitt, Malek Ben Sliman, and Yanyan Li (2021), “R2M Index 1.0: Assessing the Practical Relevance of Academic Marketing Articles,” *JOURNAL OF MARKETING*, 85 (5), 22–41.

Keller, Carmen and Michael Siegrist (2006), “Investing in stocks: The influence of financial risk attitude and values-related money and stock market attitudes,” *Journal of Economic Psychology*, 27 (2), 285–303.

Keller, Kevin Lane (1993), “Conceptualizing, Measuring, and Managing Customer-Based Brand Equity,” *Journal of Marketing*, 57 (1), 1–22.

Keller, Kevin Lane (2021), “The Future of Brands and Branding: An Essay on Multiplicity, Heterogeneity, and Integration,” *Journal of Consumer Research*, 48 (4), 527–40.

Kirvan, Paul (2022), “What is a widget? – Definitions from TechTarget.com,” *WhatIs*, (accessed April 16, 2025), [available at https://www.techtarget.com/whatis/definition/widget].

Kohli, Ajay K. and Michael Haenlein (2021), “Factors affecting the study of important marketing issues: Implications and recommendations,” *International Journal of Research in Marketing*, 38 (1), 1–11.

Krug, Steve (2014), *Don’t make me think, revisited: a common sense approach to Web [and mobile] usability*, Berkley: New Riders.

Krugmann, Jan Ole and Jochen Hartmann (2024), “Sentiment Analysis in the Age of Generative AI,” *Customer Needs and Solutions*, 11 (1), 3.

Kumar, Ashish, Ram Bezawada, Rishika Rishika, Ramkumar Janakiraman, and P.K. Kannan (2016), “From Social to Sale: The Effects of Firm-Generated Content in Social Media on Customer Behavior,” *Journal of Marketing*, 80 (1), 7–25.

Laghaie, Arash and Thomas Otter (2023), “Measuring Evidence for Mediation in the Presence of Measurement Error,” *Journal of Marketing Research*, 60 (5), 847–69.

Lambie, Glenn W. and Jaimie Stickl Haugen (2019), “Understanding greed as a unified construct,” *Personality and Individual Differences*, 141, 31–39.

Lange, Donald, Peggy M. Lee, and Ye Dai (2011), "Organizational Reputation: A Review," *Journal of Management*, 37 (1), 153–84.

Lee, Chan Jean and Eduardo B. Andrade (2011), "Fear, Social Projection, and Financial Decision Making," *Journal of Marketing Research*, 48, S121–29.

Lee, Chan Jean and Eduardo B. Andrade (2015), "Fear, excitement, and financial risk-taking," *Cognition and Emotion*, 29 (1), 178–87.

Li, Peiyao, Noah Castelo, Zsolt Katona, and Miklos Sarvary (2024), "Frontiers: Determining the Validity of Large Language Models for Automated Perceptual Analysis," *Marketing Science*, 43 (2), 254–66.

Liadeli, Georgia, Francesca Sotgiu, and Peeter W.J. Verlegh (2023), "A Meta-Analysis of the Effects of Brands' Owned Social Media on Social Media Engagement and Sales," *Journal of Marketing*, 87 (3), 406–27.

Liadeli, Giorgia, Francesca Sotgiu, and Peeter W.J. Verlegh (2022), "A Meta-Analysis of the Effects of Brand Owned Social Media on Social Media Engagement and Sales."

Lilien, Gary L. (2011), "Bridging the Academic–Practitioner Divide in Marketing Decision Models," *Journal of Marketing*, 75 (4), 196–210.

Liu, Yukun and Aleh Tsyvinski (2021), "Risks and Returns of Cryptocurrency," *The Review of Financial Studies*, 34 (6), 2689–2727.

Long, Huaigang, Adam Zaremba, Ender Demir, Jan Jakub Szczygielski, and Mikhail Vasenin (2020), "Seasonality in the Cross-Section of Cryptocurrency Returns," *Finance Research Letters*, 35, 101566.

Luangrath, Andrea Webb, Yixiang Xu, and Tong Wang (2023), "Paralanguage Classifier (PARA): An Algorithm for Automatic Coding of Paralinguistic Nonverbal Parts of Speech in Text," *Journal of Marketing Research*, 60 (2), 388–408.

Martin, Brett A. S., Polymeros Chrysochou, and Carolyn Strong (2024), "Effects of dispositional greed and need for cognition on consumer judgments of cryptocurrency and stocks," *Journal of Consumer Behaviour*, 23 (5), 2650–59.

Mathmann, Frank, Polymeros Chrysochou, Antonios Tiganis, and Brett A. S. Martin (2025), "What motivates crypto likes? How assessment regulatory mode increases social media engagement with crypto assets," *Psychology & Marketing*, 42 (1), 80–96.

Matthe, Maximilian, Daniel M. Ringel, and Bernd Skiera (2023), "Mapping Market Structure Evolution," *Marketing Science*, 42 (3), 589–613.

Mayer, Jorg H., Markus Esswein, Reiner Quick, and Sanjar Sayar (2020), "KNOWING WHICH APPS MANAGERS LIKE: A MANAGER APP PORTAL ACCOMMODATING THEIR USER PREFERENCES," in *Proceedings of the 13 th IADIS International Conference Information Systems 2020*, IADIS Press, 19–28.

McGee, Suzanne, Hannah Lang, and Hannah Lang (2024), "The decade-long journey to a US spot bitcoin ETF," *Reuters*.

McShane, Blakeley B. and Ulf Böckenholt (2022), "Meta-analysis of Studies with Multiple Contrasts and Differences in Measurement Scales," *Journal of Consumer Psychology*, 32 (1), 23–40.

Mercadante, Eric J., Jessica L. Tracy, and Friedrich M. Götz (2023), "Greed communication predicts the approval and reach of US senators' tweets," *Proceedings of the National Academy of Sciences*, 120 (11), e2218680120.

Mohammad, Saif (2018), "Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, I. Gurevych and Y. Miyao, eds., Melbourne, Australia: Association for Computational Linguistics, 174–84.

Molich, Rolf and Jakob Nielsen (1990), "Improving a human-computer dialogue," *Commun. ACM*, 33 (3), 338–48.

Moreno, Roxana and Richard Mayer (2007), "Interactive Multimodal Learning Environments," *Educational Psychology Review*, 19 (3), 309–26.

Mussel, Patrick, Andrea M. F. Reiter, Roman Osinsky, and Johannes Hewig (2015), "State- and trait-greed, its impact on risky decision-making and underlying neural mechanisms," *Social Neuroscience*, 10 (2), 126–34.

Nalley, Stephen (2024), "A Personal Journey Through Cryptocurrency's Ups and Downs," *Kiplinger.com*, (accessed November 14, 2024), [available at https://www.kiplinger.com/kiplinger-advisor-collective/a-personal-journey-through-cryptocurrencys-ups-and-downs].

Nam, Daye, Andrew Macvean, Vincent Hellendoorn, Bogdan Vasilescu, and Brad Myers (2024), "Using an LLM to Help With Code Understanding," in *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering*, Lisbon Portugal: ACM, 1–13.

Nam, Hyoryung, Yogesh V. Joshi, and P.K. Kannan (2017), "Harvesting Brand Information from Social Tags," *Journal of Marketing*, 81 (4), 88–108.

Nejjar, Mohamed, Luca Zacharias, Fabian Stiehle, and Ingo Weber (2025), "LLMs for science: Usage for code generation and data analysis," *Journal of Software: Evolution and Process*, 37 (1), e2723.

Netemeyer, Richard G., John G. Lynch, Donald R. Lichtenstein, and David Dobolyi (2024), "Financial Education Effects on Financial Behavior and Well-Being: The Mediating Roles of Improved Objective and Subjective Financial Knowledge and Parallels in Physical Health," *Journal of Public Policy & Marketing*, 07439156241228197.

Ohanian, Roobina (1990), "Construction and Validation of a Scale to Measure Celebrity Endorsers' Perceived Expertise, Trustworthiness, and Attractiveness," *Journal of Advertising*, 19 (3), 39–52.

OpenAI (2025a), "Text generation and prompting - OpenAI API," (accessed April 26, 2025), [available at https://platform.openai.com].

OpenAI (2025b), "Pricing," (accessed April 28, 2025), [available at https://openai.com/api/pricing/].

Opoku, Robert, Russell Abratt, and Leyland Pitt (2006), "Communicating brand personality: Are the websites doing the talking for the top South African Business Schools?," *Journal of Brand Management*, 14 (1), 20–39.

Packard, Grant and Jonah Berger (2017), "How Language Shapes Word of Mouth's Impact," *Journal of Marketing Research*, 54 (4), 572–88.

Packard, Grant and Jonah Berger (2024), "The Emergence and Evolution of Consumer Language Research," *Journal of Consumer Research*, 51 (1), 42–51.

Packard, Grant, Sarah G. Moore, and Jonah Berger (2023), "Consumer insights from text analysis," *Journal of Consumer Psychology*, 33 (4), 615–20.

Pamuksuz, Utku, Joseph T. Yun, and Ashlee Humphreys (2021), "A Brand-New Look at You: Predicting Brand Personality in Social Media Networks with Machine Learning," *Journal of Interactive Marketing*, 56 (1), 1–15.

Papies, Dominik, Peter Ebbes, and Harald J. Van Heerde (2017), “Addressing Endogeneity in Marketing Models,” in *Advanced Methods for Modeling Markets*, International Series in Quantitative Marketing, P. S. H. Leeflang, J. E. Wieringa, T. H. A. Bijmolt, and K. H. Pauwels, eds., Cham: Springer International Publishing, 581–627.

Peres, Renana, Eitan Muller, and Vijay Mahajan (2010), “Innovation diffusion and new product growth models: A critical review and research directions,” *International Journal of Research in Marketing*, 27 (2), 91–106.

Peres, Renana, Martin Schreier, David A. Schweidel, and Alina Sorescu (2022), “Blockchain meets marketing: Opportunities, threats, and avenues for future research,” *International Journal of Research in Marketing*, S0167811622000635.

Peres, Renana, Martin Schreier, David Schweidel, and Alina Sorescu (2023), “On ChatGPT and beyond: How generative artificial intelligence may affect research, teaching, and practice,” *International Journal of Research in Marketing*, 40 (2), 269–75.

Petrin, Amil and Kenneth Train (2010), “A Control Function Approach to Endogeneity in Consumer Choice Models,” *Journal of Marketing Research*, 47 (1), 3–13.

Pezzuti, Todd, James M. Leonhardt, and Caleb Warren (2021), “Certainty in Language Increases Consumer Engagement on Social Media,” *Journal of Interactive Marketing*, 53 (1), 32–46.

Pikulina, Elena, Luc Renneboog, and Philippe N. Tobler (2017), “Overconfidence and investment: An experimental approach,” *Journal of Corporate Finance*, 43, 175–92.

Porter, M.F. (1980), “An algorithm for suffix stripping,” *Program*, 14 (3), 130–37.

Reisenbichler, Martin, Thomas Reutterer, and David A Schweidel (2023), “Applying Large Language Models to Sponsored Search Advertising,” *MSI Working Paper*.

Reisenbichler, Martin, Thomas Reutterer, David A. Schweidel, and Daniel Dan (2022), “Frontiers: Supporting Content Marketing with Natural Language Generation,” *Marketing Science*, 41 (3), 441–52.

Richards, Mark and Neal Ford (2020), *Fundamentals of software architecture: an engineering approach*, Sebastopol, CA: O’Reilly Media, Inc.

Ringel, Daniel M. (2023), “Multimarket Membership Mapping,” *Journal of Marketing Research*, 60 (2), 237–62.

Roberts, John H., Ujwal Kayande, and Stefan Stremersch (2014), "From academic research to marketing practice: Exploring the marketing science value chain," *International Journal of Research in Marketing*, 31 (2), 127–40.

Rocklage, Matthew D., Sharlene He, Derek D. Rucker, and Loran F. Nordgren (2023), "Beyond Sentiment: The Value and Measurement of Consumer Certainty in Language," *Journal of Marketing Research*, 60 (5), 870–88.

Rodrigues, Johannes, Patrick Ruthenberg, Patrick Mussel, and Johannes Hewig (2023), "Never mind losing the pound… still got the penny! The influence of trait greed on risky decision behavior in a mixed and gain only BART," *Current Psychology*, 42 (28), 24399–409.

Rogers, Everett M. (1983), *Diffusion of innovations*, New York : London: Free Press ; Collier Macmillan.

Rust, Roland T., Katherine N. Lemon, and Valarie A. Zeithaml (2004), "Return on Marketing: Using Customer Equity to Focus Marketing Strategy," *Journal of Marketing*, 68 (1), 109–27.

Rust, Roland T., William Rand, Ming Hui Huang, Andrew T. Stephen, Gillian Brooks, and Timur Chabuk (2021), "Real-Time Brand Reputation Tracking Using Social Media," *Journal of Marketing*, 85 (4), 21–43.

Rutz, Oliver J. and George F. Watson (2019), "Endogeneity and marketing strategy research: an overview," *Journal of the Academy of Marketing Science*, 47 (3), 479–98.

Salesforce (2025), "Connect Web3 with the world's #1 CRM.," (accessed April 7, 2025), [available at https://www.salesforce.com/eu/commerce/web3-platform/].

Sarstedt, Marko, Susanne J. Adler, Lea Rau, and Bernd Schmitt (2024), "Using large language models to generate silicon samples in consumer and marketing research: Challenges, opportunities, and guidelines," *Psychology & Marketing*, 41 (6), 1254–70.

Schauerte, Nico, Maren Becker, Monika Imschloss, Julian R. K. Wichmann, and Werner J. Reinartz (2023), "The managerial relevance of marketing science: Properties and genesis," *International Journal of Research in Marketing*, 40 (4), 801–22.

Schoenmueller, Verena, Oded Netzer, and Florian Stahl (2022), "Frontiers: Polarized America: From Political Polarization to Preference Polarization," *Marketing Science*, 42 (1), 48–60.

Sepehri, Amir, Mitra Sadat Mirshafiee, and David M. Markowitz (2023), "PassivePy: A tool to automatically identify passive voice in big text data," *Journal of Consumer Psychology*, 33 (4), 714–27.

Sergeenkov, Andrey (2024), "How Volatile Is Bitcoin Compared To Other Assets," *Forbes*, (accessed November 14, 2024), [available at https://www.forbes.com/sites/digital-assets/2024/08/16/how-volatile-is-bitcoin-compared-to-other-assets/].

Seuntjens, Terri G., Niels van de Ven, Marcel Zeelenberg, and Anna van der Schors (2016), "Greed and adolescent financial behavior," *Journal of Economic Psychology*, 57, 1–12.

Seuntjens, Terri G., Marcel Zeelenberg, Seger M. Breugelmans, and Niels van de Ven (2015), "Defining greed," *British Journal of Psychology*, 106 (3), 505–25.

Seuntjens, Terri G., Marcel Zeelenberg, Niels Van De Ven, and Seger M. Breugelmans (2015), "Dispositional greed.," *Journal of Personality and Social Psychology*, 108 (6), 917–33.

Shapiro, Debra L. and Bradley Kirkman (2018), "It's Time to Make Business School Research More Relevant," *Harvard Business Review*.

Shefrin, Hersh (2002), *Beyond Greed and Fear: Understanding Behavioral Finance and the Psychology of Investing*, Oxford University Press.

Shin, Laura (2022), *The cryptopians: idealism, greed, lies, and the making of the first big cryptocurrency craze*, New York: PublicAffairs.

Stäbler, Samuel and Michael Haenlein (2025), "The unheard voice of marketing research: Breaking through to news and social media," *Journal of the Academy of Marketing Science*, 53 (1), 105–28.

Stremersch, Stefan, Jorge Gonzalez, Albert Valenti, and Julian Villanueva (2023), "The value of context-specific studies for marketing," *Journal of the Academy of Marketing Science*, 51 (1), 50–65.

Stromberg, Malik, Wendy W Moe, Thomas Reutterer, and David A Schweidel (2025), "Blind Spots in Broad Strokes: Caveats for the Use of LLMs in Marketing Research."

Tash, Moein Shahiki, Olga Kolesnikova, Zahra Ahani, and Grigori Sidorov (2024), "Psycholinguistic and emotion analysis of cryptocurrency discourse on X platform," *Scientific Reports*, 14 (1), 8585.

The White House (2025), "Fact Sheet: President Donald J. Trump Establishes the Strategic Bitcoin Reserve and U.S. Digital Asset Stockpile," (accessed April 7, 2025), [available at https://www.whitehouse.gov/fact-sheets/2025/03/fact-sheet-president-donald-j-trump-establishes-the-strategic-bitcoin-reserve-and-u-s-digital-asset-stockpile/].

Tirunillai, Seshadri and Gerard J. Tellis (2012), "Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance," *Marketing Science*, 31 (2), 198–215.

Tretina, Kat and John Schmidt (2021), "How To Buy Cryptocurrency," *Forbes Advisor*, (accessed November 14, 2024), [available at https://www.forbes.com/advisor/investing/cryptocurrency/how-to-buy-cryptocurrency/].

Trusov, Michael, Randolph E Bucklin, and Koen Pauwels (2009), "Effects of Word-of-Mouth versus Traditional Marketing: Findings from an Internet Social Networking Site," *Journal of Marketing*, 73 (5), 90–102.

Vasile, Iulia (2023), "Top 11 Crypto Communities To Join in 2024," *BeInCrypto*, (accessed December 2, 2024), [available at https://beincrypto.com/learn/crypto-communities/].

Villarroel Ordenes, Francisco, Stephan Ludwig, Ko De Ruyter, Dhruv Grewal, and Martin Wetzels (2017), "Unveiling What Is Written in the Stars: Analyzing Explicit, Implicit and Discourse Patterns of Sentiment in Social Media," *Journal of Consumer Research*, 875–94.

Villarroel Ordenes, Francisco, Grant Packard, Jochen Hartmann, and Davide Proserpio (2025), "Using Traditional Text Analysis and Large Language Models in Service Failure and Recovery," *Journal of Service Research*, 10946705241307678.

Villarroel Ordenes, Francisco and Rosaria Silipo (2021), "Machine learning for marketing on the KNIME Hub: The development of a live repository for marketing applications," *Journal of Business Research*, 137, 393–410.

Wang, Jying-Nan, Hung-Chun Liu, and Yuan-Teng Hsu (2024), "A U-shaped relationship between the crypto fear-greed index and the price synchronicity of cryptocurrencies," *Finance Research Letters*, 59, 104763.

Wang, Long and J. Keith Murnighan (2011), "On Greed," *Academy of Management Annals*, 5 (1), 279–316.

Warren, Nooshin L., Matthew Farmer, Tianyu Gu, and Caleb Warren (2021), "Marketing Ideas: How to Write Research Articles that Readers Understand and Cite," *Journal of Marketing*, 85 (5), 42–57.

Weiss-Cohen, Leonardo, Philip Newall, Yakov Bart, Mircea Zloteanu, Mike Peacey, Peter Ayton, and Iain Clacher (2023), "When Vegas Comes to Wall Street: Associations Between Stock Price Volatility and Trading Frequency Among Gamblers," SSRN Scholarly Paper, Rochester, NY.

"What Is Agentic AI, and How Will It Change Work?" (n.d.), *Harvard Business Review*.

YouGov (2025), "US Investment Trends Report 2025," YouGov.

Zeelenberg, Marcel and Seger M. Breugelmans (2022), "The good, bad and ugly of dispositional greed," *Current Opinion in Psychology*, 46, 101323.

Zhang, Kunpeng and Wendy Moe (2021), "Measuring Brand Favorability Using Large-Scale Social Media Data," *Information Systems Research*, 32 (4), 1128–39.

APPENDIX A

Table W.1. Marketing Journals in the Social Sciences Citation Index

Advances in Consumer Research, Asia Pacific Journal of Marketing and Logistics, Consumption Markets & Culture, Electronic Markets, European Journal of Marketing, Industrial Marketing Management, International Journal of Advertising, International Journal of Bank Marketing, International Journal of Consumer Studies, International Journal of Market Research, International Journal of Research in Marketing, International Journal of Sports Marketing & Sponsorship, International Marketing Review, Journal of Advertising, Journal of Advertising Research, Journal of Brand Management, Journal of Business & Industrial Marketing, Journal of Business-to-Business Marketing, Journal of Consumer Psychology, Journal of Consumer Research, Journal of Destination Marketing & Management, Journal of Fashion Marketing and Management, Journal of Global Scholars of Marketing Science, Journal of Hospitality Marketing

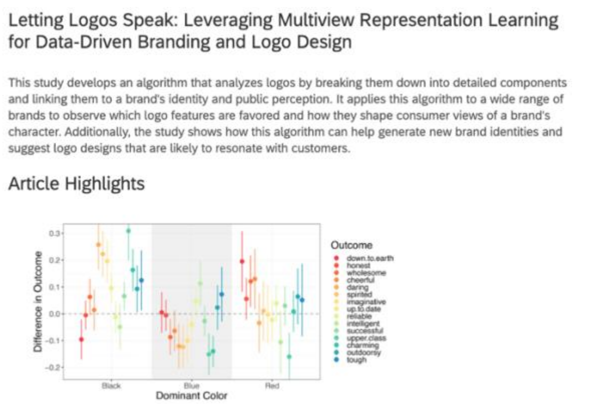
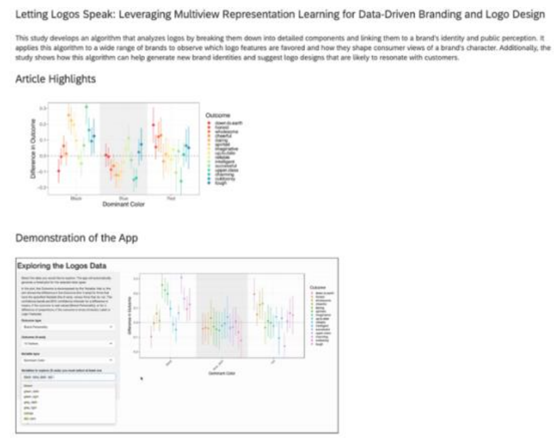
& Management, Journal of Interactive Marketing, Journal of International Marketing, Journal of Marketing, Journal of Marketing for Higher Education, Journal of Marketing Management, Journal of Marketing Research, Journal of Product and Brand Management, Journal of Public Policy & Marketing, Journal of Research in Interactive Marketing, Journal of Retailing, Journal of Services Marketing, Journal of Social Marketing, Journal of Strategic Marketing, Journal of the Academy of Marketing Science, Journal of Travel & Tourisam Marketing, Journal of Vacation Marketing, Marketing Intelligence & Planning, Marketing Letters, Marketing Science, Marketing Theory, Psychology & Marketing, QME - Quantitative Marketing and Economics, Retail and Marketing Review, Sport Marketing Quarterly

Table W.2. Keywords for literature review

Keywords for Literature Review
natural language processing, online search, relevance, text analysis, field experiments, image analysis, machine learning/AI, mediation models, text mining, abstract vs. concrete mindsets, automated text analysis, Bayes factor, Bayesian estimation, between-study variation, branding, causal direction, causal model identification, citations, Cohen's d, competitive market structure visualization, conditional independence, construal level, consumer decision journey, deep learning, experimental methods, experiments, goal progress, grammatical voice, group sequential designs, heterogeneity, image processing, imbalance, information retrieval, interpretable machine learning, lab experiments, language, language models, logos, machine learning, mapping, market evolution, market structure analysis, measurement, mediation, mediation analysis, meta-analysis, meta-science, methods, mind perception, mixture model,

multilevel, multiview learning, natural languageprocessing, neural product embedding, online reviews, overlapping clustering, passive voice, polarity, power analysis, processing fluency, random effects, readability, representation learning, search queries, self-selection, sentiment analysis, smart objects, standardized mean difference, statistics, structural heterogeneity, syntactic surprise, syntax, text mining, theory of mind, topic modeling, trajectories, user-generated content, writing

Figure W.2. Stimuli for experiment with managers

Control (No App)	Treatment (App present)
	

Experiment with marketing managers - Questions

Interactivity

To what extent do you think the research paper is interactive? (Not interactive at all = 1, Very interactive = 7)

Interestingness

In your opinion, how interesting, useful, and easy is it to implement the findings from this research? (Not at all = 1, Very = 7)

Interesting

Useful

Easy to implement

Relevance

How relevant do you find this research paper? (Not at all relevant = 1, Very relevant = 7)

For companies in general

For your own industry

For your own company

For your own position

Table W.2

Ordered Logistic Regression: Interestingness

	Dependent variable: Interestingness	
	Coefficient	SE
Condition		
App	.345**	(.145)
App Type		
Predictor	.828***	(.203)
Explorer	-.380*	(.205)
Optimizer	.519**	(.206)
Work Experience: >5 years	.017	(.228)
Education		
Secondary or lower	-.224	(.407)
University	-.508	(.366)
Vocational/Technical	-.424	(.411)
Observations	584	
Note:	*p<0.1; **p<0.05; ***p<0.01 Baseline for App: Converter	

Table W.3

Ordered Logistic Regression: Relevance

	Dependent variable: Relevance	
	Coefficient	SE
Condition		
App	.316**	(.145)
App Type		
Predictor	.585***	(.205)
Explorer	-.029	(.202)
Optimizer	.421**	(.205)
Work Experience: >5 years	-.067	(.227)
Education		
Secondary or lower	-.340	(.433)
University	-.527	(.393)
Education: Vocational/Technical	-1.084**	(.437)
Observations	584	
Note:	*p<0.1; **p<0.05; ***p<0.01 Baseline for App: Converter	

Experiment with marketing students – Questions

Please clearly mark your answers on the 7-Point scales below.

- In your opinion, how interesting is it to implement the findings from this research?
- In your opinion, how useful is it to implement the findings from this research?
- How relevant do you find this research paper?

Code W.4. Example Code for Interactive Research-Driven App in Streamlit

```
# FRONTEND PAGE 'app.py' This code can be found on https://github.com/research-driven-app/br

# This is the page that is loaded when Streamlit is initialized, therefore the starting point
# It is helpful to familiarize with the basic functionalities of streamlit (documentation can be found here https://docs.streamlit.io/)

# We import the relevant packages

# We import streamlit
import streamlit as st

# Since there are data operations that need to be done, we also import pandas
import pandas as pd

# Since there is a backend part, we also import the backend. This is important for referencing the backend later
import backend as bk
```

```python
# This package will help us with the data buffering later

from io import BytesIO


# Default options are given here. In our case, we define a default text-mining dictionary,
that we keep in the folder /data/

default_dictionary = pd.read_excel("data/default_dict.xlsx")

rename = False

user_input_integer = 1000

current_pattern = "ISO8601"

current_new_drive = None

df_add = None

df = pd.read_csv("data/default_clean.csv")


# This next line helps us with layouting, we define three columns.

# The one on the left takes 70%, then there is a small gap of 5%, and then the rest
takes 20%

headLeft, headSpacer, headRight = st.columns([0.7,0.05,0.2])


# We want to push the content of the columns down, therefore we set two whitespaces

headRight.markdown(" ")
```

```
headRight.markdown(" ")

# This is the title of our app, "Brand Reputation App", it is set on the left on top
"headLeft.title("name of the app goes here")"
headLeft.title("Brand Reputation App")

# Link to the next page, this will execute the script
st.page_link("pages/edit.py", label="Compute and Visualize Brand Reputation",
icon="▶️")
st.markdown("---")
st.markdown(" ")

# Editing of the settings on the first page is triggered (checkbox: Edit)
if headRight.checkbox("Edit"):
    rename = True
    # Create three columns (larger column for file upload, smaller column for settings)
    col1, spacer, col2 = st.columns([1.2,0.1,0.8])
    # Create a user input widget in the first column, setting the sample size, along with
the minimum and default value
    uploaded_file = col1.file_uploader("Choose a CSV file", type=("csv"))
```

```
    user_input_integer = col1.number_input('Sample Size', min_value=100, value=1000,
step=1)




    if uploaded_file is not None:
        df = pd.read_csv(uploaded_file)
        col1.write("Tweets Data Uploaded")

    # Add a horizontal line
    col1.markdown("---")

    col11, spacer11, col12 = col1.columns([1,0.1,1])




    # Add a download button for an Excel file
    with open('svg_question_mark.txt', 'r') as file:
        svg_string = file.read()
```

```
    info_string = "You can upload an XLSX file with 3 columns and the same amount of rows of the current dictionary. &#013; Download the current dictionary to generate such data."
    svg_string = svg_string.replace("INSERT_HERE", info_string)

    col11.markdown(svg_string, unsafe_allow_html=True)

    # Create a BytesIO buffer for the Excel file
    output = BytesIO()
    # Write the dictionary to this Excel buffer
    default_dictionary.to_excel(output, index=False)
    # Seek to the beginning of the stream

    output.seek(0)

    # add a download button for the default dictionary (defined above) - widget

    col12.download_button(
        label="Default Dictionary",
        data=output.read(),
        file_name="default_dict.xlsx",
```

```
        mime="application/vnd.openxmlformats-officedocument.spreadsheetml.sheet",
    )

    # Add a string input to the right column with a default value for format
    current_pattern = col2.text_input("Timestamp Format", "ISO8601")

    # Add three column selections to the right column with default values (text in the first column, ID in the second, timestamp in the third column)
    col_selection1 = col2.selectbox("Select ID", df.columns, index=1)
    col_selection2 = col2.selectbox("Select Timestamp", df.columns, index=2)
    col_selection3 = col2.selectbox("Select Text", df.columns, index=0)

    new_col_names = [col_selection1, col_selection2, col_selection3]

    no_error_renaming = True
    if len(set(new_col_names)) != 3:
        no_error_renaming = False
        col2.markdown('<span style="color: red;"><b>Error</b>: Column names must be unique</span>', unsafe_allow_html=True)

    st.markdown("---")
```

```
if rename and no_error_renaming:
    st.markdown('<span style="color: green;"><b>Note</b>: Edits to the tweets table are shown in the table below</span>', unsafe_allow_html=True)
    df.rename(columns={
        new_col_names[0]: "tweet_id-32876tjkdhsba",
        new_col_names[1]: "created_at-32876tjkdhsba",
        new_col_names[2]: "text-32876tjkdhsba"
    }, inplace=True)

    df.rename(columns={
    "tweet_id-32876tjkdhsba": "tweet_id",
    "created_at-32876tjkdhsba": "created_at",
    "text-32876tjkdhsba": "text"
    }, inplace=True)

user_input_integer = min(len(df), user_input_integer)
df_to_cache = df.head(user_input_integer)

#Shows a preview of the table
st.table(df_to_cache.head())
```

```python
# By using session states, we can save the values of the variables chosen, so that, if we re-run the script, the original values are kept
# for example, the timestamp format it stays saved in the app during that session

st.session_state['timestamp_pattern'] = current_pattern
st.session_state['cached_df'] = df_to_cache
st.session_state['cached_dictionary'] = default_dictionary
# BACKEND PAGE 'backend.py'

# This is where the data is text-mined and aggregated

# General Libraries
import pandas as pd
import numpy as np
import os

# Text Mining libraries
import nltk

# Download NLTK data (needed for text-mining)
```

```python
nltk.download('punkt')
nltk.data.path.append('nltk_data')
from nltk.stem import PorterStemmer
from sklearn.feature_extraction.text import CountVectorizer

# Detect Language (to later be able to filter out non-english tweets)
from langdetect import detect
from langdetect.lang_detect_exception import LangDetectException

from collections import Counter

import streamlit as st

def check_columns_for_neg_suffix(df_check, string):
    # Iterate through each column name in the DataFrame
    for column in df_check.columns:
        # Check if '_neg' is a substring of the column name
        if string in column:
            # Return False if '_neg' is found
            return False
```

```
    # Return True if '_neg' is not found in any column name
    return True

def detect_language(text):

    # Define a function using langdetect

    try:
        return detect(text)
    except LangDetectException as e:
        print(f"Error detecting language: {e}")
        return None

def stem_sentence(sentence):

    # Initialize the PorterStemmer
    stemmer = PorterStemmer()

    ##Define the stemming function
    words = nltk.word_tokenize(sentence)
```

```
    stemmed_words = [stemmer.stem(word) for word in words]

    return ' '.join(stemmed_words)


def change_time_columns(df, format_arg=None, dayfirst=True, errors='coerce'):

    """

    Process datetime columns in a DataFrame by converting them and extracting
components.


    Parameters:

        df (pd.DataFrame): The input DataFrame.

        format_arg (str or None): The datetime format to use for parsing. If None, the
format is inferred.

        dayfirst (bool): Whether to interpret the day as the first part of the date when
inferring.

        errors (str): How to handle errors ('raise', 'coerce', or 'ignore').


    Returns:

        pd.DataFrame: The processed DataFrame with additional datetime components.

    """

    # Ensure we are working with a copy of the DataFrame to avoid
SettingWithCopyWarning

    df = df.copy()
```

```python
try:
    # Convert 'created_at' to datetime
    df['created_at'] = pd.to_datetime(
        df['created_at'],
        format=format_arg,
        dayfirst=dayfirst,
        errors=errors
    )

    # Extract year, quarter, month, and week only for valid datetime values
    valid_dates = df['created_at'].notna()
    df.loc[valid_dates, 'year'] = df.loc[valid_dates, 'created_at'].dt.year
    df.loc[valid_dates, 'quarter'] = df.loc[valid_dates, 'created_at'].dt.quarter
    df.loc[valid_dates, 'month'] = df.loc[valid_dates, 'created_at'].dt.month
    df.loc[valid_dates, 'week'] = df.loc[valid_dates, 'created_at'].dt.isocalendar().week

except Exception as e:
    print(f"Error while processing datetime column: {e}")
```

```
    return df


# Tokenize the text and add absolute counter

def tokenize_and_count(text):

    tokens = nltk.word_tokenize(text.lower())

    # Convert to lowercase for consistency

    return Counter(tokens)


@st.cache_data

def preprocess(df):


    # Create a filter for Tweets starting with RT*

    df['retweet'] = df['text'].str.match(r'^RT')


    # Filter out those that are False on the retweet ,here '~' represents those that
evaluate to false

    df = df[~df['retweet']]


    # Use a function to drop duplicates over two columns

    df = df.drop_duplicates(subset=['text', 'tweet_id'])
```

```
# Fill empty spaces

df['text'] = df['text'].fillna('').astype(str)


# Apply language detection and put it into a new column

df['language'] = df['text'].apply(detect_language)


# For english only filter

df = df[df['language']== 'en']


# Convert to lower case

df['text'] = df['text'].str.lower()


# Apply stemming


df['text_stemmed'] = df['text'].apply(stem_sentence)


# Apply the tokenization function to the 'text' column

df['term_freq'] = df['text'].apply(tokenize_and_count)


# Expand the DataFrame so each word gets its own row (Bag of Words)
```

```
    expanded_rows = []
    for _, row in df.iterrows():
        for term, freq in row['term_freq'].items():
            expanded_rows.append([row['tweet_id'], row['text'], row['created_at'], term, freq])

    result_df = pd.DataFrame(expanded_rows, columns=['tweet_id', 'text', 'created_at',
'term', 'frequency'])

    return result_df


@st.cache_data
def join_and_multiply_data(df_join, dictionary, old_df,
timestamp_format_macro='ISO8601', extra_dict=None):

    if extra_dict is not None:
        dictionary = pd.merge(dictionary, extra_dict, on='term', how='inner')

    joined_df = pd.merge(df_join, dictionary, on='term', how='inner')

    drives_list = dictionary.columns.tolist()
    drives_list.remove('term')
```

```
    # Multiply the term frequency with each column of the dictionary
    for col in drives_list:
        joined_df[col] = joined_df['frequency'] * joined_df[col]

    joined_df = joined_df[drives_list + ['tweet_id']]
    aggregated_df = joined_df.groupby('tweet_id').sum().reset_index()


    old_df = change_time_columns(old_df,  format_arg = timestamp_format_macro)


    merged_df = pd.merge(old_df, aggregated_df, on='tweet_id', how='inner')


    return merged_df


def prepare_aggregate_dict(list_of_drives):
    aggregate_dict = {}
    aggregate_dict['tweet_id'] = 'count'
    for drive in list_of_drives:
        aggregate_dict[drive+"_pos"] = 'sum'
        aggregate_dict[drive+"_neg"] = 'sum'
```

```
    return aggregate_dict


@st.cache_data
def compute_drives(final_df, granularity, extra_driver=None):
    dict_columns_ids = ["price", "squality", "goodsq", "cool", "exciting", "innov", "socresp",
"comm", "friendly", "personalrel", "trust"]
    if extra_driver is not None:
        dict_columns_ids.append(extra_driver)

    dict_to_aggregate = prepare_aggregate_dict(dict_columns_ids)

    dict_to_aggregate["created_at"] = 'first'

    grouped_df = final_df.groupby(granularity).agg(dict_to_aggregate).reset_index()

    grouped_df = grouped_df.rename(columns={'tweet_id': 'total_tweets'})

    dict_of_drivers = {}
    dict_of_drivers["Value"] = ["price", "squality", "goodsq"]
    dict_of_drivers["Brand"] = ["cool", "exciting", "innov", "socresp"]
    dict_of_drivers["Relationship"] = ["comm", "friendly", "personalrel", "trust"]
```

```python
    if extra_driver is not None:
        dict_of_drivers[extra_driver] = [extra_driver]

    driver_columns = []

    for driver in dict_of_drivers:

        for value in dict_of_drivers[driver]:

            new_column = value+"_net"
            grouped_df[new_column] = grouped_df[value+"_pos"] - 
grouped_df[value+"_neg"]
            driver_columns.append(new_column)

        computed_nets = [value+"_net" for value in dict_of_drivers[driver]]
        new_column_driver = driver+"_Driver"
        grouped_df[new_column_driver] = grouped_df[computed_nets].mean(axis=1)
        driver_columns.append(new_column_driver)
```

```
    grouped_df['Brand Reputation'] = grouped_df['Value_Driver'] +
grouped_df['Brand_Driver'] + grouped_df['Relationship_Driver']
    if extra_driver is not None:
        grouped_df['Brand Reputation'] = grouped_df['Brand Reputation'] +
grouped_df[extra_driver+"_Driver"]

    # Perform z-score normalization: 1) first means, then standard deviation 2) substract
the mean from the value and divide by standard deviation
    for column in driver_columns:
        mean = grouped_df[column].mean()
        std = grouped_df[column].std()
        grouped_df[column] = (grouped_df[column] - mean) / std

    driver_columns = driver_columns + ['Brand Reputation']

    grouped_df.sort_values(by='created_at', inplace=True)
    grouped_df.set_index("created_at", inplace=True)
    grouped_df = grouped_df[driver_columns]
    return grouped_df

# Settings and Visualization '/pages/edit.py'
```

```python
# Import relevant libraries

import streamlit as st

import pandas as pd

import backend as bk

from io import BytesIO


# Define the session states for the dictionaries

df = st.session_state['cached_df']

dictionary = st.session_state['cached_dictionary']

pattern = st.session_state['timestamp_pattern']


df_preprocessed = bk.preprocess(df)


df_joined = bk.join_and_multiply_data(df_preprocessed,

                                       dictionary,

                                       df,

                                       timestamp_format_macro=pattern)


# Set the title of the page
```

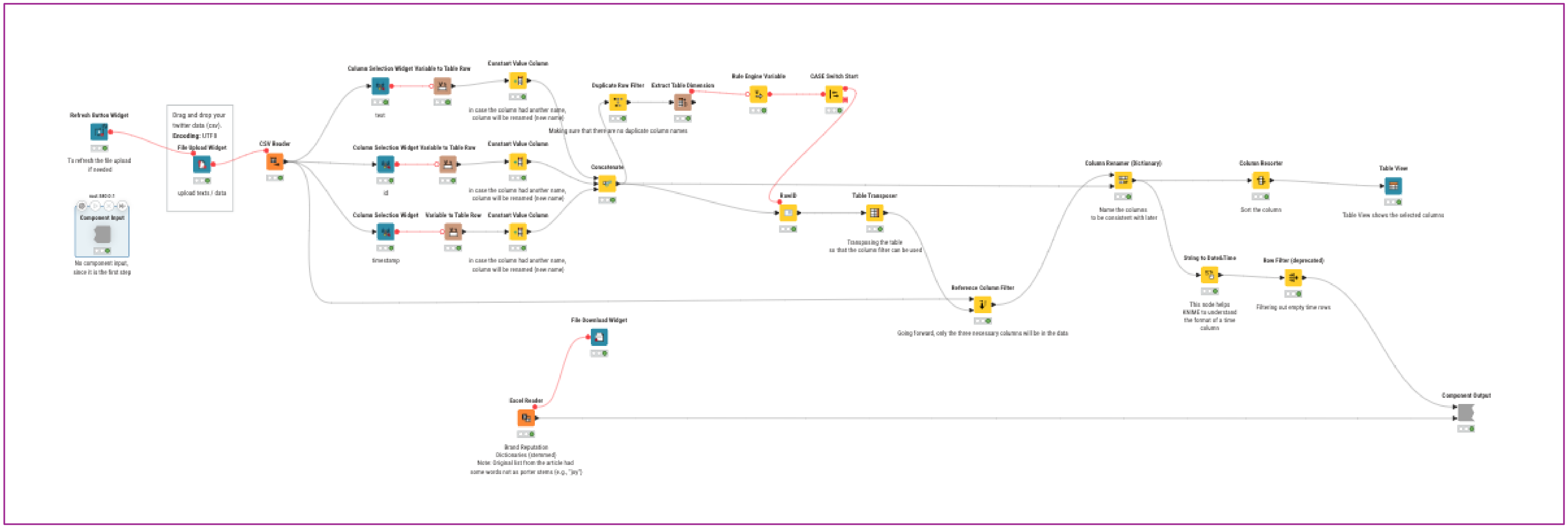
```
st.title("Brand Reputation App")

st.page_link("app.py", label="Settings", icon="🔙")
st.markdown("---")
st.markdown(" ")

# Create three columns (first column with 40% width, 10% space and third column 50% width)
col1, spacer, col2 = st.columns([0.4,0.1,0.5])

# Create options for the column selector widget
options = ['Weekly', 'Monthly', 'Quarterly', 'Yearly']

# Create the selector widget in the first column
selection = col1.selectbox('Granularity', options, index=0)

# Weekly aggregation, monthly aggregation, quarterly aggregation, yearly aggregation. we save the selection in "gran"
if selection == 'Weekly':
  gran = ['year', 'quarter', "month", "week"]
```

```
elif selection == 'Monthly':
  gran = ['year', 'quarter', "month"]

elif selection == 'Quarterly':
  gran = ['year', 'quarter']

else:
  gran = ['year']

# From the backend, apply the function "compute drives", taking in the text-mined
dataframe and the aggregation
df_drivers = bk.compute_drives(df_joined, gran)

# Create a BytesIO buffer for the Excel file
output = BytesIO()
# Write the DataFrame to this buffer (to be downloaded later, if wanted)
df_drivers.to_excel(output, index=False)
# Seek to the beginning of the stream
output.seek(0)
```

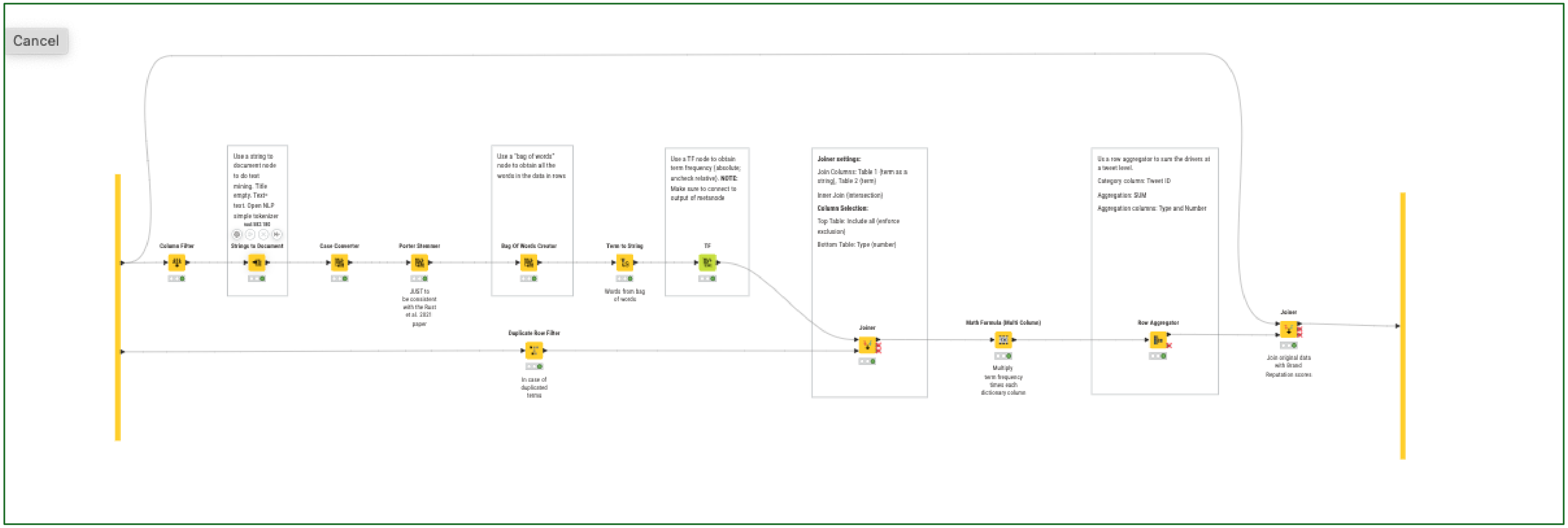
```python
# in the first column, set a download button to download drivers
col1.download_button(
    label="Download All Drivers",
    data=output.read(),
    file_name="computed_drivers.xlsx",
    mime="application/vnd.openxmlformats-officedocument.spreadsheetml.sheet",
)


# Define a list with all columns from our driver dataframe
columns = list(df_drivers.columns)


# Let user select which columns to plot
all_columns = df_drivers.columns.tolist()
selection2 = st.multiselect("Select drivers to plot:", all_columns, default='Brand
Reputation')


# Filter the dataframe to show only selected columns
if selection2:
    st.line_chart(df_drivers[selection2])
else:
    st.warning("Please select at least one driver to display the plot.")
```

Workflows KNIME

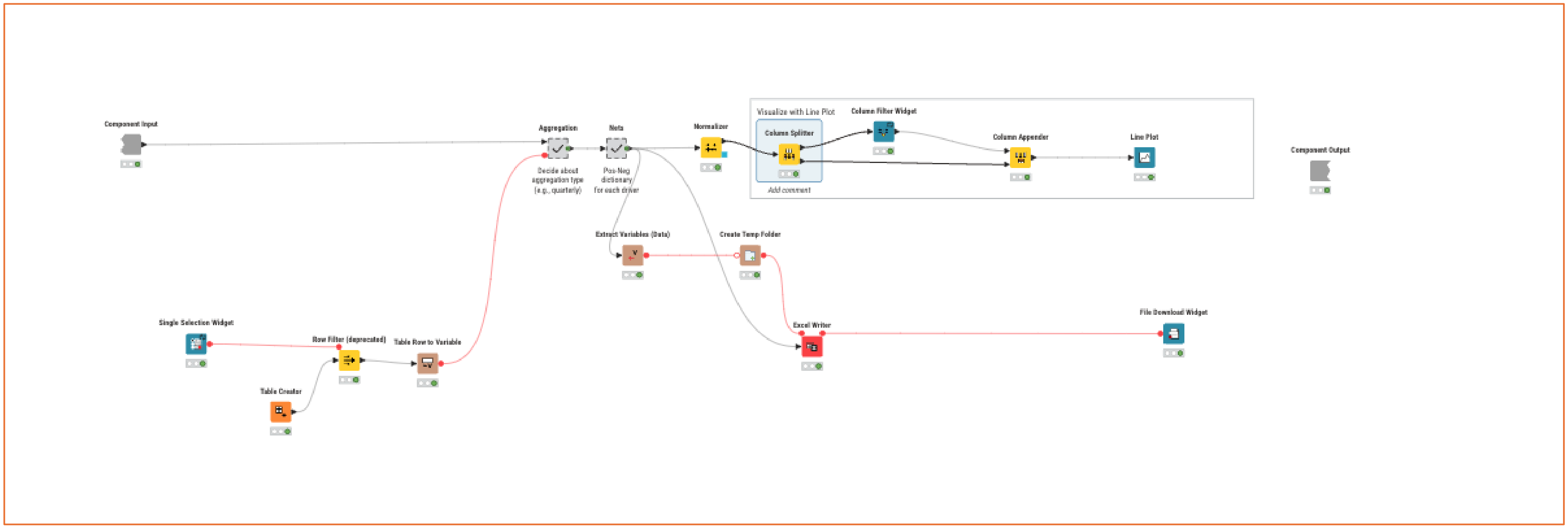
Uploading Data and choosing columns



Dictionary-based text-mining is applied to the data



Output table with brand reputation scores is calculated and visualized



APPENDIX B

W.3.1 *Dictionary from Rust et. al 2021*

Driver	Subdriver	Dictionary	Keywords
Value	Price	Positive	Cheap, afford, inexpens, deal, low, bargain, thrifti, reason, econom, frugal, joy, discount, pleas, sale
		Negative	Expens, pricey, costly, overpr, unfair, rich, excess, extravag, high, exclus, outrag
	Service quality	Positive	Help, great, fast, knowledge, attent, understand, easi, polit, patient, respect, prompt, compet
		Negative	Rude, frustrat, terribl, slow, careless, incompet, disrespect, aw, lazi, irrit, horribl, angri
	Goods quality	Positive	Quality, durabl, function, excel, perfect, use, beauty, strong, valu, sturdi, luxuri, worth, long-last, best, satisfi, impress, uniqu, clean
		Negative	Junk, bad, poor, wast, ugli, breakabl, worthless, flimsi, useless, disappoint, shoddi, mediocr, garbag, short-liv
Brand	Cool	Positive	Trendi, hip, awesom, cool, modern, stylish, current, sexi
		Negative	Ordinari, lame, ancient, averag
	Exciting	Positive	Fun, excit, inspir, happi, thrill, stimul, live, interest
		Negative	Bore, dull, uninspir, tire, bland
	Innovative	Positive	New, smart, invent, advanc, cut, futurist, intellig, progress, innov, technolog, creative, novel, cutting-edg
		Negative	Old, old-fashion, tradit, uninterest, outdate
	Social responsibility	Positive	Benevol, give, benefici
		Negative	Greedi, uncar, irrespons, evil, profit
Relationship	Community	Positive	Famili, involv, commun, social, togeth, harmoni
		Negative	Cold, sad, selfish
	Friendly	Positive	Nice, friendli, pleasant, kind, warm, welcom, trustworthi, open, accommod
		Negative	Mean, unpleas, unhelp, unfriendli, aloof, nasti, arrog
	Personal relationships	Positive	Connect, special, person, intim, close, profession, comfort
		Negative	Cold, distant, imperson, disconnect
	Trustworthy	Positive	Honest, reliable, good, depend, trust, transpar, safe, honesti, principl, hono
		Negative	Dishonest, unreli, cheat, shadi, untrustwo, deceit, decept, lie