Brand Activism: A Case Study using Sentiment Analysis

Sabrina Hegner[a]*, A. Varmaz[a], L. Köhler[a], & J. Schaumann[b]

[a]*International School of Business, City University of Applied Science Bremen, Germany*

[b]*Department of Design, Production and Management, University of Twente, Enschede, the Netherlands*

Correspondence Details:

Sabrina Hegner, Werderstr. 73, 28199 Bremen, Germany, email: Sabrina.hegner@hs-bremen.de

Brand Activism: A Case Study using Sentiment Analysis

Abstract

Brand activism has emerged as a powerful yet controversial strategy, allowing companies to take a public stance on socio-political issues. This study examines a real-world case of brand activism in Germany: Rewe's decision to terminate its long-standing partnership with the German Football Association in response to FIFA's stance on the "One Love" armband during the World Cup. Using sentiment analysis and qualitative content analysis, we analyze 3,551 first-level Instagram comments on Rewe's announcement. Findings reveal an overwhelmingly positive response (92%), with integrity and authenticity emerging as key themes. Users assessed Rewe's alignment between corporate values and actions, its consistency in implementing those values, and the clarity of its communication. Notably, while negative reactions were more fragmented, positive responses were unified, underscoring the impact of values and their protection. This study contributes to brand activism research by offering empirical insights based on actual consumer reactions, moving beyond experimental and survey-based approaches.

Keywords: brand activism, authenticity, sentiment analysis, content analysis

1. Introduction

„Brand activism is perhaps the most dramatic and surprising business trend of the past decade" (Korschun, 2021, p. 11). At first, the concept of brand activism appears promising: brands can take a public stance on socio-political issues, actively shape public discourse, and simultaneously differentiate themselves from competitors and foster a more personal connection with consumers by advocating for issues that matter to them (Koch, 2020). However, the reality is more complex. Depending on the topic and positioning, a political statement carries the risk of alienating a part of society and, consequently, potential customers. This risk explains why many brands are generally reluctant to abandon the perceived safety of neutrality.

Nevertheless, several major brands have emerged as pioneers of brand activism. Nike's case illustrates the wide spectrum of public reactions such campaigns can elicit. In the fall of 2018, Nike selected American football player Colin Kaepernick as a brand ambassador; an otherwise routine marketing decision. However, Kaepernick had previously sparked a national debate by publicly protesting against police violence and racial discrimination in the United States. While critics of the partnership went as far as burning Nike products, the campaign was simultaneously applauded. Notably, within days of the campaign's launch, Nike's stock price reached an all-time high (Thomas, 2018).

This example demonstrates that even when public reactions are polarized, the overall outcomes can still be positive. However, it is understandable that companies approach brand activism with caution. Several studies have identified a frequent negative net effect of brand activism on companies (Bhagwat et al., 2020; Hydock et al., 2019; Mukherjee & Althuizen, 2020).

Although valuable insights into brand activism exist, it remains a relatively young phenomenon, and research on the topic is still evolving (Cammarota et al., 2023; Pimentel et al., 2024). A critical gap in the literature lies in the limited alignment between theoretical findings and real-

world case studies. Many existing studies rely on experiments or survey data, while research based on actual brand activism cases is scarce.

To contribute to this underexplored area, we explore a real-world case study from Germany. In November 2022, the retail corporation Rewe announced the immediate termination of its long-standing partnership with the German Football Association (DFB). The decision was made in response to the DFB's choice to comply with FIFA's directive and forgo the "One Love" captain's armband during the FIFA World Cup in Qatar. The armband was intended as a symbol against all forms of discrimination. In its official press release, Rewe stated that "following FIFA's recent decisions and the statements made by FIFA President Infantino, [the brand felt] compelled to clearly distance itself from FIFA's position and renounce its advertising rights" (Rewe Media Center, 2022). With this move, Rewe ended a partnership that had been in place since 2008 (DFB, 2018). The present research aims to analyze consumer reactions to this instance of brand activism in shaping public sentiment.

2. Methodology

This study analyzes public reactions to Rewe's Instagram announcement terminating its sponsorship of the German Football Association (DFB) (see Figure 1). We combine manual coding, LLM-based sentiment classification, and qualitative content analysis to assess sentiment and identify key thematic patterns (Braun & Clarke, 2006). At data collection, the Instagram post had 57,255 likes and 4,111 comments. To ensure alignment with direct public responses, we restrict analysis to first-level comments (n = 3,551), excluding replies within comment threads. Comments were exported into a structured dataset, preserving punctuation, hashtags, mentions, and emojis.

-- Figure 1 here --

Sentiment was evaluated using three complementary approaches: (i) manual coding by the research team, (ii) LLM-based classification via the OpenAI API (GPT), and (iii) automated

sentiment analysis using ATLAS.ti. Triangulation across methods follows best practices in computational text analysis, emphasizing validation and robustness (Humphreys & Wang, 2018; Grimmer et al., 2022). The LLM was used as a zero-shot classifier, assigning each comment a sentiment score on a [-1,1] scale without task-specific training. This approach is increasingly adopted for large-scale social media analysis due to its strong performance relative to human annotators (Gilardi et al., 2023). To ensure reliability, two robustness checks were implemented: (1) Human–LLM agreement analysis with qualitative inspection of disagreements to identify systematic errors; and (2) Cross-tool comparison with ATLAS.ti's dictionary-based classification. Given the prevalence of emojis and context-dependent language in Instagram discourse, LLMs are better suited than rule-based tools for capturing pragmatic meaning. Observed divergence supports the methodological choice of LLMs for this domain. All LLM outputs were stored in a structured results file for quantitative summaries. While LLMs may reflect biases from training data (Bender et al., 2021), conclusions are based on aggregated, validated results rather than individual predictions. This transparent, multi-method approach enhances methodological rigor and aligns with emerging standards in AI-assisted qualitative research (Salganik, 2019; Grimmer et al., 2022). Appendix 1 gives a detailed description of the employed method.

3. Results

Results indicate an overwhelmingly positive consumer response, with 92% of comments expressing support for Rewe's decision. Figure 2 shows strong alignment between manual coding and GPT-based sentiment classification, confirming the reliability of the automated approach. In contrast, ATLAS.ti's accuracy was low (29% agreement with manual coding), particularly in classifying emoji-heavy comments often labeling them as neutral despite clear positive sentiment.

- Figure 2 here -

Qualitative content analysis corroborates the sentiment findings (see Table 1). Of the 3,551 comments, 90% (3,196) were assigned to one of eight positive categories, while 8% (279) fell into ten negative categories. The remaining 2% (63) were unclassifiable due to ambiguity, and 13 were labeled “Bag” as they referenced only the rainbow-colored bag in the post, not the campaign. Negative reactions were more fragmented, distributed across multiple categories with fewer comments each suggesting diverse criticisms. In contrast, positive responses were more unified, reflecting broad consensus on what was valued.

- Table 1 here –

The most prominent category ‘extremely positive reactions’ comprises 3,159 comments (89%), reflecting dominant public endorsement. The largest subgroup, General Approval, includes 2,170 comments, with 47% (1,018) consisting solely of positive emojis: clapping hands (👏), red hearts (💚), and rainbow flags (🌈). This widespread use of emojis suggests a collective, performative expression of solidarity, where visual symbols function as “digital applause,” signaling approval without detailed justification. The remaining 1,147 comments in this category contained textual content but offered minimal critique. Responses such as “I want to hug you!! But how do you embrace a corporation?! ♥” convey enthusiasm but lack critical engagement, indicating a superficial, affect-driven form of support. A more substantive layer emerged in a category of 379 comments (11%), where users explicitly expressed gratitude. Many praised Rewe for moving beyond symbolic gestures such as displaying rainbow flags toward concrete, potentially costly actions, particularly in light of the company’s prior history of verbal advocacy without tangible follow-through. One user noted: “Every time I saw the rainbow flag, I thought: ‘Symbols, okay. But we queers need actions.’ Now you’ve delivered—on time and in full.” This reflects a broader public demand for substance over symbolism, reinforcing the perception that the campaign marked a meaningful shift from rhetoric to action. Another key theme identified by approximately 9% of users was the integrity and authenticity

of Rewe's values. Users emphasized that the company's willingness to act, even at financial risk, enhanced its credibility. This aligns with theoretical expectations that audiences reward consistency between values and behavior. Furthermore, 30 users praised Rewe for demonstrating courage and backbone, a quality they felt was absent in institutions like the German Football Association (DFB). The metaphor of "having balls" (German: having eggs) was used in 75 comments, often playfully and deliberately, blending humor with moral endorsement: "You've got eggs (balls) —and not just in your product range!!!" This linguistic reclamation illustrates how users co-construct narratives of corporate courage. The exemplary nature of Rewe's behavior was another recurring theme, with 99 comments highlighting its potential to inspire other firms. 46% (45 of 99) expressed hope that other companies would follow suit. In 24 instances, users tagged competitors' social media accounts (e.g., @adidas_de, @cocacola_de) to publicly call for similar action, demonstrating how digital platforms enable peer pressure and normative influence. Six users also urged the DFB and captain Manuel Neuer to emulate Rewe, reinforcing its image as a first-mover and pioneer. Finally, 16 employees, 5 partners, and 2 creators expressed pride in their association with Rewe. Notably, a comment from 1. FC Köln (@fckoeln), a key partner of Rewe, generated 4,048 likes and 35 replies, underscoring how strategic alliances can amplify legitimacy and foster collective identity when values align.

The category of 'predominantly positive reactions' comprises 37 comments, reflecting a mixed but positive stance. The first theme (21 comments) acknowledges the campaign's commercial motivations, particularly its role in PR or marketing, yet argues that positive intent and strategic communication are not mutually exclusive. One user states: "Well done! ... We've known for ages companies live off PR. But what matters is the how." This reflects a critical yet constructive engagement, indicating that audiences are increasingly capable of distinguishing between performative messaging and genuine ethical commitment. The second theme (16 comments) offers a conditional endorsement, affirming the campaign's values while calling for

deeper, more concrete actions. A majority demand ethical consistency across Rewe's broader operations, particularly in its product portfolio. Users call for the removal of products from companies like Nestlé and Coca-Cola, perceived as contradicting the campaign's values: "@rewe next step: @nestle ban". This reflects a growing expectation of value congruence—the alignment between a brand's public statements and its supply chain practices. Additionally, four users explicitly request financial contributions to human rights organizations, framing this as essential to demonstrate sincerity: "If you now also sponsor a human rights org, I'll be with you... 💚" This demand signals a shift from symbolic to material action, reflecting a broader public expectation that moral statements must be backed by tangible investment.

The category of 'predominantly negative reactions' comprises 104 comments (3%), reflecting a critical stance despite mixed sentiment. The first theme centers on timing and authenticity. Rewe had announced in October it would not renew its partnership with the DFB which was before the World Cup in Qatar started. Yet users interpret the campaign as strategic opportunism, arguing it retroactively assigns moral significance to a pre-existing business decision. One user states: "You were already not going to continue with the DFB. Pretending that a political stance underlies this move is simply deliberate misinformation." Rewe responded by citing FIFA's recent decisions, but users criticized the repetitive, formulaic nature of its replies: "@rewe 300 times the same text…" The second theme focuses on inappropriate timing, given that Qatar's human rights record had been widely known for years. Users argue the campaign was reactive, not proactive, driven by media momentum rather than sustained ethical commitment. The third theme represents perceived double standards. 23% of users criticize the continued sale of Nestlé products, while two highlight the inconsistency in promoting travel to Qatar and other contested countries. 18 users emphasize the need for internal reform, citing homophobic and racist incidents involving Rewe employees, which undermine the credibility of its public values. Finally, 5 users raise a comparative critique, questioning why Rewe did not act during the 2018 World Cup in Russia or the 2022 Winter

Olympics in China, both marked by serious human rights concerns. This fuels perceptions of selective morality. The final theme (20 comments) represents a philosophical critique of brand activism in sport. Users argue that politics and sexuality should not be intertwined with sport, viewing the campaign as unnecessary and potentially harmful to athletes.

The most critical category comprises 175 comments (5%), reflecting a distinct and emotionally charged opposition. The first theme critiques the purpose and proportionality of the campaign: eight users call it meaningless, asserting "the next sponsor will simply knock on the door anyway," while five accuse Rewe of overstepping. A second theme challenges authenticity, framing the campaign as a purely instrumental marketing strategy. Eleven users cite the pre-announced contract end as evidence of strategic timing. Six others emphasize the absence of sincere intent: "In the end, it's all about money for Rewe." The remaining 15 comments use common discursive markers: #virtuesignaling "That's what you can call woke-washing. But the PR is working." "#pinkwashing rubbish!" These terms function as symbolic shorthand for accusations of superficiality and hypocrisy. A further 35 comments reflect a broad rejection of the campaign's values. The first theme involves opposition to LGBTQIA+ visibility, with users expressing discomfort with rainbow symbolism and claiming to feel "excluded as a heterosexual." One asks: "Should heterosexuals no longer be allowed to shop here?" The second theme centers on cultural relativism, arguing that Western values should not be imposed on other cultures. One user critiques the "moral superiority" of the West: "We think we must bring civilization to the barbarians—claiming to know what is right or wrong." Instead, the argument calls for respecting the values of Islamic and Muslim-majority nations. The boycott intention category (22 comments) represents the negative counterpart to supportive consumer behavior. While some users state they were never Rewe customers and now feel proud of their non-consumption, others express absolute and irreversible withdrawal: "I will never shop at Rewe again." This contrasts with the more temporary shifts in the positive counterpart, suggesting that moral offense can trigger radical, enduring disengagement. Finally, 73

comments in the General Criticism category include 18 emojis (clown, thumbs down, vomiting) and textual expressions of deep skepticism: “Market of shame”, “Rewe, no thanks.”, “I thought you dealt with groceries. Don’t act like a moral authority.” These rhetorical questions underscore a perceived boundary violation: the belief that commercial enterprises should not assume the role of ethical arbiters.

4. Discussion & Conclusion

This case study demonstrates that brand activism can generate strong public support when perceived as authentic, consistent, and value-driven. The overwhelmingly positive response to Rewe’s campaign evidenced by over 3,000 comments reflects a public increasingly attuned to the alignment between corporate values and actions. However, the success of such initiatives hinges critically on perceived authenticity, a multidimensional construct shaped by value-practice alignment, consistency across decisions, and strategic communication timing. These findings resonate with existing research (Luna-Amador et al., 2025; Mirzaei et al., 2022; Vredenburg et al., 2020). The prominence of themes such as courage, material commitment, and internal organizational integrity underscores a shift toward value-based consumerism (Sen & Bhattacharya, 2001; Grewal et al., 2021). Audiences expect brands to act in accordance with their stated principles not only in high-visibility campaigns but across supply chains, employee practices, and travel offerings. Moreover, the use of humor, public tagging, and partner endorsement illustrates how audiences actively co-construct narratives of ethical leadership, transforming corporate actions into cultural moments. The viral engagement with 1. FC Köln’s comment (receiving 4,048 likes) demonstrates how strategic alliances can amplify legitimacy when values align. Yet this also reveals a risk: affective responses, while powerful, may be short-lived. The dominance of emojis in the General Approval category suggests that digital platforms enable low-effort, high-impact solidarity but this may coexist with limited critical reflection.

Conversely, negative reactions reveal a public equally sophisticated in its critique. Skepticism toward timing, perceived double standards, and cultural imperialism highlights a demand for systemic consistency. The criticism of Rewe's travel offerings, employee conduct, and selective activism (e.g., inaction during the 2018 World Cup) underscores that audiences hold brands to a higher standard: one of moral coherence across time, space, and context.

The emotional intensity of responses evident in emojis, hyperbolic language, and calls for boycott suggests that brand activism can trigger affective polarization, particularly when perceived as performative or ideologically driven. Terms like #virtuesignaling, #pinkwashing, and woke-washing function as symbolic shorthand for distrust, reflecting the entrenchment of critical discourses in public debate.

This study contributes to the field by offering a real-world, empirically grounded case based on public social media discourse moving beyond experimental or survey-based approaches. Its focus on a German retail brand also diversifies the predominantly U.S.-centric literature, enriching the global understanding of brand activism.

In conclusion, while this case offers compelling evidence for the potential of brand activism, its generalizability is limited by its single-brand, single-country scope. Nevertheless, the findings underscore a clear imperative: authenticity is not a marketing tactic but a strategic necessity. Brands that align values with behavior, communicate transparently, and act consistently across all domains are more likely to earn lasting trust. Future research should explore how affective engagement translates into sustained behavioral change and how authenticity is perceived across different cultural and institutional contexts.

References

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21) (pp. 610–623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922

Bhagwat, Y., Warren, N. L., Beck, J. T., & Watson IV, G. F. (2020). Corporate Sociopolitical Activism and Firm Value. *Journal of Marketing, 84*(5), 1–21. https://doi.org/10.1177/0022242920937000

Cammarota, A., D'Arco, M., Marino, V., & Resciniti, R. (2023). Brand activism: A literature review and future research agenda. *International Journal of Consumer Studies, 47*(5), 1669-1691.

DFB (2018). BEI REWE: DFB-SAMMELKARTEN ZUR WM. *DFB News*. retrieved 20. April 2024, https://www.dfb.de/news/detail/bei-rewe-dfbsammelkarten-zur-wm-186317/

Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. Proceedings of the National Academy of Sciences of the United States of America, 120(30), e2305016120. https://doi.org/10.1073/pnas.2305016120

Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). Text as Data: A New Framework for Machine Learning and the Social Sciences. Princeton University Press.

Humphreys, A., & Wang, R. J.-H. (2018). Automated text analysis for consumer research. Journal of Consumer Research, 44(6), 1274–1306.

Hydock, C., Paharia, N., & Weber, T. J. (2019). The Consumer Response to Corporate Political Advocacy: A Review and Future Directions. *Customer Needs and Solutions, 6*(3–4), 76–83. https://doi.org/10.1007/s40547-019-00098-x

Koch, C. H. (2020). Brands as activists: The Oatly case. *Journal of Brand Management, 27*(5), 593–606. https://doi.org/10.1057/s41262-020-00199-2

Korschun, D. (2021). Brand Activism Is Here to Stay: Here's Why. *NIM Marketing Intelligence Review, 13*(2), 10–17. https://doi.org/10.2478/nimmir-2021-0011

Novak, P., Smailović, J., Sluban, B., & Mozetič, I. (2015). Sentiment of Emojis. PLoS ONE, 10(12), e0144296. https://doi.org/10.1371/journal.pone.0144296

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. Journal of Finance, 66(1), 35–65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Luna-Amador, J., Caraballo-Payares, A., & Escobar-Espinoza, A. (2025). Brand activism: Research trends and cluster analysis. European Research on Management and Business Economics, 31(2), 100279.

Mirzaei, A., Wilkie, D. C., & Siuki, H. (2022). Woke brand activism authenticity or the lack of it. *Journal of Business Research, 139*, 1–12. https://doi.org/10.1016/j.jbusres.2021.09.044

Mukherjee, S., & Althuizen, N. (2020). Brand activism: Does courting controversy help or hurt a brand? *International Journal of Research in Marketing*, *37*(4), 772–788. https://doi.org/10.1016/j.ijresmar.2020.02.008

Pimentel, P.C., Bassi-Suter, M. & Didonet, S.R. (2024) Brand activism as a marketing strategy: an integrative framework and research agenda. Journal of Brand Management, *31*, 212–234. https://doi.org/10.1057/s41262-023-00335-8

Rewe Media Center (2022). REWE beendet Kooperation mit dem DFB. retrieved 09. February 2026, https://mediacenter.rewe.de/pressemitteilungen/dfb-kooperation-beendet

Salganik, M. J. (2019). Bit by Bit: Social Research in the Digital Age. Princeton University Press.

Thomas, L. (2018). *Nike's Kaepernick ad draws record likes on social media, sends stock to all-time high amid sales boost*. CNBC, retrieved 09. February 2026, https://www.cnbc.com/2018/09/14/nikes-kaepernick-ad-should-fuel-sales-as-retailer-knows-its-consumer.html#:~:text=%22Indeed%2C%20%5BNike%5D%20management,a%20share

%20on%20the%20stock.

Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. Journal of Marketing Research, 51(4), 463–479. https://doi.org/10.1509/jmr.12.0106

van Atteveldt, W., Trilling, D., & Arcíla Calderón, C. (2022). Computational Analysis of Communication: A Practical Introduction to the Analysis of Texts, Networks, and Images with Code Examples in Python and R. John Wiley & Sons. https://cssbook.net

Vredenburg, J., Kapitan, S., Spry, A., & Kemper, J. A. (2020). Brands taking a stand: Authentic brand activism or woke washing? *Journal of Public Policy & Marketing*, *39*(4), 444–460. https://doi.org/10.1177/0743915620947359

Figure 1. Instagram post announcing the termination (text: “our values – not for sale”)



Figure 2. Results of the sentiment analysis

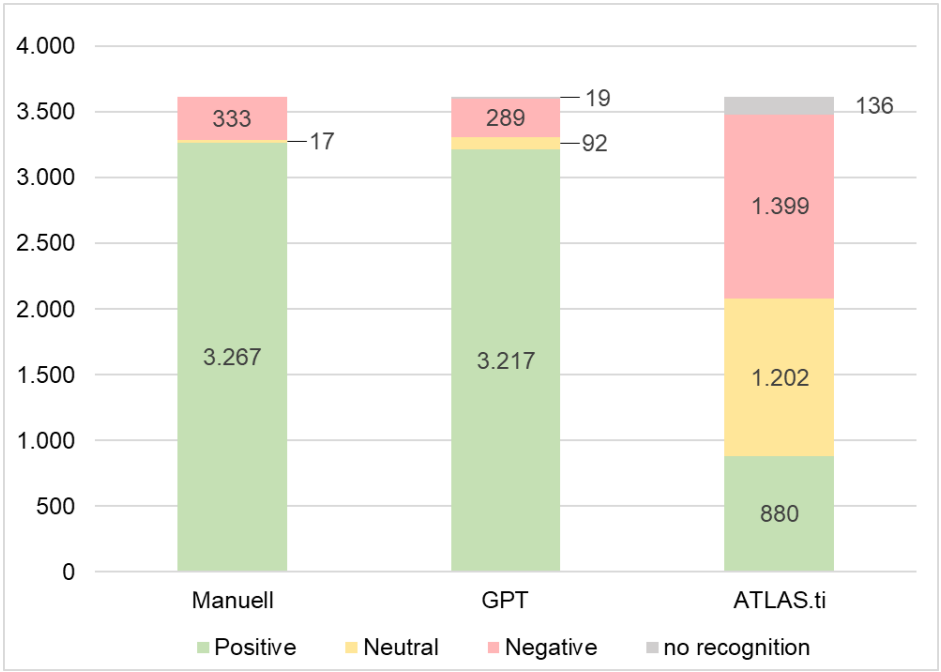

Table 1. Results of the qualitative analysis

Category	Sentiment	Sentiment-Score (Mean)	Frequency	Frequency (%)
General approval	very positive	0,85	2170	61%
Gratitude		0,85	379	11%
Integrity / Value stability		0,79	354	10%
Leading by example		0,72	99	3%
Buycott		0,83	134	4%
Pride		0,84	23	1%
Marketing for the right reason	positive	0,78	21	1%
Approval and further demands		0,48	16	0%
Lacking transparency about termination date	negative	-0,27	24	1%
Termination date / delayed response		-0,29	9	0%
Double standards / Whataboutism		-0,55	51	1%
Politics in sport		-0,46	20	1%
Critique sense / Appropriateness	very negative	-0,49	13	0%
Marketing exploitation		-0,47	32	1%
Rejection of values		-0,59	35	1%
Boycott		-0,71	22	1%
General critique		-0,59	73	2%
Bag	-	0,51	13	0%
not classified	-	-0,11	63	2%
			3551	100%

Appendix 1: Sentiment analysis LLM approach

Sentiment analysis approaches and triangulation

To quantify overall evaluative reactions to the brand-activism announcement, we employ three sentiment classification approaches: (i) manual coding by the research team, (ii) LLM-based sentiment ratings using the OpenAI API (GPT), and (iii) the sentiment function implemented in ATLAS.ti. Triangulating across methods follows established best practices in automated text analysis: algorithmic outputs should be validated against human judgment and checked for robustness to alternative operationalizations before being used to support substantive inferences (Humphreys & Wang, 2018; Grimmer, Roberts, & Stewart, 2022).

Manual sentiment coding serves as the primary benchmark and was conducted using a shared codebook defining positive, neutral, and negative reactions. Disagreements were resolved through discussion. The LLM-based approach is described in detail below. ATLAS.ti provides an off-the-shelf benchmark frequently used in qualitative research. The ATLAS.ti performance helps illustrate how conventional tools handle short, emoji-heavy social-media text.

LLM-based sentiment classification using the OpenAI API

The LLM-based sentiment classification treats the model as a zero-shot text annotator. Zero-shot classification is increasingly used when researchers seek to label large volumes of text without training task-specific models, and recent evidence suggests that LLMs can match or exceed crowd-worker and even trained-annotator performance on several annotation tasks (Gilardi, Alizadeh, & Kubli, 2023).

We used a Python pipeline that calls the OpenAI API for each comment and returns a sentiment label. Each comment is processed independently and stored with its predicted label in a structured results file. This file constitutes the basis for all quantitative summaries and robustness checks reported in the Results section.

The prompt instructs the model to assess the overall sentiment expressed toward Rewe's announcement. We use an ordinal sentiment scale and map model outputs into categories

consistent with the manual codebook. Compared to lexicon/dictionary-based tools, LLMs can incorporate context, pragmatic cues, and emoji usage in a single judgment, which is particularly relevant on Instagram (Novak et al., 2015).

We treat the LLM as a measurement instrument rather than an oracle. LLM outputs reflect a probabilistic judgment conditioned on the prompt, model version, and sampling settings, and they can inherit biases from their training data (Bender, Gebru, McMillan-Major, & Shmitchell, 2021). Consequently, substantive conclusions should not depend on a single specification. This motivates the sensitivity analysis described next (Salganik, 2019; Grimmer et al., 2022).

Sensitivity analysis assesses whether the substantive findings are robust to plausible variations in how the LLM is used and how its outputs are mapped into analytic categories. In the present context, sensitivity concerns arise less from parameter estimation (as in classical econometrics) and more from (i) borderline or mixed-sentiment comments, (ii) emoji-only comments and other ultra-short texts, and (iii) aggregation rules that collapse fine-grained labels into positive/neutral/negative classes. We implement the following robustness checks.

(1) Human–LLM agreement and error inspection. We quantify agreement between manual sentiment labels and the LLM's labels and report this correspondence in Figure 2. Beyond aggregate agreement, we inspect disagreement cases qualitatively to diagnose systematic failure modes (e.g., irony, sarcasm, or context-dependent slang). This aligns with guidance that validation should combine quantitative reliability metrics with close reading of edge cases (Humphreys & Wang, 2018; Grimmer et al., 2022).

(2) Alternative label aggregation. We re-map the LLM's ordinal outputs into coarser positive/neutral/negative classes using alternative aggregation rules (e.g., treating weakly positive as neutral, or collapsing neutral into the nearest polarity). The goal is to ensure that the central descriptive conclusion, an overwhelmingly supportive reaction, does not hinge on an arbitrary threshold choice. The importance of robustness to dictionary and mapping choices is well documented in sentiment applications (Loughran & McDonald, 2011).

(3) Cross-tool comparison (ATLAS.ti). We compare LLM outputs against ATLAS.ti classifications. Systematic differences between tools are interpreted as evidence about construct validity in this setting. In particular, we expect dictionary-like tools to underperform on short, emoji-heavy content, while LLMs should better incorporate pragmatic meaning; observing divergence therefore helps justify the methodological choice of an LLM for Instagram discourse.

Collectively, these checks operationalize a conservative principle: the LLM is used to scale the analysis, but conclusions are accepted only if they are robust to alternative mappings and remain broadly consistent with human judgment. This approach is consistent with emerging best practices in computational social science that emphasize transparency, validation, and robustness when using automated text analysis to support substantive claims (Salganik, 2019; Grimmer et al., 2022).