Decoding Digital Trust:

A Multi-dimensional Analysis of Tech Influencer Credibility on YouTube

Submitted by: Kevin Floether, CM

Department of Communications Studies and Multimedia

Faculty of Humanities, McMaster University

Course: COMMGMT 740 Professional Project

Supervisor: Dr. Alex Sévigny, Ph.D., APR

Acknowledgements

First and foremost, I would like to express my gratitude to my wife for her patience throughout this journey as she gracefully endured significant competing demands for my time and the occasional academic meltdown when I hit a wall in my study. Through it all, she remained encouraging and supportive. I cannot thank her enough. Additionally, I'd like to thank my study partner, our 9-year-old cockapoo, Charlie, who dutifully sat with me daily as I worked.

I am also profoundly grateful to my parents for their constant support; in particular, my mother, whose willingness to discuss and refine ideas with me was instrumental to this work.

I owe a significant debt of gratitude to my supervisor, Dr. Alex Sévigny, whose guidance was instrumental in my development during the Master of Communications Management (MCM) program. His insights and mentorship through this program have been irreplaceable. I am equally grateful to the program chair, Dr. Terry Flynn, for his guidance throughout the program and for serving as my second reader. His written feedback went above and beyond the responsibilities of his role and was sincerely appreciated.

I also thank my colleagues and coworkers at Chartered Business Valuators Institute for supporting me and allowing me to pursue my studies while working full time. Their understanding and flexibility made it possible to balance work and academics.

I would also like to thank my supportive MCM cohort, whose camaraderie made the MCM journey enriching and life changing. Special thanks to Amy Kennedy for her assistance with coding at a critical juncture when I was facing challenges I couldn't overcome alone, and to Anne Locke and Julia Wilkinson for talking me through problems on multiple occasions.

To everyone who has been part of this journey, thank you for making this achievement possible.

Abstract

This study examines trust dynamics in technology influencer marketing on YouTube, employing a multi-faceted approach to analyze trust indicators in comments, content similarity between influencer and consumer reviews, and media portrayal. The research focuses on 10 prominent tech influencers, using natural language processing techniques including sentiment analysis and lemmatization. Findings reveal significant correlations between trust scores derived from YouTube comments and the similarity of influencer content to consumer reviews on Amazon. Media coverage analysis uncovered unexpected positive correlations between trust scores and controversy-related keywords, particularly when framed in terms of integrity. The study contributes to understanding trust formation in digital spaces and offers insights for effective and ethical influencer marketing practices. Limitations include reliance on keyword analysis and sentiment scores, and the cross-sectional nature of the research. Future studies could benefit from longitudinal approaches and cross-cultural comparisons to further explore trust dynamics in influencer-audience relationships.

Keywords: Tech influencers, trust measurement, social media marketing, content analysis, sentiment analysis, media portrayal, YouTube, influencer credibility, consumer opinions, digital trust

Table of Contents

Introduction

In the rapidly evolving landscape of digital media, technology (tech) influencers have emerged as significant voices shaping consumer opinions and purchasing decisions. Evidence of that can be found in a McKinsey article that estimates the value of the influencer marketing economy at $21.1 billion USD in 2023 (McKinsey & Company, 2023). Sprout Social's 2024 influencer marketing report noted that "49% of all consumers make daily, weekly, or monthly purchases because of influencer posts, with 30% trusting influencers more today than they did just six months ago" (2024, para. 1).

However, there are also risks for brands when it comes to navigating this influencer-rich business environment. From influencer misconduct, to inauthentic partnerships, there are several potential missteps for brands when leveraging influencer marketing – just ask Google (Song, 2024). This study investigates the complex dynamics of trust in the realm of tech influencers on YouTube, exploring how trust indicators in audience interactions correlate with consumer opinions and media portrayals. By examining the multifaceted nature of trust in this context, the researcher aims to develop a comprehensive trust rating system for tech influencers, contributing to a deeper understanding of influencer credibility and its implications for consumer behaviour.

The research employs a multi-stage data collection and analysis process, drawing on a diverse range of sources, including YouTube comments, Amazon product reviews, and media coverage. Using Grunig and Hon's (1999) trust components - integrity, dependability, and competence - as a theoretical framework, this study applies advanced natural language processing techniques and statistical analysis to measure trust levels and explore their relationships with various factors.

Central to this investigation is the analysis of trust indicators in YouTube comments on tech influencers' product reviews. Through an iterative approach combining keyword analysis, sentiment analysis, and content similarity comparisons, the researcher examines how these indicators align with consumer opinions expressed in Amazon reviews for the same products. This methodology provides insights into the predictive power of influencer trustworthiness on consumer sentiment.

Furthermore, this research explores the relationship between influencer trust levels and media portrayal, analyzing how tech influencers are represented in news articles and investigating any legal actions against them. This multifaceted approach allows for a nuanced understanding of how trust in tech influencers is constructed, maintained, and potentially challenged across different platforms and contexts.

For the purposes of this study, we define a "tech influencer" as a content creator on YouTube who regularly produces technology product reviews and has a significant following (over 500,000 users). *Trust* is conceptualized using Grunig and Hon's (1999) components, operationalized through the presence and frequency of specific keywords and sentiment in audience interactions.

Rationale

This research addresses a critical gap in the understanding of trust dynamics within the realm of technology influencer marketing on YouTube. The necessity for this research is evidenced by recent critiques of social media's impact on society and individual psychology. Fisher (2022) presents a compelling argument for how social media platforms have rewired our cognitive processes and social interactions. In light of these broad societal changes, understanding the nuances of trust formation in digital spaces becomes increasingly crucial. By focusing on tech influencers, this study not only contributes to marketing and public relations literature but also provides insights into the larger question of how digital platforms shape our perceptions and decision-making processes. As Fisher (2022) notes, the algorithms driving social media can amplify certain voices and perspectives. Thus, understanding how trust is built and maintained by influencers within this system is essential for a comprehensive grasp of modern digital communication dynamics.

As digital platforms continue to shape consumer behaviours and purchasing decisions, the role of tech influencers has become increasingly significant. However, there is a lack of comprehensive studies that examine the multifaceted nature of trust in this context, particularly ones that integrate various data sources and analytical approaches.

According to Hubspot, there are over 3.3 billion digital video viewers worldwide, and 62% of consumers have watched a product review or unboxing video to learn more about a product or a brand (Hubspot, 2024). Additionally, over 90% of YouTube users say they find new products on the platform (Weinstein, 2019).

The relevance of this research is underscored by the rapidly evolving landscape of technology and digital media. As highlighted in the 2024 Edelman Trust Barometer, there is a growing dispersion of authority in society, with peers often being as trusted as traditional experts when it comes to information about new technologies. This shift in trust dynamics necessitates a deeper understanding of how influencers, who often occupy a space between peer and expert, cultivate and maintain trust with their audiences.

This study's unique approach combines sentiment analysis of YouTube comments, comparative analysis of influencer and consumer reviews, and examination of media portrayal, to offer a holistic view of trust that has been largely absent from existing literature. By developing a trust rating system based on Grunig and Hon's (1999) components of trust, this research provides a framework for evaluating influencer trustworthiness in the technology sector.

The findings of this study will benefit multiple stakeholders across academia, industry, and society. For the academic community, this research contributes to the fields of public relations, digital marketing, consumer behaviour, and trust studies by providing empirical evidence on the factors that influence trust in technology influencers. It also offers a methodological framework that can be adapted for studying trust in other digital contexts. In the marketing industry, professionals collaborating with technology influencers will gain insights into the elements that contribute to influencer trustworthiness and how those elements have an impact on consumer opinions. As such, this research will help inform more effective influencer selection and partnership strategies.

Technology companies stand to benefit from understanding how influencer reviews align with consumer sentiments, allowing them to better anticipate market reactions to their products and adjust their marketing and product development strategies accordingly. Consumers, in turn, can be empowered by this research, as it sheds light on the dynamics of influencer trustworthiness, potentially leading to more informed decision-making when engaging with influencer content. Furthermore, policymakers concerned with consumer protection and digital literacy can use these findings to inform regulations and educational initiatives as digital platforms and the practice of influencer marketing both continue to grow.

The originality of this research lies not only in its comprehensive approach to measuring trust but also in its focus on technology influencers specifically. While influencer marketing has been studied broadly, the unique position of tech influencers as intermediaries between complex technological innovations and everyday consumers warrants dedicated attention in our current knowledge and technology driven economy. Moreover, by examining the relationship between influencer trust and media portrayal, this study offers insights into the broader media ecosystem surrounding technology influencers. This aspect is particularly valuable given the increasing scrutiny of digital platforms concerning the spread of information (and misinformation) online.

This research addresses a significant gap in our understanding of trust in the digital age, with specific focus on the influential yet understudied domain of technology influencers on YouTube. Its findings have the potential to impact both academic discourse and industry practices, contributing to a more nuanced and effective approach to influencer marketing in the technology sector. By providing a comprehensive analysis of trust dynamics in this context, this study aims to enhance our understanding of digital influence and its implications for consumer behaviour, technological adoption, and the broader digital media landscape.

Literature Review

Trust & Trust Measurement

Trust is a fundamental concept in public relations that has grown in importance in recent years. As Paine (2013) notes, trust has an explicit impact on the financial health of organizations, and a lack of trust can have severe consequences. Further, "trust is important because it helps consumers overcome perceptions of uncertainty and risk and engage in 'trust-related behaviors' with Web-based vendors, such as… making purchases" (Choudhury et al., 2002, p. 355).

Public relations theorists conceptualize trust as a multi-dimensional construct. Grunig and Hon (1999) identify three key dimensions of trust in organization-public relationships: integrity (the belief the organization is fair and just); dependability (the belief the organization will do what it says); and competence (the belief the organization can do what it says). Trust is also recognized as being multi-level (existing between individuals, teams, and organizations), culturally-rooted, communication-based, dynamic, and multi-dimensional (Choudhury et al., 2002; Fleischmann et al., 2007; Paine, 2013). In addition, Beldad et al. (2010) also comment on the importance of context in trust formation. The authors note that the determinants of online trust may vary depending on the specific context of the online transaction and the parties involved.

Building on this understanding, Forde et al. (2015) distinguish between trust and distrust as separate constructs in their study of user experiences with websites. Their research highlights how different website characteristics influence trust and distrust, emphasizing that these two constructs have unique antecedents. Specifically, distrust is more often triggered by poor graphic design, usability issues, and privacy concerns, while trust is fostered by strong content design, visible security signals, and social proof, such as positive user reviews and recommendations from friends. This aligns well with insights from Scott Baradell in his 2021 book, *Trust Signals: Brand Building in a Post-Truth World*. Baradell writes, "trust signals are evidence points that inspire confidence in a brand online" (2022, p. 35). While the term *trust signal* is most closely associated with e-commerce sites, the author argues that it applies to all brands today. Further, he argues that brands must focus on building trust with their audiences through authentic, consistent, and credible trust signals (Baradell, 2022).

The trust components proposed by Grunig and Hon (1999) offer a solid foundation for assessing trust in various contexts. Adapting these components and employing a combination of research methods can provide valuable insights into the relationship between trust and consumer opinion.

Fleishmann et al. (2007) developed a model of trust in digital information. The authors integrated research on trust from behaviour and social sciences with research on information quality and human-computer interaction. Within their model, trust is positioned in a central mediating role between information quality, and information usage. In short, if trust is lacking, the provided information will not be used by the recipient.

Measuring trust is complicated but has been discussed in depth in public relations literature. Paine writes that “trust measurement is a way of giving a result a precise dimension, generally by comparison to some standard or baseline, and usually is done in a quantifiable manner” (2013, p. 6). Paola Pascual-Ferra (2020; 2021) published two scholarly articles that survey the literature on trust measurement. One of the most significant takeaways from this comprehensive systematic review is that surveys are a very common approach, which indicates a lack of experimental variety in how trust is measured.

When measuring trust, Ghafari et al. (2020) highlight significant challenges, particularly the lack of user-specified trust relations in online social networks and the context-dependent nature of trust. They note that on social media, the density of a standard trust network is below 0.01, and that predicting trust relations with such limited observed links is daunting. However, despite these challenges, recent research demonstrates the prevalence and utility of computational methods for trust measurement. Computational sentiment analysis of social media content has proven effective as a proxy for trust.

Chandio and Sah (2020) show how sentiment analysis of tweets can reflect public trust in political contexts. Additionally, Sabatini and Sarracino (2019) employed sophisticated econometric techniques, including instrumental variables, to analyze how social media usage affects different types of trust. Their approach demonstrates that computational methods, when combined with careful statistical analysis, can yield valuable insights into trust dynamics. Building on the computational approach of Chandio and Sah, and Sabatini and Sarracino, Choudhury et al. (2002) developed and validated a model for measuring trust in e-commerce contexts. The authors, having validated their model across multiple dimensions, provide a framework through which various digital environments can be examined for the presence and strength of trust. Their work is particularly valuable for understanding how trust can be operationalized and measured in online contexts, where direct interpersonal interactions are minimal. Trust in such environments is often established through factors like website quality, security assurances, and user experience.

Additional depth comes from Rawlins (2008), who draws connections between transparency and trust. Greater transparency by an organization (conceived as public participation, substantial information sharing, accountability, and reduced secrecy) tends to instill more trust, while less transparency diminishes it. Tschannen-Moran and Hoy's (2000) review further explains the meaning and dynamics of trust across disciplines. The authors distill trust into the facets of willingness to risk vulnerability, confidence, benevolence, reliability, competence, honesty, and openness. They examine how trust ebbs and flows throughout a relationship based on cognitive appraisals and emotional responses. Sustaining trust relies on supportive organizational attributes and collective norms.

Social Media

The rise of social media has profoundly changed the communication landscape, providing new ways for organizations to engage with their stakeholders and build brand relationships. This has important implications for how organizational credibility and trust are established and maintained in the digital age.

The relationship between social media use and trust has been a subject of debate among scholars. A systematic literature review by Håkansson and Witmer (2015) synthesized existing research on this topic, revealing a complex picture. Their analysis found that the majority of studies (8 out of 10) indicated a positive relationship between social media use and trust. For instance, some studies found that digital skills were associated with increased generalized trust, particularly among certain ethnic groups (Marichal & Monforti, 2014). Others observed that social media facilitated information sharing and communication, which in turn built trust (Liss, 2011). However, the review also highlighted important nuances. The positive effects of social media on trust were often mediated by factors such as social resource motivation (Beaudoin, 2008) and the type of online activity involved (Shah et al., 2001). Importantly, Håkansson and Witmer (2015) noted that most studies relied on cross-sectional data, making it difficult to establish causality between social media use and trust. They concluded that more research is needed, particularly using longitudinal data and qualitative methods, to better understand the causal mechanisms and contextual factors influencing the relationship between social media and trust.

One key aspect in the establishment of organizational trust is the shift towards more interactive, two-way dialogue enabled by social media platforms. As Kim and Brown (2015) argue, social media platforms have allowed publics to actively contribute to meaning creation with organizations, rather than being passive recipients of one-way messages. Meaning creation refers to the process of organizations and publics jointly developing understanding and ideas through two-way dialogue and interaction on social media. Organizations that embrace this dialogic potential, inviting conversation and valuing individuals' perspectives, can enhance their credibility in the eyes of publics. The authors identify four key dimensions of credibility in social media spaces: personable interaction, expertise, invitational rhetoric, and trustworthiness. In the authors' view, organizations can build credibility by personally interacting with publics, demonstrating expertise in content and presentation, inviting dialogue, and cultivating trust. Further, the authors study found that "trustworthiness, defined as the … brand's social media being perceived as honest, reliable, and with integrity, is at the core of social media credibility" (Kim & Brown, 2015, p. 11).

However, the participatory nature of social media also presents challenges for organizations seeking to manage their reputation. As Valentini (2015) notes, the participatory challenges in social media include the difficulty of fostering genuine engagement due to the prevalence of one-way promotional content, the potential backlash from audiences when content is perceived as inappropriate or offensive, and the increased public skepticism and distrust resulting from unethical practices and privacy violations by organizations. This calls for active reputation monitoring and management through strategic social media engagement. Valentini (2015) argues for a critical re-examination of traditional public relations models considering the social media environment, suggesting that Grunig and Hunt's (1984) two-way symmetrical model, with its emphasis on mutual understanding over persuasion, is particularly well-suited to building authentic socially mediated relationships.

The dynamic, real-time nature of social media also opens new possibilities for measuring and tracking trust. Boertjes et al. (2012) propose a formal model for gauging trust based on sentiment analysis of social media posts, considering both expressed sentiment and source authority. Source authority in this context refers to how influential or authoritative the source of the post is considered to be. In this study, the researchers used *number of Twitter followers* as a simple measure of source authority. By fitting this model to user feedback data, they demonstrate the feasibility of a social media-based trust monitor as an alternative to traditional survey methods. Such tools can support organizations in assessing current trust levels by detecting shifts in public opinion, and simulating the impact of interventions.

Boertjes et al. also argue for the appropriateness of sentiment analysis as a way to determine trust on social media, defining the technique as a technology that “attempts to detect opinions and subjective utterances, labelling subjective, opinion-bearing utterances with polarity scores such as ’negative’, ’positive’, or points on a metric scale” (2012, p. 253).

Recent research has emphasized the growing sophistication of sentiment analysis methods in social media contexts. Casánez-Ventura et al. (2023) conducted an extensive review of over 2,300 academic publications, revealing that while traditional techniques like lexicons and natural language processing remain prevalent, newer transformer-based methods are emerging but not yet widely adopted. Their analysis found that temporal dimensions of sentiment are particularly valuable for understanding emotional dynamics, especially when combined with spatial analysis. The authors also highlighted a significant disparity between academic research and industry applications, noting that the number of patents filed in this domain (over 8,000 in five years) far exceeds academic publications, suggesting substantial unreported advances in commercial applications of sentiment analysis.

While social media platforms offer rich opportunities for building credibility and trust, organizations must also navigate the complexities of this environment with care. As Song et al. (2023) observe in the context of higher education, institutions face increasing costs and market pressures that necessitate effective social media engagement to build brand image and loyalty. Their study highlights the importance of personable interaction and information sharing in driving relationship quality and brand performance outcomes, something with which Kim and Brown (2015) would likely agree with. This underscores the need for organizations and brands to develop social media strategies that prioritize authentic dialogue and value co-creation with publics.

The emergence of social media influencers has added another layer of complexity to trust and credibility in digital spaces. Freberg et al. (2011) conducted an early study exploring public perceptions of influencer personalities, providing a foundation for understanding the characteristics that make influencers effective in social media contexts. The authors defined social media influencers as “a new type of independent third party endorsers who shape audience attitudes through blogs, tweets, and the use of other social media” (2011, p. 90). Building on this, Arora et al. (2019) developed an index using multiple machine-learning techniques to measure social media influencers' impact across Facebook, X (Twitter), and Instagram, offering insights into the factors that contribute to influencer effectiveness and audience engagement. This is particularly important in an age where the influencer community is exercising increasingly significant power over brand perceptions (Childers et al., 2019).

A systematic review by Jagani et al. (2023) of 214 articles revealed that social media influencer marketing has seen exponential growth since 2018, coinciding with the rise of visually-focused platforms like Instagram. Their analysis identified four major themes in influencer research: parasocial interactions and relationships, sponsorship transparency, marketing authenticity, and audience engagement. The authors found that influencer credibility is built through multiple factors including perceived authenticity, expertise, and trustworthiness, with smaller influencers often demonstrating stronger trust-similarity correlations than those with larger followings. Importantly, their review highlighted that while content-related factors were the most researched antecedents of trust (36 studies), followed closely by influencer-related factors (34 studies), comparative factors examining influencer-follower relationships remained underexplored (6 studies), suggesting a significant gap in understanding these critical relational dynamics.

Empirical research by Hew et al. (2023) provides insight into how trust dynamics operate in influencer marketing contexts. The author's large-scale study of social media influencer marketing revealed that while influencer credibility strongly affects consumer attitudes and trust formation, the relationship between interactivity and trust is more complex than previously theorized. The authors found that increased interaction between influencers and followers can actually have a negative impact on trust when the quality of engagement is compromised. This is particularly evident with larger influencers who struggle to maintain meaningful, personalized interactions with their growing audience base. Their findings suggest that the mechanics of trust building in digital spaces depend not just on the frequency of interactions, but on their quality and authenticity. The authors emphasize that traditional metrics like follower count may be less reliable indicators of potential influence than measures of engagement quality and authentic communication.

The effectiveness of influencer marketing is closely tied to the perceived credibility and trustworthiness of the influencers themselves. Lou and Yuan (2019) examined how influencer credibility and message value affect consumer trust in branded content on social media. Their findings highlight the importance of informative content and specific aspects of influencer credibility in building trust with audiences and driving marketing outcomes. The study revealed that the informative value of influencer-generated content positively affects followers' trust in branded posts and their purchase intentions. Regarding influencer credibility, the authors found that trustworthiness, attractiveness, and perceived similarity to followers positively influenced trust in branded posts. Interestingly, influencer expertise did not significantly impact followers' trust in sponsored content. The study also demonstrated that followers' trust in influencer-generated branded posts had the strongest effect on purchase intentions compared to other content and source-related factors.

Similarly, Casaló et al. explored how influencers' perceived originality and uniqueness affect their credibility and followers' behavioural intentions on Instagram. The authors found that "instead of perceived quality or quantity, perceived originality and uniqueness of the posts on an Instagram account are the key factors that lead a poster to be perceived as an opinion leader" (2020, p. 516).

The effectiveness of influencer marketing can also vary depending on the type of influencer and the product being promoted. Cauberghe et al. (2017) investigated how the number of followers and product divergence affect influencer marketing effectiveness on Instagram. The study found that a high number of followers positively affects an influencer's likeability, mainly due to perceptions of popularity. However, this doesn't necessarily translate into true influence or opinion leadership. Importantly, they discovered that for products with divergent designs – items that have unique, unconventional, or distinctive aesthetic features that set them apart from more typical or standard products in their category – endorsement by influencers with extremely high follower counts can actually decrease perceptions of brand uniqueness and, consequently, brand attitudes. This suggests that while a large follower base can be beneficial, it's not always the optimal choice for every product type or marketing goal. The authors emphasize the importance of considering the product type, the influencer's content focus, and their audience's interests rather than solely relying on follower count when selecting influencers for marketing campaigns.

Jin et al. (2019) investigated the effectiveness of Instagram-based influencer marketing, comparing traditional celebrities with Instagram celebrities. Their study revealed that Instagram celebrities generate higher levels of social presence, which in turn leads to greater trustworthiness, more positive brand attitudes, and increased envy among followers. The authors found that this effect is particularly pronounced for luxury brands and is moderated by consumers' appearance-related self-discrepancy. This means that the influence of Instagram celebrities on trustworthiness, brand attitudes, and envy is stronger or weaker depending on how much consumers perceive a difference between their actual appearance and their ideal appearance. Individuals with higher appearance-related self-discrepancy (i.e., those who perceive a larger gap between their actual and ideal appearance) may be more susceptible to the influence of Instagram celebrities, especially when it comes to luxury brands. These findings highlight the unique power of social media influencers in creating relatable and engaging content that resonates with their audience, potentially making them more effective than traditional celebrities in certain digital marketing contexts.

Luoma-aho et al. (2019) investigated the effectiveness of YouTube vlog endorsements, revealing several key factors that influence their success. The authors' study resulted in several interesting, interrelated findings. First, higher audience participation (e.g., likes, comments, shares) leads to stronger parasocial relationships between viewers and vloggers. Second, these stronger parasocial relationships enhance the perceived credibility of the vlogger. Third, increased vlogger credibility positively influences viewers' attitudes towards endorsed brands. Fourth, the audience's initial attitude toward the vlogger plays a role; positive initial attitudes strengthen the relationship between participation and parasocial relationships. These results show the importance of audience engagement and relationship building in YouTube influencer marketing.

While Luoma-aho et al. (2019) provide insights into the mechanisms of effective vlog endorsements, it is also crucial to consider the broader trends in product promotion on YouTube. In this context, Schwemmer and Ziewiecki (2018) analyzed product promotion trends on popular German YouTube channels, revealing a significant increase in commercialization over time. Their study found that referral links in video descriptions increased by 310% from 2009 to 2017, while content related to product promotion, particularly in fashion and beauty categories, showed consistent growth. As the authors note, this raises genuine concerns about the difference between genuine, unpaid content and paid promotion.

Recent research has expanded the use of social media data for competitive analysis. Liu et al. (2019) propose a method for assessing a product's competitive advantages using user-generated content. Their approach addresses the limitation of analyzing products in isolation by incorporating competitor identification through comparative text analysis. The method employs machine learning techniques to efficiently process large volumes of social media data. Their case study in the automotive industry demonstrated the method's ability to reveal competitive insights that might be overlooked by traditional market analyses.

Product Reviews

In the era of e-commerce and social media, online product reviews have become a crucial source of information for consumers, shaping their attitudes and purchasing decisions. The trust that consumers place in these reviews is a critical factor in their influence on public opinion and behaviour. Research has explored various aspects of online product reviews, including the role of user-generated content, the impact of sentiment and valence, and the factors that influence perceived credibility and trust in reviewers.

One key aspect is the interplay between expert and user-generated ratings in shaping consumer trust. Flanagin and Metzger (2013) investigated how the volume and valence of user-generated ratings, as well as consumer characteristics, influence trust in online ratings. They found that the volume of user-generated ratings had a significant impact on trust, with higher volumes leading to greater trust. Additionally, the valence of ratings (positive or negative) interacted with consumer characteristics such as prior knowledge and experience to influence trust. This suggests that the aggregate sentiment expressed in user-generated reviews can be a powerful driver of consumer opinion and trust. While the volume and valence of reviews play a crucial role in shaping trust, the identity of the reviewer also has a significant impact on perceived credibility, which is another critical factor in the influence of product reviews.

Bronner et al. (2012) explored how source identification affects the perceived credibility of online product reviewers. Interestingly, they found an ironic effect where reviewers who disclosed their identities were perceived as less credible than anonymous reviewers. Further, the study noted that “the presence of an ironic effect depended on the way an expert source was demarcated from a layperson” (Bronner et al., 2012, p. 27). If an expert was self-identified, rather than identified by a third party, that had an impact on their perceived trustworthiness: those who were self-identified were perceived as less credible than those who were identified by a third party. This suggests that the relationship between source identification and trust in online contexts is complex and may be influenced by factors such as the perceived motives of the reviewer. Building on this, Chan-Olmsted et al. found that on YouTube, "source credibility, argument quality, and source attractiveness are significant predictors of influencer marketing credibility" (2018, p. 188).

Further, the credibility of influencers and product reviewers plays a crucial role in their effectiveness. Lou and Yuan found that "influencer-generated posts' informative value and some components of influencer credibility can positively affect followers' trust in influencer-generated branded posts, which in turn affects brand awareness and purchase intentions" (2019, p. 67). This underscores the importance of not just the content itself, but also the perceived credibility of the source in shaping consumer trust and, ultimately, purchase intentions. As the credibility of reviewers influences consumer trust, it is important to consider how different types of reviews, particularly negative ones, affect consumer attitudes.

The impact of negative reviews on consumer attitudes has been a particular focus of research. Han et al. (2008) examined how negative online reviews influence consumers' attitudes towards products, drawing on an information-processing perspective. They found that the proportion and quality of negative online reviews had a significant negative effect on consumer attitudes, and that this effect was moderated by the consumer's involvement and prior knowledge. This highlights the potent role that negative sentiment in online reviews can play in shaping consumer opinion and trust. Beyond the impact of negative reviews, the nature of the content itself - whether emotional or informational - can also shape consumer responses.

Cervellon and Galipienzo (2015) compared the effects of emotional versus informational content on consumer responses to Facebook posts about a luxury hotel. They found that informational content led to a higher perception of quality and a more positive attitude toward the hotel chain. However, neither emotional nor informational content significantly influenced booking intent or the intention to follow the hotel on Facebook. This suggests that while informational content may enhance brand perception, the type of content does not significantly impact consumer behaviour intentions.

In the context of social media influencers, Casaló et al. (2020) found that perceived originality and uniqueness of posts, rather than quality or quantity, are key factors in establishing opinion leadership on Instagram. Their study revealed that opinion leadership influences consumer behaviour by increasing intentions to interact with the account, recommend it to others, and follow fashion advice. This influence is particularly strong when content aligns with the consumer's personality and interests, highlighting the importance of authenticity and creativity in building trust and influence in the fashion industry. While these studies focus on general product reviews and social media posts, it is also crucial to examine how these principles apply to celebrity endorsements and influencer marketing.

While much of the research on online reviews has focused on consumer products, the principles of trust and influence extend to other domains such as celebrity endorsements on social media. Chung and Cho (2017) investigated how parasocial relationships with celebrities on social media influence trust and endorsement effectiveness. They found that parasocial relationships, which are one-sided relationships that consumers develop with media personalities, positively influenced trust in the celebrity and the perceived credibility of their endorsements. This suggests that the emotional connections and perceived intimacy fostered by social media interactions can be powerful drivers of trust and influence. However, being a celebrity may not be enough to be an effective product influencer. Djafarova and Rushworth note that " non-traditional celebrities such as bloggers, YouTube personalities and 'Instafamous' profiles are more powerful, as participants regard them as more credible and are able to relate to these, rather than more traditional, celebrities" (2017, p. 1). This shift in consumer trust has significant implications for how brands approach their marketing strategies.

The impact of influencers on brand attitudes is another crucial aspect of product review effectiveness. Cauberghe et al. (2017) explored the challenges in influencer marketing, particularly in identifying the right influencers. They found that an influencer's number of followers affects consumers' attitudes towards them, mainly through perceptions of popularity. However, they caution that a high follower count may not always translate into true influence. Their study contributes to the ongoing debate about opinion leadership and its identification, suggesting that while follower numbers are often used as a metric for potential reach, they may not be a definitive indicator of an influencer's impact on consumer behaviour and brand attitudes.

Moving beyond influencers and celebrities, it's important to consider how trust operates in more specialized contexts, such as product review blogs. Ghazisaeedi et al. investigated the source trustworthiness of product review blogs among online consumers. They found that "younger consumers exhibit higher levels of source trustworthiness" towards product review blogs compared to older consumers (2012, p. 7498). Additionally, they discovered that "there is a positive and significant relationship between source trustworthiness and both the frequency with which respondents access blogs and the number of blogs accessed" (Ghazisaeedi et al., 2012, p. 7506). This suggests that familiarity and engagement with blogs is associated with higher levels of trust.

The role of reviewer credibility in building trust has also been examined in depth. Au Yeung and Iwata analyzed large datasets from Epinions and Ciao to investigate the relationship between trust networks and product ratings. They found that "in general, users who trust each other tend to have smaller differences in their ratings as time passes, giving support to the theories of homophily and social influence" (2011, p. 495). However, they also discovered that "a trust relation does not guarantee that two users have similar preferences, implying that personalized recommendations based on trust relations do not necessarily produce more accurate predictions" (2011, p. 495). This highlights the complex relationship between trust, similarity, and influence in online review communities. These varied studies collectively paint a complex picture of trust in online reviews and influencer marketing.

When discussing written product reviews, there is existing research that speaks to collection and analysis methods. For example, the challenge of processing large volumes of customer reviews was addressed by Hu and Liu (2004), who developed a feature-based summarization approach. Their method identifies product features mentioned in reviews, extracts relevant opinion sentences, and determines their sentiment orientation. This technique provides a more detailed and structured analysis of customer feedback, enabling both consumers and manufacturers to gain insights into specific product attributes and their reception in the market.

Recent work by Haque et al. (2018) has advanced sentiment analysis techniques for large-scale Amazon product reviews. Their study employed a combination of pool-based active learning for dataset labeling and multiple feature extraction methods, including bag-of-words and TF-IDF with Chi-square. The author’s approach, yielding accuracies more than 93% for all categories, demonstrates the potential of combining active learning, sophisticated feature extraction, and appropriate classifier selection for effective large-scale sentiment analysis of product reviews.

Recent advancements in analyzing product reviews have leveraged artificial intelligence and machine learning techniques. He et al. (2024) employed large language models to analyze sentiment in e-commerce customer feedback, revealing a higher-order functional relationship between star ratings and emotional sentiments in reviews. Their research demonstrated that customer emotions expressed in reviews can significantly influence customer-product relationships over time, with negative reviews showing particularly strong propagation patterns through social networks. Supporting these findings, Al-Gasawneh et al. (2023) found that AI-powered sentiment analysis tools can efficiently process vast quantities of customer feedback in real-time, enabling businesses to identify emerging trends and issues proactively. Their research emphasized that while automated sentiment analysis achieves high accuracy rates—with Wang et al. reporting 90% accuracy in sentiment prediction—human interpretation remains crucial for understanding broader context and maintaining analysis quality.

Expanding on the significance of online written reviews, Mudambi and Schuff (2010) conducted a study on what makes customer reviews helpful on Amazon.com. Their research examined 1,587 reviews across six products and found that review extremity, review depth, and product type significantly affect the perceived helpfulness of reviews. Importantly, they discovered that product type (search vs. experience goods) moderates the effect of review extremity and depth on helpfulness. The authors define search goods as products for which consumers can easily obtain information on quality prior to purchase, with key attributes being objective and easily comparable (e.g., digital cameras, cell phones). In contrast, experience goods are products for which it is relatively difficult to obtain quality information before use, with key attributes being subjective or difficult to compare, often requiring the use of one's senses to evaluate (e.g., music CDs, MP3 players). For experience goods, reviews with extreme ratings were less helpful than those with moderate ratings. Additionally, review depth had a greater positive effect on helpfulness for search goods compared to experience goods.

The behaviour of consumers in online review platforms is not static but evolves over time as internet usage becomes more widespread and social media platforms mature. Chen et al. (2011) provide empirical evidence of these temporal dynamics in the context of automobile reviews. Their study, comparing data from 2001 and 2008, reveals significant changes in the relationships between marketing variables and review behaviour. For instance, the impact of price on review volume shifted from a negative relationship to a U-shaped one, while the positive effect of luxury branding on review volume decreased over time. These changes reflect a broader shift in the motivations driving consumer review behaviour, from demonstrating expertise and status to expressing satisfaction or dissatisfaction. Such findings highlight the importance of considering the temporal aspect when studying online review behaviour. As social media platforms continue to evolve and user behaviour changes, the dynamics observed in online reviews are likely to keep shifting. This temporal perspective adds an important dimension to our understanding of social media and online review behaviour, suggesting that marketers and researchers need to regularly reassess the factors influencing consumer engagement in these platforms.

Taken together, this body of research underscores the central role of trust in the influence of online product reviews and user-generated content on public opinion and consumer behaviour. The sentiment, volume, and perceived credibility of reviews interact with consumer characteristics to shape attitudes and decisions. As such, understanding the factors that build or erode trust in online reviewers and influencers is critical for predicting and managing their impact on public sentiment.

Hypothesis & Research Questions

H1

Trust indicators (integrity, dependability, and competence) in the comments section of tech influencers' YouTube product reviews are predictive of consumer opinion as expressed in Amazon reviews of the same products.

H0

There is no significant relationship between the sentiment of trust indicators in the comments on tech influencers' product review videos and the sentiment of Amazon reviews for the same products.

H2

The level of trust in tech influencers, as determined by the content analysis of comments on their product reviews, is associated with the nature of media coverage mentioning and quoting the influencer, as well as any legal actions against them.

H0

There is no significant relationship between the level of trust in tech influencers and the nature of media coverage or legal actions against them.

RQ1: How and to what extent do the comments on tech influencers' product reviews reflect trust indicators (integrity, dependability, and competence), and what is the overall sentiment for each trust component?

This research question directly addresses the measurement of trust in tech influencers by analyzing comments on their product reviews. By identifying and categorizing trust indicators in the comments and conducting a content analysis, the researcher will attempt to determine the level of trust in the influencers. The data will be used to test both hypotheses by providing a trust metric that can be compared to measurements of consumer opinion (Amazon reviews) and media coverage, as well as an analysis of legal actions against the influencers.

RQ2: How closely do the contents of tech influencers' product reviews resemble the contents of Amazon reviews for the same products? Does the similarity between Amazon

reviews and tech influencers' product reviews vary based on the level of trust in the influencer?

This research question examines the relationship between tech influencers' product reviews and consumer opinion, as expressed in Amazon reviews. By comparing the content of influencers' reviews and Amazon reviews using a bag of words analysis and other content analysis metrics, the researcher will attempt to determine if the influencers' opinions are predictive of consumer sentiment. This directly tests Hypothesis 1, which posits that trust in influencers is predictive of consumer opinion. The inclusion of trust levels as a variable in this analysis will provide insights into how the relationship between influencer reviews and consumer sentiment is related to the level of trust in the influencer.

RQ3: How are tech influencers portrayed in media coverage that mentions and quotes them, and is there a relationship between the level of trust in the influencer (as determined by RQ1) and the nature of the media coverage they receive? Additionally, are there any instances of legal actions against the influencers?

This research question investigates the relationship between trust in tech influencers and the media coverage they receive, as well as any legal actions against them. By qualitatively examining how influencers are portrayed in media coverage and comparing this to the level of trust determined by the sentiment analysis of comments, the researcher can test Hypothesis 2. This hypothesis suggests that the level of trust in an influencer is associated with the nature of media coverage they receive and any legal actions against them. The inclusion of legal actions as a variable in this analysis will provide a more comprehensive understanding of the potential consequences of varying levels of trust in tech influencers, both in terms of media portrayal and legal repercussions.

Methodology

Data Collection & Cleaning

This section delineates the multi-stage data collection and cleaning process employed to gather and prepare diverse types of data related to tech influencers, their product reviews, consumer opinions, media coverage, and potential legal actions. The methodology encompasses data collection and cleaning from YouTube, Amazon US, news outlets, and legal databases, using various web scraping, API techniques, and database searches. A standardized cleaning approach was taken for YouTube transcripts, comments, and Amazon reviews. This was crucial in preparing the datasets for a comparative analysis, ensuring that any differences observed in the subsequent analysis would be due to genuine variations in content rather than inconsistencies in data format or quality.

Selection of Influencers & Products

To begin the data collection process, the researcher employed a purposive sampling approach (Stacks, 2017) to select a sample of 10 prominent tech influencers based on their YouTube follower count. The selection criteria ensured that the chosen influencers had a significant following, regularly produced English-language technology product review content, and had reviewed at least three products within the past year. An additional requirement was that the influencers had reviewed products for which equivalent items could be found on Amazon US, facilitating comparison between influencer reviews and consumer opinions. For each influencer, the researcher identified three recent product reviews, focusing on products released within the past year to ensure relevance and comparability.

The sample size of 10 influencers was determined based on several methodological considerations. First, this number provided sufficient depth for meaningful analysis while remaining manageable for the intensive multi-modal data collection and analysis required by the study's design. Each influencer generated multiple data points across three product reviews, yielding 30 distinct cases for analysis. The selection of 10 influencers also allowed for representation across different subscriber tiers (from 500,000 to over 24 million followers), enabling examination of potential relationships between audience size and trust indicators. Furthermore, this sample size facilitated the thorough collection and analysis of approximately 13,890 comments (averaging 463 comments per product per influencer), providing robust data for trust analysis while remaining within the scope of a master's-level research project.

Table 1

Overview of Influencers, Followers, and Products

Influencer	Followers	Product 1	Product 2	Product 3
Unbox Therapy	24,300,000	Nothing Phone 2A	Insta360 X4	Google Pixel 8A
Marques Brownlee	19,200,000	M3 MacBook Air	M4 iPad Pro	OnePlus Watch
Mrwhosetheboss	19,100,000	Nothing Phone 2A	M4 iPad Pro	Samsung S24 Ultra
iJustine	7,090,000	Insta360 X4	M4 iPad Pro	Phone 15 Pro
Austin Evans	5,550,000	HP Laptop	M3 MacBook Pro	M3 iMac
UrAvgConsumer	3,250,000	M3 Macbook Air	M4 iPad Pro	S24 Ultra
The Tech Chap	1,480,000	Insta360 GO 3S	M4 iPad Pro	Nothing Phone 2A
MrMobile	1,250,000	Google Pixel 8A	Nothing Phone 2A	Samsung S24 Ultra
Tech Spurt	1,170,000	Google Pixel 8A	Nothing Phone 2A	Samsung S24
Created by Ella	509,000	M4 iPad Pro	Samsung S24 Ultra	Samsung S24

YouTube Data Collection & Cleaning

Transcript Collection & Cleaning. Video transcripts were collected using the YouTube Transcript API, a Python library that facilitates the retrieval of automatically generated transcripts from YouTube videos. A custom script was developed to extract video IDs from given YouTube URLs, fetch the corresponding transcripts, and handle potential errors during the process. The script successfully retrieved transcripts for all 30 selected videos. The retrieved transcript data was subsequently saved to individual CSV files for each video, enabling further textual analysis (see Appendix A for the full script).

To clean the transcript data, a uniform cleaning process was applied using a custom Python script. The script loaded the CSV files containing raw transcript data, retained the 'text' column, and removed any rows with missing values. Metadata (platform, product, and source) was extracted from each input filename for consistent labelling. The cleaned transcript data was then saved to new CSV files in a designated output folder, with filenames structured to include the extracted metadata. This script can be seen in Appendix D.

Comment Collection & Cleaning. To gather user comments from the YouTube videos, the study employed the YouTube Data API v3 in conjunction with the Selenium WebDriver for Python. The developed script navigated to specified YouTube video pages, implemented scrolling functionality to load available comments, and extracted relevant information including the author, comment text, and like count.

The script was designed to collect comments from below each of the 30 selected videos. The number of comments collected per video varied depending on factors such as the video's popularity, age, and YouTube's comment display algorithm. However, on average, the script returned 463 comments per product per influencer. The script also converted abbreviated like counts (e.g., "1.5K") to numerical values. The collected comment data was stored in a structured format and saved to CSV files for subsequent analysis. (See Appendix B for the full script.)

The cleaning process for comments mirrored that of the transcripts. The cleaning script retained the 'text' column, removed rows with missing values, extracted metadata from filenames, and saved the cleaned data to new CSV files in the output folder. Again, this script can be seen in Appendix D.

Amazon Review Collection & Cleaning

To facilitate a comparison between influencer opinions and consumer sentiment, Amazon US reviews were collected for the same products featured in the YouTube videos. The focus was Amazon US because of the larger size of the market, which provided a larger comment base to draw from for the selected products.

A web scraping script was developed using the Selenium WebDriver for Python. This script navigated to specified Amazon product pages, handled potential CAPTCHA challenges or login requirements, extracted review data (including title, rating, review text, username, date, and helpful votes), implemented pagination to collect reviews from multiple pages, and stored the collected review data in CSV files, with separate files for each product.

The script was configured to collect reviews for each of the 30 products corresponding to the YouTube reviews. A total of 100 reviews were collected for each product to ensure there was an even number to compare against across influencers and products. This dataset enabled a comparison between influencer reviews and general consumer sentiment. The full script for this process is available in Appendix C.

The cleaning process for Amazon reviews was similar to that of YouTube data, with the addition of retaining the 'title' column where applicable. The cleaning script loaded the raw data, selected relevant columns ('text' and 'title'), removed rows with missing values in the 'text' column, extracted metadata from filenames, and saved the cleaned data to new CSV files in the output folder. Once again, this script can be seen in Appendix D.

News Article Collection & Cleaning

To gather media coverage data, a three-step process was implemented using a custom Python script. This process involved searching for articles, scraping their content, and cleaning the collected data.

First, the script used the Google Custom Search API to search for news articles related to each of the selected tech influencers. The search queries were constructed using the influencers' names and the term "YouTube" to ensure relevance. The script was configured to retrieve up to 100 articles per influencer, adhering to the limitations of the Google Search API key.

In the second step, the script attempted to scrape the content of each article identified in the search phase. This involved extracting the article title, author, publication name, publication date, and body content. Of the 599 articles initially identified, 536 were successfully scraped, representing a success rate of 89.5%.

The final step involved cleaning and filtering the scraped data. This process included removing HTML tags, decoding HTML entities, eliminating non-printable characters, and ensuring the content was in English. Additionally, any blank entries or articles that were incorrectly scraped were removed from the dataset. After this cleaning process, the final dataset contained 417 articles, representing 77.8% of the successfully scraped articles and 69.6% of the initially identified articles.

The number of cleaned articles varied considerably across influencers. Influencer iJustine had the highest number of usable articles at 99, while no usable articles were returned for Created by Ella.

Legal Case Collection

To investigate any potential legal actions involving the selected tech influencers, the researcher conducted a comprehensive search using Nexis Uni, a reputable legal and news database. Each of the 10 influencers was individually searched by name and user handle within the database's legal case records. The search parameters were set to cover all available years to ensure a thorough examination of any historical or current legal proceedings.

The Nexis Uni search yielded no legal cases directly involving any of the selected tech influencers. This absence of legal records suggests that, as of the time of the search, none of the influencers in the study had been involved in significant legal actions that were recorded in the Nexis Uni database.

Data Analysis

This section will describe the specific analytical techniques and statistical methods used to process and interpret the diverse dataset, which includes keyword and sentiment analysis of comments, comparative analysis of product reviews, and qualitative content analysis of media coverage and legal *actions*.

RQ1: How and to what extent do the comments on tech influencers' product reviews reflect trust indicators (integrity, dependability, and competence), and what is the overall sentiment for each trust component?

In addressing Research Question 1, the researcher employed an iterative approach to analyze the comments on tech influencers' product reviews. This analysis focused on how the comments reflected trust indicators, specifically examining the components of integrity, dependability, and competence, as well as the overall sentiment for each trust component. The researcher developed and refined a Python script through multiple iterations, ultimately producing seven versions, each addressing different aspects of the analysis.

The script provided (see Appendix E), which represented version 5 of 7, marked a significant milestone in the analytical process. This version primarily focused on the presence of keywords associated with trust components, without considering the broader sentiment context. The researcher used lists of trust-related keywords for each component (integrity, dependability, and competence), as well as their inverse counterparts. Additionally, the script incorporated influencer and review-related keywords to ensure that the analysis focused on comments specifically addressing the influencer or their review.

This version of the script featured a keyword-based analysis that counted the occurrence of positive and negative trust-related keywords for each trust component. The researcher calculated an Adjusted Trust Score using logarithmic scaling to account for the volume of trust-related keywords while balancing positive and negative indicators. Furthermore, Net Keyword Scores were computed for each trust component, providing a measure of the relative prevalence of positive versus negative trust indicators. To enhance the relevance of the analysis, the script filtered comments to include only those referring directly to the influencer or the review itself.

The iterative development process allowed the researcher to refine the approach through multiple script versions, each adjusting key variables to optimize the analysis. Across seven iterations, the researcher systematically modified three primary variables: the weighting of keywords, the weighting of the average sentiment, and the weighting of the sentiment ratio in the calculation of the trust score. This process of continuous adjustment aimed to achieve the highest possible correlation between the trust scores derived from the comment analysis and the results obtained from Research Question 2, which examined the similarity between influencer reviews and Amazon consumer reviews.

To validate the findings from the Python script, a random sample of 10% of the comments across influencers and products were selected for manual review. Content was then reviewed across six columns, and each presence of a keyword/inverse keyword was noted. These columns were then summed to provide a final value. This validation count was compared to a modified version of the analysis script that was adjusted to output granular data for validation purposes. The modifications included: (1) adding explicit count tracking for each trust component category; (2) implementing detailed logging of keyword matches with their surrounding context, and; (3) generating an itemized report showing the frequency and location of each keyword occurrence. Instead of producing the final aggregated trust scores, this validation version of the script produced line-by-line keyword identification and category-specific counts that could be directly compared with manual coding results. The high level of agreement between the manual coding and the modified script's output (89% accuracy) provided strong evidence for the reliability of the primary analysis script.

By fine-tuning these variables, the researcher sought to balance the influence of keyword presence with sentiment analysis, ensuring that the final analysis accurately captured the complexities of measuring trust in online comments while aligning with real-world indicators of influencer trustworthiness.

RQ2: How closely do the contents of tech influencers' product reviews resemble the contents of Amazon reviews for the same products? Does the similarity between Amazon reviews and tech influencers' product reviews vary based on the level of trust in the influencer?

To address Research Question 2, the researcher employed a two-part approach: calculating similarity scores between influencer reviews and Amazon consumer reviews, and then comparing these scores to the trust scores obtained from Research Question 1.

The process of calculating similarity scores involved the development and refinement of a Python script. The researcher created two iterations of this script, with the first version using both stemmed and non-stemmed analysis, while the second version incorporated lemmatization instead. This iterative approach allowed for a more nuanced comparison of textual similarities between the influencer reviews and consumer reviews.

The initial script used the sklearn library to implement TF-IDF (Term Frequency-Inverse Document Frequency) vectorization and cosine similarity calculations. The researcher preprocessed the text data from both YouTube transcripts and Amazon reviews, removing punctuation, converting to lowercase, and eliminating stop words. The script then calculated similarity scores for each product, producing both stemmed and non-stemmed versions of the analysis. See Appendix F for the full script.

In the second version, the researcher altered the script by incorporating lemmatization, a more sophisticated form of text normalization compared to stemming. This version used the NLTK (Natural Language Toolkit) library for text preprocessing, including tokenization, stop word removal, and lemmatization. See Appendix G for the full script.

Following the calculation of similarity scores, the researcher transitioned to the second part of the analysis. This involved importing the similarity scores into an Excel spreadsheet to facilitate a correlation analysis between these scores and the trust scores derived from Research Question 1. The researcher used Excel's correlation function to systematically explore the relationship between various metrics of similarity and trust.

This correlation analysis served as a critical step in addressing the research question, as it allowed the researcher to identify which aspects of similarity between influencer reviews and consumer reviews were most strongly associated with the perceived trustworthiness of the influencers. By iteratively adjusting the variables and metrics used in both the similarity score calculations and the trust score derivations, the researcher aimed to uncover the most significant relationships between these two dimensions of influencer effectiveness.

The iterative nature of this process, involving multiple script versions and careful correlation analysis, demonstrates the researcher's commitment to thoroughness and precision in addressing Research Question 2. This approach allowed for a comprehensive exploration of the relationship between the content similarity of reviews and the perceived trustworthiness of tech influencers, contributing valuable insights to the overall study objectives.

RQ3: How are tech influencers portrayed in media coverage that mentions and quotes them, and is there a relationship between the level of trust in the influencer (as determined by RQ1) and the nature of the media coverage they receive? Additionally, are there any instances of legal actions against the influencers?

In addressing Research Question 3, the researcher developed and implemented a Python script to analyze media coverage of tech influencers (see Appendices H and I). The first script, Appendix H, acted as a cleaning and formatting script to ensure all media articles were formatted consistently. This script detects encoding and reads CSV files containing news articles about different influencers, then extracts and cleans specified columns (e.g., text and title/headline) and saves the cleaned data in a new CSV file for each influencer in a designated output folder.

The second script, Appendix I, employed a multi-faceted approach to examine the portrayal of influencers in media articles and investigate potential relationships between media representation and the trust levels determined in Research Question 1.

The script used the Hugging Face transformers library, specifically the 'nlptown/bert-base-multilingual-uncased-sentiment' model, to conduct sentiment analysis on the media articles. This approach allowed for a nuanced evaluation of sentiment, providing a foundation for understanding the overall tone of media coverage for each influencer.

Central to the analysis were several key components defined by the researcher. These included a dictionary that mapped influencer handles to their full names, ensuring accurate identification across various naming conventions in articles. The researcher also incorporated trust component keywords – categorized into integrity, dependability, and competence – building upon the trust framework established in earlier parts of the study. Additionally, the script included thematic categories with associated keywords, covering areas such as product reviews, industry insights, controversies, sponsorships, personal life, consumer behaviour impact, platform-specific content, and legal/ethical issues.

The script processed each media article through a series of analytical steps. It began with sentiment analysis using the Hugging Face model, determining overall sentiment on a scale from -1 (negative) to 1 (positive). The script then conducted a trust component analysis, identifying the presence of trust-related keywords and scoring each component on a scale of 0-3. Thematic categorization followed, assessing the prevalence of different themes in the article, also on a 0-3 scale. The script also performed a quotation analysis, detecting the presence of direct quotes from the influencer, and evaluated article prominence based on factors such as influencer title mentions, early mentions, and overall mention frequency. To validate the findings from the automated media analysis process, a random sample of 10% of the articles were selected reviewed for validity. Content was reviewed across 15 columns, and each presence of a theme/keyword was noted. These manual codings were then compared to a similar output from the Python script. The validation process yielded an accuracy rate of 94%, indicating high reliability in the automated thematic and keyword analysis. This robust validation outcome supports the trustworthiness of the automated analysis approach, while simultaneously acknowledging the importance of human verification in computational methods.

After processing all articles for each influencer, the script calculated average scores across these dimensions, producing metrics for each article and aggregated scores for each influencer. These results were then imported into Excel for further statistical analysis. The researcher conducted correlation analyses between each variable derived from the media analysis and the trust scores obtained from Research Question 1. This step was crucial in testing for relationships between media portrayal and perceived trustworthiness of the influencers.

The correlation analysis in Excel enabled the researcher to identify which themes and aspects (i.e., trust keywords, presence of influencer quotations) of media coverage were most strongly associated with the trust levels of the influencers. This approach provided insights into potential links between how influencers are portrayed in the media and how trustworthy they are perceived to be based on their product reviews and audience engagement.

By combining sophisticated natural language processing techniques with statistical analysis, the researcher provided a comprehensive answer to Research Question 3. This method allowed for a nuanced understanding of the complex relationship between media portrayal and influencer trustworthiness, contributing valuable insights to the overall study of tech influencer credibility and impact.

Results

RQ1: How and to what extent do the comments on tech influencers' product reviews reflect trust indicators (integrity, dependability, and competence), and what is the overall sentiment for each trust component?

To address this research question, trust scores were calculated for each influencer and their individual product reviews. These scores were derived from the analysis of comments on the influencers' product review videos, focusing on indicators of trust such as integrity, dependability, and competence.

Table 2 presents the overall trust scores for each influencer across different versions of the analysis (V1 to V7). Table 3 shows the trust scores for individual product reviews by each influencer.

Table 2

Influencer Trust Scores

Influencer	Trust Score Version						
	V1	V2	V3	V4	V5	V6	V7
Unbox Therapy	0.429704251	0.08603333	0.63090121	0.75925926	0.63090121	0.226467574	0.226467574
Marques Brownlee	0.374043217	0.2044	1.66086261	0.83357499	1.56876077	0.296625163	0.296625163
Mrwhosetheboss	0.392352536	0.2021	1.67322103	0.8419316	1.61851211	1.618512109	0.278378757
iJustine	0.504884031	0.24966667	1.1130892	0.89090909	1.1130892	1.113089195	0.264350914
Austin Evans	0.348130274	0.11586667	1.2003931	0.82661583	1.05347639	1.053476387	0.302910053
UrAvgConsumer	0.389178048	0.1178	1.04019131	1	1.04019131	1.04019131	0.297550777
The Tech Chap	0.483341066	0.05646667	0.93398041	0.76767677	0.87301587	0.873015869	0.282413031
MrMobile	0.543563039	0.32446667	1.45945944	0.96747967	1.3800849	1.380084903	0.271323914
Tech Spurt	0.382623947	0.20206667	0.99584024	0.93333333	0.95355726	0.953557259	0.246964177
Created by Ella	0.56136274	0.3247	1.35493732	0.86137956	1.32881469	1.32881469	0.375667557

Table 3

Product-Specific Trust Scores

Influencer	Product	Trust Scores						
		V1	V2	V3	V4	V5	V6	V7
Unbox Therapy	Nothing Phone 2A	0.352124467	0.0695	0.698970004	1	0.698970004	0.108695652	0.982395459
Unbox Therapy	Insta360 X4	0.377185437	0.04	0.742188618	0.777777778	0.742188618	0.388888889	0.8
Unbox Therapy	Google Pixel 8A	0.559802848	0.1486	0.451544993	0.5	0.451544993	0.181818182	0.473360311
Marques Brownlee	M3 MacBook Air	0.35469701	0.1873	1.826065498	0.881355932	1.654478429	0.362962963	1.778933744
Marques Brownlee	M4 iPad Pro	0.417675218	0.1588	1.982496168	0.917241379	1.982496168	0.344559585	2.058352391
Marques Brownlee	OnePlus Watch	0.349757424	0.2671	1.174026156	0.70212766	1.069307724	0.182352941	1.341029806
Mrwhosetheboss	Nothing Phone 2A	0.377151305	0.2141	1.456609016	0.80952381	1.411078105	0.182481752	1.670230379
Mrwhosetheboss	M4 iPad Pro	0.34661645	0.0848	1.590940368	0.818181818	1.519966396	0.223642173	1.740873007
Mrwhosetheboss	Samsung S24 Ultra	0.453289852	0.3074	1.972113701	0.898089172	1.924491826	0.429012346	2.026119745
iJustine	Insta360 X4	0.384313432	0.3133	0.852048561	0.818181818	0.852048561	0.346153846	0.940931528
iJustine	M4 iPad Pro	0.495418586	0.3436	1.487219026	0.854545455	1.487219026	0.219626168	1.52247199
iJustine	iPhone 15 Pro	0.634920074	0.0921	1	1	1	0.227272727	1.204119983
Austin Evans	HP Laptop	0.365236717	0.0435	1.372752495	0.803921569	1.12383516	0.302521008	1.385862647
Austin Evans	MacBook Pro M3	0.364289729	0.0757	1.325336819	0.925925926	1.24042117	0.444444444	1.359176003
Austin Evans	M3 iMac	0.314864377	0.2284	0.903089987	0.75	0.796172831	0.161764706	1.183099323
UrAvgConsumer	M3 Macbook Air	0.311936226	0.2859	1.079181246	1	1.079181246	0.387096774	0.230448921
UrAvgConsumer	M4 iPad Pro	0.447793713	0.121	0.698970004	1	0.698970004	0.2	0.954242509
UrAvgConsumer	S24 Ultra	0.407804206	-0.0535	1.342422681	1	1.342422681	0.305555556	1.414973348
The Tech Chap	Insta360 GO 3S	0.460152478	-0.0619	0.25938375	0.333333333	0.25938375	0.142857143	0.318080836
The Tech Chap	M4 iPad Pro	0.52149943	0.2386	1.764406241	0.96969697	1.581512606	0.431654676	1.721239191
The Tech Chap	Nothing Phone 2A	0.468371288	-0.0073	0.77815125	1	0.77815125	0.272727273	0.8
Mr. Mobile	Google Pixel 8A	0.447381882	0.4259	1.447158031	1	1.361542964	0.275510204	1.469841251
Mr. Mobile	Nothing Phone 2A	0.73332104	0.1955	1.204119983	1	1.204119983	0.205128205	1.462397998
Mr. Mobile	Samsung S24 Ultra	0.449986194	0.352	1.727100306	0.902439024	1.574591763	0.333333333	1.697998018
Tech Spurt	Google Pixel 8A	0.377193398	0.0966	0.8	0.8	0.8	0.222222222	1.019279091

Influencer	Product	Trust Scores						
		V1	V2	V3	V4	V5	V6	V7
Tech Spurt	Nothing Phone 2A	0.403269809	0.1147	1.041392685	1	1.041392685	0.305555556	1.113943352
Tech Spurt	Samsung S24	0.367408633	0.3949	1.146128036	1	1.019279091	0.213114754	1.306129244
Created by Ella	M4 iPad Pro 13	0.690082247	0.3781	1.799110226	0.971830986	1.799110226	0.644859813	1.915985538
Created by Ella	Samsung S24 Ultra	0.623794276	0.2854	1.286104808	0.92	1.286104808	0.261363636	1.361750993
Created by Ella	Samsung S24	0.370211696	0.3106	0.979596933	0.692307692	0.901229037	0.220779221	1.080904566

These tables present the quantitative results of the trust analysis for each influencer and their product reviews. The scores across different versions (V1 to V7) represent iterations of the researcher's analytical approach, with each version refining the method of calculating trust based on comment analysis and optimizing for correlation with similarity scores between influencer and consumer reviews.

V1 established the baseline approach using a weighted combination of sentiment analysis (54.27%), sentiment ratio (10.68%), and trust indicator keywords (35.04%). These weightings emerged from initial correlation testing between trust indicators and consumer opinions. A trust score of 0.5 in V1 indicates an equal balance of positive and negative trust signals, while scores above 0.5 suggest predominantly positive trust indicators. For example, Unbox Therapy's V1 score of 0.429704251 indicates slightly more negative than positive trust signals in the comments.

V2 introduced optimization techniques using the Pearson correlation coefficient to refine these weights, resulting in generally lower but more precise scores. The lower scores in V2 (such as Unbox Therapy's 0.08603333) reflect a more stringent measurement approach that emphasized strong correlations with consumer opinions.

V3 modified the trust calculation to include specific trust components (integrity, dependability, and competence) with equal weighting, producing scores that could exceed 1.0 when multiple trust components were strongly present. Marques Brownlee's V3 score of 1.66086261 indicates substantial presence of multiple trust components in audience comments.

V4 introduced logarithmic scaling to account for varying comment volumes while maintaining the component-based approach. Scores in V4 are normalized between 0 and 1, with scores closer to 1 (such as UrAvgConsumer's 1.0) indicating consistently high trust signals across a large volume of comments.

V5 marked a significant methodological shift by removing sentiment analysis and sentiment ratio calculations entirely, focusing purely on the net difference between positive and negative trust keywords without volume adjustments.

V6 reintroduced logarithmic scaling but with an expanded set of inverse trust keywords to better capture negative sentiment. The scores in this version (such as Created by Ella's 1.32881469) represent a more nuanced measurement of trust that accounts for both positive and negative indicators while controlling for comment volume.

V7 normalized the trust scores by the number of relevant comments rather than total comments, addressing potential bias from irrelevant or spam comments. This produced more conservative scores (generally between 0.2 and 0.4) that better reflected the proportion of trust-relevant engagement in the comments.

In practical terms, these trust scores represent varying levels of audience trust signals. Scores below 0.3 indicate limited trust signals or mixed sentiment in audience engagement. Those falling between 0.3 and 0.6 suggest moderate trust levels with balanced positive and negative indicators. When examining versions 1-6, scores above 0.6 indicate strong trust signals from the audience, while in V7, scores above 0.3 represent similarly strong trust indicators. The higher scores observed in later versions (particularly V5-V7) correlate more strongly with similarity to consumer opinions. The variation in scores across versions reflects the iterative process of refining the analytical approach to best capture the nuances of trust expression in online comments.

RQ2: How closely do the contents of tech influencers' product reviews resemble the contents of Amazon reviews for the same products? Does the similarity between Amazon reviews and tech influencers' product reviews vary based on the level of trust in the influencer?

To address this research question, similarity scores were calculated between the content of tech influencers' product reviews and corresponding Amazon consumer reviews. These similarity scores measure how closely the language used in an influencer's review match the language found in consumer reviews on Amazon - essentially, how often influencers and consumers discuss the same features, use similar descriptions, or reach comparable conclusions about a product. Three distinct text processing methods were employed: stemming, unstemming (leaving as-is), and lemmatization. The results are presented in Table 4 for overall influencer scores and Table 5 for product-specific scores.

The analysis revealed that lemmatized similarity scores (V2 Lemmatized) were consistently higher than both stemmed and unstemmed scores across all influencers and products. The average lemmatized similarity score was 0.35, compared to 0.29 for stemmed and 0.24 for unstemmed scores.

Table 4

Overall Influencer Similarity Scores

Influencer	Similarity Scores		
	V1 Stemmed	V1 Unstemmed	V2 Lemmatized
Unbox Therapy	0.283421667	0.230248	0.2982
Marques Brownlee	0.285925	0.249202667	0.36553333
Mrwhosetheboss	0.197942	0.157362333	0.42276667
iJustine	0.201699	0.155757667	0.28776667
Austin Evans	0.220010333	0.186722333	0.3005
UrAvgConsumer	0.339353667	0.290031667	0.32823333

The Tech Chap	0.360512333	0.299056667	0.4043
Mr. Mobile	0.281718	0.226263	0.37646667
Tech Spurt	0.33003	0.292072667	0.33236667
Created by Ella	0.382799333	0.332295667	0.39053333

Among individual influencers, Created by Ella received the highest average similarity score (0.37) across all versions. Conversely, iJustine recorded the lowest average similarity score of 0.22 across all similarity score versions. Created by Ella maintained the highest average similarity score across all three versions, indicating a consistent alignment between this influencer's content and consumer reviews across different text processing methods.

Table 5

Product-Specific Similarity Score

Influencer	Product	Similarity Scores		
		V1 Stemmed	V1 Unstemmed	V2 Lemmatized
Unbox Therapy	Nothing Phone 2A	0.242382	0.197971	0.2565
Unbox Therapy	Insta360 X4	0.351033	0.27101	0.3189
Unbox Therapy	Google Pixel 8A	0.25685	0.221763	0.3192
Marques Brownlee	M3 MacBook Air	0.319653	0.289591	0.418
Marques Brownlee	M4 iPad Pro	0.304331	0.254982	0.3349
Marques Brownlee	OnePlus Watch	0.233791	0.203035	0.3437
Mrwhosetheboss	Nothing Phone 2A	0.169613	0.131348	0.4662
Mrwhosetheboss	M4 iPad Pro	0.226076	0.189752	0.3546
Mrwhosetheboss	Samsung S24 Ultra	0.198137	0.150987	0.4475
iJustine	Insta360 X4	0.227493	0.154038	0.3228
iJustine	M4 iPad Pro	0.242946	0.192159	0.3414
iJustine	Phone 15 Pro	0.134658	0.121076	0.1991
Austin Evans	HP Laptop	0.194453	0.155575	0.3148
Austin Evans	M3 MacBook Pro	0.250071	0.223359	0.2888
Austin Evans	M3 iMac	0.215507	0.181233	0.2979

Influencer	Product	Similarity Scores		
		V1 Stemmed	V1 Unstemmed	V2 Lemmatized
UrAvgConsumer	M3 Macbook Air	0.372994	0.318639	0.356
UrAvgConsumer	M4 iPad Pro	0.320608	0.278493	0.2908
UrAvgConsumer	S24 Ultra	0.324459	0.272963	0.3379
The Tech Chap	Insta360 GO 3S	0.370377	0.278487	0.4484
The Tech Chap	M4 iPad Pro	0.458116	0.410164	0.4401
The Tech Chap	Nothing Phone 2A	0.253044	0.208519	0.3244
Mr. Mobile	Google Pixel 8A	0.303622	0.244743	0.3725
Mr. Mobile	Nothing Phone 2A	0.260536	0.202585	0.4025
Mr. Mobile	Samsung S24 Ultra	0.280996	0.231461	0.3544
Tech Spurt	Google Pixel 8A	0.369045	0.326452	0.3309
Tech Spurt	Nothing Phone 2A	0.266943	0.212559	0.3862
Tech Spurt	Samsung S24	0.354102	0.337207	0.28
Created by Ella	M4 iPad Pro	0.446308	0.405976	0.4378
Created by Ella	Samsung S24 Ultra	0.313789	0.283648	0.4088
Created by Ella	Samsung S24	0.388301	0.307263	0.325

To examine the relationship between trust scores derived from RQ1 and similarity scores, correlation analyses were conducted. A heat map visualization (Tables 6 & 7) revealed several notable correlations between different versions of trust and similarity scores.

Table 6

Trust Score / Similarity Score Correlation Analysis by Product

	V1 Stemmed	V1 Unstemmed	V2 Lemmatized
Version 1	0.092691208	0.118411788	0.21770409
Version 2	0.154523281	0.188015118	0.21759569
Version 3	0.020613828	0.090697262	0.41045212
Version 4	0.002153031	0.093176513	-0.1625712

Version 5	0.023294	0.086846158	0.42201843
Version 6	0.393207595	0.411526787	0.39320759
Version 7	-0.107070888	-0.030093626	0.32076118

The Table 6 heat map illustrates that the Version 5 trust score showed a correlation of 0.42 with the V2 Lemmatized similarity score, representing the strongest relationship observed. Additionally, the Version 7 trust score correlated at 0.41 with the V1 unstemmed similarity score, while the Version 3 trust score demonstrated a 0.41 correlation with the V2 lemmatized similarity score.

Table 7

Correlation Analysis Between Trust Score Versions and Similarity Score Methods Across Aggregated Influencer Data

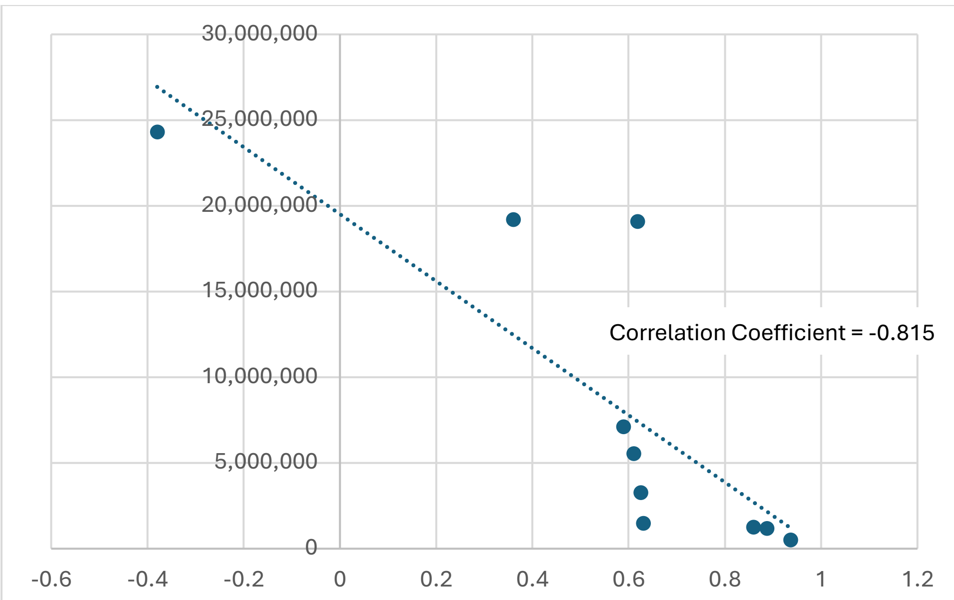
	V1 Stemmed	V1 Unstemmed	V2 Lemmatized
V1	0.311686521	0.227179356	0.263554939
V2	-0.030652683	-0.022652596	0.236571012
V3	-0.287147123	-0.247293082	0.573467451
V4	0.100341677	0.118568267	-0.095333325
V5	-0.266153182	-0.231577182	0.581818289
V6	-0.169468406	-0.18288392	0.424841098
V7	0.378490557	0.408718316	0.389586481

When all influencer product trust scores and similarity scores were averaged to create a genuine "influencer trust score," (as in Table 7) rather than individual product scores, two strong, positive correlations were observed: first, between V3 trust score and V2 Lemmatized similarity score, and second, between V5 trust score and V2 lemmatized similarity score. The former returned a correlation coefficient of 0.57, the latter of 0.58.

An additional trend in the data emerged when examining correlations between trust and similarity scores for individual influencers: the strength of correlation increased as the influencer's follower count decreased. This relationship is illustrated in the accompanying scatter plot below (Figure 3) and discussed in the Discussion section under RQ3.

Figure 3

Influencer Trust/Similarity Score Correlations vs Follower Size



RQ3: How are tech influencers portrayed in media coverage that mentions and quotes them, and is there a relationship between the level of trust in the influencer (as determined by RQ1) and the nature of the media coverage they receive? Additionally, are there any instances of legal actions against the influencers?

The analysis of media coverage for the selected tech influencers revealed several noteworthy patterns and relationships. A total of 417 news articles were analyzed across nine of the ten influencers, with one influencer (Created by Ella) having no articles found for analysis. The distribution of articles varied considerably among influencers, ranging from 99 articles for iJustine to 6 articles for The Tech Chap. Table 8 presents the criteria used to analyze the articles.

Table 8

Media Analysis Criteria Definitions

Analysis Criterion	Description
Overall Sentiment	Measures the emotional tone of articles using VADER sentiment analysis. Scores are categorized as: Positive (>0.05), Negative (<-0.05), or Neutral (between -0.05 and 0.05).
Trust - Integrity	Quantifies mentions of keywords related to moral character, honesty, and ethical behaviour (e.g., 'honest', 'transparent', 'ethical'). Scored 0-3 based on keyword frequency.
Trust - Dependability	Measures references to reliability and consistency (e.g., 'consistent', 'reliable', 'trustworthy'). Scored 0-3 based on keyword frequency.
Trust - Competence	Assesses mentions of expertise and professional capability (e.g., 'expert', 'knowledgeable', 'proficient'). Scored 0-3 based on keyword frequency.
Quotation Presence	Binary measure (0/1) indicating whether the article contains direct quotes from the influencer, verified by analyzing text preceding quotation marks.
Article Prominence	Three-tier scoring system (0-2): Primary focus (2): Influencer mentioned in title, first 100 words, and >3 times total; Secondary focus (1): Mentioned in title/early and >1 time; Mentioned in passing (0): Other cases.
Product Reviews	Binary measure (0/1) indicating presence of product review-related content based on keywords (e.g., 'review', 'unboxing', 'hands-on').
Industry Insights	Binary measure (0/1) for presence of broader technology industry analysis (e.g., 'market trends', 'innovation', 'future tech').
Controversies	Binary measure (0/1) tracking mentions of problematic events or issues (e.g., 'controversy', 'scandal', 'criticism').
Sponsorships/Partnerships	Binary measure (0/1) identifying discussion of commercial relationships (e.g., 'sponsored', 'partnership', 'brand deal').
Personal Life	Binary measure (0/1) indicating coverage of non-professional aspects (e.g., 'lifestyle', 'family', 'background').
Impact on Consumer Behaviour	Binary measure (0/1) tracking discussion of influence on purchasing decisions (e.g., 'influence', 'buying decisions', 'consumer trends').
Platform Specific	Binary measure (0/1) for content about social media platforms and strategies (e.g., 'YouTube', 'algorithm', 'content strategy').
Legal/Ethical Issues	Binary measure (0/1) identifying coverage of regulatory or moral concerns (e.g., 'lawsuit', 'ethics', 'FTC guidelines').

The sentiment analysis of the news articles showed that two influencers, UrAvgConsumer and The Tech Chap, received exclusively positive coverage, with an overall sentiment score of 1. This indicates that the media portrayal for these influencers was uniformly favourable within the analyzed articles. Full results are available in Table 9.

The media article analysis revealed several correlations between V5 trust scores (derived from RQ1) and various themes present in the news articles, as are evident from Table 10 below. Trust scores showed a moderate positive correlation (0.44) with the prevalence of "controversy" keywords in news articles, suggesting that as trust scores increased, so did the frequency of controversy-related terms in media coverage. Trust scores also demonstrated a positive correlation (0.24) with the presence of "integrity" keywords. Notably, a strong positive correlation (0.82) was observed between the presence of "controversy" keywords and "integrity" keywords in the news articles.

Weaker positive correlations were found between trust scores and themes related to personal life (0.19), impact on consumers (0.20), and the total number of articles found for each influencer (0.21).

Further analysis of the relationships between different themes in the news articles revealed additional correlations. The presence of integrity-related keywords correlated positively with the presence of quotations (0.56) and article prominence (0.33). Dependability and competence themes showed a moderate positive correlation (0.55), while competence-related keywords correlated strongly with industry insights (0.67).

A strong positive correlation (0.83) was observed between integrity and controversy themes. Conversely, a weak negative correlation (-0.02) was found between trust scores and the presence of ethical and legal issues in the news articles.

The "impact on consumers" theme, which included keywords related to influence on buying decisions, consumer trends, and market impact, showed a slight positive correlation (0.20) with trust scores.

Finally, no legal actions against the influencers were identified during the research process.

Table 9

News Media Analysis Results by Influencer

				Trust Components								Themes			
Influencer	Total Articles	Overall Sentiment	Integrity	Dependability	Competence	Quotation Presence	Article Prominence	Industry Reviews	Industry Insights	Controversies	Sponsorships/Partnerships	Personal Life	Impact on Consumer Bahviour	Platform Specific	Legal/Ethical Issues
Unbox	63	0.746031746	0.93650794	0.61904762	0.88888889	0.49206349	0.50793651	0.80952381	0.06349206	0.26984127	0.47619048	0.11111111	0.03174603	0.79365079	0.07936508
MB	80	0.875	1.1625	0.6375	1.0375	0.625	0.65	0.6625	0.0125	0.375	0.4	0.2	0.0375	0.8625	0.0875
Boss	40	0.8	1.325	0.625	1.275	0.425	0.125	0.9	0.125	0.6	0.375	0.2	0.05	0.825	0.125
iJustine	99	0.919191919	0.87878788	0.60606061	1.24242424	0.24242424	0.16161616	0.75757576	0.01010101	0.16161616	0.35353535	0.25252525	0.12121212	0.83838384	0.11111111
AE	39	0.846153846	0.97435897	0.61538462	1.1025641	0.66666667	0.15384615	0.69230769	0.05128205	0.51282051	0.51282051	0.23076923	0.15384615	0.79487179	0.28205128
AvgCons	10	1	0.6	0.9	2.1	0.3	0	1	0.1	0.1	0	0.5	0	0.7	0
TechChap	6	1	0.83333333	0.66666667	0.66666667	0.33333333	0.33333333	0.83333333	0	0.16666667	0.66666667	0	0	1	0
MrMobile	70	0.828571429	1.01428571	1.02857143	0.88571429	0.42857143	0.14285714	0.71428571	0.04285714	0.52857143	0.6	0.27142857	0.08571429	0.55714286	0.07142857
TechSpurt	10	0.8	0.7	0.3	0.3	0.3	0.2	0.9	0	0.2	0.5	0	0	0.8	0.3

Table 10

Correlation Heat Map of Variables Tested for RQ3

	Overall Sentiment	Integrity	Dependability	Competence	Quotation Presence	Article Prominence	Industry Reviews	Industry Insights	Controversies	Sponsorships/Partnerships	Personal Life	Impact on Consumer Behaviour	Platform Specific	Legal/ Ethical Issues
Overall Sentiment	1	-0.46573082	0.3300711	0.44901027	-0.3935011	-0.250304844	0.232900083	-0.167921073	-0.559931744	-0.327851156	0.33811648	-0.1929731	0.28488513	-0.5419026
Integrity	-0.46573082	1	0.04195973	-0.0841057	0.55650119	0.333551372	-0.478084581	0.283916955	0.826797919	0.283865022	-0.1084088	0.33895683	0.07315519	0.04011306
Dependability	0.330071101	0.041959732	1	0.55296542	0.03150571	-0.259020462	-0.074207986	0.370604744	0.188567412	-0.202518492	0.6786145	0.12296096	-0.5939016	-0.6665518
Competence	0.449010273	-0.084105742	0.55296542	1	-0.0658519	-0.410819741	0.326927547	0.665156501	-0.090710803	-0.861460743	0.92081992	0.09221305	-0.2713169	-0.4513234
Quotation Presence	-0.393501125	0.556501192	0.03150571	-0.0658519	1	0.487683491	-0.64677126	0.103199708	0.637809354	0.242112965	-0.0114381	0.38622513	-0.0204752	0.273164
Article Prominence	-0.250304844	0.333551372	-0.2590205	-0.4108197	0.48768349	1	-0.507974154	-0.413882239	-0.015283845	0.350256389	-0.4736824	-0.2059053	0.42883823	-0.1271021
Industry Reviews	0.232900083	-0.478084581	-0.0742078	0.32692755	-0.6467713	-0.507974154	1	0.463084436	-0.44085974	-0.538556181	0.15412929	-0.6511017	0.00439919	-0.1883228
Industry Insights	-0.167921073	0.283916955	0.37060474	0.6651565	0.10319971	-0.413882239	0.46308436	1	0.392327359	-0.540564449	0.56616468	-0.0045526	-0.3598734	-0.1853226
Controversies	-0.559931744	0.826797919	0.18856741	-0.0907108	0.63780935	-0.015283845	-0.44085974	0.392327359	1	0.330459503	0.02542306	0.49316578	-0.3117678	0.29701202
Sponsorships/ Partnerships	-0.327851156	0.283865022	-0.2025185	-0.8614607	0.24211296	0.350256389	-0.538556181	-0.540564449	0.330459503	1	-0.7851135	0.17363151	0.2296997	0.24574998
Personal Life	0.338116485	-0.108408797	0.6786145	0.92081992	-0.0114381	-0.4736824	0.154129286	0.566164678	0.025423064	-0.785113534	1	0.23744463	-0.5806	-0.3304445
Impact on Consumer Bahviour	-0.192973073	0.338956826	0.12296096	0.09221305	0.38622513	-0.205905294	-0.651101718	-0.004552629	0.493165784	0.173631506	0.23744463	1	-0.2245528	0.38829951
Platform Specific	0.284885132	0.073155187	-0.5939016	-0.2713169	-0.0204752	0.428838228	0.004399191	-0.359873395	-0.311767823	0.2296997	-0.5806	-0.2245528	1	-0.0300407
Legal/Ethical Issues	-0.541902622	0.040113062	-0.6665518	-0.4513234	0.273164	-0.127102143	-0.188322821	-0.185322605	0.297012021	0.245749983	-0.3304445	0.38829951	-0.0300407	1

Discussion

The following discussion section is centred around key research findings and corresponding theoretical frameworks. The main insights and theories have been laid out below in Table 11.

Table 11

Theoretical Framework and Research Findings Integration

Finding	Related Theory	Theoretical Support	Research Implications
Trust scores show significant positive correlation (0.58) with lemmatized similarity scores	(Grunig & Hon, 1999); (Fleischmann et al., 2007); (Pascual-Ferra, 2020, 2021); (Choudhury et al., 2002); (Au Yeung & Iwata, 2011)	Trust is multi-dimensional, incorporating integrity, dependability, and competence; trust mediates between information quality and information usage; trust measurement requires multiple methodological approaches; trust is multidimensional and context dependent; trust relations indicate homophily and social influence	Validates the use of trust components as predictive measures of content alignment with consumer opinion
Trust indicators in comments predict consumer sentiment on Amazon	(Beldad et al., 2010); (Boertjes et al., 2012); (Bronner et al., 2012)	Trust determinants vary by context; sentiment analysis can measure trust; source identification affects credibility	Demonstrates the value of comment analysis for predicting broader consumer opinion
Smaller influencers demonstrate stronger trust-similarity correlations than those with larger followings	(Luoma-aho et al., 2019); (Casaló et al 2020); (Cauberghe et al., 2007); (Chung & Cho, 2017); (Djafarova & Rushworth, 2017); (Ghazisaeedi et al., 2012); (Valentini, 2015);	Non-traditional celebrities and smaller influencers are often perceived as more credible; perceived originality and uniqueness of posts matter; product type, content focus, and audience interest matter more than followers; younger consumers exhibit higher levels of source trustworthiness; genuine engagement matters more than reach; two-way symmetrical communication builds authentic relationships	Challenges conventional influencer selection metrics based primarily on follower count
Positive correlation (0.44) between controversy coverage and trust scores	(Rawlins, 2008); (Valentini, 2015)	Greater transparency builds trust; social media requires genuine dialogue	Suggests transparency in addressing controversies may enhance rather than diminish influencer credibility
Strong correlation (0.82) between integrity keywords and controversy coverage in media analysis	(Valentini, 2015); (Rawlins, 2008)	Social media requires authentic engagement; greater transparency builds trust	Suggests integrity/transparency in handling controversies may strengthen opinion leadership position
Strong correlation (0.56) between integrity-related keywords and presence of quotations in media coverage	(Valentini, 2015); (Kim & Brown, 2015)	Two-way communication fosters authentic relationships and transparent dialogue; direct discourse enhances credibility by demonstrating expertise and enabling authentic voice	Direct quotations in media coverage serve as mechanisms for two-way communication and demonstrations of expertise, contributing to influencer credibility through authentic voice and transparent dialogue

RQ1: How and to what extent do the comments on tech influencers' product reviews reflect trust indicators (integrity, dependability, and competence), and what is the overall sentiment for each trust component?

The analysis of comments on tech influencers' product reviews revealed significant insights into how trust indicators are reflected in audience responses. By examining the presence of keywords related to integrity, dependability, and competence, the study provided a nuanced understanding of trust dynamics in the context of technology product reviews on platforms like YouTube.

The research methodology involved multiple iterations of trust score calculations, reflecting the complexity of quantifying trust in online interactions. This iterative approach aligns with Pascual-Ferra's (2021; 2020) observations on the challenges of trust measurement in digital contexts and the need for adaptive research methods.

A notable finding from the analysis was the substantial variation in trust scores across different influencers. For instance, in the V5 iteration of trust score calculation, Unbox Therapy received a score of 0.63090121, while Mrwhosetheboss achieved a score of 1.61851211. The actual difference between these two scores is 0.9876109, which represents a 156.5% increase from Unbox Therapy's score to Mrwhosetheboss's score.

This significant difference in trust scores between Unbox Therapy and Mrwhosetheboss reveals a substantial disparity in how audiences perceive and express trust indicators in their comments. The 156.5% higher score for Mrwhosetheboss suggests that the comments on this influencer's product reviews contain a much higher prevalence of trust-related keywords associated with integrity, dependability, and competence.

This stark contrast in trust scores underscores the variability in audience perceptions of trustworthiness among tech influencers, even within the same content niche. It suggests that Mrwhosetheboss has been more successful in cultivating expressions of trust from his audience, as reflected in the comments on his product reviews. The magnitude of this difference highlights the importance of individual influencer characteristics, content strategies, and audience engagement in building and maintaining trust in the digital sphere. It also reinforces the notion that trust in tech influencers is not uniform across the industry but can vary dramatically based on factors specific to each content creator. In other words, expressions of trust are context-specific (Beldad et al., 2010).

These findings align with Choudhury et al.'s (2002) assertion that trust is multi-dimensional and context-dependent. The variability in trust scores across different influencers underscores the complex nature of trust in digital environments, particularly in the realm of technology product reviews where technical expertise and unbiased assessments are highly valued by audiences. Moreover, the results suggest that trust in tech influencers is not a monolithic concept but rather a nuanced construct that varies significantly across different content creators. This variation highlights the importance of individual influencer characteristics and content strategies in building and maintaining audience trust.

The methodology employed in this study, which focused on identifying and categorizing trust indicators in comments, represents a novel approach to trust measurement in influencer contexts. While it diverges from traditional survey-based methods prevalent in public relations literature (Pascual-Ferra, 2020), it offers a more scalable approach to analyzing large volumes of user-generated content and provides insights into naturally occurring expressions of trust in audience interactions.

The trust components proposed by Grunig and Hon (1999) - integrity, dependability, and competence - proved to be valuable frameworks for analyzing trust in tech influencer content. The presence of these keywords in viewer comments suggests that audiences actively engage with these trust dimensions when evaluating influencer credibility, even if not explicitly articulating them. This will become particularly evident when examining the findings of research question 2.

RQ2: How closely do the contents of tech influencers' product reviews resemble the contents of Amazon reviews for the same products? Does the similarity between Amazon reviews and tech influencers' product reviews vary based on the level of trust in the influencer?

The results of this analysis reveal significant correlations between certain trust score versions and the lemmatized similarity scores, providing insights into the relationship between influencer trustworthiness and the alignment of their content with consumer reviews. Specifically, the V3 trust score showed a correlation of 0.57 with the V2 lemmatized similarity score, while the V5 trust score demonstrated an even stronger correlation of 0.58. These findings support the hypothesis that trust in tech influencers, as measured by indicators in the comments section of their videos, is predictive of consumer opinion as expressed in Amazon reviews of the same products.

The strong correlations observed between trust scores and lemmatized similarity scores align with several theories and findings from previous research. For instance, Choudhury et al. (2002) emphasized the multi-dimensional nature of trust, including its basis in communication. The results of this current research suggest that influencers who are perceived as more trustworthy tend to produce content that more closely aligns with consumer opinions, potentially indicating effective communication of product features and benefits.

The strong correlations between trust scores and similarity scores validate the use of sentiment analysis in measuring trust, as proposed by Boertjes et al. (2012). Their model for gauging trust based on sentiment analysis of social media posts aligns with the approach used in this study, specifically trust score V3, which combined sentiment analysis with keyword identification and frequency analysis. However, the findings suggest that while sentiment analysis can be an important component in measuring trust, it is not always necessary and might not even be the strongest measure. The V5 trust score version, which showed the highest correlation with similarity scores, incorporated only keyword frequency analysis.

This observation adds nuance to the existing literature on trust measurement in digital spaces. While previous studies, such as Chandio and Sah (2020), have demonstrated the effectiveness of sentiment analysis in reflecting public trust, this research suggests that a combination of methods, including keyword identification and frequency analysis, can provide a more accurate measure of trust in influencers.

The effectiveness of the trust measurement approach aligns with the multi-dimensional construct of trust proposed by Grunig and Hon (1999), which includes integrity, dependability, and competence. The keyword-based analysis, which focused on these trust components, appears to capture important aspects of influencer trustworthiness that correlate with the similarity between their content and consumer opinions.

Furthermore, the strong correlation between trust scores and similarity scores supports the findings of Lou and Yuan (2019), who found that influencer credibility positively affects followers' trust in branded posts and purchase intentions. The results suggest that influencers who are perceived as more trustworthy tend to produce content that aligns more closely with consumer opinions, potentially indicating a higher level of credibility and influence.

The effectiveness of the lemmatized approach in capturing content similarities also reflects the importance of informative content in building trust (Lou & Yuan, 2019). The lemmatization process may better preserve the informative value of the content, allowing for more accurate comparisons between influencer reviews and consumer opinions.

These findings also have implications for the concept of trust signals, as discussed by Baradell (2022). The alignment between influencer content and consumer opinions, as measured by the similarity scores, could be seen as a form of trust signal. Influencers whose content more closely matches consumer sentiments may be perceived as more authentic and credible, thereby building trust with their audience.

The relationship between trust scores and similarity scores also relates to the work of Fleishmann et al. (2007) on trust in digital information. Their model positions trust as a mediating factor between information quality and information usage. The results of this current study suggest that influencers with higher trust scores produce content that more closely aligns with consumer opinions, potentially indicating higher information quality and greater likelihood of information usage by consumers.

The analysis also revealed an interesting trend in the relationship between trust scores, similarity scores, and influencer follower count. As illustrated in Figure 3, there appears to be a negative correlation between V5 trust scores correlative relationship with similarity scores and follower count, with the correlation declining as follower count increases. This inverse relationship presents a paradox in influencer marketing: while conventional wisdom might suggest that larger followings indicate greater credibility, the findings reveal a more nuanced reality. Influencers with smaller followings demonstrate stronger correlations between trust indicators and content similarity with consumer opinions.

This phenomenon could be explained through several theoretical lenses and aligns with findings from previous studies. First, Djafarova and Rushworth (2017) noted that non-traditional celebrities, such as bloggers and YouTube personalities, are often perceived as more credible and relatable than traditional celebrities with larger followings. As influencers' audiences scale up, they may be seen more as celebrities rather than relatable personalities, potentially diminishing perceived authenticity.

The role of parasocial relationships provides another framework for understanding this relationship. Smaller influencers may foster stronger parasocial relationships with their followers due to their ability to maintain more personal connections with their audience. This aligns with the findings of Chung and Cho (2017), who found that parasocial relationships positively influence trust in celebrities and the perceived credibility of their endorsements. Similarly, Luoma-aho et al. (2019) suggest that smaller influencers can develop stronger parasocial interactions, leading to higher-quality engagement and more genuine trust relationships.

Furthermore, this finding supports Casaló et al. (2020), who discovered that perceived originality and uniqueness of posts, rather than follower count, are key factors in establishing opinion leadership on social media platforms. The authors stressed authenticity and creativity in building trust, which may be reflected in the higher correlations observed for smaller influencers. These content creators may focus more on delivering accurate, detailed product analyses that resonate with actual consumer experiences, rather than prioritizing broader appeal or entertainment value.

Lastly, this supports the work of Cauberghe et al. (2017), who emphasized the importance of considering the product type, the influencer's content focus, and their audience's interest rather than solely relying on follower count when selecting influencers for campaigns. The stronger correlation between trust and similarity scores for smaller influencers suggests that follower count alone is not a reliable indicator of true influence.

The effectiveness of the lemmatized approach in producing higher similarity scores is noteworthy. Lemmatization, which reduces words to their base or dictionary form, appears to capture the semantic similarities between influencer reviews and consumer opinions more accurately than stemming or unstemmed approaches. This finding suggests that future research in this field should consider using lemmatization techniques for more precise content comparison.

The strong correlations observed between trust scores and lemmatized similarity scores align with recent research by He et al. (2024) and Al-Gasawneh et al. (2023)a, who found that emotional sentiments in reviews demonstrate predictable patterns that can be accurately analyzed using advanced language processing techniques. He et al.'s (2024) finding that negative reviews propagate rapidly through social networks, combined with Al-Gasawneh et al.'s (2023) emphasis on real-time sentiment monitoring, adds context to the observation of the relationship between trust indicators and consumer opinions. This suggests that the alignment between influencer content and consumer sentiments may be particularly important when managing negative feedback or controversies. Furthermore, both studies reinforce this study's methodological approach of combining automated analysis with human interpretation to ensure accurate and contextually appropriate results.

The researcher's findings regarding the inverse relationship between follower count and trust-similarity correlations align with the comprehensive review by Jagani et al. (2023), who analyzed 214 articles on social media influencer marketing. Their review revealed that while content-related factors and influencer characteristics are heavily researched as antecedents of trust, the comparative factors examining influencer-follower relationships remain underexplored. This gap in the literature is particularly relevant to the finding that smaller influencers demonstrate stronger correlations between trust indicators and content similarity with consumer opinions. Jagani et al.'s identification of parasocial interactions and relationships as a major theme in influencer research provides theoretical grounding for why smaller influencers may be more effective at building trust – through their ability to maintain more personal connections and authentic engagement with their audience.

The inverse relationship between follower count and trust-similarity correlations found in the study aligns with recent empirical findings from Hew et al. (2023), whose comprehensive investigation into social media influencer marketing offers important context for understanding this dynamic. Their research revealed that while influencer credibility positively affects consumer attitudes, interactivity between influencers and followers can have a negative impact when not properly managed. They found that influencers with larger followings often struggle to maintain high-quality engagement with their audience, as they "do not have enough time to communicate with their followers intimately" (Hew et al., 2023, p. 10). This helps explain the finding that smaller influencers demonstrate stronger correlations between trust indicators and content similarity with consumer opinions. The mechanics underlying this relationship appear to be rooted in the ability of smaller influencers to dedicate more time to meaningful interactions with their followers, leading to more authentic relationships and stronger trust signals. Hew et al.'s work reinforces the finding that marketers should look beyond surface-level metrics like follower count when evaluating influencers, as the quality of engagement and authenticity of interactions are more reliable predictors of trust and influence. Their research particularly validates the observation that trust in tech influencers is not uniform across the industry but varies significantly based on factors specific to each content creator's ability to maintain authentic engagement with their audience.

The analysis of the relationship between trust scores and content similarity scores provides valuable insights into the dynamics of influencer trustworthiness and its impact on content alignment with consumer opinions. The strong correlations observed, particularly with lemmatized similarity scores, validate the approach to measuring trust while also highlighting the complex interplay between various factors such as follower count, content authenticity, and audience engagement. These findings contribute to the growing body of research on influencer marketing and trust in digital spaces, offering both theoretical insights and practical implications for future studies in this field.

RQ3: How are tech influencers portrayed in media coverage that mentions and quotes them, and is there a relationship between the level of trust in the influencer (as determined by RQ1) and the nature of the media coverage they receive? Additionally, are there any instances of legal actions against the influencers?

The analysis of media coverage for the selected tech influencers reveals complex relationships between trust scores, media sentiment, and various thematic elements in news articles. These findings offer valuable insights into the relationship between influencer trustworthiness and media portrayal, contributing to our understanding of trust dynamics in the digital age.

The moderate positive correlation (0.44) between trust scores and the prevalence of "controversy" keywords in news articles initially appears counterintuitive. However, this relationship can be better understood through Rawlins' (2008) emphasis on transparency in trust-building. This researcher posits that the media's discussion of controversies, particularly when framed in terms of integrity (as evidenced by the 0.82 correlation between "controversy" and "integrity" keywords), may serve to reinforce trust by demonstrating transparency. This suggests that when influencers are involved in controversies, the way these issues are addressed and discussed in the media can potentially enhance rather than diminish their perceived trustworthiness, provided the coverage emphasizes integrity and transparency.

This interplay between controversy and integrity in media coverage aligns with Grunig and Hon's (1999) multi-dimensional construct of trust. Their inclusion of integrity as a key component of trust is reflected in the weak but positive correlation (0.24) between trust scores and the presence of "integrity" keywords in news articles. Valentini (2015) would likely argue that this focus on integrity in media coverage contributes to the two-way symmetrical communication model by fostering transparency and open dialogue. When media coverage emphasizes an influencer's integrity, or transparency (Rawlins, 2008), for example, it encourages a more balanced and honest exchange of information between the influencer and their audience. This approach aligns with the two-way symmetrical model's emphasis on mutual understanding and reciprocal communication, rather than one-sided messaging. By highlighting integrity, media coverage can prompt influencers to engage in more authentic and transparent communication with their audience, thereby strengthening relationships and trust in the social media environment.

The correlation between integrity-related keywords and the presence of quotations (0.56) in news articles brings together the perspectives of Valentini (2015) and Kim and Brown (2015).

Valentini would likely emphasize this as an example of two-way communication fostering authentic relationships. This interpretation is supported by several factors.

First, quotations allow influencers to speak directly to their audience through media, providing unfiltered thoughts and opinions. This direct communication helps create a sense of transparency and authenticity. Second, by including influencer quotes, media outlets are effectively facilitating a form of dialogue between influencers and their audience. This can be seen as an extension of the two-way communication model into traditional media spaces. Third, quotes can often provide context and nuance to discussions, allowing influencers to explain their positions more fully. This depth of communication can contribute to building more authentic and understanding relationships with the audience. Fourth, when influencers are directly quoted, they are held accountable for their words, which can increase perceived integrity and trustworthiness. Finally, quotes can spark further discussion and engagement from the audience, potentially leading to more direct interactions on social media platforms. This ongoing engagement is crucial for maintaining and deepening relationships between influencers and their followers.

In contrast, Kim and Brown (2015) might focus on how direct quotations contribute to the perceived expertise of influencers, which they identify as a key dimension of credibility in social media spaces. Quotations can demonstrate an influencer's knowledge, articulation skills, and thought leadership, all of which contribute to their perceived expertise and, by extension, their credibility.

This dual perspective highlights how the presence of quotations in media coverage can simultaneously foster authentic relationships through two-way communication and enhance the perceived expertise and credibility of influencers. The strong correlation observed in this study between integrity-related keywords and the presence of quotations (0.56) supports both interpretations, suggesting that media coverage incorporating direct quotes from influencers may play a significant role in building trust and credibility.

The weaker positive correlations found between trust scores and themes related to personal life (0.19), impact on consumers (0.20), and the total number of articles (0.21) speak to the multifaceted nature of influencer trust. Freberg et al. found that social media influencers are perceived as “verbal, smart, ambitious, productive, and poised” (2011, p. 91). They are also seen as likely to be sought out for advice and to give advice. These findings align with the study results, suggesting that various aspects of an influencer's public persona contribute to their perceived trustworthiness. The correlation with "impact on consumers" that we observed supports Freberg et al.'s (2011) finding that social media influencers are perceived as advice-givers, further indicating how social media usage affects different types of trust, including trust in influencers as sources of consumer information. Sabatini and Sarracino (2019) would likely see the correlation with "impact on consumers" as indicative of how social media usage affects different types of trust, including trust in influencers as sources of consumer information.

The strong positive correlation (0.55) between dependability and competence keywords in news articles reinforces Grunig and Hon's (1999) multi-dimensional view of trust. This finding suggests that media coverage tends to discuss these trust components in tandem, potentially reinforcing the overall perception of an influencer's trustworthiness. It is interesting to note that integrity does not present a notable correlation with either of the other trust keywords, suggesting that these may be distinct components that are discussed separately but all contribute to trust. Furthermore, the strong correlation (0.67) between competence-related keywords and industry insights in news articles aligns with Kim and Brown's (2015) emphasis on expertise as a crucial factor in social media credibility.

The varying distribution of articles among influencers, ranging from 99 for iJustine to 6 for The Tech Chap, aligns with Flanagin and Metzger's (2013) findings on the impact of information volume on trust in online contexts. Their research suggests that higher volumes of user-generated content led to greater trust, which could imply that iJustine, with the highest number of articles, might be perceived as more trustworthy due to the sheer volume of media coverage. However, the cases of UrAvgConsumer and The Tech Chap present a counterpoint to this volume-based trust hypothesis.

Despite their relatively low article counts, UrAvgConsumer and The Tech Chap received uniformly positive sentiment scores in their media coverage. This scenario offers a potential counter-case to the volume-trust relationship proposed by Flanagin and Metzger (2013). Beldad et al. (2010) argue that the determinants of online trust may vary depending on the specific context of the online transaction and the parties involved. In this case, the consistently positive coverage of UrAvgConsumer and The Tech Chap, even in the absence of high volume, might contribute significantly to trust formation.

This phenomenon aligns with the concept of trust as multi-dimensional (Choudhury et al., 2002; Grunig & Hon, 1999; Paine, 2013). While volume plays a role, the uniformly positive sentiment in the limited coverage of UrAvgConsumer and The Tech Chap may compensate for the lack of volume by consistently reinforcing a positive narrative around the influencer. This scenario suggests that the quality and consistency of media coverage might, in some cases, outweigh the quantity in terms of trust formation.

In synthesizing these various theoretical perspectives, a complex picture emerges of how media coverage interacts with and potentially influences influencer trust. The findings suggest that controversies, when framed in terms of integrity and transparency, may not necessarily diminish trust and could even enhance it in some cases. This nuanced understanding challenges simplistic views of trust formation and maintenance in the digital influencer landscape.

The multi-faceted nature of trust, as reflected in the correlations between different themes in media coverage, underscores the complexity of trust dynamics in digital spaces. It suggests that trust in influencers is not solely a function of positive sentiment or high visibility, but rather a complex interplay of factors including perceived integrity, dependability, and competence.

Conclusion

This study on trust in tech influencers reveals a complex interplay of factors that contribute to the perceived trustworthiness and impact of these digital content creators. The research, which employed a multi-faceted approach to analyze trust indicators in YouTube comments, content similarity with consumer reviews, and media portrayal, provides valuable insights into the dynamics of trust in the digital influencer landscape.

The analysis of trust indicators in YouTube comments demonstrated significant variations in trust scores across different influencers, highlighting the individualized nature of trust-building in online, digital environments. This finding aligns with previous research emphasizing the context-dependent nature of trust in digital spaces.

The examination of content similarity between influencer reviews and consumer opinions on Amazon revealed intriguing correlations with trust scores. Notably, the study found that lemmatized similarity scores showed the strongest correlations with certain versions of trust scores, suggesting that the semantic alignment between influencer content and consumer opinions may be a meaningful indicator of trustworthiness. However, the research also uncovered an unexpected negative correlation between follower count and trust scores, challenging assumptions about the relationship between popularity and perceived credibility.

The analysis of media coverage provided complexity and nuance to the understanding of influencer trust. The observed positive correlations between trust scores and the prevalence of controversy-related keywords in news articles, particularly when framed in terms of integrity, suggest that media portrayal of controversies may not necessarily diminish trust and could potentially enhance it in some cases. This finding underscores the importance of considering how issues are framed and discussed in media coverage, in particular when assessing their impact on influencer trustworthiness.

These findings have several significant implications for marketing practitioners and anyone concerned with digital trust formation and maintenance. The strong correlation between trust scores and content similarity suggests that in some situations, practitioners should prioritize influencers whose content demonstrates alignment with genuine consumer experiences, rather than solely focusing on reach metrics. This represents a substantial shift from traditional influencer selection methods that often emphasize audience size over content quality.

The research provides empirical support for a more nuanced approach to influencer selection, particularly in the technology sector. The data suggests that marketing professionals would benefit from implementing trust scoring mechanisms similar to those developed in this study when evaluating potential influencer partnerships. Such mechanisms, incorporating analysis of integrity, dependability, and competence components, could provide a more sophisticated framework for assessing influencer effectiveness than current methods that primarily rely on follower counts and engagement metrics.

Furthermore, the finding that lemmatized similarity scores show the strongest correlations with expressions of trust indicates that practitioners should evaluate the substance of influencer content rather than just surface-level metrics. This suggests the need for more sophisticated content analysis tools in influencer evaluation processes. The research also demonstrates that transparent handling of controversies might enhance rather than diminish trust, as evidenced by the strong positive correlation between transparent controversy handling and trust scores. This challenges conventional industry wisdom and suggests that practitioners should reconsider traditional crisis management approaches in influencer marketing, potentially viewing controversies as opportunities for trust building rather than purely as reputational threats.

The implications extend beyond brand managers to platform operators and tech influencers themselves. For platforms, the research demonstrates the need for more nuanced success metrics that better reflect genuine audience trust. For influencers, the findings about smaller creators' effectiveness in building trust relationships suggests that maintaining authenticity and alignment with consumer opinions may be more valuable for long-term success than rapid follower growth. This is evidenced by the stronger trust-similarity correlations found among influencers with smaller, more engaged audiences. Including such creators in influencer portfolios might offer higher trust-to-reach ratios, potentially providing better returns on investment in terms of building genuine consumer trust and influencing purchasing decisions.

While these results offer valuable insights, several limitations of the study must be acknowledged. The reliance on keyword analysis and sentiment scores for trust measurement, while providing quantifiable data, may not fully capture the nuanced ways in which trust is expressed and perceived in online interactions.

Furthermore, the cross-sectional nature of the study does not account for the dynamic and evolving nature of trust over time. A longitudinal approach could provide more robust insights into how trust in influencers develops and changes in response to various factors, including content consistency, controversies, and shifts in media portrayal.

The sample size for certain aspects of the study, particularly the number of media articles analyzed for some influencers, was relatively small. This limitation may have introduced bias into the results and restricted the ability to draw definitive conclusions about the relationship between media coverage and influencer trust.

Future research could address these limitations by employing a mixed-methods approach that combines quantitative analysis with qualitative methods such as in-depth interviews or focus groups with audience members. This approach could provide a more comprehensive understanding of how trust is built, maintained, and potentially eroded in influencer-audience relationships. For example, researchers could conduct semi-structured interviews with viewers who frequently engage with tech influencer content to understand how their trust assessments evolve over time and what specific factors trigger changes in perceived trustworthiness.

A longitudinal study tracking trust indicators and consumer alignment over an extended period (e.g., 12-24 months) would enable researchers to examine how trust dynamics shift in response to various factors. Such research could reveal patterns in influencer growth and their impact on trust metrics, examine the effect of controversies and their resolution on long-term trust scores, and assess how changes in content strategy relate to trust indicators. This approach would also allow for analysis of how sponsorship disclosure practices evolve and impact audience trust over time.

Cross-cultural analysis represents another promising direction for future research and could also yield valuable insights into how trust dynamics in influencer marketing may vary across different cultural contexts. For example, these studies could reveal whether the trust components identified in this study (integrity, dependability, and competence) have universal applicability or require cultural adaptation. This research could examine comparative trust indicators across multiple language markets, cultural variations in controversy handling, and regional differences in how sponsorship disclosures are interpreted and received.

Platform-specific research could explore how technological affordances and algorithmic systems impact trust formation and maintenance. As social media platforms continue to evolve, understanding how these technological factors interact with human perceptions of trust could provide valuable insights for both researchers and practitioners in the field of influencer marketing. This could include examining the role of platform-specific features in trust-building, the impact of recommendation algorithms on trust distribution across influencer tiers, and how platform monetization models affect perceived trustworthiness. Understanding these technical aspects could provide valuable insights into the structural factors influencing trust in digital spaces.

Future studies could also explore the potential of advanced machine learning techniques for trust analysis. The development of more sophisticated natural language processing models for trust component identification, implementation of multimodal analysis incorporating video and audio content, and creation of real-time trust monitoring systems could enhance our ability to measure and understand trust in digital contexts.

Market-specific studies could examine how trust dynamics vary across different product categories beyond technology. Understanding how trust indicators manifest in various sectors, how product complexity affects the relationship between trust and consumer alignment, and how price points influence trust requirements in influencer recommendations would provide valuable insights for both researchers and practitioners.

These research directions would not only address the limitations of the current study but also contribute to a more comprehensive understanding of trust dynamics in digital influence. As the influencer marketing landscape continues to evolve, understanding these various dimensions of trust will become increasingly crucial for both academic research and industry practice.

In conclusion, this research contributes to the growing body of literature on trust in digital spaces by providing empirical evidence of the complex relationships between various factors that influence perceived trustworthiness of tech influencers. The findings underscore the multifaceted nature of trust in online environments and highlight the need for nuanced approaches to measuring and understanding trust in the context of influencer marketing.

The importance of this research lies in its potential to inform more effective and ethical practices in influencer marketing. By shedding light on the factors that contribute to trust-building and the potential pitfalls that may erode trust, this study can guide influencers, marketers, and platforms in fostering more authentic and trustworthy relationships with audiences. As the digital landscape continues to evolve, understanding the dynamics of trust in influencer-audience relationships will remain crucial for navigating the challenges and opportunities of this increasingly significant form of digital communication.

Generative AI Disclosure

This research project incorporated generative artificial intelligence (AI) tools as part of its methodology. The integration of generative AI assistance reflects an effort to leverage advanced technologies while maintaining the integrity and rigour of traditional academic research methodologies. Generative AI was employed in two primary capacities: as a research assistant and for code development.

In its role as a research assistant, generative AI aided in the identification and retrieval of relevant academic articles and sources. This AI-assisted literature review process helped to broaden the scope of examined research and ensured a comprehensive foundation for the study.

However, it is crucial to note that all selected sources were manually reviewed, vetted, and, crucially, read by the human researcher to ensure their relevance, credibility, and appropriateness for inclusion in the final paper.

Generative AI was also extensively used in the drafting and refinement of Python scripts utilized throughout the data collection and analysis phases of this study. The AI assistance in code development facilitated the creation of more efficient and robust scripts for tasks such as web scraping, data cleaning, sentiment analysis, and statistical computations. As with the literature review process, all generated code was thoroughly reviewed, tested, and modified as necessary by the human researcher to ensure its accuracy and suitability for the specific research requirements.

While generative AI played a significant role in the research process, it is important to emphasize that all final decisions regarding research design, data interpretation, and conclusions drawn were made solely by the human researcher. The use of generative AI in this project served to enhance the efficiency and scope of the research process, but did not replace critical thinking, analysis, or academic judgment.

This disclosure is made in the interest of transparency and to acknowledge the evolving role of AI tools in academic research. As the academic community continues to grapple with the implications of generative AI in research processes, it is imperative that researchers remain transparent about their use of these tools and maintain a clear distinction between AI-assisted processes and human-driven analysis and interpretation.

References

Al-Gasawneh, J. A., Almaududi Ausat, A. M., Diantoro, K., Khootimah Azzaakiyyah, H., & Sudirjo, F. (2023). Application of ChatGPT in improving customer sentiment analysis for businesses. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *5*(3), 283–288. https://doi.org/10.47233/jteksis.v5i3.871

Arora, A., Aswani, R., Bansal, S., Dwivedi, Y., & Kandpal, C. (2019). Measuring social media influencer index- insights from facebook, Twitter and Instagram. *Journal of Retailing and Consumer Services*, *49*, 86–101. https://doi.org/10.1016/j.jretconser.2019.03.012

Au Yeung, C., & Iwata, T. (2011). Strength of social influence in trust networks in product review sites. *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, 495–504. https://doi.org/10.1145/1935826.1935899

Baradell, S. (2022). Trust signals: Brand building in a post-truth world. In *Trust signals: Brand building in a post-truth world* (1st ed.). Lioncrest Publishing.

Beaudoin, C. E. (2008). Explaining the relationship between internet use and interpersonal trust: Taking into account motivation and information overload. *Journal of Computer-Mediated Communication*, *13*(3), 550–568. https://doi.org/10.1111/j.1083-6101.2008.00410.x

Beldad, A., Jong, M. de, & Steehouder, M. (2010). How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust. *Computers in Human Behavior*, *26*(5), 857–869. https://doi.org/10.1016/j.chb.2010.03.013

Boertjes, E., Gerrits, B., Kooij, R., van Maanen, P.-P., Raaijmakers, S., & de Wit, J. (2012). Towards a Social Media-Based Model of Trust and Its Application. In M. D. Hercheui, D. Whitehouse, W. McIver, & J. Phahlamohlaka (Eds.), *ICT Critical Infrastructures and*

Society (pp. 250–263). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33332-3_23

Bronner, F., Neijens, P. C., & Willemsen, L. M. (2012). The Ironic Effect of Source Identification on the Perceived Credibility of Online Product Reviewers: The Ironic Effect of Source Identification. *Journal of Computer-Mediated Communication*, *18*(1), 16–31. https://doi.org/10.1111/j.1083-6101.2012.01598.x

Casaló, L. V., Flavián, C., & Ibáñez-Sánchez, S. (2020). Influencers on Instagram: Antecedents and consequences of opinion leadership. *Journal of Business Research*, *117*, 510–519. https://doi.org/10.1016/j.jbusres.2018.07.005

Casánez-Ventura, A., Castejón-Mateos, F., Cuenca-Jiménez, P.-M., & Rodríguez-Ibánez, M. (2023). A review on sentiment analysis from social media platforms. *Expert Systems with Applications*, *223*, 119862-. https://doi.org/10.1016/j.eswa.2023.119862

Cauberghe, V., De Veirman, M., & Hudders, L. (2017). Marketing through Instagram influencers: The impact of number of followers and product divergence on brand attitude. *International Journal of Advertising*, *36*(5), 798–828. https://doi.org/10.1080/16522354.2018.1501146

Cervellon, M.-C., & Galipienzo, D. (2015). Facebook Pages Content, Does it Really Matter? Consumers' Responses to Luxury Hotel Posts with Emotional and Informational Content. *Journal of Travel & Tourism Marketing*, *32*(4), 428–437. https://doi.org/10.1080/10548408.2014.904260

Chandio, M. M., & Sah, M. (2020). Brexit Twitter Sentiment Analysis: Changing Opinions About Brexit and UK Politicians. In *Intelligent Computing Paradigm and Cutting-Edge*

Technologies (Vol. 9, pp. 1–11). Springer International Publishing AG. https://doi.org/10.1007/978-3-030-38501-9_1

Chan-Olmsted, S., Wang, R., & Xiao, M. (2018). Factors affecting YouTube influencer marketing credibility: A heuristic-systematic model. *Journal of Media Business Studies*, *15*(3), 188–213. https://doi.org/10.1080/16522354.2018.1501146

Chen, Y., Fay, S., & Wang, Q. (2011). The role of marketing in social media: How online consumer reviews evolve. *Journal of Interactive Marketing*, *25*(2), 85–94. https://doi.org/10.1016/j.intmar.2011.01.003

Childers, C. C., Hoy, M. G., & Lemon, L. L. (2019). Sponsored #Ad: Agency perspective on influencer marketing campaigns. *Journal of Current Issues and Research in Advertising*, *40*(3), 258–274. https://doi.org/10.1080/10641734.2018.1521113

Cho, H., & Chung, S. (2017). Fostering parasocial relationships with celebrities on social media: Implications for celebrity endorsement. *Psychology & Marketing*, *34*(4), Article 4. https://doi.org/10.1002/mar.21001

Choudhury, V., McKnight, D. H., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, *13*(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

Djafarova, E., & Rushworth, C. (2017). Exploring the credibility of online celebrities' Instagram profiles in influencing the purchase decisions of young female users. *Computers in Human Behavior*, *68*, 1–7. https://doi.org/10.1016/j.chb.2016.11.009

Fisher, M. (2022). The chaos machine: The inside story of how social media rewired our minds and our world. In *The chaos machine: The inside story of how social media rewired our minds and our world* (First edition.). Little, Brown and Company.

Flanagin, A. J., & Metzger, M. J. (2013). Trusting expert- versus user-generated ratings online: The role of information volume, valence, and consumer characteristics. *Computers in Human Behavior*, *29*(4), 1626–1634. https://doi.org/10.1016/j.chb.2013.02.001

Fleischmann, K. R., Kelton, K., & Wallace, W. A. (2007). Trust in digital information. *Journal of the American Society for Information Science and Technology*, *59*(3), 363–374. https://doi.org/10.1002/asi.20722

Forde, S., Heinz, S., Opwis, K., Seckler, M., & Tuch, A. N. (2015). Trust and distrust on the web: User experiences and website characteristics. *Computers in Human Behavior*, *45*, 39–50. https://doi.org/10.1016/j.chb.2014.11.064

Freberg, K., Freberg, L. A., Graham, K., & McGaughey, K. (2011). Who are the social media influencers? A study of public perceptions of personality. *Public Relations Review*, *37*(1), 90–92. https://doi.org/10.1016/j.pubrev.2010.11.001

Ghafari, S. M., Beheshti, A., Joshi, A., Paris, C., Mahmood, A., Yakhchi, S., & Orgun, M. A. (2020). A Survey on Trust Prediction in Online Social Networks. *IEEE Access*, *8*, 144292–144309. https://doi.org/10.1109/ACCESS.2020.3009445

Ghazisaeedi, M., Steyn, P., & van Heerden, G. (2012). Trustworthiness of product review blogs: A source trustworthiness scale validation. *African Journal of Business Management*, *6*(25), 7498-. https://doi.org/10.5897/AJBM12.079

Grunig, J. E., & Hon, L. C. (1999). *Guidelines for Measuring Relationships in Public Relations*. Institute for Public Relations. https://www.instituteforpr.org/wp-content/uploads/Guidelines_Measuring_Relationships.pdf

Grunig, J. E., & Hunt, T. (1984). Managing public relations. In *Managing public relations*. Holt, Rinehart and Winston.

Håkansson, P., & Witmer, H. (2015). Social media and trust—A systematic literature review. *Journal of Business and Economics*, *6*(3), 517–524. https://doi.org/10.15341/jbe(2155-7950)/03.06.2015/010

Han, I., Lee, J., & Park, D.-H. (2008). The effect of negative online consumer reviews on product attitude: An information processing view. *Electronic Commerce Research and Applications*, *7*(3), 341–352. https://doi.org/10.1016/j.elerap.2007.05.004

Haque, T. U., Saber, N. N., & Shah, F. M. (2018). Sentiment analysis on large scale Amazon product reviews. *2018 IEEE International Conference on Innovative Research and Development (ICIRD)*, 1–6. https://doi.org/10.1109/ICIRD.2018.8376299

He, S., Wang, Z., Yan, H., Zhu, Y., & Zhu, Z. (2024). *LLM for Sentiment Analysis in E-commerce: A Deep Dive into Customer Feedback*. https://doi.org/10.5281/ZENODO.12730477

Hew, J.-J., Lee, V.-H., Leong, L.-Y., Lim, A.-F., Ooi, K.-B., & Tan, G. W.-H. (2023). Social media influencers: An effective marketing approach? *Journal of Business Research*, *160*, 113773. https://doi.org/10.1016/j.jbusres.2023.113773

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168–177. https://doi.org/10.1145/1014052.1014073

Hubspot. (2024, July 29). *The video marketing playbook*. https://blog.hubspot.com/marketing/how-video-consumption-is-changing

Jagani, K., Joshi, Y., Kumar, S., & Lim, W. M. (2023). Social media influencer marketing: Foundations, trends, and ways forward. *Electronic Commerce Research*. https://doi.org/10.1007/s10660-023-09719-z

Jin, S. V., Muqaddam, A., & Ryu, E. (2019). Instafamous and social media influencer marketing. *Marketing Intelligence & Planning*, *37*(5), 567–579. https://doi.org/10.1108/MIP-09-2018-0375

Kim, C. M., & Brown, W. J. (2015). Conceptualizing Credibility in Social Media Spaces of Public Relations. *Public Relations Journal (2007)*, *9*(4). https://www.researchgate.net/publication/325808045_Conceptualizing_Credibility_in_Social_Media_Spaces_of_Public_Relations/link/5b25a568a6fdcc6974695ce8/download?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19

Liss, J. (2011). Negotiating the marcellus: The role of information in building trust in extractive deals. *Negotiation Journal*, *27*(4), 419–446. https://doi.org/10.1111/j.1571-9979.2011.00317.x

Liu, Y., Jiang, C., & Zhao, H. (2019). Assessing product competitive advantages from the perspective of customers by mining user-generated content on social media. *Decision Support Systems*, *123*, 113079. https://doi.org/10.1016/j.dss.2019.113079

Lou, C., & Yuan, S. (2019). Influencer marketing: How message value and credibility affect consumer trust of branded content on social media. *Journal of Interactive Advertising*, *19*(1), 58–73. https://doi.org/10.1080/15252019.2018.1533501

Luoma-aho, V., Maity, D., Munnukka, J., & Reinikainen, H. (2019). “Thanks for watching”. The effectiveness of YouTube vlogendorsements. *Computers in Human Behavior*, *93*, 226–234. https://doi.org/10.1016/j.chb.2018.12.014

Marichal, J., & Monforti, J. L. (2014). The role of digital skills in the formation of generalized trust among latinos and African Americans in the United States. *Social Science Computer Review*, *32*(1), 3–17. https://doi.org/10.1177/0894439313497469

McKinsey & Company. (2023, April 10). *What is influencer marketing?* http://ceros.mckinsey.com/quarterly-digital-promo

Mudambi, S., & Schuff, D. (2010). Research note: What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly*, *34*(1), 185. https://doi.org/10.2307/20721420

Paine, K. D. (2013). *Guidelines for Measuring Trust in Organizations*. The Institute for Public Relations Commission on PR Measurement and Evaluation. https://instituteforpr.org/wp-content/uploads/Guidelines-for-Measuring-Trust-KDP-4-13.pdf

Pascual-Ferrá, P. (2020). The Measurement of Trust in Communication Research. *Communication Research Trends*, *39*(4), 4–34. http://libaccess.mcmaster.ca/login?url=https://www.proquest.com/scholarly-journals/measurement-trust-communication-research/docview/2474294321/se-2?accountid=12347

Pascual-Ferra, P. (2021). The Measurement of Trust in Communication Research: Part 2. *Communication Research Trends*, *40*(1), 4–36. http://libaccess.mcmaster.ca/login?url=https://www.proquest.com/scholarly-journals/measurement-trust-communication-research-part-2/docview/2504561742/se-2?accountid=12347

Rawlins, B. L. (2008). Measuring the relationship between organizational transparency and employee trust. *Public Relations Journal (2007)*, *2*(2).

https://www.researchgate.net/publication/239919667_Measuring_the_relationship_between_organizational_transparency_and_employee_trust

Sabatini, F., & Sarracino, F. (2019). Online Social Networks and Trust. *Social Indicators Research*, *142*(1), 229–260. https://doi.org/10.1007/s11205-018-1887-2

Schwemmer, C., & Ziewiecki, S. (2018). Social media sellout: The increasing role of product promotion on YouTube. *Social Media + Society*, *4*(3), 205630511878672-. https://doi.org/10.1177/2056305118786720

Shah, D. V., Kwak, N., & Holbert, R. L. (2001). 'Connecting' and 'Disconnecting' with civic life: Patterns of internet use and the production of cocial capital. *Political Communication*, *18*(2), 141–162. https://doi.org/10.1080/105846001750322952

Song, B. L., Lee, K. L., Liew, C. Y., & Subramaniam, M. (2023). The role of social media engagement in building relationship quality and brand performance in higher education marketing. *International Journal of Educational Management*, *37*(2), 417–430. https://doi.org/10.1108/IJEM-08-2022-0315

Song, V. (2024, August 16). *Google threatened tech influencers unless they 'preferred' the Pixel—The Verge*. The Verge. https://www.theverge.com/2024/8/16/24221755/google-team-pixel-reviews-influencers

Sprout Social. (2024, April 25). *New research reveals influencers significantly drive purchasing decisions*. https://investors.sproutsocial.com/news/news-details/2024/New-Research-Reveals-Influencers-Significantly-Drive-Purchasing-Decisions/default.aspx

Stacks, D. W. (2017). *Primer for public relations research* (Third edition.). Guilford Press.

Tschannen-Moran, M., & Hoy, W. K. (2000). A Multidisciplinary Analysis of the Nature, Meaning, and Measurement of Trust. *Review of Educational Research*, *70*(4), 547. https://doi.org/10.2307/1170781

Valentini, C. (2015). Is using social media “good” for the public relations profession? A critical reflection. *Public Relations Review*, *41*(2), 170–177. https://doi.org/10.1016/j.pubrev.2014.11.009

Weinstein, D. (2019, January). *Online video’s role in the purchase funnel*. Think with Google. https://www.thinkwithgoogle.com/marketing-strategies/video/video-purchase-funnel/

Appendix A

This Python script automates the collection of YouTube video transcripts using the YouTube Transcript API. The script accepts a YouTube URL as input, extracts the video ID, fetches the auto-generated transcript, and saves it to a CSV file. It includes error handling for invalid URLs and missing transcripts. The script processes the raw transcript data and organizes it into a structured format, with each row containing the text content. The script is as follows:

```
from youtube_transcript_api import YouTubeTranscriptApi
import pandas as pd

def extract_video_id(url):
    if "youtube.com/watch?v=" in url:
        return url.split("v=")[1].split("&")[0]
    elif "youtu.be/" in url:
        return url.split("youtu.be/")[1]
    else:
        raise ValueError("Invalid YouTube URL")


def fetch_transcript(url):
    try:
        video_id = extract_video_id(url)
        transcript = YouTubeTranscriptApi.get_transcript(video_id)
```

```
        return transcript
    except Exception as e:
        print(f"An error occurred: {e}")
        return None



def save_to_csv(transcript, filename):
    if transcript:
        df = pd.DataFrame(transcript)
        df.to_csv(filename, index=False, encoding='utf-8')
        print(f"Saved transcript to {filename}")
    else:
        print("No transcript data to save.")



def main():
    url = input("Please enter the YouTube URL: ")
    transcript = fetch_transcript(url)

    if transcript:
        video_id = extract_video_id(url)
        filename = f"YT_{video_id}_Transcript.csv"
        save_to_csv(transcript, filename)
```

```
    else:
        print("Failed to fetch transcript.")


if __name__ == "__main__":
    main()
```

Appendix B

This Python script automates the collection of YouTube comments using Selenium WebDriver and the YouTube Data API v3. It processes multiple video URLs, handling dynamic page loading to capture comments beyond the initially visible set. For each comment, the script extracts the author, comment text, and like count, converting abbreviated metrics (e.g., "1.5K") to numerical values. The script includes error handling and rate limiting to manage YouTube's access constraints. Data is saved in CSV format with consistent structure across all processed videos. The script is as follows:

```
from selenium import webdriver
from selenium.webdriver.chrome.service import Service
from selenium.webdriver.common.by import By
from selenium.webdriver.common.keys import Keys
from selenium.webdriver.support.ui import WebDriverWait
from selenium.webdriver.support import expected_conditions as EC
import time
import pandas as pd

# Updated list of tuples containing (URL, output_filename)
VIDEOS = [
    ("https://www.youtube.com/watch?v=qie7cfjnLAY",
"YT_M3MacbookPro_DLee_Comments.csv"),
```

```
        ("https://www.youtube.com/watch?v=CREM-mFuyyo",
"YT_iPhone15Pro_DLee_Comments.csv"),
        ("https://www.youtube.com/watch?v=f-nV6DHVgLo",
"YT_M4iPadPro_DLee_Comments.csv"),
    ]
    def convert_likes(like_count):
      if not like_count:
        return 0
      if 'K' in like_count:
        return int(float(like_count.replace('K', '')) * 1000)
      elif 'M' in like_count:
        return int(float(like_count.replace('M', '')) * 1000000)
      else:
        return int(like_count.replace(',', '').strip())


    def fetch_comments(driver, url):
      try:
        print(f"Navigating to {url}")
        driver.get(url)


        print("Waiting for comments to load")
        WebDriverWait(driver, 30).until(
```

```
        EC.presence_of_element_located((By.TAG_NAME, "ytd-comment-thread-
renderer"))
    )
    print("Comments loaded successfully")

    print("Scrolling to load more comments")
    last_height = driver.execute_script("return
document.documentElement.scrollHeight")
    while True:
        driver.execute_script("window.scrollTo(0,
document.documentElement.scrollHeight);")
        time.sleep(2)
        new_height = driver.execute_script("return
document.documentElement.scrollHeight")
        if new_height == last_height:
            break
        last_height = new_height
        print("Scrolled, loading more comments...")

    print("Finished scrolling")

    print("Extracting comments")
```

```
        comment_elements = driver.find_elements(By.TAG_NAME, "ytd-comment-thread-
renderer")
        comments = []
        for i, element in enumerate(comment_elements, 1):
            try:
                author = element.find_element(By.ID, "author-text").text.strip()
                comment_text = element.find_element(By.ID, "content-text").text.strip()
                try:
                    like_count = element.find_element(By.ID, "vote-count-middle").text.strip()
                except:
                    like_count = "0"
                like_count = convert_likes(like_count)

                comment = {
                    'author': author,
                    'text': comment_text,
                    'likes': like_count
                }
                comments.append(comment)
                print(f"Extracted comment {i}/{len(comment_elements)}")
            except Exception as e:
                print(f"Error extracting comment {i}: {e}")
```

```
        print(f"Successfully extracted {len(comments)} comments")
        return comments

    except Exception as e:
        print(f"An error occurred: {e}")
        return []

def save_to_csv(comments, filename):
    df = pd.DataFrame(comments)
    df.to_csv(filename, index=False, encoding='utf-8')
    print(f"Saved {len(comments)} comments to {filename}")

def main():
    service = Service('/Users/floether/Documents/MCM/MCM 740
Capstone/Coding/Chrome Driver/chromedriver')
    options = webdriver.ChromeOptions()
    options.add_argument("--start-maximized")
    driver = webdriver.Chrome(service=service, options=options)

    try:
        for url, filename in VIDEOS:
            print(f"\nStarting comment extraction for {url}")
            comments = fetch_comments(driver, url)
```

```
            save_to_csv(comments, filename)

            print(f"Completed processing for {url}")

        finally:

            print("Closing browser")

            driver.quit()


    print("All videos processed")


if __name__ == "__main__":

    main()
```

Appendix C

This Python script automates the collection of Amazon product reviews using Selenium WebDriver. It manages complex web scraping challenges including CAPTCHA verification and pagination handling. For each review, the script extracts multiple data points: title, rating, review text, username, date, and helpful vote count. The script implements intentional delays and error handling to ensure reliable data collection within Amazon's access constraints. Using BeautifulSoup for HTML parsing, it processes 100 reviews per product, maintaining consistent data structure across all product categories. The script is as follows:

```
from selenium import webdriver
from selenium.webdriver.common.by import By
from selenium.webdriver.chrome.service import Service
from selenium.common.exceptions import NoSuchWindowException,
WebDriverException
from bs4 import BeautifulSoup
import csv
import time
import random
import os

# List of URLs for Amazon product reviews
URLS = [
```

```
    ("https://www.amazon.com/Apple-iPad-Pro-13-Inch-Intelligence/product-reviews/B0D3J98W75", "AZ_M4iPadPro_DLee_Comments"),
]

def fetch_page(driver, url):
    try:
        driver.get(url)

        # Pause for manual interaction (CAPTCHA or login)
        input(f"Press Enter after you have solved the CAPTCHA or logged in for {url}...")

        time.sleep(5)  # Wait for the page to fully load

        if driver.window_handles:
            print(f"Browser window is still open for {url}, proceeding with scraping...")
            return driver.page_source
        else:
            print(f"Browser window was closed or lost focus for {url}. Exiting...")
            return None

    except WebDriverException as e:
        print(f"WebDriver error for {url}: {e}")
        return None
```

```
def extract_reviews(html):
    if not html:
        return []

    soup = BeautifulSoup(html, 'html.parser')
    reviews = []

    review_elements = soup.find_all('div', {'data-hook': 'review'})

    for element in review_elements:
        title_element = element.find('a', {'data-hook': 'review-title'})
        title = title_element.find_all('span')[-1].text.strip() if title_element else "No Title"

        rating_element = element.find('i', {'data-hook': 'review-star-rating'})
        rating = rating_element.find('span', {'class': 'a-icon-alt'}).text.strip() if
rating_element else "No Rating"

        review_text_element = element.find('span', {'data-hook': 'review-body'})
        review_text = review_text_element.text.strip() if review_text_element else "No
Review Text"

        username_element = element.find('span', {'class': 'a-profile-name'})
```

```
        username = username_element.text.strip() if username_element else "Anonymous"

        date_element = element.find('span', {'data-hook': 'review-date'})
        date = date_element.text.strip() if date_element else "No Date"

        helpful_votes_element = element.find('span', {'data-hook': 'helpful-vote-statement'})
        helpful_votes = helpful_votes_element.text.strip() if helpful_votes_element else "0"

        review = {
            'title': title,
            'rating': rating,
            'text': review_text,
            'username': username,
            'date': date,
            'helpful_votes': helpful_votes
        }
        reviews.append(review)

    return reviews

def save_to_csv(reviews, filename):
    with open(filename, 'w', newline='', encoding='utf-8') as file:
```

```
        writer = csv.DictWriter(file, fieldnames=['title', 'rating', 'text', 'username', 'date',
'helpful_votes'])
        writer.writeheader()
        writer.writerows(reviews)


def scrape_reviews(driver, url, csv_name, max_reviews=100):
    all_reviews = []
    page = 1


    while len(all_reviews) < max_reviews:
        if page == 1:
            page_url = url
        else:
            # Check if the URL already has parameters
            if '?' in url:
                page_url = f"{url}&pageNumber={page}"
            else:
                page_url = f"{url}?pageNumber={page}"


        html = fetch_page(driver, page_url)
        if html:
            reviews = extract_reviews(html)
            if not reviews:
```

```
            print(f"No reviews found on page {page} for {csv_name}. Stopping.")
            break
        all_reviews.extend(reviews)
        print(f"Scraped {len(reviews)} reviews from page {page} for {csv_name}")

        time.sleep(random.uniform(1, 3))  # Delay to avoid anti-bot measures

        if len(all_reviews) >= max_reviews:
            break
    else:
        print(f"Failed to fetch page {page} for {csv_name}. Stopping.")
        break

    page += 1

all_reviews = all_reviews[:max_reviews]  # Ensure we only keep the top 100 reviews
return all_reviews

def main():
    os.makedirs("reviews", exist_ok=True)
    os.chdir("reviews")
```

```
    service = Service('/Users/floether/Documents/MCM/MCM 740 Capstone/Coding/Chrome Driver/chromedriver')
    options = webdriver.ChromeOptions()
    options.add_argument("--start-maximized")
    driver = webdriver.Chrome(service=service, options=options)

    # Dictionary to store scraped data for each unique URL
    scraped_data = {}

    try:
        for url, csv_name in URLS:
            try:
                print(f"\nProcessing {csv_name}")
                if url in scraped_data:
                    print(f"Reusing data for {url}")
                    reviews = scraped_data[url]
                else:
                    print(f"Scraping new data for {url}")
                    reviews = scrape_reviews(driver, url, csv_name, 100)
                    scraped_data[url] = reviews

                # Save the reviews to a new CSV file with the current csv_name
                filename = f"{csv_name}.csv"
```

```
            save_to_csv(reviews, filename)

            print(f"Saved {len(reviews)} reviews to '{filename}'")


        except Exception as e:

            print(f"An error occurred while processing {csv_name}: {e}")

        finally:

            time.sleep(5)  # Add a delay between processing each URL

    except Exception as e:

        print(f"A critical error occurred: {e}")

    finally:

        driver.quit()


    print("All products processed")


if __name__ == "__main__":

    main()
```

Appendix D

This Python script automates the cleaning and standardization of data files used in the research project. It processes files from an input folder, retaining relevant columns such as 'text' and 'title', while removing rows with missing values. The script also extracts metadata from filenames, including platform, product, and source type. It applies a consistent cleaning process across YouTube transcripts, comments, and Amazon reviews, ensuring uniformity in the resulting datasets. By saving the cleaned data to new CSV files in a designated output folder, the script maintains data integrity and comparability across various sources. The script is as follows:

```
import os
import pandas as pd

# Set correct paths
input_folder = '/content'
output_folder = '/content/cleaned'

def clean_and_save_file(input_file, output_folder):
    print(f"Processing file: {input_file}")
    df = pd.read_csv(os.path.join(input_folder, input_file))

    columns_to_keep = ['text']
    if 'title' in df.columns:
        columns_to_keep.insert(0, 'title')
```

```
    df = df[columns_to_keep]
    df = df.dropna(subset=['text'])

    # Extract platform, product, and source from filename
    parts = input_file.split('_')
    platform = parts[0]
    product = parts[1]
    source = parts[3].split('.')[0]

    # Create generic output filename
    output_filename = f"{platform}_{product}_{source}_cleaned.csv"
    output_path = os.path.join(output_folder, output_filename)
    df.to_csv(output_path, index=False)
    print(f"Cleaned and saved: {output_path}")

def process_influencer_files(influencer, products):
    print(f"Processing files for influencer: {influencer}")

    os.makedirs(output_folder, exist_ok=True)

    content_files = os.listdir(input_folder)
```

```
        influencer_files = [f for f in content_files if influencer in f and any(product in f for product in products)]

        print(f"Files found: {len(influencer_files)}")
        for file in influencer_files:
            clean_and_save_file(file, output_folder)

    influencers = {
        "Unbox": ["GooglePixel8A", "Insta360X4", "NothingPhone2A"],
        "MB": ["OnePlusWatch", "M4iPadPro", "M3MacbookAir"],
        "Boss": ["SamsungS24Ultra", "NothingPhone2A", "M4iPadPro"],
        "iJustine": ["M4iPadPro", "iPhone15Pro", "Insta360X4"],
        "AE": ["MacbookProM3", "iMacM3", "HPLaptop"],
        "AvgCons": ["SamsungS24Ultra", "M4iPadPro", "M3MacbookAir"],
        "TechChap": ["NothingPhone2A", "M4iPadPro", "Insta360GO3S"],
        "MrMobile": ["SamsungS24Ultra", "NothingPhone2A", "GooglePixel8A"],
        "TechSpurt": ["SamsungS24", "NothingPhone2A", "GooglePixel8A"],
        "Ella": ["SamsungS24Ultra", "M4iPadPro", "SamsungS24"],
    }

    process_influencer_files("Unbox", influencers["Unbox"])

    print("\nContents of the 'cleaned' folder:")
```

```
cleaned_files = os.listdir(output_folder)
for file in cleaned_files:
    print(f"  {file}")
```

Appendix E

This Python script in performs an analysis of trust indicators in YouTube comments on tech influencer product reviews. The script identifies and categorizes trust-related keywords based on Grunig and Hon's (1999) trust components: integrity, dependability, and competence. It also incorporates inverse trust keywords to capture negative sentiment. The analysis filters comments to focus on those directly referencing the influencer or review, ensuring relevance. The script is as follows:

```
import pandas as pd
import nltk
import os
import glob

# Download necessary NLTK resources
nltk.download('punkt')

# Define input and output folders
input_folder = '/content/cleaned'  # Adjust this path as necessary
output_folder = '/content/RQ1'

# Create the output folder if it doesn't exist
if not os.path.exists(output_folder):
```

```
        os.makedirs(output_folder)

    # Expanded Trust Component Keywords (same as before)
    trust_keywords = {
        'integrity': [
            'honest', 'truth', 'truthful', 'transparent', 'transparency', 'sponsor', 'sponsorship',
'authentic',
            'authenticity', 'disclosure', 'integrity', 'ethical', 'morals', 'trust', 'trustworthiness',
'sincere',
            'sincerity', 'candid', 'frank', 'upfront', 'genuine', 'reliable', 'straightforward', 'credible',
'dependable',
            'fair', 'loyal', 'open', 'consistent', 'principled', 'just', 'honorable', 'upright', 'fair-minded',
'veracity'
        ],
        'dependability': [
            'consistent', 'consistency', 'reliable', 'reliability', 'trusted', 'trust', 'confidence', 'follow-
through',
            'dependable', 'commitment', 'faithful', 'steadfast', 'trustworthy', 'predictable',
'punctual', 'timely',
            'regular', 'steady', 'unwavering', 'conscientious', 'loyal', 'repeat', 'routine',
            'persistent', 'dedicated', 'assurance', 'security', 'fidelity', 'stability', 'devoted', 'stable',
'diligent',
            'accurate', 'faith', 'constant'
```

```
        ],
        'competence': [
            'expert', 'expertise', 'knowledgeable', 'knowledge', 'detail', 'detailed', 'thorough',
'thoroughness',
            'explain', 'explanation', 'comprehensive', 'depth', 'in-depth', 'proficiency', 'skillful',
'adept',
            'capable', 'competent', 'mastery', 'proficient', 'accurate', 'accuracy', 'insight',
'insightful',
            'qualified', 'professional', 'articulate', 'clarity',
            'intelligent', 'savvy', 'experienced', 'technical', 'innovative', 'analytical', 'critical
thinker',
            'clever', 'apt', 'gifted', 'resourceful', 'proficient'
        ]
    }

    # Inverse (negative) trust keywords for each trust component
    inverse_trust_keywords = {
        'integrity': [
            'dishonest', 'deceitful', 'untrustworthy', 'unreliable', 'inconsistent', 'fraudulent',
'corrupt', 'misleading',
            'unethical', 'shady', 'untruthful', 'opaque', 'hypocritical', 'suspicious', 'secretive', 'liar',
'lying',
            'false', 'insincere', 'fake', 'paid', 'fraud'
```

```
    ],
    'dependability': [
        'unreliable', 'inconsistent', 'undependable', 'untrustworthy', 'unpredictable',
'unfaithful', 'disloyal',
        'erratic', 'irresponsible', 'fickle', 'unsteady', 'capricious', 'unstable', 'half-hearted',
        'careless', 'irregular', 'forgetful', 'inconstant', 'lazy'
    ],
    'competence': [
        'incompetent', 'unskilled', 'ignorant', 'unqualified', 'unprofessional', 'amateur',
'clueless', 'careless',
        'inept', 'inaccurate', 'sloppy', 'inefficient', 'unfit', 'untrained', 'ineffective', 'incoherent',
        'unfocused', 'weak', 'disorganized'
    ]
}

# Expanded Influencer and Review-related Keywords
influencer_keywords = {
    'reviewer', 'vlogger', 'channel', 'influencer', 'host', 'he', 'she', 'they', 'his', 'her', 'content
creator',
    'youtuber', 'critic', 'blogger', 'video', 'presentation', 'podcaster', 'commentator', 'media'
}

# Keywords referring to the review itself
```

```
review_keywords = {
    'this review', 'review', 'opinion', 'analysis', 'feedback', 'critique', 'perspective', 'summary',
'assessment',
    'evaluation', 'breakdown', 'overview', 'reflection', 'impression', 'take', 'thoughts',
'judgment', 'discussion'
}

# Combine influencer and review-related keywords
influencer_review_keywords = influencer_keywords | review_keywords

# Function to check if a comment refers to the influencer or the review
def refers_to_influencer_or_review(text):
    words = nltk.word_tokenize(text.lower())
    return any(word in influencer_review_keywords for word in words)

# Function to calculate Net Trust Score without volume adjustments
def calculate_trust_score(data):
    positive_counts = {'integrity': 0, 'dependability': 0, 'competence': 0}
    negative_counts = {'integrity': 0, 'dependability': 0, 'competence': 0}

    for _, row in data.iterrows():
        text = row['text']
```

```
        # Check if the comment refers to the influencer or the review
        if refers_to_influencer_or_review(text):
            tokens = nltk.word_tokenize(text.lower())

            for component in trust_keywords:
                # Count positive trust indicators
                count = sum(1 for word in tokens if word in trust_keywords[component])
                positive_counts[component] += count

                # Count negative (inverse) trust indicators
                inverse_count = sum(1 for word in tokens if word in
inverse_trust_keywords[component])
                negative_counts[component] += inverse_count
        else:
            continue  # Skip comments that do not refer to the influencer or review

    # Sum total positive and negative counts across all components
    total_positive = sum(positive_counts.values())
    total_negative = sum(negative_counts.values())
    total_keywords = total_positive + total_negative

    if total_keywords > 0:
        net_trust_score = (total_positive - total_negative) / total_keywords
```

```
    else:
        net_trust_score = 0  # No trust-related keywords found

    # Calculate Net Keyword Scores for each component
    net_keyword_scores = {}
    for component in trust_keywords:
        positive = positive_counts[component]
        negative = negative_counts[component]
        total = positive + negative
        if total > 0:
            net_score = (positive - negative) / total
        else:
            net_score = 0
        net_keyword_scores[component] = net_score

    return net_trust_score, net_keyword_scores, positive_counts, negative_counts

# Function to extract product name from filename
def extract_product_name(filename):
    # Assuming the filename format is: platform_product_sourcetype_cleaned.csv
    base_name = os.path.basename(filename)
    base_name = os.path.splitext(base_name)[0]  # Remove the file extension
    parts = base_name.split('_')
```

```
    if len(parts) >= 3:
        product_name = parts[1]
        return product_name
    else:
        return 'Unknown Product'

# Get all files in the input folder that contain 'Comments' in their name
file_pattern = os.path.join(input_folder, '*_Comments_cleaned.csv')
comment_files = glob.glob(file_pattern)

# Check if any comment files were found
if not comment_files:
    print("No comment files found in the input folder.")
else:
    # Initialize list to store results
    results = []

    # Process each comment file
    for file_path in comment_files:
        # Extract product name from filename
        product_name = extract_product_name(file_path)

        # Read the comments data
```

```
        try:
            comments_data = pd.read_csv(file_path)
        except Exception as e:
            print(f"Error reading {file_path}: {e}")
            continue

        # Check if 'text' column exists
        if 'text' not in comments_data.columns:
            print(f"Column 'text' not found in {file_path}. Skipping this file.")
            continue
        # Calculate Net Trust Score
        net_trust_score, net_keyword_scores, positive_counts, negative_counts =
calculate_trust_score(comments_data)

        result = {
            'Product': product_name,
            'Net Trust Score': net_trust_score,
            'Total Positive Keywords': sum(positive_counts.values()),
            'Total Negative Keywords': sum(negative_counts.values()),
            'Integrity Net Keyword Score': net_keyword_scores['integrity'],
            'Integrity Positive Keywords': positive_counts['integrity'],
            'Integrity Negative Keywords': negative_counts['integrity'],
            'Dependability Net Keyword Score': net_keyword_scores['dependability'],
```

```
        'Dependability Positive Keywords': positive_counts['dependability'],
        'Dependability Negative Keywords': negative_counts['dependability'],
        'Competence Net Keyword Score': net_keyword_scores['competence'],
        'Competence Positive Keywords': positive_counts['competence'],
        'Competence Negative Keywords': negative_counts['competence']
    }

    # Add the result to the results list
    results.append(result)

    # Save the result to CSV in the output folder
    result_df = pd.DataFrame([result])
    output_file = os.path.join(output_folder, f"{product_name}_Trust_Score.csv")
    result_df.to_csv(output_file, index=False)
    print(f"Results saved to {output_file}")

# Optionally, save all results to a single CSV file
all_results_df = pd.DataFrame(results)
all_results_file = os.path.join(output_folder, "All_Products_Trust_Scores.csv")
all_results_df.to_csv(all_results_file, index=False)
print(f"\nAll results saved to {all_results_file}")

# Print the Net Trust Scores for each product
```

```
print("\nNet Trust Scores for each product:")
for res in results:
    print(f"{res['Product']}:")
    print(f"  Net Trust Score: {res['Net Trust Score']}")
    print(f"  Total Positive Keywords: {res['Total Positive Keywords']}")
    print(f"  Total Negative Keywords: {res['Total Negative Keywords']}\n")
```

Appendix F

This Python script calculates similarity scores between the content of tech influencers' product reviews and corresponding Amazon consumer reviews. It employs text preprocessing techniques, including tokenization, stopword removal, and stemming, to normalize the text data. The script utilizes the sklearn library to perform TF-IDF vectorization and cosine similarity calculations, quantifying the degree of content alignment between influencer reviews and consumer opinions. It processes the text data using both stemmed and unstemmed approaches, providing a comprehensive comparison of textual similarities. The resulting similarity scores are saved to CSV files for each product. The script is as follows:

```
import pandas as pd
import nltk
import os
import string
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity

# Download necessary NLTK data
nltk.download('punkt')
nltk.download('stopwords')

# Define custom stopwords
```

```
file_stopwords = {"nothingphone2a", "insta360x4", "googlepixel8a", "unbox", "reviews",
"comments", "transcript", "n't", "also"}
custom_stopwords = {"'s", "phone", "camera", "laptop", "computer", "tablet"} |
file_stopwords

# Preprocess the text (tokenization, stopword removal, stemming)
def preprocess_text(text):
    if isinstance(text, str):
        words = nltk.word_tokenize(text.lower())
        words = [word for word in words if word not in string.punctuation]
        stop_words = set(nltk.corpus.stopwords.words('english'))
        words = [word for word in words if word not in stop_words and word not in
custom_stopwords]
        stemmer = nltk.PorterStemmer()
        words = [stemmer.stem(word) for word in words]
        return ' '.join(words)
    else:
        return ''

# Preprocess the text without stemming
def preprocess_text_no_stem(text):
    if isinstance(text, str):
        words = nltk.word_tokenize(text.lower())
```

```
        words = [word for word in words if word not in string.punctuation]
        stop_words = set(nltk.corpus.stopwords.words('english'))
        words = [word for word in words if word not in stop_words and word not in
custom_stopwords]
        return ' '.join(words)
    else:
        return ''

# Function to preprocess transcripts
def preprocess_transcript(df):
    full_text = df.apply(lambda row: ' '.join(row.dropna().astype(str)), axis=1)
    return full_text.apply(preprocess_text)

# Function to preprocess transcripts without stemming
def preprocess_transcript_no_stem(df):
    full_text = df.apply(lambda row: ' '.join(row.dropna().astype(str)), axis=1)
    return full_text.apply(preprocess_text_no_stem)

# Load the cleaned transcripts and reviews data
transcript_nothing_phone =
pd.read_csv('/content/cleaned/YT_NothingPhone2A_Transcript_Cleaned.csv', header=None)
transcript_insta360 =
pd.read_csv('/content/cleaned/YT_Insta360X4_Transcript_Cleaned.csv', header=None)
```

```
transcript_pixel =
pd.read_csv('/content/cleaned/YT_GooglePixel8A_Transcript_Cleaned.csv', header=None)

amazon_reviews_nothing_phone =
pd.read_csv('/content/cleaned/Amazon_NothingPhone2a_Reviews_Cleaned.csv')
amazon_reviews_insta360 =
pd.read_csv('/content/cleaned/Amazon_Insta360X4_Reviews_Cleaned.csv')
amazon_reviews_pixel =
pd.read_csv('/content/cleaned/Amazon_GooglePixel8A_Reviews_Cleaned.csv')

# Preprocess the text in transcripts and reviews (with stemming)
transcript_nothing_phone['processed_text'] =
preprocess_transcript(transcript_nothing_phone)
transcript_insta360['processed_text'] = preprocess_transcript(transcript_insta360)
transcript_pixel['processed_text'] = preprocess_transcript(transcript_pixel)

amazon_reviews_nothing_phone['processed_text'] =
amazon_reviews_nothing_phone['text'].apply(preprocess_text)
amazon_reviews_insta360['processed_text'] =
amazon_reviews_insta360['text'].apply(preprocess_text)
amazon_reviews_pixel['processed_text'] =
amazon_reviews_pixel['text'].apply(preprocess_text)
```

```
# Preprocess the text in transcripts and reviews (without stemming)
transcript_nothing_phone['processed_text_no_stem'] =
preprocess_transcript_no_stem(transcript_nothing_phone)
transcript_insta360['processed_text_no_stem'] =
preprocess_transcript_no_stem(transcript_insta360)
transcript_pixel['processed_text_no_stem'] =
preprocess_transcript_no_stem(transcript_pixel)

amazon_reviews_nothing_phone['processed_text_no_stem'] =
amazon_reviews_nothing_phone['text'].apply(preprocess_text_no_stem)
amazon_reviews_insta360['processed_text_no_stem'] =
amazon_reviews_insta360['text'].apply(preprocess_text_no_stem)
amazon_reviews_pixel['processed_text_no_stem'] =
amazon_reviews_pixel['text'].apply(preprocess_text_no_stem)

# Function to calculate cosine similarity between transcripts and reviews
def calculate_similarity(transcripts, reviews, use_stem=True):
    # Choose the appropriate processed text column
    text_column = 'processed_text' if use_stem else 'processed_text_no_stem'

    # Combine all the processed text into a single document for comparison
    transcript_text = ' '.join(transcripts[text_column].tolist())
    review_text = ' '.join(reviews[text_column].tolist())
```

```
    # Vectorize the text using TF-IDF
    vectorizer = TfidfVectorizer()
    vectors = vectorizer.fit_transform([transcript_text, review_text])

    # Calculate cosine similarity
    similarity_matrix = cosine_similarity(vectors)
    similarity_score = similarity_matrix[0, 1]  # This is the similarity between the
transcript and reviews

    return similarity_score

# Create the RQ2 folder if it doesn't exist
output_folder = '/content/RQ2'
if not os.path.exists(output_folder):
    os.makedirs(output_folder)

# Calculate similarity for each product
results = []

for product_name, transcripts, reviews in zip(
    ["Nothing Phone 2(a)", "Insta360 X4", "Google Pixel 8A"],
    [transcript_nothing_phone, transcript_insta360, transcript_pixel],
```

```
    [amazon_reviews_nothing_phone, amazon_reviews_insta360, amazon_reviews_pixel]
):
    similarity_score_stem = calculate_similarity(transcripts, reviews, use_stem=True)
    similarity_score_no_stem = calculate_similarity(transcripts, reviews, use_stem=False)

    result = {
        'Product': product_name,
        'Similarity Score (Stemmed)': similarity_score_stem,
        'Similarity Score (No Stem)': similarity_score_no_stem,
    }

    results.append(result)

# Results DataFrame
results_df = pd.DataFrame(results)

# Save the results to the RQ2 folder
results_df.to_csv(f'{output_folder}/Unbox_Similarity_Scores.csv', index=False)

# Display the results
print(f"Results saved to {output_folder}/Unbox_Similarity_Scores.csv")
print(results_df.to_string(index=False))
```

Appendix G

This Python script in Appendix G enhances the similarity score calculation process between tech influencer product reviews and Amazon consumer reviews by incorporating lemmatization, a more advanced text normalization technique compared to stemming. It utilizes the NLTK library for text preprocessing, including tokenization, stopword removal, and lemmatization, to effectively normalize the text data while preserving the base or dictionary form of words. The script processes a comprehensive list of products, automatically detecting and loading the corresponding YouTube transcripts and Amazon reviews from a designated data folder. It employs the TF-IDF vectorization and cosine similarity calculation from the sklearn library to quantify the content alignment between influencer reviews and consumer opinions. The resulting similarity scores for each product are displayed in the console output. The script is as follows:

```
import os
import pandas as pd
import numpy as np
import re
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity
import nltk
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
```

```
# Ensure NLTK resources are available
nltk.download('stopwords')
nltk.download('punkt')
nltk.download('wordnet')

# Define the path to the cleaned data
data_path = '/content/cleaned'

# List of products
products = [
    "GooglePixel8A", "Insta360X4", "NothingPhone2A", "OnePlusWatch",
    "M4iPadPro", "M3MacbookAir", "SamsungS24Ultra", "iPhone15Pro",
    "MacbookProM3", "iMacM3", "HPLaptop", "Insta360GO3S", "SamsungS24",
    "SamsungS24Ultra", "M3MacbookAir"
]

# Keep track of files found and processed
files_found = 0
files_processed = 0

# Initialize stop words and lemmatizer
stop_words = set(stopwords.words('english'))
lemmatizer = WordNetLemmatizer()
```

```
def preprocess_text(text):
    # Lowercase
    text = text.lower()
    # Remove special characters and emojis
    text = re.sub(r'[^a-zA-Z\s]', '', text)
    # Tokenize
    tokens = nltk.word_tokenize(text)
    # Remove stop words and lemmatize
    tokens = [lemmatizer.lemmatize(word) for word in tokens if word not in stop_words]
    # Rejoin
    text = ' '.join(tokens)
    return text

for product in products:
    yt_filename = f'YT_{product}_Transcript_cleaned.csv'
    az_filename = f'AZ_{product}_Reviews_cleaned.csv'

    yt_filepath = os.path.join(data_path, yt_filename)
    az_filepath = os.path.join(data_path, az_filename)

    if os.path.isfile(yt_filepath) and os.path.isfile(az_filepath):
        files_found += 2
```

```
    # Read YT transcript
    yt_df = pd.read_csv(yt_filepath)
    yt_df['text'] = yt_df['text'].astype(str)
    yt_text = yt_df['text'].apply(preprocess_text).str.cat(sep=' ')

    # Read AZ reviews
    az_df = pd.read_csv(az_filepath)
    az_df['title'] = az_df['title'].fillna('').astype(str)
    az_df['text'] = az_df['text'].fillna('').astype(str)
    az_df['combined_text'] = az_df['title'] + ' ' + az_df['text']
    az_text = az_df['combined_text'].apply(preprocess_text).str.cat(sep=' ')

    # Vectorize texts
    vectorizer = TfidfVectorizer()
    tfidf_matrix = vectorizer.fit_transform([yt_text, az_text])

    # Compute cosine similarity
    similarity = cosine_similarity(tfidf_matrix[0:1], tfidf_matrix[1:2])[0][0]

    print(f"Similarity score for {product}: {similarity:.4f}")
    files_processed += 2
else:
```

```
        if not os.path.isfile(yt_filepath):
            print(f"File not found: {yt_filepath}")
        if not os.path.isfile(az_filepath):
            print(f"File not found: {az_filepath}")


print(f"\nFiles found: {files_found}")
print(f"Files processed: {files_processed}")
```

Appendix H

This Python script automates the cleaning and formatting of news articles related to tech influencers. It processes CSV files containing news articles from a designated input folder, automatically detecting the file encoding to handle various text formats. The script identifies the relevant text and title columns, extracts and cleans the specified data, and saves the cleaned articles in a new CSV file for each influencer in a designated output folder. It also renames the columns for consistency across all files. The script is as follows:

```
import os
import pandas as pd
import re
import chardet

# Set correct paths
input_folder = '/content'
output_folder = '/content/cleaned'

def detect_encoding(file_path):
    with open(file_path, 'rb') as f:
        result = chardet.detect(f.read(100000))
    return result['encoding']

def clean_and_save_news_file(influencer):
```

```
        input_file = f"News_{influencer}.csv"
        input_path = os.path.join(input_folder, input_file)

        if not os.path.exists(input_path):
            print(f"News file not found for influencer: {influencer}")
            return

        print(f"\nProcessing news file for influencer: {influencer}")

        # Detect file encoding
        encoding = detect_encoding(input_path)
        print(f"Detected encoding for {input_file}: {encoding}")

        try:
            df = pd.read_csv(input_path, encoding=encoding)
        except Exception as e:
            print(f"An error occurred while reading {input_file} with encoding {encoding}:
{e}")
            return

        # Possible columns that contain text data
        text_column_candidates = ['text', 'content', 'body']
        text_column = None
```

```
# Find the text column
for col in text_column_candidates:
    if col in df.columns:
        text_column = col
        break

if not text_column:
    print(f"No text column found in {input_file}. Skipping this file.")
    return

columns_to_keep = [text_column]

# Optionally include 'title' or 'headline' column
if 'title' in df.columns:
    columns_to_keep.insert(0, 'title')
elif 'headline' in df.columns:
    columns_to_keep.insert(0, 'headline')

df = df[columns_to_keep]
df = df.dropna(subset=[text_column])

# Rename columns for consistency
```

```
    df.rename(columns={text_column: 'text', 'headline': 'title'}, inplace=True)

    # Create output filename
    output_filename = f"News_{influencer}_cleaned.csv"
    output_path = os.path.join(output_folder, output_filename)
    df.to_csv(output_path, index=False)
    print(f"Cleaned and saved: {output_path}")

# Create the output folder if it doesn't exist
os.makedirs(output_folder, exist_ok=True)

# Automatically detect all 'News_*.csv' files in the input_folder
all_files = os.listdir(input_folder)
news_files = [f for f in all_files if f.startswith('News_') and f.endswith('.csv')]

# Extract influencer names from filenames
influencers = []
for file_name in news_files:
    match = re.match(r'News_(.*)\.csv', file_name)
    if match:
        influencer = match.group(1)
        influencers.append(influencer)
```

```
# Remove duplicates in case there are any
influencers = list(set(influencers))

# Clean and save the news files for each influencer
for influencer in influencers:
    clean_and_save_news_file(influencer)

print("\nContents of the 'cleaned' folder:")
cleaned_files = os.listdir(output_folder)
for file in cleaned_files:
    print(f"  {file}")
```

Appendix I

This Python script performs an analysis of media coverage related to tech influencers. It uses the Hugging Face transformers library, specifically the 'nlptown/bert-base-multilingual-uncased-sentiment' model, to conduct sentiment analysis on the news articles. The script incorporates a dictionary that maps influencer handles to their full names, ensuring accurate identification across various naming conventions. It also includes trust component keywords and thematic categories with associated keywords to assess the presence and prevalence of trust-related and thematic elements in the articles. The script processes each article, performing sentiment analysis, trust component analysis, thematic categorization, quotation analysis, and article prominence evaluation. It calculates average scores across these dimensions for each influencer, providing a comprehensive assessment of media portrayal. The script saves the analysis results for each influencer in separate CSV files and generates a combined results file for all influencers. The script is as follows:

```
import pandas as pd
import nltk
from nltk.sentiment import SentimentIntensityAnalyzer
import os
import re

# Download necessary NLTK resources
nltk.download('vader_lexicon', quiet=True)
nltk.download('punkt', quiet=True)
```

```
# Initialize VADER sentiment analyzer
sia = SentimentIntensityAnalyzer()

# Dictionary to map influencer handles to full names
influencer_names = {
    "Unbox": ["Unbox Therapy", "Lewis Hilsenteger"],
    "MB": ["MKBHD", "Marques Brownlee"],
    "Boss": ["Mrwhosetheboss", "Arun Maini"],
    "iJustine": ["iJustine", "Justine Ezarik"],
    "AE": ["AustinEvans", "Austin Evans"],
    "AvgCons": ["UrAvgConsumer", "Judner Aura"],
    "TechChap": ["The Tech Chap", "Tom Honeyands"],
    "MrMobile": ["Mr. Mobile", "Michael Fisher"],
    "TechSpurt": ["Tech Spurt", "Chris Barraclough"],
    "Ella": ["Created by Ella", "Ella"]
}

# Trust component keywords (unchanged)
trust_keywords = {
    'integrity': [
        'honest', 'truth', 'truthful', 'transparent', 'transparency', 'sponsor', 'sponsorship',
'authentic',
```

```
        'authenticity', 'disclosure', 'integrity', 'ethical', 'morals', 'trust', 'trustworthiness', 'sincere',
        'sincerity', 'candid', 'frank', 'upfront', 'genuine', 'reliable', 'straightforward', 'credible', 'dependable',
        'fair', 'loyal', 'open', 'consistent', 'principled', 'just', 'honorable', 'upright', 'fair-minded', 'veracity'
    ],
    'dependability': [
        'consistent', 'consistency', 'reliable', 'reliability', 'trusted', 'trust', 'confidence', 'follow-through',
        'dependable', 'commitment', 'faithful', 'steadfast', 'trustworthy', 'predictable', 'punctual', 'timely',
        'regular', 'steady', 'unwavering', 'conscientious', 'loyal', 'repeat', 'routine',
        'persistent', 'dedicated', 'assurance', 'security', 'fidelity', 'stability', 'devoted', 'stable', 'diligent',
        'accurate', 'faith', 'constant'
    ],
    'competence': [
        'expert', 'expertise', 'knowledgeable', 'knowledge', 'detail', 'detailed', 'thorough', 'thoroughness',
        'explain', 'explanation', 'comprehensive', 'depth', 'in-depth', 'proficiency', 'skillful', 'adept',
```

```
                'capable', 'competent', 'mastery', 'proficient', 'accurate', 'accuracy', 'insight',
'insightful',
                'qualified', 'professional', 'articulate', 'clarity',
                'intelligent', 'savvy', 'experienced', 'technical', 'innovative', 'analytical', 'critical
thinker',
                'clever', 'apt', 'gifted', 'resourceful', 'proficient'
            ]
        }

        # Thematic categories with keywords (unchanged)
        thematic_categories = {
            'Product Reviews': [
                'review', 'hands-on', 'unboxing', 'first impressions', 'comparison', 'versus', 'pros and
cons',
                'performance', 'features', 'specifications'
            ],
            'Industry Insights': [
                'market trends', 'innovation', 'future tech', 'industry analysis', 'predictions', 'emerging
technologies',
                'competitor analysis', 'market share', 'product strategy', 'tech landscape'
            ],
            'Controversies': [
                'controversy', 'scandal', 'criticism', 'backlash', 'debate', 'issue', 'problem', 'dispute',
```

```
        'conflict', 'allegations'
    ],
    'Sponsorships/Partnerships': [
        'sponsored', 'partnership', 'collaboration', 'brand deal', 'affiliate', 'endorsement',
'advertisement',
        'promoted', 'sponsor', 'paid promotion'
    ],
    'Personal Life': [
        'personal', 'lifestyle', 'family', 'hobbies', 'background', 'education', 'childhood',
'relationships',
        'off-camera', 'private life'
    ],
    'Impact on Consumer Behavior': [
        'influence', 'buying decisions', 'consumer trends', 'purchasing habits', 'product
popularity',
        'market impact', 'consumer opinion', 'sales influence', 'recommendation impact',
'consumer trust'
    ],
    'Platform Specific': [
        'YouTube', 'Instagram', 'TikTok', 'Twitter', 'Twitch', 'algorithm', 'content strategy',
'channel growth',
        'viewer engagement', 'platform features'
    ],
```

```
    'Legal/Ethical Issues': [
        'lawsuit', 'legal action', 'ethics', 'disclosure', 'FTC guidelines', 'copyright', 'fair use',
        'privacy concerns', 'terms of service', 'community guidelines'
    ]
}

def extract_influencer_handle(filename):
    return filename.split('_')[1].split('.')[0]

def categorize_article(text, title, influencer_handle, influencer_full_names):
    influencer_identifier, influencer_full_name = influencer_full_names

    # Overall sentiment
    sentiment_scores = sia.polarity_scores(text)
    if sentiment_scores['compound'] > 0.05:
        sentiment = 1  # Positive
    elif sentiment_scores['compound'] < -0.05:
        sentiment = -1  # Negative
    else:
        sentiment = 0  # Neutral

    # Trust components
    trust_scores = {component: 0 for component in trust_keywords}
```

```
for component, keywords in trust_keywords.items():
    for keyword in keywords:
        if re.search(r'\b' + re.escape(keyword) + r'\b', text.lower()):
            trust_scores[component] += 1

# Normalize trust scores
for component in trust_scores:
    trust_scores[component] = min(trust_scores[component], 3)

# Thematic categories
themes = {theme: 0 for theme in thematic_categories.keys()}
combined_text = (text + ' ' + title).lower()
for theme, keywords in thematic_categories.items():
    if any(keyword.lower() in combined_text for keyword in keywords):
        themes[theme] = 1

# Quotation analysis
quotation_pattern = r'"([^"]*)"|“([^”]*)”'
quotes = re.findall(quotation_pattern, text)
quotation_presence = 0
for quote in quotes:
    quote = ' '.join(quote).strip()  # Join tuple elements and strip whitespace
    text_before_quote = text.lower()[:text.lower().index(quote.lower())]
```

```
        if quote and (influencer_handle.lower() in text_before_quote or
                      influencer_identifier.lower() in text_before_quote or
                      influencer_full_name.lower() in text_before_quote or
                      any(verb in text_before_quote
                          for verb in ['said', 'stated', 'mentioned', 'noted', 'explained',
'commented'])):
            quotation_presence = 1
            break

    # Article prominence
    word_count = len(text.split())
    influencer_mentions = (text.lower().count(influencer_handle.lower()) +
                           text.lower().count(influencer_identifier.lower()) +
                           text.lower().count(influencer_full_name.lower()))
    title_mention = 1 if (influencer_handle.lower() in title.lower() or
                          influencer_identifier.lower() in title.lower() or
                          influencer_full_name.lower() in title.lower()) else 0
    first_100_words = ' '.join(text.split()[:100]).lower()
    early_mention = 1 if (influencer_handle.lower() in first_100_words or
                          influencer_identifier.lower() in first_100_words or
                          influencer_full_name.lower() in first_100_words) else 0

    if title_mention and early_mention and influencer_mentions > 3:
```

```
        prominence = 2  # Primary focus
    elif (title_mention or early_mention) and influencer_mentions > 1:
        prominence = 1  # Secondary focus
    else:
        prominence = 0  # Mentioned in passing

    return {
        'Sentiment': sentiment,
        'Trust_Integrity': trust_scores['integrity'],
        'Trust_Dependability': trust_scores['dependability'],
        'Trust_Competence': trust_scores['competence'],
        'Quotation_Presence': quotation_presence,
        'Article_Prominence': prominence,
        **themes
    }

def process_file(file_path, influencer_handle, influencer_full_names):
    try:
        # Attempt to read the CSV file with default utf-8 encoding
        df = pd.read_csv(file_path, encoding='utf-8')
    except UnicodeDecodeError:
        # If utf-8 fails, fallback to 'latin1' or 'ISO-8859-1'
        df = pd.read_csv(file_path, encoding='ISO-8859-1')
```

```
    # Fill NaN values in text and title columns with empty strings
    df['text'] = df['text'].fillna("")
    df['title'] = df['title'].fillna("")

    results = []

    for _, row in df.iterrows():
        text = str(row['text'])  # Convert text to string in case of any non-string types
        title = str(row['title'])  # Convert title to string in case of any non-string types
        result = categorize_article(text, title, influencer_handle, influencer_full_names)
        results.append(result)

    return pd.DataFrame(results)

def calculate_average_scores(df):
    # Calculate the average sentiment and trust scores per article
    num_articles = len(df)

    if num_articles == 0:
        return {
            'Average_Sentiment': 0,
            'Average_Trust_Integrity': 0,
```

```
        'Average_Trust_Dependability': 0,
        'Average_Trust_Competence': 0,
        'Quotation_Presence': 0,
        'Article_Prominence': 0,
        'Total_Articles_Analyzed': 0,
        **{f'Average_{theme}': 0 for theme in thematic_categories.keys()}
    }

average_scores = {
    'Average_Sentiment': df['Sentiment'].mean(),
    'Average_Trust_Integrity': df['Trust_Integrity'].mean(),
    'Average_Trust_Dependability': df['Trust_Dependability'].mean(),
    'Average_Trust_Competence': df['Trust_Competence'].mean(),
    'Quotation_Presence': df['Quotation_Presence'].mean(),
    'Article_Prominence': df['Article_Prominence'].mean(),
    'Total_Articles_Analyzed': num_articles
}

for theme in thematic_categories.keys():
    average_scores[f'Average_{theme}'] = df[theme].mean()

return average_scores
```

```
def main():
    input_folder = '.'  # Current directory
    output_folder = './output'
    os.makedirs(output_folder, exist_ok=True)

    combined_results = []

    for file in os.listdir(input_folder):
        if file.startswith('News_') and file.endswith('.csv'):
            input_path = os.path.join(input_folder, file)
            output_path = os.path.join(output_folder, f'Analyzed_{file}')

            influencer_handle = extract_influencer_handle(file)
            influencer_full_names = influencer_names.get(influencer_handle,
[influencer_handle, influencer_handle])

            results_df = process_file(input_path, influencer_handle, influencer_full_names)
            results_df.to_csv(output_path, index=False)
            print(f"Processed {file} and saved results to {output_path}")

            # Calculate and store average scores
            average_scores = calculate_average_scores(results_df)
            average_scores['Influencer'] = influencer_full_names[1]
```

```
            combined_results.append(average_scores)


            print(f"\nAverage Scores for {file}:")
            for category, score in average_scores.items():
                print(f"{category}: {score}")
            print("\n" + "-" * 50 + "\n")


        # Save combined results for all influencers
        combined_df = pd.DataFrame(combined_results)
        combined_output_path = os.path.join(output_folder,
'Combined_Analysis_Results.csv')
        combined_df.to_csv(combined_output_path, index=False)
        print(f"Combined analysis results saved to {combined_output_path}")


    if __name__ == "__main__":
        main()
```