

The Impact of AI-Generated Content on Trust and Authenticity in Social Media Communication: A Study of Three Australian Case Studies

Zhaoyang Song

Media and Communication, Monash University, Melbourne, Australia
zson0043@student.monash.edu

Abstract. The robust growth of generative artificial intelligence (AI), especially large language models and image-generation systems, has greatly shaped communication practices on social media. In this context, AI-generated content has become widely adopted by businesses, media organizations, influencers, and individual users to efficiently create text, images, videos, and interactive content. While these technologies boost creativity and communication productivity, they also introduce ethical concerns regarding transparency, authenticity, misinformation, and public trust. This study investigates how AI-generated content influences trust and authenticity in social media communication through a qualitative interpretive case study approach. Three representative cases, the Georgie Purcell AI image controversy, the AI-generated "Aboriginal Steve Irwin" incident, and the Jeanswest AI-generated advertising backlash, are investigated using secondary sources including academic literature, news reports, organizational statements, and public responses. Based on this analytical framework, AI ethics principles of transparency, accountability, fairness, and human oversight are applied to guide the analysis. The findings suggest that generative AI has reshaped social media communication by improving efficiency and personalization while disrupting authenticity assessment. Therefore, public trust relies on transparent AI disclosure, ethical governance, and responsible practices. Moreover, sustainable AI integration requires stronger governance, platform transparency, and human oversight to maintain communication credibility.

Keywords: Generative AI, Social Media Communication, Trust, Authenticity, AI Ethics

1. Introduction

Generative artificial intelligence (AI) has become an important part of digital communication. Unlike earlier AI systems that primarily classified or predicted information, generative AI can produce text, images, audio, and video. As a result, tools such as ChatGPT and Google Gemini allow organizations and individuals to create content quickly and at relatively low cost, while social media enables faster production, personalized communication, and audience interaction [1]. However, existing research has identified risks associated with synthetic media, including misinformation, manipulated visual content, algorithmic bias, and public distrust [2]. To address

these concerns, this study uses a qualitative interpretive analysis of three Australian cases. In particular, it investigates the transformation of social media communication via generative AI, the effects of AI-generated content on trust and authenticity, and the governance practices that may reduce these risks. By comparing cases from news, cultural representation, and advertising, the study shows how transparency interacts with accountability, fairness, and human oversight. Importantly, the analysis shifts attention from whether AI should be used to how its use is communicated, supervised, and interpreted by audiences.

2. Literature review

2.1. AI-generated content in social media communication

From automation and prediction-oriented systems to technologies enabling complex content generation across industries, AI has undergone remarkable advancement [3]. In social media communication, generative AI can draft marketing copy, create advertising images, assist customer service, produce multilingual material, and support journalistic workflows. As a result, these applications enable real-time adaptation and empower organizations to produce and personalize content at greater efficiency [1]. However, the more important change is not simply technical efficiency. Rather, generative AI alters the communication process by reshaping authorship and editorial control between humans and AI systems. As a result, a social media post may now be conceived by a human, generated by an AI system, modified through automated tools, reviewed by another person, and distributed through algorithmic recommendations. This process makes the origin of communication more difficult for audiences to identify, complicating conventional ideas of authorship and responsibility.

Such hybrid production can also change information credibility. When audiences cannot determine whether an image, statement, or persona is human-produced, AI-assisted, or entirely synthetic, familiar cues like originality and authorship become less dependable. Moreover, research on AI-generated text detection suggests that separating AI-produced text from human-written text is becoming increasingly challenging [4]. This uncertainty makes disclosure and source identification increasingly important within digital communication. Therefore, generative AI should be understood as a transformation of communication structures rather than merely a content-production technology.

2.2. Trust and transparency in AI-mediated communication

Trust is central to communication because audiences evaluate information through perceived credibility, honesty, expertise, and reliability. However, AI complicates this process because its contribution to a message may be invisible. In response to these challenges, Australian AI ethics frameworks identify transparency, accountability, fairness, and human oversight as important principles for responsible AI [5, 6]. Among these principles, transparency is the key analytical focus because it determines whether audiences can understand content production processes and evaluate them appropriately.

More specifically, transparency does not require that every minor automated process be explained, but significant AI involvement should not be hidden when it materially influences representation or meaning. In this regard, disclosure can reduce the gap between what audiences believe they are seeing and how content was actually created. Moreover, accountability complements transparency: even when an output is automatically generated, the organization publishing it remains responsible for checking its accuracy and appropriateness. Similarly, fairness

becomes particularly important where AI reproduces cultural or social stereotypes, while human oversight helps to identify and address such issues before release. These principles serve as an integrated framework: transparency allows audiences to recognize AI involvement; accountability clarifies responsibility; fairness assesses representational impacts; and human oversight helps prevent or address harmful outputs. Therefore, they demonstrate why trust can be undermined even when AI errors occur unintentionally.

2.3. Authenticity and representation in AI-generated content

From a communication perspective, authenticity refers to whether social media content is perceived as genuine, credible, and consistent with the identity or intentions behind it. Nevertheless, generative AI complicates this notion because synthetic images and text can appear realistic without being grounded in direct human experience. For example, research on AI-generated images shows that users can have difficulty distinguishing synthetics from authentic visual material, while studies of virtual influencers similarly connect disclosure and perceived authenticity with consumer trust [7, 8].

Beyond factual accuracy, a synthetic image may be visually convincing but culturally inauthentic, while an AI-generated advertisement may accurately display clothing yet conflict with consumers' expectations of a brand's identity or creative practices. In journalism, undisclosed automated alteration can also undermine confidence in the evidentiary role of images. Public trust in AI-mediated communication therefore depends partly on whether the mode of production matches audience expectations and whether significant AI intervention is made visible [9]. The literature suggests that authenticity should not be reduced to a simple human-versus-AI distinction. AI-assisted content can still be perceived as legitimate when its use is transparent and appropriately supervised. The key problem is the mismatch between synthetic production and audience expectations of representation, authorship, or organizational responsibility.

3. Method

3.1. Research design and case selection

This study uses a qualitative interpretive case study design because the research questions concern how trust, authenticity, and responsibility are understood in specific communication contexts rather than how frequently particular attitudes occur [10]. Additionally, a comparative case approach makes it possible to identify similarities and differences across different forms of AI-mediated communication.

Based on this research approach, cases were selected according to three criteria: identifiable AI-generated or AI-assisted public communication, documented public or media concerns about trust or authenticity, and distinct ethical risks. The cases reflect diverse AI-related challenges. The Georgie Purcell image incident, in which Nine News used an automated image-processing function that altered the Victorian MP's appearance, represents editorial accountability and human-oversight risk [11]. The AI-generated "Aboriginal Steve Irwin" character represents cultural authenticity and fairness concerns because criticism focused on stereotypical representation of Indigenous identity [12]. Jeanswest's AI-generated advertising represents brand authenticity and transparency, particularly audience reactions to synthetic models and the replacement of human creative labor [13]. These cases enable comparison of trust mechanisms across news, cultural representation, and commercial communication.

3.2. Data collection and analytical framework

The study relies on secondary data, including peer-reviewed research, government guidance, reputable news coverage, organizational explanations, and publicly reported audience responses. The selection of sources focused on materials that directly documented the case, explained the relevant AI practice, or provided evidence about trust, authenticity, transparency, or public reaction. Established news and organizational sources were prioritized where possible, while isolated social media comments were treated as illustrative responses rather than representative measures of public opinion.

The analysis examined each case through three stages. Each case was initially assessed for the role played by AI in producing or altering content. The resulting communication issues were analyzed via four ethical dimensions: transparency, accountability, fairness, and human oversight [5, 6]. The cases were then compared to examine how these dimensions influenced trust and authenticity across different contexts. This comparative approach ensures consistency in analysis while preserving context-specific differences.

4. Findings and analysis

4.1. AI use and changes in social media content production

Through these cases, generative AI reshapes social media communication by redefining responsibility, transparency, and authorship. The Purcell case reveals an accountability gap in editorial workflows. Although the image alteration resulted from automation, the controversy focused on insufficient human review and unclear responsibility for the final representation [11]. This demonstrates that AI-assisted production can weaken traditional accountability structures when human oversight fails to accompany automation. In addition, the Jeanswest case reveals a different challenge related to transparency and authenticity. While AI-generated imagery increased production efficiency, audience reactions focused less on the technology itself and more on whether the brand openly communicated its use of synthetic models and how these affected perceptions of creative authenticity [13]. This shows that AI involvement becomes a communication issue when audiences perceive a gap between the presented content and its actual production process. Moreover, the "Aboriginal Steve Irwin" case further demonstrates how AI-generated content can create representational risks. The controversy was not primarily about technical capability but about the use of culturally significant identity markers without sufficient contextual understanding or sensitivity [12]. And this highlights how generative AI can amplify existing concerns about fairness and cultural representation. From these cases, it becomes clear that generative AI creates hybrid communication systems where creative decisions, editorial judgments, and representational choices are increasingly shared between humans and algorithms. While this can improve efficiency and accessibility, it also requires clearer disclosure, stronger accountability, and continued human oversight. The central transformation is therefore not simply automation, but a shift in how responsibility and trust are negotiated in AI-mediated communication.

4.2. Trust and authenticity effects of AI-generated content

The three cases reveal different ways in which AI-generated content influences public trust through transparency, cultural representation, and brand authenticity. In the Purcell case, the main concern was the lack of transparency and editorial accountability. Audiences initially encountered the altered

image as ordinary news content and were unaware that automated processing had changed the representation. Although the alteration was later explained, the incident exposed a gap between AI involvement and expectations of editorial verification [11]. This gap weakened trust because the news organization responsible for accuracy did not identify or disclose the modification before publication. By contrast, the issue shifts from transparency to cultural representation in the "Aboriginal Steve Irwin" case. The controversy was not primarily about factual accuracy but about the use of culturally significant markers to construct a synthetic Aboriginal identity without sufficient contextual understanding [12]. In this context, realistic appearance does not necessarily create authentic representation. This distinction arises as AI can reproduce recognizable cultural elements while lacking the social and historical knowledge required for responsible representation. However, authenticity concerns are not limited to cultural representation. For example, a different form of trust concern appears in the Jeanswest case, where the focus moves to brand authenticity. The criticism was not related to factual errors but to audience expectations about human creativity, models, and brand identity [13]. Specifically, the use of AI-generated advertising created uncertainty about whether the brand's presentation aligned with its actual production practices. Viewed across these contexts, AI-generated content affects trust when audiences cannot clearly understand the role of AI or when AI use conflicts with expectations of authenticity and responsibility. More broadly, transparency helps clarify AI involvement, while accountability, fairness, and contextual awareness influence whether such use is accepted.

4.3. Transparency and human oversight in AI communication

In practice, the cases clarify the importance of transparency and human oversight in AI communication. In the Purcell incident, stronger human review could have identified the altered image before broadcast. Moreover, clearer internal responsibility could have reduced the possibility of treating automated processing as merely a technical problem [11]. In the culturally sensitive "Aboriginal Steve Irwin" case, human review required contextual and cultural judgement rather than only technical accuracy, as the main concern involved the appropriateness of representation rather than factual correctness [12]. In the commercial context, the Jeanswest case further demonstrates that transparency becomes important when synthetic people or scenes are presented in ways that audiences may reasonably perceive as real. Therefore, disclosure of AI involvement can help audiences better understand the relationship between synthetic content and brand communication [13]. These cases show that responsible AI communication depends on both transparency and meaningful human involvement. Whereas transparency helps clarify AI-generated content, human oversight allows organizations to assess broader issues beyond technical review. The cases suggest that clear disclosure and human judgement are essential when AI-generated content is used in public communication [14, 15].

5. Discussion

At a broader level, generative AI changes the formation of trust by making the production process itself an object of judgement. In conventional communication, audiences often assess credibility via visible cues such as journalists, organizations, models, or creators. However, hybrid human-AI communication weakens these cues because the apparent author or subject may no longer fully represent the process behind the content [16]. As a result, trust increasingly depends on whether organizations make AI involvement and production practices understandable to audiences.

This shift also reshapes perceptions of authenticity. The cases indicate that authenticity is not simply determined by the absence of AI. Instead, audiences judge whether the content, its production process, and the communicator's claimed identity or values remain consistent. For example, the Purcell case raised concerns about documentary trust, the Indigenous representation case highlighted cultural authenticity issues, and the Jeanswest case revealed tensions around brand identity. These cases reveal that the same technology can create different authenticity concerns depending on the social context in which it is used.

In this context, transparency helps manage uncertainty, but disclosure alone is not sufficient. A label indicating that content is AI-generated does not automatically make a stereotypical representation appropriate, nor does it remove responsibility for inaccurate or harmful outputs. Rather, transparency provides the basis for accountability by allowing audiences to understand when AI has influenced the communication process and assess whether its use is appropriate. Meanwhile, these findings challenge the idea that AI simply replaces human communication. Instead, AI-mediated communication involves a redistribution of agency: humans select tools, design workflows, provide prompts, and determine publication practices, while AI systems influence the form of the resulting content. Therefore, human oversight remains necessary because AI communication involves collaboration, not full automation.

These findings answer the research questions by showing that generative AI transforms social media through hybrid production, affects trust when AI uses authenticity, and requires transparency, responsibility, and human review. For Gen Z audiences, distrust of traditional media and reliance on social media for news verification make clear disclosure and responsible AI use important for digital credibility [17].

6. Conclusion

This study examined how AI-generated content affects trust and authenticity in social media through three Australian case studies. The analysis shows that generative AI transforms communication not only by increasing production efficiency but by redistributing authorship and editorial agency between humans and automated systems. As a result, this makes the origin and responsibility of content more difficult for audiences to evaluate. Through the cases, trust is most vulnerable when AI use fails to align with audience expectations. The Georgie Purcell case revealed weaknesses in editorial oversight, the "Aboriginal Steve Irwin" case highlighted limits in synthetic cultural representation, and the Jeanswest case showed how AI production influences brand authenticity. Transparency emerged as a common principle, but disclosure alone is insufficient without accountability, fairness, and human review. This study has several limitations. It relies on three Australian cases and secondary sources, which may not capture the full range of audience responses. Moreover, publicly reported reactions do not represent all audience attitudes, and the cases reflect a rapidly changing technological environment. Future research could therefore use interviews, surveys, or cross-cultural comparisons to examine how disclosure practices influence trust across different platforms and communication contexts. Overall, the study suggests that sustainable AI communication depends less on hiding or rejecting AI and more on making its role understandable, accountable, and appropriate within the social contexts where communication takes place.

References

- [1] Gulati, S., & Grover, A. (2026). AI-generated content and its effects on social media engagement. *Lex Localis - Journal of Local Self-Government*, 43–51. <https://doi.org/10.52152/fb28s185>

- [2] Hussain, S. A., Georges, H., & Askar, N. A. R. (2025). Illusion or reality: Ethical navigation in times of AI-enhanced visual communication. *Social Sciences & Humanities Open*, 12, 102228. <https://doi.org/10.1016/j.ssaho.2025.102228>
- [3] Ogoo, L. (2026). The history of artificial intelligence (AI): Origins, evolution, global power, risks, and the future (1950-2026). pp. 2-19.
- [4] Fetzer, M. (2024). Q&A: The increasing difficulty of detecting AI- versus human-generated text | Penn State University. In *Psu.edu*. <https://www.psu.edu/news/information-sciences-and-technology/story/qa-increasing-difficulty-detecting-ai-versus-human>
- [5] Australian Government Department of Industry, Science and Resources. (2019). Australia's AI ethics principles. <https://www.industry.gov.au/publications/australias-ai-ethics-principles>
- [6] Sanderson, C., Douglas, D., Lu, Q., Schleiger, E., Whittle, J., Lacey, J., ... & Hansen, D. (2023). AI ethics principles in practice: Perspectives of designers and developers. *IEEE Transactions on Technology and Society*, 4(2), 171-187.
- [7] Miskolczi, M. (2025). The illusion of reality: How AI-generated images (AIGIs) are fooling social media users. *Computers in Human Behavior*, 176, 108876. <https://doi.org/10.1016/j.chb.2025.108876>
- [8] Khalfallah, D., & Keller, V. (2025). Authenticity, ethics, and transparency in virtual influencer marketing: A cross-cultural analysis of consumer trust and engagement: A systematic literature review. *Acta Psychologica*, 260, 105573. <https://doi.org/10.1016/j.actpsy.2025.105573>
- [9] Salehi, A., Motaghi, H., Hashemi, M. S., & Fahim, Y. (2026). Public trust and legitimacy in the AI era: A systematic review of public media organizations. *Journal of Communication and Management*, 5(1), 72-84.
- [10] Lim, W. M. (2024). What is qualitative research? An overview and guidelines. *Australasian Marketing Journal (AMJ)*, 33(2), 199-229. <https://doi.org/10.1177/14413582241264619>
- [11] Kolovos, B., & Taylor, J. (2024). Victorian MP Georgie Purcell criticises Nine News for 'sexist' image editing to make outfit more revealing. *The Guardian*. <https://www.theguardian.com/media/2024/jan/30/georgie-purcell-mp-photoshop-automated-image-nine-news-apology-victoria-animal-justice-party>
- [12] Collard, S. (2026). It's AI blackface': Social media account hailed as the Aboriginal Steve Irwin is an AI character created in New Zealand. *The Guardian*, 15.
- [13] Skipper, M. (2026). Australian retailer lashed over AI ads, which experts warn are on the rise. *Nine.com.au*. <https://www.nine.com.au/australia-news/jeanswest-ai-advertising-social-media-expert-reactions-20260116-p5nula.html>
- [14] Australian Government: National AI center. (2026). *Be clear about AI-generated content: Understand when and how to disclose your use of artificial intelligence*. <https://www.ai.gov.au/staying-safe-and-responsible/essential-ai-practices/be-clear-about-ai-use>
- [15] Melendez, S. (2025). AI won't make the call: Why human judgment still drives innovation. In *Harvard Business School*. <https://www.hbs.edu/bigs/artificial-intelligence-human-jugment-drives-innovation>
- [16] Guettaf, S. (2026). Structural transformations in the communication system from human interaction to algorithmic mediation-critical analytical study. *International Journal of Social Communication*, 13(1), 26-42.
- [17] Trang, T. T. N., Thang, P. C., Nguyen, G. T. T., & Nguyen, H. T. M. (2025). Factors driving Gen Z's news engagement on TikTok: A hybrid analysis through CB-SEM and PLS-SEM. *Computers in Human Behavior Reports*, 18, 100645.