Artificial Authenticity: The Transparency Paradox in SME Marketing

Krisztina Finta, Luca Utassy, Sarolta Ács, Csaba Dezső Dér
Budapest Metropolitan University, Faculty of Business, Communication and Tourism, Budapest, Hungary
kfinta@metropolitan.hu, luca.utassy@gmail.com, saci.acs@gmail.com, cder@metropolitan.hu

This study examines how disclosure of Generative AI usage in SME marketing creates a "Transparency Paradox" that simultaneously demands and undermines consumer trust. An integrative literature review synthesizes empirical studies on AI disclosure effects with Signaling Theory and the Integrative Model of Organizational Trust, focusing on customer-facing communications where source authenticity underpins SMEs' competitive advantage. The analysis identifies a core tension: while customers demand transparency, explicit AI labeling triggers a "Word-of-Machine" effect and a Personalization Paradox, reducing perceived value and emotional connection in hedonic contexts; conversely, concealing AI usage constitutes "AI-washing," risking reputational damage upon exposure. For SMEs relying on relational proximity as a core asset, this paradox is damaging. Given limited longitudinal evidence, findings draw on case studies and cross-contextual inference. The research demonstrates that technical transparency is insufficient; SMEs require 'Artificial Authenticity' strategies emphasizing human-in-the-loop oversight, linking ethics washing to psychological rejection of automated agency in SME marketing.

DOI
https://doi.org/10.18690/um.epf.7.2026.34

ISBN
978-961-299-166-1

Keywords:
AI-washing, transparency paradox, consumer trust, SME marketing, generative AI, artificial authenticity

University of Maribor Press

1 Introduction

Generative AI is rapidly reshaping how firms create and personalize marketing content, promising efficiency gains and new forms of value creation across industries (Davenport et al., 2020). For small and medium-sized enterprises (SMEs) - which account for a substantial share of global employment - these models significantly enhance productivity and customer engagement (Rajaram & Tinguely, 2024). This shift extends beyond marketing outputs, as the integration of artificial intelligence directly influences overall management productivity and operational efficiency (Garg et al., 2024). However, as Sarker et al. (2025) emphasize, this technological shift also amplifies critical concerns regarding algorithmic opacity and the erosion of stakeholder trust (Dwivedi et al., 2023).

In parallel, regulators and platforms increasingly frame transparency as the primary remedy. The EU AI Act and major platforms such as Meta and TikTok are moving toward mandatory labeling, making the disclosure of AI use a legal and infrastructural constraint rather than a voluntary ethical choice (Koning & Voorveld, 2025). Under the EU AI Act, which entered into force in August 2024 with transparency obligations applying from August 2026, AI systems that interact directly with consumers - including customer-facing chatbots and AI-generated marketing content - are subject to mandatory disclosure requirements (European Parliament and Council of the European Union, 2024, Art. 50). Providers must inform users that they are interacting with an AI system, and synthetic or AI-manipulated content must be labeled as such. While the Act includes proportionality provisions that acknowledge SMEs' more limited resources (European Parliament and Council of the European Union, 2024, Art. 55; Recital 91), smaller firms must nonetheless implement disclosure mechanisms and document their AI use without the dedicated compliance infrastructure typically available to larger organizations (Rajaram & Tinguely, 2024). Consequently, resource-constrained SMEs are forced to decide not whether, but how to integrate generative AI and signal its use to their audiences.

Recent empirical evidence, however, indicates that transparency itself is double-edged. Experimental work shows that AI disclosures in generative AI-created advertisements increase consumers' conceptual AI knowledge and activate persuasion knowledge, which in turn reduce trust in both the ad and the organization, even as perceived appropriateness can modestly increase trust (Koning

& Voorveld, 2025). Similarly, studies on social media content find that brands' adoption of generative AI - especially for full automation of content creation - undermines perceived brand authenticity and deteriorates attitudes, electronic word-of-mouth intentions, and loyalty, despite the fact that users often cannot perceptually distinguish AI- from human-generated content; the negative shift is triggered by disclosure (Brüns & Meißner, 2024). These findings point to a transparency paradox, in which making AI use visible can simultaneously uphold ethical ideals and erode consumer trust (Koning & Voorveld, 2025; Brüns & Meißner, 2024).

This paradox is particularly acute for SMEs, whose competitive advantage often rests on relational closeness, perceived genuineness, and locally grounded authenticity in their marketing interactions (Sarker et al., 2025). Yet previous research on AI disclosures and generative AI in marketing has largely focused on large brands, generic consumer contexts, or platform-level policy, paying limited attention to the specific vulnerabilities and trust dynamics of SME marketing.

To address this gap, the present paper introduces the notion of "artificial authenticity" and investigates the transparency paradox in SME marketing. Artificial Authenticity is a strategic communication framework in which an organization openly acknowledges its use of generative AI while simultaneously foregrounding the human expertise, curation, and oversight that guided the process, thereby maintaining perceived source authenticity, relational trust, and brand integrity without resorting to either concealment (AI-washing) or reductionist binary labeling ('AI-generated').

Specifically, this review addresses the following research questions:

- RQ1: How can "AI-Washing" be defined from the perspective of consumer trust, specifically regarding organizational integrity and benevolence?
- RQ2: How do the "Transparency Paradox" and the "Word-of-Machine" effect manifest in the context of SME marketing?
- RQ3: What strategic communication practices can SMEs employ to restore and maintain brand authenticity?

2 Theoretical Background

2.1 Defining AI-Washing in the SME Context (RQ1)

AI-washing represents the latest iteration in a documented family of corporate "washing" phenomena. The foundational analogy lies in greenwashing, defined by Delmas and Burbano (2011) as misleading consumers about a firm's environmental practices. De Jong, Huluba, and Beldad (2020) advance this framework by demonstrating that greenwashing operates along a spectrum: their experimental study distinguishes between outright lies, half-lies, and taking credit for legal obligations. Half-lies produce the same reputational damage as complete lies - consumers judge deception normatively, not proportionally (De Jong et al., 2020). According to cognitive dissonance theory (Festinger, 1957), this finding illustrates how consumers react to deceptive signaling: they resolve the tension between corporate claims and reality by questioning the organization's integrity, regardless of the severity of the discrepancy.

The transition from environmental to technological deception is mediated by the concept of "ethics washing." Bietti (2020) argues that technology companies instrumentalize ethics as a self-legitimizing strategy to deflect regulation, while Wagner (2019) positions ethical frameworks as deliberate substitutes for binding legal obligations - reducing AI governance to voluntary, unenforceable commitments. Seele and Schultz (2022) formalize this bridge by introducing the term "machinewashing," defined as "a strategy that organizations adopt to engage in misleading behavior about ethical AI/algorithmic systems" (p. 1128). Their model identifies the same exploitation of information asymmetry that characterizes greenwashing, applied specifically to algorithmic operations.

The concept gained regulatory teeth when the U.S. Securities and Exchange Commission charged two investment advisory firms with making false statements about their AI use, imposing $400,000 in combined penalties and explicitly labeling the practice "AI washing" (U.S. SEC, 2024). In this paper, we conceptualize AI-washing as the strategic misrepresentation of AI use within customer-facing communications - encompassing both the exaggeration or fabrication of AI capabilities to project unwarranted technological sophistication (De Jong et al., 2020; U.S. SEC, 2024), and the deliberate concealment of AI use to maintain a false impression of human authorship - exploiting the information asymmetry inherent in

algorithmic systems (Seele & Schultz, 2022). It should be noted that a parallel strand of the literature uses the same term to describe organizational governance failures: the adoption of symbolic AI ethics policies without substantive reform in regulated deployment contexts (Bietti, 2020; Seele & Schultz, 2022). The present study focuses exclusively on the consumer-facing dimension of this misrepresentation.

2.2 The Psychology of Trust and Signaling

The psychological impact of AI-washing is understood through four interconnected theoretical lenses: organizational trust, machine agency perception, signaling theory, and brand authenticity.

Trust. Following Mayer et al. (1995), we understand trust as the willingness to accept vulnerability based on positive expectations. Colquitt et al. (2007) demonstrate that this willingness relies on three distinct dimensions of trustworthiness: ability, benevolence, and integrity. Furthermore, organizational success and competitiveness increasingly depend on maintaining a balance between interpersonal relationships and impersonal trust in the firm's systems and processes (Michalec et al., 2024). AI-washing threatens this entire framework: misrepresenting capabilities exposes a lack of actual competence, the inherent deception violates ethical principles, and the underlying strategy signals self-interest rather than genuine consumer welfare.

Machine agency. Sundar's (2020) Human-AI Interaction - Technology Is Mediating Experience (HAII-TIME) model explains how consumers process AI claims through a "machine heuristic" - a mental shortcut attributing both positive qualities (precision, objectivity) and negative ones (coldness, lack of empathy) to algorithmic systems. This creates an inherent tension for SMEs: AI labels can signal competence through automation bias (the tendency to overtrust machines) but simultaneously undermine the warmth and personal connection central to the SME's relational advantage. Moreover, algorithm aversion - the tendency to reject algorithmic systems entirely after a single observed failure (Sundar, 2020) - means that discovery of AI-washing can trigger a loss of trust.

Signaling. Signaling theory holds that effective signals must be costly to produce, ensuring only genuinely capable firms can send them (Connelly et al., 2011). Ching, Gerab, and Toste (2017) apply this framework to sustainability reporting, finding no

significant association between reporting quality and financial performance - suggesting that the signal alone is insufficient without alignment to underlying reality. In SME marketing, humanness functions as a costly signal of authenticity: human-crafted content requires time, skill, and personal investment that cannot be easily scaled. AI-washing decouples this signal from its substance, mirroring the reporting paradox identified by Ching et al. (2017) - projecting the appearance of skilled human effort in cases where the content was, in reality, produced automatically, at the push of a button.

Brand authenticity. The cumulative effect of these dynamics is captured by Morhart, Malär, Guèvremont, Girardin, and Grohmann's (2015) perceived brand authenticity (PBA) construct, comprising four dimensions: credibility (delivering on promises), integrity (genuine moral values), symbolism (identity expression), and continuity (narrative consistency). AI-washing directly undermines credibility (exaggerated claims), integrity (deception signals self-interest, paralleling Colquitt et al.'s trust dimension), and continuity (a relational brand suddenly signaling algorithmic automation). The Transparency Paradox examined in the following sections emerges from this convergence: in the short term, AI labels exploit automation bias to signal ability (Sundar, 2020), but the resulting signal-substance gap destabilizes brand authenticity (Morhart et al., 2015), and exposure triggers normative deception judgments (De Jong et al., 2020) compounded by algorithm aversion (Sundar, 2020) - producing a trust collapse disproportionate to the perceived offense.

3 Methodology

3.1 Research Design

Given the rapid development of Generative AI in SME marketing, limited empirical evidence exists that directly examines the disclosure paradox in this specific context. The latest systematic reviews on AI in marketing and Generative AI (Chan & Choi, 2025; Prasanna & Kushwaha, 2025) highlight a rapidly expanding but fragmented body of work that lacks an integrated perspective on consumer trust and authenticity. This study therefore adopts an integrative literature review methodology (Snyder, 2019) to synthesize fragmented findings across marketing, consumer psychology, and information systems. Unlike traditional narrative reviews that primarily serve as background for primary research, an integrative review treats the literature itself as data, enabling the construction of a novel conceptual

framework where primary data collection would be premature (Snyder, 2019). The primary objective is to identify recurring patterns, paradoxes, and theoretical tensions in AI disclosure strategies, ultimately linking them to dimensions of organizational trust and brand authenticity in SME contexts.

3.2 Scope and Boundaries

To ensure analytical depth, the research scope is narrowed to the intersection of SME marketing communications and consumer trust, with a focus on customer-facing interactions where source authenticity is essential for SMEs' relational competitive advantage. The analysis concentrates on three overlapping domains of customer-facing Generative AI use. The first is generative content creation - the use of large language models for copywriting, blog posts, and personalized messaging, where AI can augment or replace the human voice of the brand (Wahid et al., 2023). The second is conversational AI, specifically the deployment of chatbots and virtual assistants in customer service and conversational marketing contexts, where AI serves as the direct interface between brand and consumer (Israfilzade & Sadili, 2024). The third is attribution and authorship claims - the strategic use or omission of labels such as "AI-powered" or "Human-made" in marketing collateral, and the misrepresentation of human versus AI authorship in brand communications (Kirk & Givi, 2025). The review explicitly excludes technical performance evaluations of AI models (e.g., accuracy, latency, model architectures), internal operational AI applications (e.g., HR automation, inventory optimization) that do not directly affect consumer perception, and B2B procurement and supply-chain analyses where end-consumer trust is only indirectly involved.

3.3 Literature Search and Selection

A systematic search was conducted in Scopus, Web of Science, and Google Scholar covering the period from January 2020 to December 2025 for literature specifically addressing Generative AI, algorithm aversion, and algorithmic disclosure. Meanwhile, foundational theories regarding organizational trust, signaling theory, and corporate washing phenomena were drawn from seminal literature without date restrictions. The search strategy employed keyword combinations such as: ("AI disclosure" OR "algorithm aversion" OR "AI transparency" OR "generative AI") AND ("consumer trust" OR "brand authenticity" OR "transparency paradox") AND ("SME" OR "small business" OR "small enterprise"). Additional targeted

searches were performed for key constructs including "Word-of-Machine effect", "AI-washing", and "ethics washing" to capture adjacent literatures on automated agency, moral disgust, and deceptive signaling. Inclusion criteria were: (1) peer-reviewed empirical or conceptual papers, (2) focus on customer-facing AI contexts where source authenticity is salient, (3) outcomes related to trust, authenticity, moral evaluation, or consumer perception, and (4) English language. Exclusion criteria eliminated: (1) purely technical AI performance studies, (2) internal, non-customer-facing AI use cases, and (3) studies focused exclusively on large enterprises without SME-relevant implications. The search yielded 42 relevant papers after screening, supplemented by regulatory and policy documents (e.g., EU AI Act) and seminal theoretical contributions on Signaling Theory and organizational trust.

3.4 Data Extraction, Theoretical Lens, and Analysis

Data extraction and thematic synthesis followed the integrative review protocol outlined by Snyder (2019). Each included source was systematically coded along four dimensions: (1) AI disclosure strategy (explicit labeling, concealment, partial or conditional transparency), (2) affected trust dimensions, applying Colquitt et al.'s (2007) framework, (3) contextual factors (hedonic vs. utilitarian consumption, SME vs. large-firm setting, type of AI application), and (4) observed outcomes (changes in perceived authenticity, consumer response, reputational or regulatory risk). A multi-theoretical lens guided the synthesis. Signaling Theory was employed to interpret AI labels (or their absence) as market signals of quality and authenticity, building on prior work in sustainability reporting and authentic communication (Ching et al., 2017; Morhart et al., 2015) to explain why SMEs might engage in deceptive signaling ("washing"). By integrating these signaling dynamics with Colquitt's trust dimensions, the analysis illuminated specific contexts where an intended signal of efficiency conflicts with a perceived signal of inauthenticity. Subsequent cross-case comparative analysis identified convergent and divergent findings across studies, yielding three dominant thematic clusters: the Transparency Paradox, the Word-of-Machine Effect, and the Personalization Paradox. Ultimately, the resulting empirical patterns were theoretically integrated into the proposed "Artificial Authenticity" framework, connecting AI-washing and disclosure strategies to brand authenticity in SME marketing.

4 Results (The Paradoxes)

The following three sections examine the conditions under which AI use in SME marketing communications undermines, rather than increases, perceived authenticity and trust. The synthesized literature reveals three vulnerabilities based on the authorship of the message (the Transparency Paradox), the context of the message (the Word-of-Machine effect), and the depth of profiling behind the message (the Personalization Paradox).

Table 1: Summary of AI-Integration Vulnerabilities in SME Marketing

Paradox / Vulnerability	Core Mechanism	Impact on SME Marketing	Key Literature
Transparency Paradox	Disclosing AI authorship reduces perceived authenticity; concealing it leads to trust collapse upon exposure.	Creates a double bind where both proactive disclosure and reactive exposure erode relational authenticity.	Delmas & Burbano (2011); Morhart et al. (2015); Wagner (2019); Bietti (2020); Luo et al. (2019); Seele & Schultz (2022); Brüns & Meißner (2024); U.S. SEC (2024); Kirk & Givi (2025); Koning & Voorveld (2025)
Word-of-Machine Effect	Consumers systematically discount AI recommendations in hedonic, symbolic, or identity-relevant contexts.	Penalizes SMEs for using AI in affective categories where emotional resonance and human effort are essential.	Castelo et al. (2019); Sundar (2020); Longoni & Cian (2020); Granulo et al. (2021); Bellaiche et al. (2023)
Personalization Paradox	Scalable personalization mimics intimate knowledge but crosses into intrusiveness and feelings of surveillance.	Deep algorithmic profiling degrades perceived benevolence, converting relational advantage into strategic exploitation.	Mayer et al. (1995); Colquitt et al. (2007); Aguirre et al. (2015); Lefkeli et al. (2023); Abrokwah-Larbi (2023); Hennighausen et al. (2025)
Strategic Solutions (Artificial Authenticity)	Utilizing Human-in-the-Loop (HITL) integration and transparent signaling to mitigate algorithm aversion.	Restores relational proximity and trust, allowing SMEs to leverage AI efficiency without sacrificing authenticity.	Connelly et al. (2011); Rajaram & Tinguely (2024); Haupt et al. (2024); Sarker et al. (2025)

While Section 2.1 already addressed RQ1 by establishing the conceptualization of AI-washing, these three vulnerabilities directly correspond to the paper's remaining research questions: the Transparency Paradox addresses RQ2 by examining how AI disclosure manifests in SME marketing; the Word-of-Machine effect further elaborates RQ2 by delineating the contextual boundaries of algorithm aversion; and the Personalization Paradox rounds out RQ2 by identifying the third dimension of the structural trap. Collectively, the 'Artificial Authenticity' framework proposed in Section 5 responds to RQ3.

Table 1 synthesizes the three core vulnerabilities associated with AI integration in SME marketing and outlines 'Artificial Authenticity' as a strategic solution to mitigate these algorithmic aversions.

4.1 The Transparency Paradox (RQ2)

The Transparency Paradox describes a structural dilemma in which both disclosure and non-disclosure of AI use in marketing communications can erode consumer trust - leaving SMEs with no straightforwardly safe option. Disclosing AI authorship can reduce the perceived authenticity of a message and trigger negative consumer responses (Kirk & Givi, 2025); yet concealing AI use constitutes a form of deception that, once uncovered, causes equal or greater reputational damage (Luo et al., 2019; Kirk & Givi, 2025). To illustrate: consider a local artisan bakery that uses a generative AI tool to write its weekly newsletter to loyal customers. If the bakery discloses that its warm, personally addressed messages were AI-generated, subscribers may feel the emotional connection was simulated. Conversely, if the bakery omits this information and AI involvement is later revealed - through a journalist's investigation, a regulatory audit, or mandatory labeling under the EU AI Act - customers may interpret the prior silence as deliberate dishonesty, damaging the relationship more severely than the original disclosure would have. This double bind is the Transparency Paradox.

To understand the mechanics of this paradox, it is necessary to examine how the act of disclosure itself triggers consumer backlash. Developing work on generative AI in marketing shows that disclosure of AI use is far from a simple "more transparency is always better" story for firms. Across seven preregistered experiments, Kirk and Givi (2025) demonstrate a robust "AI-authorship effect": when consumers believe

that emotional marketing communications are written by an AI rather than a human, they perceive the message as less authentic, experience greater moral disgust, generate less positive word of mouth (PWOM), and show lower customer loyalty.

However, this effect applies specifically to emotional content; it is attenuated when the communication is factual, when AI only edits rather than authors the message, when the AI signs the communication directly, and when consumers believe that most marketers already use AI. In other words, disclosure that an emotional message is AI-written reliably hurts authenticity and downstream behavioural responses, whereas disclosure in purely factual contexts may be comparatively harmless.

At the same time, regulatory and ethical pressures increasingly push firms toward more transparency. Kirk and Givi (2025) document how legislators and regulators (e.g., the EU AI Act) begin to require clear AI labels for generative AI communications, while consumers themselves express a desire for transparency about AI use in marketing. This aligns with Rai's (2020) commentary on "explainable AI", which frames the core challenge as a move from "black box" to "glass box" systems: explanations and disclosures are needed to build trust, expose biases, and meet fairness and accountability requirements, especially in high-stakes domains. Explainability and disclosure, however, can also surface the very aspects of AI decision-making that users find normatively troubling, thereby reducing trust and acceptance rather than enhancing it.

When applied to emotional marketing, this means that making the AI visible can both satisfy an ethical/regulatory requirement and trigger a negative authenticity judgement. The paradox for SMEs follows directly: if an SME openly discloses that its emotional communications are AI-generated, the AI-authorship effect predicts lower perceived authenticity, increased moral disgust, and reduced PWOM and loyalty. Yet if the same SME conceals AI use, and this use is later revealed through regulation, media exposure, or improved AI-detection tools, consumers interpret the previous non-disclosure as dishonesty, again eliciting moral disgust and avoidance responses. This creates a "transparency paradox": transparency about AI use in emotional communications undermines authenticity by revealing the machine author, whereas non-transparency undermines authenticity once hidden AI use is uncovered.

Evidence from adjacent AI-disclosure contexts reinforces this dilemma. Luo et al. (2019) show in an e-commerce experiment that when customers are made aware that they are interacting with an AI chatbot rather than a human operator, their purchase likelihood decreases, even though the functional performance of the chatbot may be adequate. Disclosing machine involvement suppresses conversion in a commercial setting.

SMEs face a double bind: they cannot safely hide AI use, but they also cannot straightforwardly signal it without incurring authenticity and trust penalties when communications are emotional, relational, or identity-relevant. Both proactive disclosure and reactive exposure erode consumer trust in emotional marketing contexts, rendering conventional disclosure mechanisms insufficient in relational communication.

4.2 The Word-of-Machine Effect

A second, closely related paradox arises when considering not only who is disclosed as the author, but what kind of advice or content AI is perceived to provide. Longoni & Cian (2020) introduce the "word-of-machine" effect by contrasting AI advice with traditional "word-of-mouth". In multiple studies, they show that consumers systematically discount recommendations when they believe they come from an AI rather than a human, even when the informational content is held constant. This AI aversion is strongly context-dependent: AI is more accepted in utilitarian, functional decision domains (e.g., where accuracy and efficiency are paramount), but is resisted in hedonic domains, where consumers seek identity-relevant or emotionally rich benefits. The central role of these settings is further emphasized in retail environments, where pleasure-driven shopping motivation - driven by the pursuit of experiential and sensorial stimuli (Tyrväinen et al., 2020) - is a primary catalyst for driving positive emotional and cognitive customer experiences. Because creating an enjoyable and entertaining shopping environment is key to generating authentic word-of-mouth and long-term customer loyalty (Tyrväinen et al., 2020), replacing human touchpoints with algorithmic communication in these contexts threatens the core mechanism of brand connection. In such affective categories, AI advice feels less appropriate and less trustworthy, leading to lower compliance, lower choice share, and more negative affective responses compared to human advice.

Huang and Rust (2021) provide a strategic framework that helps position this effect in a broader AI-in-marketing landscape. They distinguish "mechanical", "thinking" and "feeling" AI, and argue that content creation for marketing communications increasingly sits in the intersection of thinking and feeling AI: it requires balancing data-driven optimization with emotional resonance and relationship-building. Their framework also reviews evidence that consumers resist AI in domains where identity, intimacy and emotional exchange are central - for example, in medical decisions, identity-laden consumption, and anthropomorphised service robots in symbolic consumption contexts. This aligns with Longoni & Cian's (2020) finding that the very contexts in which emotional content is most valuable (hedonic and symbolic) are precisely those where "word-of-machine" is least acceptable.

The reluctance to accept AI-generated or AI-recommended experiential content generalises beyond classic marketing settings into aesthetic domains. Bellaiche et al. (2023) examine how people evaluate artworks that are in fact all created by AI, but are randomly labelled as "human-created" or "AI-created". In two large-sample, within-subjects studies, they find that artworks labelled as human-created receive higher ratings on liking, beauty, profundity, and perceived worth than identical artworks labelled as AI-created. Additional measures reveal that the human label increases perceived narrativity (story), effort, and meaning, and that these more "communicative" judgements mediate the label effect for deeper evaluations such as profundity and worth. In other words, knowing (or believing) that a human created a piece makes it feel more effortful, meaningful and narrative-rich, even when the sensory stimulus is unchanged. Consistent with the marketing evidence, consumers in these symbolic contexts value not only what is said or shown, but the perceived human experience behind it.

Related work on algorithm aversion and preferences for human versus automated labour further clarifies when and why "word-of-machine" is problematic for marketers. Castelo et al. (2019) show that aversion to algorithms is task-dependent: consumers are more willing to rely on algorithms for tasks that are objective, quantifiable and impersonal, and are more reluctant when tasks require intuition, moral judgement or an understanding of human uniqueness. Granulo et al. (2021) find that preferences for human labour over robotic labour are particularly strong in symbolic consumption contexts - situations where products and services serve as identity signals or carriers of meaning, rather than purely functional solutions. When marketing communications are framed as human, symbolic, or morally laden acts

(e.g., expressing gratitude, pride, sympathy, or inspiration), consumers systematically devalue AI involvement.

The devaluation of AI involvement in symbolic contexts stems from a perceived lack of human effort and narrative depth. Bellaiche et al. (2023) demonstrate that human-authored labels increase the perceived meaning and profundity of a message, elements that algorithms cannot replicate. This evidence grounds Longoni and Cian's (2020) word-of-machine effect in Huang and Rust's (2021) framework: the penalty for AI is highest when it acts as a 'feeling' agent. Attempting to emulate relational warmth triggers expectations of internal states that machines inherently lack, transforming authentic engagement into a devalued output even when the content quality remains identical.

4.3 The Personalization Paradox and the Convergence of Vulnerabilities

While the Transparency Paradox concerns the act of disclosure (who authored the content) and the Word-of-Machine effect dictates the contextual boundaries (hedonic vs. utilitarian), a third distinct vulnerability manifests regarding the depth of profiling: the Personalization Paradox. Generative AI promises SMEs scalable personalization, yet this creates a backlash when personalization becomes "too good" and crosses into intrusiveness.

This paradox operates through a different mechanism than transparency. While the Transparency Paradox damages trust because the firm outsourced its voice to a machine, the Personalization Paradox damages trust because the machine appears to know too much, triggering feelings of vulnerability. As Aguirre et al. (2015) established, highly tailored content backfires when consumers infer covert data collection, shifting their perception from being "understood" (relational proximity) to being "surveilled" (exploitation).

For SMEs, this tension is particularly dangerous. Abrokwah-Larbi (2023) posits that generative AI, fueled by deep learning and smart data, enables real-time, context-oriented personalization at scale. Firms increasingly rely on such deep personalization to drive user engagement in highly interactive digital environments, which often yield stronger consumer responses than traditional formats (Szeberényi et al., 2025). However, when an SME uses GAI to mimic intimate customer knowledge, it risks converting its relational advantage into a perceived instrument of

surveillance. Hennighausen et al. (2025) demonstrate that when AI-mediated communication is recognized in personalized contexts, it specifically degrades perceived benevolence and integrity; customers infer that the artificial intimacy is driven by strategic efficiency rather than genuine care. These three vulnerabilities converge to form a strict strategic trap. When SMEs combine deep algorithmic profiling with emotional or hedonic messaging, any subsequent disclosure of AI authorship actively dismantles the very relational authenticity the firm attempted to scale.

5 Discussion

Building on the paradoxes identified in Section 4, this discussion proposes 'Artificial Authenticity' as a strategic framework for SMEs navigating the transparency dilemma.

The integration of Generative AI into SME marketing disrupts the relational trust mechanisms on which these firms have traditionally relied, by introducing a trade-off between operational scaling and perceived source authenticity.

5.1 The Trust Implications of AI-Washing (RQ1)

Viewed through foundational theories of organizational behavior, 'AI-washing' fundamentally undermines consumer trust in SME marketing. Based on Colquitt et al.'s (2007) framework, AI-washing simultaneously threatens all core dimensions of organizational trust. Maintaining the false appearance of personal human effort - when, in reality, an AI is doing the work - not only reveals the absence of genuine professional investment and exposes a lack of true competence, but the accompanying misrepresentation violates ethical integrity and reflects the firm's self-interest rather than genuine consumer welfare. Furthermore, based on Sundar's (2020) machine agency framework, algorithm aversion triggers a severe loss of trust upon the discovery of such deception. As a consequence, the signal-substance gap inherent in AI-washing directly destabilizes perceived brand authenticity by undermining the firm's credibility, integrity and narrative continuity (Morhart et al., 2015).

5.2 The Transparency Paradox, Word-of-Machine, and Personalization (RQ2)

The integration of generative AI into SME marketing creates a severe double bind that transcends mere regulatory compliance. Applying Signaling Theory, the convergence of the Transparency Paradox, the Word-of-Machine effect, and the Personalization Paradox creates a structural trap that conventional disclosure mechanisms cannot resolve. When an SME uses generative AI to create highly personalized, emotionally resonant content, it aims to maximize the relational signal; however, it simultaneously maximizes the risk of consumer backlash. If the SME explicitly discloses this AI usage, it triggers the "AI-authorship effect" (Kirk & Givi, 2025), where the disclosure immediately strips the message of its perceived authenticity. This is further compounded by the "Word-of-Machine" effect (Longoni & Cian, 2020), as consumers tend to devalue AI-generated content in the very emotional and experiential contexts that SMEs rely on for differentiation.

Conversely, if the firm conceals the AI to maintain the illusion of a personal touch, it engages in AI-washing, risking a total collapse of trust upon exposure. This tension is greatly exacerbated by the Personalization Paradox. While AI enables mass personalization, the explicit recognition of algorithmic profiling shifts the consumer's perception from being relationally "understood" to being strategically "surveilled" (Hennighausen et al., 2025). Conventional transparency mechanisms fail SMEs in affective and deeply personal contexts. To address this impasse, this study proposes 'Artificial Authenticity' as a strategic framework. This approach shifts the firm's focus from declaring what the machine generated to demonstrating how human expertise guided the process, mitigating the authenticity costs associated with algorithmic communication.

Figure 1 summarizes the mechanism through which both disclosure and concealment lead to a loss of consumer trust, and illustrates how the Artificial Authenticity strategy resolves this structural trap.

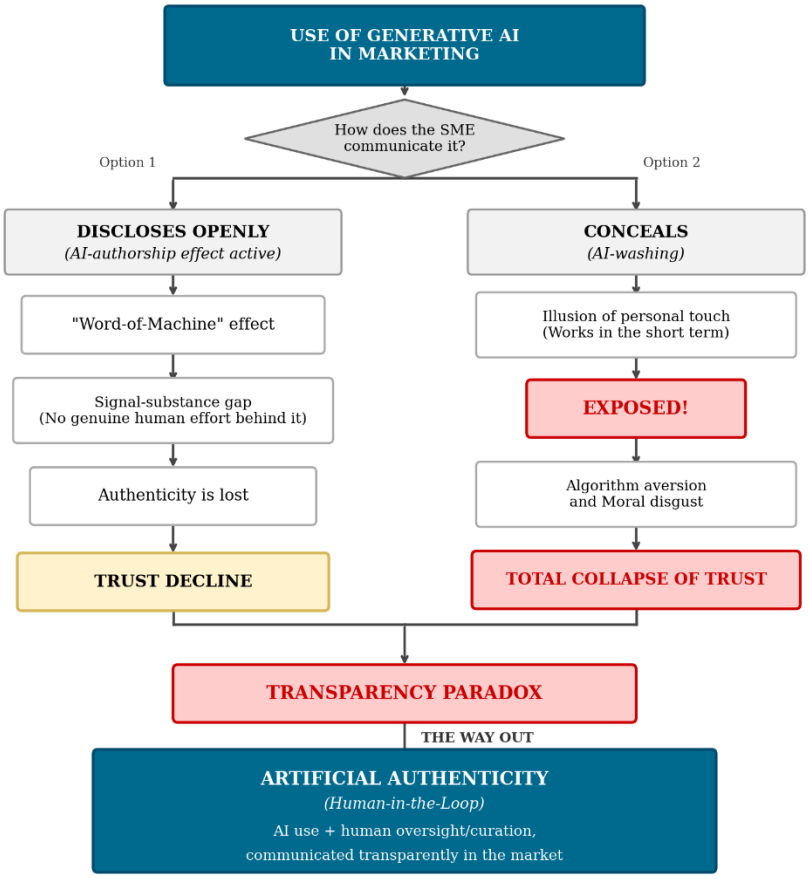

Figure 1: The Transparency Paradox and the Artificial Authenticity Framework
Source: Authors' own compilation

5.3 Restoring Brand Authenticity through Human-in-the-Loop (RQ3)

To address the Transparency Paradox and the Word-of-Machine effect (RQ3), SMEs should adopt 'Artificial Authenticity'-a strategy centered on human-in-the-loop (HITL) integration. AI integration yields the strongest outcomes when it augments human oversight rather than substituting for it (Davenport et al., 2020). Fully autonomous operations can diminish perceived responsibility and impair trust by heightening a sense of consumer exploitation (Lefkeli et al., 2023).

Empirical evidence confirms that human-AI collaboration alleviates negative consumer responses when the firm indicates explicit human control (Haupt et al., 2024). By demonstrating that human experts guide and validate AI-assisted content, SMEs can leverage generative efficiency while maintaining the empathy and ethical oversight required for relational proximity (Ismail et al., 2025).

To illustrate how this resolves the tension outlined in Section 4.1, we return to the artisan bakery example. If the bakery relies on a binary compliance label (e.g., *"This newsletter was generated by AI"*), it triggers the AI-authorship effect, stripping the message of its emotional resonance. Conversely, completely concealing the tool risks AI-washing and subsequent trust collapse. By applying the Artificial Authenticity framework, the bakery shifts the disclosure focus from the machine's autonomous output to the human's curation process. A strategic disclosure such as, *"Drafted with AI assistance, but carefully curated and approved by our head baker,"* explicitly signals human-in-the-loop oversight. This nuanced reframing satisfies ethical transparency requirements while demonstrating the narrative effort and relational warmth that define the SME's competitive advantage.

6 Conclusions

6.1 Theoretical and Practical Contributions

This research demonstrates that the adoption of Generative AI forces SMEs into a fundamental dilemma: prioritizing operational efficiency directly threatens the relational proximity that forms their core competitive advantage. Our synthesis indicates that converging vulnerabilities - the Transparency Paradox, the Word-of-Machine effect, and the Personalization Paradox - render binary disclosure methods inadequate in emotional and symbolic contexts.

Beyond individual firm strategies, Artificial Authenticity informs both marketing practice and regulatory policy. As policymakers and digital platforms increasingly mandate algorithmic transparency (e.g., the EU AI Act), strict compliance inadvertently punishes SMEs by stripping their communications of perceived genuineness. Therefore, industry standards and future regulations must evolve to recognize and accommodate collaborative human-AI authorship, rather than enforcing a reductionist, binary "human versus machine" labeling system.

6.2 Limitations and Future Research

This study is limited by its reliance on cross-contextual inference in the absence of SME-specific primary data; the integrative review methodology, while appropriate given the fragmented state of the field, cannot substitute for experimental or longitudinal validation of the proposed Artificial Authenticity framework.

Ultimately, consumer trust in an era of AI-mediated communication is not built by pretending the machine does not exist, nor by allowing it to operate unchecked, but by transparently showcasing the human expertise that guides it. Future empirical research should focus on testing the boundary conditions of these collaborative signaling strategies across different sectors and regulatory environments to further refine the governance of algorithmic communication.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used Google Gemini for structuring the manuscript, Gemini and Perplexity Pro to assist in integrating sources into the text, and Claude AI for structural review and critical feedback. After using these tools, the author(s) thoroughly reviewed and edited the content as needed and take(s) full responsibility for the final content and integrity of the publication.

Acknowledgment

This study was supported by the S.M.A.R.T. International Research Group.

References

Abrokwah-Larbi, K. (2023). The role of generative artificial intelligence (GAI) in customer personalisation (CP) development in SMEs: a theoretical framework and research propositions. *Industrial Artificial Intelligence*, *1*(1). https://doi.org/10.1007/s44244-023-00012-4

Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K., & Wetzels, M. (2015). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of Retailing*, *91*(1), 34-49. https://doi.org/10.1016/j.jretai.2014.09.005

Bellaiche, L., Shahi, R., Turpin, M.H., Ragnhildstveit, A., Sprockett, S., Barr, N., Christensen, A., & Seli, P. (2023). Humans versus AI: Whether and Why We Prefer Human-Created Compared to AI-Created Artwork. *Cognitive Research: Principles and Implications*, *8*, 42. https://doi.org/10.1186/s41235-023-00499-6

Bietti, E. (2020). From ethics washing to ethics bashing. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 210-219. https://doi.org/10.1145/3351095.3372860

Brüns, J. D., & Meißner, M. (2024). Do You Create Your Content yourself? Using Generative Artificial Intelligence for Social Media Content Creation Diminishes Perceived Brand Authenticity. *Journal of Retailing and Consumer Services*, *79*, 103790. https://doi.org/10.1016/j.jretconser.2024.103790

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, *56*(5), 809-825. https://doi.org/10.1177/0022243719851788

Chan, H.-L., & Choi, T.-M. (2025). Using generative artificial intelligence (GenAI) in marketing: Development and practices. *Journal of Business Research*, *191*, 115276. https://doi.org/10.1016/j.jbusres.2025.115276

Ching, H., Gerab, F., & Toste, T. (2017). The Quality of Sustainability Reports and Corporate Financial Performance: Evidence From Brazilian Listed Companies. *SAGE Open*, *7*. https://doi.org/10.1177/2158244017712027

Colquitt, J., Scott, B., & Lepine, J. (2007). Trust, trustworthiness, and trust propensity: a meta-analytic test of their unique relationships with risk taking and job performance. *Journal of Applied Psychology*, *92*(4), 909-927. https://doi.org/10.1037/0021-9010.92.4.909

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling Theory: A Review and Assessment. *Journal of Management*, *37*(1), 39-67. https://doi.org/10.1177/0149206310388419

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, *48*(1), 24-42. https://doi.org/10.1007/s11747-019-00696-0

De Jong, M. D. T., Huluba, G., & Beldad, A. D. (2020). Different Shades of Greenwashing: Consumers' Reactions to Environmental Lies, Half-Lies, and Organizations Taking Credit for following Legal Obligations. *Journal of Business and Technical Communication*, *34*(1), 105065191987410. https://doi.org/10.1177/1050651919874105

Delmas, M. A., & Burbano, V. C. (2011). The Drivers of Greenwashing. *California Management Review*, *54*(1), 64-87. https://doi.org/10.1525/cmr.2011.54.1.64

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., . . . Wright, R. (2023). "So What If ChatGPT Wrote it?" Multidisciplinary Perspectives on opportunities, Challenges and Implications of Generative Conversational AI for research, Practice and Policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

Garg, S., Haralayya, B., Qudah, A. A. M., Maguluri, P. L., Szeberényi, A., & Sameen, Z. A. (2024). The impact of artificial intelligence on management productivity and efficiency. *Business Management and Economics Engineering*, *22*(1), 424-434. https://doi.org/10.2139/ssrn.5000221

Granulo, A., Fuchs, C., & Puntoni, S. (2021). Preference for human (vs. automated) labor is stronger in symbolic consumption contexts. *Journal of Consumer Psychology*, *31*(1), 72-80. https://doi.org/10.1002/jcpy.1192

Haupt, M., Freidank, J., & Haas, A. (2024). Consumer responses to human-AI collaboration at organizational frontlines: strategies to escape algorithm aversion in content creation. *Review of Managerial Science*. https://doi.org/10.1007/s11846-024-00748-y

Hennighausen, C., Navarro-Schär, V. G. Y., & Eller, E. (2025). AI-Mediated Communication in E-Commerce: Implications for Customer Trust. *International Journal of Consumer Studies*, *49*(5). https://doi.org/10.1111/ijcs.70111

Huang, M. H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, *49*(1), 30-50. https://doi.org/10.1007/s11747-020-00749-9

Ismail, I., Sabri, S. M., Hamid, N., Khushairi, N. A. M., & Shafee, M. I. K. (2025). Reshaping Strategic Corporate Communication Practices in the Digital Era: The Role of AI-Driven. *International Journal of Research and Innovation in Social Science*, *9*(22), 328. https://doi.org/10.47772/ijriss.2025.922ileiid0032

Israfilzade, K., & Sadili, N. (2024). Beyond interaction: Generative AI in conversational marketing - foundations, developments, and future directions. *Journal of Life Economics*, *11*(1), 13. https://doi.org/10.15637/jlecon.2294

Kirk, C. P., & Givi, J. (2025). The AI-authorship effect: Understanding authenticity, moral disgust, and consumer responses to AI-generated marketing communications. *Journal of Business Research*, *186*(1), 114984. https://doi.org/10.1016/j.jbusres.2024.114984

Koning, B., & Voorveld, H. A. M. (2025). Disclaimer! This Content Is AI-Generated: How AI-Disclosures Influence Trust in Advertisements and Organizations. *Journal of Interactive Advertising*, *25*(3), 240-253. https://doi.org/10.1080/15252019.2025.2554149

Lefkeli, D., Karataş, M., & Gürhan-Canlı, Z. (2023). Sharing information with AI (versus a human) impairs brand trust: The role of audience size inferences and sense of exploitation. *International Journal of Research in Marketing*, *41*(1), 138. https://doi.org/10.1016/j.ijresmar.2023.08.011

Longoni, C., & Cian, L. (2020). Artificial intelligence in utilitarian vs. hedonic contexts: The "word-of-machine" effect. *Journal of Marketing*, *86*(1), 91-108. https://doi.org/10.1177/0022242920957347

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science*, *38*(6), 937-947. https://doi.org/10.1287/mksc.2019.1192

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review*, *20*(3), 709-734. https://doi.org/10.5465/amr.1995.9508080335

Michalec, G., Hargitai, D. M., & Bencsik, A. (2024). Organizational trust as a success factor. *Journal of Business Sectors*, *2*(1), 61-67. https://doi.org/10.62222/XQQE2812

Morhart, F., Malär, L., Guèvremont, A., Girardin, F., & Grohmann, B. (2015). Brand authenticity: an Integrative Framework and Measurement Scale. *Journal of Consumer Psychology*, *25*(2), 200-218. https://doi.org/10.1016/j.jcps.2014.11.006

Prasanna, A., & Kushwaha, B. P. (2025). Transforming marketing landscapes: a systematic literature review of generative AI using the TCCM model framework. *Management Review Quarterly*. https://doi.org/10.1007/s11301-025-00486-9

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137-141. https://doi.org/10.1007/s11747-019-00710-5

Rajaram, K., & Tinguely, P. N. (2024). Generative artificial intelligence in small and medium enterprises: Navigating its promises and challenges. *Business Horizons*, *67*(5), 629. https://doi.org/10.1016/j.bushor.2024.05.008

Sarker, I. H., Janicke, H., Mohsin, A., & Maglaras, L. (2025). SME-TEAM: Leveraging Trust and Ethics for Secure and Responsible Use of AI and LLMs in SMEs. *arXiv preprint*. https://doi.org/10.48550/arxiv.2509.10594

Seele, P., & Schultz, M. D. (2022). From Greenwashing to Machinewashing: A Model and Future Directions Derived from Reasoning by Analogy. *Journal of Business Ethics*. https://doi.org/10.1007/s10551-022-05054-9

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*(1), 333-339. https://doi.org/10.1016/j.jbusres.2019.07.039

Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human-AI interaction (HAII). *Journal of Computer-Mediated Communication*, *25*(1), 74-88. https://doi.org/10.1093/jcmc/zmz026

Szeberényi, A., Soltani, A., Najafloo, A., & Balku, R. (2025). Interactive and immersive advertising in the metaverse: User perceptions and engagement compared to traditional online environments. *Acta Carolus Robertus*, *15*(Különszám), 3-19. https://doi.org/10.33032/acr.7019

Tyrväinen, O., Karjaluoto, H., & Saarijärvi, H. (2020). Personalization and hedonic motivation in creating customer experiences and loyalty in omnichannel retail. *Journal of Retailing and Consumer Services*, *57*, 102233. https://doi.org/10.1016/j.jretconser.2020.102233

U.S. Securities and Exchange Commission. (2024, March 18). *SEC charges two investment advisers with making false and misleading statements about their use of artificial intelligence* [Press release]. https://www.sec.gov/news/press-release/2024-36

Wagner, B. (2019). Ethics As An Escape From Regulation. From "Ethics-Washing" To Ethics-Shopping?. *BEING PROFILED*, 84-89. https://doi.org/10.1515/9789048550180-016

Wahid, R. M., Mero, J., & Ritala, P. (2023). Editorial: Written by ChatGPT, illustrated by Midjourney: generative AI for content marketing. *Asia Pacific Journal of Marketing and Logistics*, *35*(8), 1813-1822. https://doi.org/10.1108/apjml-10-2023-994

About the authors

Krisztina Finta. Krisztina has been working in digital marketing for more than 10 years, with a strong focus on campaign planning, data analysis, and the application of AI-based tools. As Marketing Manager of Oszkár Telekocsi, she played an active role in both strategic and operational functions. She currently works as an independent consultant and teaches Social Media and SME Marketing at Budapest Metropolitan University, where the integration of AI solutions is a key focus of her courses.

Luca Utassy. Luca has been working as a full-stack marketing professional for many years, with a primary focus on digital and performance marketing. For nearly 10 years, she has been strengthening the teams of educational institutions, while also teaching digital marketing related subjects at Budapest Metropolitan University, Budapest International College and Budapest University of Economics and Business.

Sarolta Ács. Sarolta has been working in the field of PR and communication for more than ten years, contributing to international brand campaigns and agency projects. Her research examines how artificial intelligence is transforming trust, corporate reputation, and the functioning of the PR profession. She is currently a doctoral candidate at Széchenyi István University and a lecturer at Budapest Metropolitan University.

Csaba Dezső Dér. Dezső Dér is Associate Professor and Head of the Marketing and Communication Institute at Budapest Metropolitan University, where he has taught since 2008. He earned his PhD in history from Eötvös Loránd University in 2010. His research focuses on cultural sustainability and the role of marketing communication trends in promoting cultural consumption. In practice, he has contributed to innovative cultural projects such as the National Anniversaries publication series, the National Memory Program of the Republic of Hungary, and the "We, Hungarians" visitor and educational center.