ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DEPARTMENT OF MANAGEMENT - DISA

SECOND CYCLE DEGREE

……

THE INFLUENCE OF GENERATIVE AI ON CONSUMER PERCEPTIONS AND DECISION-MAKING: A STUDY ON PERSONALIZATION, TRUST AND ETHICAL IMPLICATIONS

Supervisor

Prof. Gabriele Pizzi

Defended by

Davide Camaioni

Graduation Session / March / 2026

Academic Year 2024/2025

Abstract

The rapid diffusion of generative artificial intelligence is transforming how firms design, personalize and deliver marketing communication. While artificial intelligence (AI) promises improvements in efficiency, engagement and targeting accuracy, its impact on consumer perceptions and behavioural outcomes remains theoretically and empirically debated. This thesis investigates the influence of generative AI on marketing effectiveness, trust, authenticity, with a specific focus on email communication, besides ethical evaluations. Integrating performance metrics with consumer psychology, the study examines whether AI-driven content and humanization cues influence perceived humanness and trust, and whether these perceptions translate into behavioural intentions. A between-subjects experimental design was implemented to test the effects of typographical imperfections and non-rounded send-time cues in promotional emails. The findings indicate that forced humanization does not significantly enhance perceived authenticity or trust, while timing precision exerts no measurable effect. Moreover, the results reveal a divergence between click intention and purchase intention, suggesting that improvements in engagement do not automatically translate into downstream conversion-related intentions. By linking AI effectiveness to perceptual mechanisms and ethical considerations, this thesis contributes to a more nuanced understanding of generative AI in marketing and highlights the importance of balancing automation, transparency and consumer trust in AI-mediated communication strategies.

Table of Contents

1. Introduction

1.1. Background and Relevance

Over the last two decades, Artificial Intelligence (AI) has shifted from a passive back-office analytical aid to an active front-stage capability that shapes how firms research markets, design offers and communicate with consumers across their journey. Systematic reviews document AI's penetration across five marketing domains (integrated digital marketing, content, experiential marketing, marketing operations and market research) with scores of concrete use cases now standard practice as recommender systems, dynamic pricing, conversational agents or creative optimization (Chintalapati & Pandey, 2022). This breadth reflects AI's evolution from rules and predictive models toward generative systems that can create novel texts, images, audios and videos, enabling firms to personalize at scale and automate more or less big portions of creative work that were previously human-exclusive (Büyüksomer & Tümbek Tekeoğlu, 2024).

Conceptually, contemporary marketing AI rests on complementary logics. First, a prediction logic underpins applications in processes such as targeting, ranking and recommendations, in which machine learning through the so called "prediction machines" reduces human and financial resources and time required to forecast outcomes (Agrawal, Gans, & Goldfarb, 2022). Second, a generative logic, which is operationalized via transformers and diffusion models, extends prediction to content generation. This step forward shifts managerial attention from "who to target" toward "what to say" and "how to say it" to each individual at the right time and place, which also involves adapting content to the mean of communication, mainly social media and email marketing campaigns (Hermann & Puntoni, 2024). Frameworks in the marketing literature usually integrate three logics. For example, Huang and Rust (2020) map "mechanical", "thinking" and "feeling" AI to marketing research, strategy (which involves segmentation, targeting and positioning) and actions, clarifying where automation versus augmentation is most likely to effectively create value (Huang & Rust, 2020). Likewise, Davenport et al. (2020) detail how AI capabilities, task types and embodiment jointly determine strategic impact, foreshadowing today's surge in consumer-facing generative tools (Davenport, Guha, Grewal, & Bressgott, 2019).

With respect to the Huang and Rust (2020) classification, communication mostly fall into the "feeling" category, involving persuasion and relational execution. In this group, Generative AI

(GAI) expands the toolkit beyond optimization to creative production and interactive dialogue. A recent systematic review highlights GAI's roles in content generation across modalities, audience analysis, chatbots and virtual assistants, real-time campaign optimization and adaptive pricing. In each one of these processes, GAI enables finer-grained personalization and faster test-and-learn cycles than traditional approaches. While classic advertising perspectives already anticipated AI's leverage along the consumer journey (attention, interest, desire, action), GAI makes such leverage programmable at scale across touchpoints (Kietzmann, Paschen, & Treen, 2018). Emerging synthesis in the field argues GAI will materially reshape how firms interact with customers, create and deliver content and even conduct marketing research over the coming decade (Grewal, Satornino, Davenport, & Guha, 2024).

These shifts are behaviourally consequential. On one hand, advances in personalized persuasion suggest Large Language Model (LLM)-authored messages tailored to psychological profiles can measurably increase persuasive impact relative to non-personalized controls, indicating that GAI can scale influence mechanisms that were previously artisanal (Matz, et al., 2024). On the other hand, consumer responses to AI-generated content are not uniformly positive. Recent evidence shows so called "human favouritism" biases evaluations toward content believed to be human made, even when AI quality is objectively high. This duality amid heightened effectiveness potential and nuanced perception dynamics makes GAI's role in consumer decision-making and brand loyalty promising and theoretically rich and at the same time an evaluative asymmetry that managers must anticipate in disclosure and brand voice decisions.

Email marketing is an especially relevant context for these developments. It remains a high-ROI (Return On Investment), text-centric channel in which incremental gains in subject-line persuasion, message relevance and CTA (Call To Action) clarity translate directly into opens, clicks and conversions, areas where LLMs and personalization models can excel (Brachten, 2025). Empirically, AI-personalized email content has been associated with higher click-through and conversion rates versus generic baselines, with organizational-level personalization sometimes outperforming hyper-personalized messages when the latter triggers perceived intrusiveness, underscoring the primary importance of calibration and trust. Complementary B2B evidence indicates that GAI-assisted campaigns can lift open, click and conversion rates by generating contextually relevant variants at scale, provided that governance (privacy, disclosure) and human oversight remain in place. More broadly, the capability to auto-generate and test high-variety message portfolios aligns with marketing's longstanding emphasis on experimentation and segmentation, now executed with unprecedented speed through AI (Davenport, Guha, Grewal, & Bressgott, 2019).

Against this backdrop, existing reviews of AI in marketing highlight its widespread adoption across multiple functions and the accelerating trend toward extreme personalization. However, these studies also emphasize the lack of evidence on the long-term consequences of such practices, particularly regarding their impact on brand–consumer relationships and consumer well-being (Chintalapati & Pandey, 2022). In addition, recent overviews of GAI and consumer behaviour distinguish “AI to predict” from “AI to generate,” urging research on valuation of AI outputs, disclosure effects and human–AI collaboration in creative tasks, central topic for an effective communication and brand loyalty formation.

In sum, AI’s relevance to marketing communications lies in its capacity to, at first sight, lower the cost of accurate prediction for targeting and timing, automate and augment creative production for message-market fit at the individual level and finally, compress experimentation cycles. These capabilities are broadly visible in email marketing, where GAI systems can personalize at scale while simultaneously posing challenges for authenticity perceptions, transparency, data management and security. This intersection enables a rigorous assessment of when and how GAI enhances communication quality and loyalty and under what circumstances such gains are ethically sustainable.

1.2. Research Problems and Objectives

The rapid integration of generative artificial intelligence into marketing communication has profoundly reshaped how firms design, personalize and deliver promotional content. Automated systems are now capable of producing human-like text, adapting messages in real time and tailoring communication at an unprecedented scale. While these developments promise substantial gains in efficiency and relevance, they simultaneously generate a series of unresolved research problems. The central question is not merely whether AI can produce marketing content, but whether its use meaningfully enhances campaign effectiveness, how it reshapes consumer perceptions and what ethical tensions it introduces into the marketing ecosystem.

From a performance perspective, the diffusion of AI tools is often justified by their potential to improve measurable outcomes such as engagement, click-through rates and conversion-related behaviours. However, empirical evidence suggests that improvements in one stage of the marketing funnel do not automatically translate into gains in downstream metrics. AI-generated subject lines, personalized recommendations or automated copywriting may increase attention and click

propensity, yet conversion outcomes remain influenced by broader contextual and relational factors. This raises a fundamental research problem: to what extent can generative AI be considered an effective driver of measurable performance improvements and under which conditions are these effects most pronounced or limited?

Beyond quantitative outcomes, the adoption of AI in marketing communication transforms the psychological dynamics underlying persuasion. Marketing messages are not evaluated solely on informational relevance, but they are interpreted through social and relational lenses. Consumers form impressions about the intentions, competence and authenticity of the sender. When content is generated or mediated by AI, these inferences may shift. Generative systems can produce linguistically sophisticated and contextually coherent messages, but recipients may still question whether such outputs are genuine, human or trustworthy. In this sense, AI use does not simply alter message production. Indeed, it may significantly alter the perceived relationship between brand and consumer. Thus, trust, authenticity and perceived humanness become central mediating constructs in understanding the broader impact of AI-driven communication.

An additional layer of complexity emerges from the fact that AI involvement is not always transparent. Consumers may not consciously consider whether a message was generated by a human or by an algorithm, yet subtle cues embedded in the communication can shape their perceptions indirectly. Minor linguistic imperfections, timing patterns, stylistic nuances or contextual framing may function as signals of human presence or, conversely, algorithmic precision. This introduces an important research tension, that is whether and how such cues influence consumer evaluations, even in the absence of explicit disclosure. The question is therefore not only whether AI is used, but how its implicit or explicit presence interacts with cognitive heuristics and social expectations.

At the same time, the increasing reliance on AI in marketing intensifies longstanding ethical concerns while introducing new ones. Hyper-personalization may enhance relevance but also risks crossing perceived privacy boundaries, generating feelings of intrusiveness or manipulation. The opacity of algorithmic systems can challenge transparency norms, while biases embedded in training data may produce unfair or discriminatory outcomes. Furthermore, as AI systems become more capable of generating persuasive and human-like content, the distinction between assistance and manipulation becomes more difficult to delineate. These issues raise broader questions about governance, responsibility and the boundaries of acceptable influence in digital marketing environments.

Against this backdrop, the present work seeks to investigate the multifaceted impact of generative AI on marketing communication by integrating effectiveness, perception and ethics into a coherent analytical framework. Rather than treating performance metrics and consumer psychology as separate domains, this thesis explores how they are interrelated. It examines whether AI-driven content and personalization are associated with measurable variations in campaign outcomes, while simultaneously analysing how these variations may be mediated by consumer perceptions of trust, authenticity and humanness. Particular attention is given to the context of email marketing, where small textual and structural elements can significantly influence engagement behaviour.

Ultimately, this research aims to contribute to a more nuanced understanding of AI in marketing. It does not assume that AI is inherently beneficial or detrimental. Instead, it investigates how its effects depend on contextual, perceptual and ethical factors. By situating generative AI within both performance-oriented and relational frameworks, this work seeks to clarify not only whether AI works, but how it works, for whom and what normative constraints are involved.

1.3. Structure of the Thesis

This thesis is organised to progressively build from foundational concepts to empirical investigation and, finally, to a critical discussion of ethical, managerial and academical implications. The overall pathway is designed to move from the broader evolution of AI in marketing to the specific mechanisms through which generative AI can influence consumer decision-making and brand-related outcomes, with a particular focus on personalization, trust and authenticity.

Chapter 2 provides the conceptual and historical background required to situate generative AI within the wider trajectory of marketing technologies. It outlines how marketing has evolved from early forms of segmentation and rule-based automation toward data-driven personalization and, more recently, to generative systems capable of producing content and interacting in natural language. The chapter clarifies what distinguishes generative AI from earlier analytical and predictive approaches, and it introduces the technological and organisational conditions that enable its adoption in contemporary marketing practice.

Chapter 3 shifts from the technological evolution to the domain of marketing communication, focusing on the concrete ways AI tools are used in campaign design and execution. It discusses the main applications of AI in content creation and message adaptation across channels, with particular attention to how generative systems support copywriting, creative variation and personalization at

scale. The chapter establishes the operational context for the empirical part of the thesis by identifying how AI enters the marketing workflow and what forms of human–AI collaboration are most common.

Building on this, Chapter 4 examines consumer-facing consequences by addressing the psychological and interpretive dimensions of AI-mediated communication. It develops the theoretical foundations for understanding the most relevant perceptive shades and mechanisms in marketing and explains why authenticity becomes especially salient when content may be generated by non-human agents. The chapter reviews how consumers infer human involvement, sincerity and value alignment, and it integrates these ideas with the broader literature on trust and perceived humanness, providing the conceptual bridge between AI adoption and consumer responses.

Chapter 5 then turns to the question of short-term effectiveness and long-term brand-consumer relationship impact, focusing on whether AI use is associated with measurable variations in marketing performance. It synthesises evidence on outcome-based metrics such as engagement, click-through rates and conversions, highlighting both the potential of AI to improve attention and relevance and the reasons why downstream outcomes may be more difficult to shift. At the same time, this chapter connects performance metrics to consumer perceptions by explaining how trust, authenticity and humanness can influence concrete actions, thereby framing effectiveness not as a purely technical matter but as a perception-dependent process. The chapter also addresses comparative evaluations of AI versus human-generated content and discusses boundary conditions under which AI-driven personalization may produce negative reactions, such as perceived intrusiveness or overpersonalization.

Chapter 6 presents an empirical study in a comprehensive manner, discussing methodological design, results and their interpretation. The study examines how consumers respond to AI-mediated email marketing communication and message-level cues that may shape perceptions when AI authorship is not explicitly disclosed. Beyond methodological transparency, the chapter reports the empirical findings and interpreting them in light of the theoretical framework developed in the previous chapters. In doing so, it evaluates how message characteristics influence perceived trust, authenticity and humanness, and how these perceptual variables relate to behavioural intention indicators such as click and purchase propensity.

Chapter 7 then expands the discussion toward future perspectives, exploring how the role of generative AI in marketing is likely to evolve in technological, behavioural and institutional terms. This chapter reflects on emerging developments such as increasingly autonomous AI agents, multimodal generative systems, real-time adaptive personalization and deeper integration between

AI and customer data ecosystems. It also considers how rising consumer awareness, regulatory intervention and shifting expectations around transparency may reshape acceptance and resistance patterns. By moving beyond the present empirical context, this chapter positions the thesis within a forward-looking framework, identifying potential trajectories for managerial practice.

Chapter 8 develops the ethical dimension of the thesis by examining the dilemmas that arise when generative AI is used to influence consumer decisions. It discusses key issues such as transparency and disclosure, data privacy and surveillance-like personalization, fairness and bias, and the risk of manipulative persuasion. This chapter situates these concerns within current debates on responsible innovation and governance, framing ethical analysis as a necessary complement to performance-oriented discussions.

The final chapter synthesizes the theoretical and empirical insights of the work, outlining key contributions, managerial implications and avenues for future research.

2. Evolution of AI in Marketing

To develop an in-depth analysis of how AI is used in marketing, it is first necessary to understand how these technologies emerged and evolved over time. This chapter therefore retraces the main stages in the evolution of AI—from early rule-based systems to statistical and predictive models, up to today's generative and multimodal architectures—and links each paradigm shift to corresponding changes in marketing practice. By following this trajectory, the chapter shows how advances in AI have progressively enabled more granular personalisation, culminating in the current focus on AI-driven email marketing, where prediction and generation converge to shape the content, timing and tone of brand communications.

2.1. Historical Development of AI

The historical development of artificial intelligence can be understood as a progression through three major technological paradigms: symbolic systems, statistical machine learning and, more recently, deep learning and generative models.

In its early decades (between the last years of 50s and early 60s), AI was dominated by symbolic and expert systems, which relied on hand-crafted rules and logical reasoning to simulate and automatise decision-making processes. While these systems demonstrated that knowledge-based inference could replicate aspects of expert reasoning, they were ultimately constrained by their rigidity and inability to scale to complex, dynamic environments (Nilsson, 2009).

This limitation gave way to the rise of statistical machine learning in the 1990s and 2000s, where algorithms began to learn patterns directly from data rather than depending on predefined rules. This shift marked an important transition, as methods such as regression, decision trees and support vector machines became central to predictive analytics across industries (Domingos, 2012).

The most transformative leap, however, occurred with the advent of deep learning, a family of neural network techniques capable of automatically extracting features from unstructured data. A pivotal breakthrough came in 2012 with the success of AlexNet, a deep Convolutional Neural Network (CNN) that dramatically outperformed prior approaches to image classification challenge (Krizhevsky, Sutskever, & Hinton, 2017). This achievement not only validated the power of deep

architectures when combined with large datasets and GPU acceleration but also ignited the “deep learning revolution” that has since defined AI research and practice.

Ever since, the introduction of the Transformer architecture in 2017 represented another paradigm shift. Unlike recurrent neural networks, Transformers relied on self-attention mechanisms, enabling models to process entire sequences of text in parallel and to capture long-range dependencies efficiently (Vaswani, 2017). This innovation made possible large-scale language pretraining, which in turn gave rise to massive models trained on diverse corpora that could be adapted to numerous downstream tasks with minimal fine-tuning, so called foundation models. Among the most influential examples are GPT-3 (Brown, 2020), an autoregressive LLM capable of coherent, few-shot text generation and diffusion models (Ho, Jain, & Abbeel, 2020), which introduced a new probabilistic framework for high-quality image synthesis.

Together, these milestones illustrate how AI has evolved from rule-based reasoning toward scalable, data-driven auto-generative and learning systems. Each step in this trajectory has progressively expanded AI’s capacity to support, and increasingly automate, human communication and decision-making processes. This progression underpins contemporary applications in marketing communication, where the move “from AI to predict to AI to generate” (Hermann & Puntoni, 2024) is enabling unprecedented levels of personalization, message creation and consumer engagement at scale (Chintalapati & Pandey, 2022).

In marketing practice, adoption has moved from rule-based automation to predictive analytics and now to generative systems that produce and adapt content across touchpoints. Systematic reviews map this diffusion across five functional themes: integrated digital marketing, content, experiential marketing, operations and market research (Chintalapati & Pandey, 2022). Within this trajectory, the 2010s emphasized prediction: CRM scoring, recommendation systems, programmatic advertising and uplift modelling, as firms learnt to exploit behavioural data at scale (Koren, Bell, & Volinsky, 2009). Reviews in the Journal of Marketing characterize this period as the rise of marketing analytics for data-rich environments, where model-based targeting and optimization became foundational decision tools for communication planning (Wedel & Kannan, 2016).

Since the early 2020s, the centre of gravity has shifted “from AI to predict to AI to generate”: LLM and diffusion models create copy, images or micro-variants, enabling personalization at scale and rapid test-and-learn in consumer-facing communications as emails, ads and chats. Hermann and Puntoni synthesize this turn, outlining how generative systems alter consumer–firm interactions and identifying research questions around disclosure, valuation of AI outputs and human–AI collaboration in creative tasks (Hermann & Puntoni, 2024). Parallel systematic reviews focused on

marketing communication document concrete applications as message generation, audience analysis, chatbots/assistants, real-time campaign optimization, adaptive pricing, and argue that GAI is transforming how firms design and deliver personalized content.

In particular, email has become a primary focus for this evolution. Recent works propose structured frameworks for deploying GAI in digital marketing workflows (model selection, training or evaluation) and situates email alongside social media as a major use case for LLM-assisted content creation and experimentation. Empirical research further indicates that AI-personalized emails can improve click-through and conversion rates versus generic messages, while also revealing boundary conditions leading to over-personalization, triggering mistrust or "creepiness" and ultimately, loss of effectiveness. These findings underscore that the technological arc toward generation must be paired with attention to consumer perception and trust, truthfulness and reliability in communication (Brachten, 2025).

2.2. From Rule Based to Language Models and Generative AI

Recalling the three basic steps of AI evolution, there are: rule-based systems, predictive/statistical models and generative AI. This progression reflects mounting data availability, the demand for personalization and the need to reduce creative bottlenecks in communication and will be discussed more in detail.

2.2.1 Rule-Based Systems

Rule-based tools encode domain knowledge as if–then rules based on sequential logical chains of orders for example, "IF cart abandoned AND total value > €50 THEN send email reminder in 2 hours". This logic mechanism yields transparent and auditable decisions but also embeds limited ability of adaptation beyond predefined cases. In practice they rely on structured inputs and hard-coded heuristics (Chintalapati & Pandey, 2022).

These systems were applied in early customer relationship management triggers, heuristic segmentation, programmatic rule trees and template-based email workflows (e.g. welcome/abandonment/renewal streams), exemplifying this phase across "integrated digital

marketing," "content," and "operations" themes mapped by the majority of marketing literature review (Chintalapati & Pandey, 2022).

The introduction of these mechanisms helped satisfying needs of consistency and compliance through the rapid automation of recurring touchpoints, even though the usage was possible only where policies required deterministic behaviours (Chintalapati & Pandey, 2022).

However, as consumer journeys grew heterogeneous and data volume alongside variety exploded, rule sets became fragile and costly to scale; the literature documents a paradi□m shift from rules to data and insight driven approaches, motivating predictive learning to handle nuance and variability (Chintalapati & Pandey, 2022).

2.2.2 Predictive/Statistical Models and (Pre-Generative) Language Modelling

Moving a step forward, supervised and semi-supervised models, also identified as early machine learning algorithms, learn mappings from historical features to outcomes (open to click, convert and churn), often complemented by language models for classification of topics or sentiment detection to turn unstructured texts into features for targeting and decisioning (Chintalapati & Pandey, 2022).

Literature catalogues the widespread adoption of these systems across six functional themes of practical use in marketing which are: recommendation systems, programmatic advertising (automated advertising placement on websites), send-timing optimization, audience propensity scoring (customers ranking based on their likelihood to positively respond to a certain campaign), creative testing/selection (automatic testing and selection of different versions of the same communication based on engagement levels) and content personalization (Chintalapati & Pandey, 2022).

Undoubtfully, predictive modelling allowed to scale relevance (right person at the right time through the right channel), improved budget allocation and enable continuous test and optimization loops in digital communications. Firms used these tools to operationalize the "5Ps" of the new so-called AI-powered marketing (Planning, Production, Personalization, Promotion and Performance) and leverage through technology the value created for customers on a personal level (Chintalapati & Pandey, 2022).

Nevertheless, despite efficiency in who/when/where decision making processes, creative message generation remained a bottleneck: textual content, design and imagery, tone and style still required

human authorship, slowing down and preventing from capillary personalization. The need to personalize not only selection but also the creation of content catalysed the move to automatic generative methods (Chintalapati & Pandey, 2022).

2.2.3 Generative AI (LLMs and Multimodal Models)

The last and current step are therefore generative models. The main technologies are LLMs used for text generation, GANs (Generative Adversarial Network) for image generation and increasingly multimodal systems that are moving beyond single-modality models toward models that can integrate text, images, audio and video in a unified way. All these mechanisms are pretrained on large corpora and then adapted to produce context-appropriate outputs given prompts and constraints. This explains the sentence “from AI to predict to AI to generate,” extending automation from decisioning to content creation (Hermann & Puntoni, 2024).

Recent reviews of marketing communications testify the use of LLM for message generation, chatbots or virtual assistants, real-time campaign optimization, audience analysis and multimodal creative (text–image–audio-video) that aligns with brand voice. In the world of LLM, email emerges as a prime use case because its text-centric, measurable and amenable to at-scale experimentation nature (Büyüksomer & Tümbek Tekeoğlu, 2024).

For instance, generative systems reduce the creative bottleneck, enabling personalization at scale, rapid A/B/n testing exploration and omnichannel consistency for voice and visuals. Some proposed frameworks in literature formalize GAI deployment steps for businesses, from model selection to prompt/finetune, evaluation and governance (Islam, et al., 2024). In particular, the MARK-GEN framework outlines the integration of generative AI into marketing communication workflows, emphasizing model selection, adaptation to brand voice, content generation, knowledge integration and governance. It demonstrates how organizations can move beyond ad-hoc experimentation toward systematic and responsible deployment of generative AI in digital marketing (Islam, et al., 2024).

Early empirical work shows LLM-crafted personalized persuasion outperforms non-personalized messages across domains, underscoring the channel-agnostic potential of GenAI to lift persuasive impact when matched to psychological profiles (Matz, et al., 2024). In facts, reviews synthesize that generative integration promises higher efficiency, stronger engagement and brand-consistent personalization. However, this industrialization of content automation also raises governance

demands and questions about its disclosure, possible biases, privacy concerns and the need for human oversight (Büyüksomer & Tümbek Tekeoğlu, 2024).

2.3. The Milestone: Introduction of Personalisation

In marketing communications, the word “personalisation” refers to the adaptation of message content, timing and format to the characteristics, context and inferred preferences of the single specific recipient. As it has been previously retraced the evolution of AI, it is easy to understand why this technology has played a major role in reshaping how personalisation is carried out in recent marketing campaigns and daily management of communication with customers. Hence, the “personalisation milestone” marks a structural shift from segment-level heuristics and static templates to data-driven, continuously learning systems that tailor what is said, how it is framed (including tone, appeal, format) and when it is delivered (Patil, 2024).

Two literature streams jointly motivate AI’s centrality in personalisation: integrative reviews that document AI’s diffusion across marketing functions and the rise of fine-grained, data-enabled targeting and content adaptation (Chintalapati & Pandey, 2022); and practice-oriented analyses of generative systems in marketing communications that explain how these tools compress creative and testing cycles (Büyüksomer & Tümbek Tekeoğlu, 2024). These two lines establish personalization as a broad, long-term trend in marketing research and at the same time explain the practical mechanics of how personalization can now happen at scale, in ways that were impossible with manual content creation.

Earlier rough rule-based personalisation models depended, as discussed before, on triggers to generate automatic actions and predictive scoring mechanisms incorporated in propensity models and recommendation systems that optimized who/when/where to contact but left what and how largely to manual creative production. As the Customer Relationship Management theory developed and datasets of clickstreams, products and contextual signals grew, statistical and machine-learning tools improved optimization but still faced a so called “creative bottleneck”: the cost and time required to generate and test many content variants for many micro-audiences was too high to make individual-specific changes (Chintalapati & Pandey, 2022). The introduction of generative systems resolves this constraint by producing personalised subject lines, body copy and calls-to-action and by automating variant generation for experimentation, thereby moving personalisation from targeting to message-level adaptation (Büyüksomer & Tümbek Tekeoğlu, 2024; Islam, et al., 2024).

The opportunity to deliver high-level personalization is not the result of a single innovation itself, but rather the gradual integration of a set of interdependent resources and capabilities. At its foundation lies data and identity management. Firms require access to clean first-party data, mechanisms to resolve identities across devices and channels which provides the raw material to get good knowledge of customers and preferences (Büyüksomer & Tümbek Tekeoğlu, 2024). Building on this base, decisioning models apply predictive analytics to determine which customers should receive which content, when messages should be delivered and which products or services are most relevant to each recipient, decisions based on propensity scoring, send-time optimization and ranking algorithms (Islam, et al., 2024). Yet, data and decisioning alone are insufficient without appropriate content systems. For personalization to be scalable, creative assets need to be modular and brand-voice guidelines must be formalized so that they can be dynamically and properly adapted by those AI models. This is where GAI becomes central: LLMs are able to produce personalised subject lines, while image generation systems adapt visuals (Büyüksomer & Tümbek Tekeoğlu, 2024). Moreover, as these capabilities expand into multimodal models, the possibility to create cohesive, cross-channel personalized content increases (Islam, et al., 2024). Finally, the cycle is reinforced by experimentation frameworks, which allow organizations to validate the effectiveness of their campaigns in advance. These tools ensure that personalization efforts are not only efficient but also empirically optimized, enabling rapid discovery of which message variants resonate most strongly among diversified audiences (Büyüksomer & Tümbek Tekeoğlu, 2024).

Using AI for personalising content has surely had a positive impact on speed, easiness and effectiveness of communication processes carried out by marketers. For what concerns speed, self-learning machines convert generating and testing processes into a near-real-time loop. Instead of weeks to craft and localize copy for multiple segments, systems now produce dozens of on-brand variants in minutes, push them into experiments and learn rapidly by themselves which tones, benefits, or Call to Action methods perform the best. This compression of creative cycle time is repeatedly highlighted as a core operational benefit of GAI for marketing communications (Büyüksomer & Tümbek Tekeoğlu, 2024; Islam, et al., 2024).

As for easiness, fully human personalisation research and creation processes used to require large and highly specialized teams made up of analysts, copywriters and designers, and a complex integration of marketing software tools and platforms to plan, execute and measure campaigns was essential. Nevertheless, prompt-based workflows, low-code editors and model-assisted templates reduce skill and coordination barriers: marketers can describe the intents and the system will return options consistent with compliance constraints. Reviews note this lowering of organizational limits as a major driver of adoption in email and adjacent channels (Büyüksomer & Tümbek Tekeoğlu,

2024). Furthermore, the reduction of the skills needed to generate inputs for AI and the reduction of highly specialized workers have made marketing research and content generation and adaptation much more accessible even to smaller businesses, fostering growth and competition.

Finally, as for effectiveness of AI in marketing communication, this is most visible in its capacity to improve the fit between messages and recipients, to expand the scope of experimentation and to preserve consistency across campaigns. First, relevance effects play a central role: when generative systems tailor copy to reflect individual characteristics or contextual cues, such as highlighting specific product benefits or framing offers appropriately, the perceived alignment between message and recipient improves. This, in turn, increases the likelihood of engagement, lifting open rates, click-throughs and eventual conversions in email and other direct channels such as push notification or direct messages on social platforms (Islam, et al., 2024; Sahni, Wheeler, & Chinta□unta, 2016). Second, exploration effects stem from the ability of AI to generate and test a far greater number of content variants than manual teams could realistically create. Generative systems make it possible to run tests at scale, rapidly identifying high-performing versions that would probably remain undiscovered otherwise (Büyüksomer & Tümbek Tekeoğlu, 2024). Finally, consistency effects arise from the model-guided constraints that help ensure brand voice and stylistic coherence across large numbers of message variants and touchpoints. Whereas heavy manual editing risks diluting identity and tone, AI systems can balance controlled diversity with overall brand alignment (Büyüksomer & Tümbek Tekeoğlu, 2024).

To sum up, personalisation operationalizes the transition from AI that predicts selection to AI that generates communication, integrating selection and content into a unified, continuously learning loop. Reviews of generative AI in marketing communications identify this integration as the essence of GAI's managerial value: faster cycles, easier execution and higher response when organizations govern data use and model their behaviour with rigor.

2.4. Focus on Email Marketing

Why is email a prime context of application for AI adoption? Email remains a permission-based (based on the consent of subscribers), text-centric and highly measurable channel in which incremental improvements in subject-line appeal, content relevance, timing and call-to-action are reflected directly in performance metrics such as open rates, click-through rates and conversions;

these characteristics make this channel particularly suitable to AI-driven optimization and experimentation (Islam, et al., 2024; Sahni, Wheeler, & Chinta□unta, 2016).

Just as all the other fields impacted by AI, historically, email programs advanced from rule-based triggers, mainly automated cart-abandonment communications and welcome streams to predictive models for segmentation, sending systems optimization and product recommendation. The recent introduction of generative AI extends to producing what is said and how it is framed for each recipient thanks to LLMs (Islam, et al., 2024; Sahni, Wheeler, & Chinta□unta, 2016).

The transformation of email marketing through AI is supported by an integrated capability stack. The base certainly lies on an effective use of data. First-party behavioural and transactional records such as browsing activity, purchase histories and prior engagement patterns form the basis for audience scoring and message selection (Sahni, Wheeler, & Chintagunta, 2016). On this foundation, after decisioning models employ predictive analytics to rank audiences, forecast open and click probabilities, these models are further complemented by optimised sending algorithms that determine when reception is most likely to capture attention (Islam, et al., 2024). Furthermore, beyond targeting and timing, the creative dimension of email has been reshaped by the use of LLMs that now assist in drafting subject lines, preheaders and body copy tailored to segment- or even individual-level cues, while image generation systems adapt visuals to campaign themes (Islam, et al., 2024). As discussed, the process is further strengthened by systematic experimentation with automated testing to quickly scale up those versions that perform best (Islam, et al., 2024). Finally, the entire system is safeguarded by governance mechanisms. Brand lexicons and “do-not-say” lists enforce stylistic and ethical boundaries and human-in-the-loop review provides oversight to mitigate risks of bias, inappropriateness, or over-personalization which will be discussed later (Sahni, Wheeler, & Chintagunta, 2016).

AI makes email communication faster by speeding creation and testing cycles, easier by lowering the expertise required to produce on-brand options via prompt-based workflows and more effective by improving message–recipient fit and satisfaction (Islam, et al., 2024). In particular, LLM-assisted personalization adjusts framing, tone and CTA to audience signals, which has been associated with higher click-through and conversion relative to generic baselines in empirical settings (Sahni, Wheeler, & Chintagunta, 2016). Complementary evidence from personalized persuasion experiments indicates that AI-generated messages tailored to psychological profiles can outperform non-personalized content, suggesting pathway-level gains (Matz, et al., 2024).

Nevertheless, boundary conditions apply and studies have proved that performance improvements are not monotonic with more data or finer granularity. Instead, over-personalization can trigger

perceived intrusiveness, thus dampening trust and engagement. Several studies recommend calibrating the degree of personalization and disclosing automatic framings when authorship cues matter (Dietvorst & Bartels, 2022; Sahni, Wheeler, & Chintagunta, 2016). Related work on authenticity perception shows that audiences may evaluate content believed to be human-created more favourably than content believed to be AI-generated, even at equivalent quality, underscoring the importance of brand voice consistency and transparent communication policies (Jago, Carroll, & Lin, 2022).

Empirical studies in the email domain support these mechanisms: AI-personalised content has been shown to outperform generic baselines on click-through and conversion metrics, but they also reveal that more personalization is not always better. In fact, calibrated personalization, for instance, tailoring content at the organizational or segment level rather than inserting highly intrusive individual-level signals, can yield superior results when hyper-personalization risks crossing into perceived "creepiness." This boundary condition highlights the need for careful managerial calibration and ongoing testing to maximize effectiveness without undermining consumer trust, which remains the main concern for marketing campaigns (Dietvorst & Bartels, 2022; Jago, Carroll, & Lin, 2022; Sahni, Wheeler, & Chintagunta, 2016).

3. AI Tools in Marketing Communications

The integration of artificial intelligence into marketing communication has reshaped how messages are created, personalized and delivered to consumers. AI technologies now operate as interconnected systems that support both large-scale content production and individualized interactions, promising gains in efficiency, scalability and performance. At the same time, the effectiveness of AI-driven personalisation depends not only on technical capabilities but also on how these tools are embedded within communication strategies and perceived by audiences. This section will now examine the integration and complementary roles of various AI technologies in creating cohesive communication ecosystems and then analyse the possible factors which pull their application according to the literature.

3.1. Integration and Roles of AI Technologies in Marketing Communications

As discussed earlier, the list of technologies applied in marketing operations is very extensive. However, these technologies are not adopted in isolation from one another. Instead, in contemporary practice, marketing communications draw on a stack of AI tools that interact with one another and form part of integrated strategies that work together.

AI has become embedded across the entire spectrum of marketing communication activities, reshaping how organizations design, deliver and adapt messages to their consumers. Rather than a single technology, literature converges on the view that these components are complementary and AI now functions as an ecosystem of complementary tools ranging from language and image generation to recommendation engines determine which products or content items to feature in marketing communications, chatbots and automated response engines determine how these recommendations are expressed and so on, each one addressing a specific communicative task while maintaining coherence and efficiency in brand messaging (Büyüksomer & Tümbek Tekeoğlu, 2024; Islam, et al., 2024).

Indirect marketing communication involves mass-mediated efforts aimed at influencing a broad audience’s awareness, attitudes and preferences toward a brand, or eliciting short-term behavioural responses. It follows that when generative models are used to produce mass-mediated marketing

content, their function aligns with indirect marketing communication (Grewal, Satornino, Davenport, & Guha, 2024).

Among generative tools, LLMs automate and augment the drafting of marketing copy across channels. They allow to write subject lines and bodies in email communications, web or app copy and captions for social media promotional campaigns, brand awareness or brand engagement, enabling tone control, style transfer and audience-specific framing from prompts or brief guidelines. Foundational work demonstrates that scaling autoregressive transformers as GPT-3, affords strong few-shot performance and controllable text generation, which marketers repurpose for ideation, variant creation and on-brand rewriting. Recent reviews in marketing synthesize these applications by positioning generative AI as a content engine that supports personalization (Brown, 2020).

In parallel, diffusion-based image generators and their latent variants now underpin fast production of campaign visuals in social posts, thumbnails and creative adaptations of format, background and scene edits, lowering cost and cycle time. Recent advances in computer vision, particularly the development of “Denoising Diffusion Probabilistic Models” (DDPMs) described by Ho and their optimized “Latent Diffusion Models” (LDMs) variants as explained by Rombach, have enabled the generation of highly realistic and stylistically coherent images from the extraction of textual prompts. These models progressively transform random noise into detailed visual outputs within a compressed latent space, significantly improving efficiency and image quality. As a result, diffusion-based architectures allow practitioners to quickly produce high-fidelity visuals, perform seamless image editing (“inpainting”) and generate content that consistently aligns with brand aesthetics across campaigns (Ho, Jain, & Abbeel, 2020; Rombach, Blattmann, Lorenz, Esser, & Ommer, 2022).

Another important side of marketing communications involves direct, one-to-one and timely communication aimed to satisfy immediate customers’ communication needs. AI-powered chatbots and automated email response systems constitute forms of direct marketing communication, as they deliver personalized and interactive messages to identifiable consumers. Their purpose is to facilitate engagement and continued interaction with the brand (Adam, Wessel, & Benlian, 2021).

Automated conversational agents and chatbots for customer service and sales management handle responses to queries, guided selling and post-purchase support via webchat, messaging apps or email handoffs, thereby extending service hours and standardizing response quality. Academic research in human-computer interactions and service marketing has shown that chatbots are now key tools for providing information, assisting purchases and offering customer support. Users mainly value them for their efficiency. Although their effectiveness still depends on factors like

disclosure and perceived empathy, chatbots have become central instruments in direct, consumer-facing marketing communication (Adamopoulou & Moussiades, 2020).

Likewise, automatic email responses and assistive messaging systems generate short, context-appropriate responses and drafts, improving email throughput and consistency while preserving user control and optimization for efficiency and personalization. These systems illustrate how sequence-to-sequence and feed-forward architectures can support cheaper and timelier email communications at scale as assistive authoring tools (Kannan, et al., 2016).

Hand in hand with automatic email generation, recent research conceptualizes subject-line generation as a Natural Language Processing (NLP) task, where AI models are trained to craft informative and attention-capturing subjects based on the email body or campaign metadata. At the same time, complementary studies on send-time optimization employ predictive analytics to determine the most effective delivery moments for each recipient, aligning message timing with individual engagement patterns. Other models further refine this process by regulating the frequency of contact, ensuring that communication remains consistent and at the same time non-intrusive. Together, these AI-driven systems seek to enhance engagement by coherently matching both the content and timing of email delivery to each user's behavioural context (Zhang & Tetreault, 2019; Araújo, et al., 2022). (Araújo)

Finally, used in both direct and indirect marketing, recommendation engines determine what products and offers to communicate by learning user–item affinities from interactions and content signals. They serve as the backbone for shaping personalized emails, on-site placements, digital spaces, display points and notification triggers. Scholarly references document the methods, such as collaborative filtering, factorization, deep candidate generation and ranking, and their industrialization at scale by many multiproduct companies, multisided platforms and big digital ecosystems as Netflix, Amazon or YouTube, establishing how recommendation layers feed downstream messaging and creative selection (Ricci, Rokach, & Shapira, 2021).

3.2. The Effectiveness Logic for AI Application in Marketing Communications

As seen in the previous chapter, a growing body of research converges on several, mutually reinforcing logics that explain why marketers embed AI into communications workflows: performance optimization, acceleration of experimentation and operational efficiency and

scalability, maintaining brand coherence across channels. Together, these logics connect AI deployment to measurable outcomes such as engagement, click-through and conversion rates while laying foundations for longer-run effects on satisfaction and brand loyalty (Nesterenko & Olefirenko, 2023).

In general, by automating labor-intensive steps such as drafting and localise variants, AI expands the feasible frontier of content volume without proportionate increases in headcount or cycle time, an effect that general reviews on AI strategies highlight as central to adoption (Kshetri, Dwivedi, Davenport, & Panteli, 2024). This translates into more frequent refreshes, finer audience cuts and continuous campaign tuning, correlated with engagement and conversion lift when governed by robust testing protocols (Patil, 2024).

Also, generative systems can be constrained with style guides and guardrails so that any variation remains on-brand, reducing drift across touchpoints while still allowing contextual adaptation. This controlled diversity supports coherent experiences that contribute to recognition and trust over time and delivery methods. Reviews explicitly set this as part of the engagement logic in which consistent but customised content sustains long-term interaction quality (Patil, 2024).

Moreover, AI reduces marginal cost and time of producing variants for experimentation, enabling A/B tests among others and rapid subsequent exploitation of winners. This compresses the learn–measure–iterate cycle and turns communications into a continuous optimization process, moving value creation from static targeting to dynamic message optimization for commercial outcomes (Patil, 2024).

However, the dominant rationale is the promise of improved communication's performance realised by AI. Relying on big amount of data, AI systems infer interests and contexts from behavioural and content pieces of information to tailor what is said, what is offered, how it is framed and when it is delivered. Evidence from GAI research shows that personalized messages crafted by LLMs can yield greater attitudinal and behavioural influence than non-personalized baselines, implying direct lifts in performance metrics when personalization is well-calibrated to the audience and task (Matz, et al., 2024).

In channel-specific studies the consensus is that AI-assisted communication personalization could improve standard engagement metrics in marketing campaigns, but these gains are contingent on careful design, calibration and governance. In short, the literature anticipates possible lifts in open, click-through and conversion rates and broader engagement relative to generic content since AI raises content–recipient fit, underlining its suitability for targeting and message adaptation (Brachten, 2025).

Personalization in subject-line appeal and tone matching assisted by LLMs are linked by the literature to higher attention and intent to explore, implying potential improvements in subsequent actions. These effects are expected to be stronger when personalization cues are appropriate and non-intrusive and when copy variants have already been tested to discover audience-specific resonances (Islam, et al., 2024).

Accordingly, by aligning offer, product recommendation and call-to-action with inferred preferences and by sequencing content to reduce friction across the path to purchase, AI-generated messages could translate incremental customers' attention into concrete purchase actions. Studies on personalized persuasion techniques more broadly report that tailored messages produced by LLMs are more influential than non-personalized baselines, which supports the possibility of conversion gains when such tailoring is applied to commercial contexts (Matz, et al., 2024).

As of engagement, generative systems can maintain content freshness and contextual relevance across touchpoints (email, ads, on-site), potentially strengthening ongoing engagement (repeat opens, repeat clicks, time-on-site) by reducing message fatigue and improving perceived usefulness over time. Reviews frame this as a shift from one-off optimization to continuous learning loops, thereby supporting sustained interaction rather than episodic spikes (Büyüksomer & Tümbek Tekeoğlu, 2024).

On the other hand, as mentioned in the previous chapter, the same literature cautions that effects are moderated by perceptions of authenticity, appropriateness and trust. Over-personalization or ill-judged disclosure of AI authorship can dampen the positive effect and translate in a harm instead with lower willingness to share or reduced loyalty intentions. This insight implies that any potential uplift in actions or loyalty depends on calibrating personalization depth and marking AI involvement carefully. In other words, possible gains can only coexist with possible penalties as well, when messages feel manipulative or inauthentic (Brüns & Meißner, 2024; Dietvorst & Bartels, 2022; Kirk & Givi, 2025).

Taken together, these findings support improvements in performance metrics and synthetise the positioning of GAI as a mean to intensify consumer engagement through content-level tailoring rather than solely audience selection (Patil, 2024). Nevertheless, a prudent empirical approach is to pair AI-generated variants with rigorous testing and to measure not only response metrics but also perceived appropriateness and authenticity as moderators of performance (Brüns & Meißner, 2024).

4. Consumers Perceptions and Behavioural Impact

The increasing use of AI in personalised marketing communication is changing how consumers interpret and evaluate messages. Beyond technical performance, audiences form judgments about credibility, sincerity and intent, often relying on subtle cues that shape trust, authenticity and perceived humanness. These perceptions are particularly salient in communicative contexts that are relational or identity-relevant, where AI involvement can generate both appreciation and resistance. This chapter examines how such evaluations emerge and why consumer responses to AI-driven personalisation remain ambivalent, oscillating between efficiency gains and concerns about authenticity.

4.1. Trust in AI-Driven Personalisation

Human trust in AI can be defined as the willingness of individuals to place themselves in a position of reliance and vulnerability toward an AI system, based on the belief that the system will act in ways that fulfil their needs and expectations of performance, reliability, safety and transparency. It reflects not only the system's technical trustworthiness but also non-technical cues such as reputation, documentation and institutional accountability. In this sense, trust in AI is not merely passive acceptance, yet active readiness to depend on AI outputs (decisions, recommendations, predictions and personalised communications) without needing constant monitoring or verification. Importantly, trust is also domain-specific: different applications (healthcare, finance or marketing) require different thresholds of trust, because the stakes and vulnerabilities differ (Afroogh, Akbari, Malone, Kargar, & Alambeigi, 2024).

As demonstrated by Castelo, his research on algorithm aversion and appreciation shows that trust in AI depends strongly on the type of task. The investigation found that consumers rely relatively comfortably on algorithms for tasks and topics which they consider to be objective and analytical but trust them less for those which involve a certain degree of subjective or taste-based judgment. This turns out to be true even when algorithms are more accurate than humans (Castelo, Bos, & Lehmann, 2019).

The research first categorized 26 topics into subjective and objective as perceived by the sample's population. As an example, within the subjective category fall topics as writing news, gifts, music

and movies recommendations, while within the objective category fall topics as stock prices predictions, weather forecasts, data analysis or election results predictions (Castelo, Bos, & Lehmann, 2019).

Then, the researchers tested the respondents by using the CTR (Click Through Rate) as an estimation for evaluating their level of trust in human vs algorithmic advice provided in Facebook advertisement. The results highlighted the sample trusted human-advised advertisements more than algorithmic-advised ones (as clearly stated in each advertisement) with a CTR of 2.1% versus 0.6% respectively, with a high level of statistical significance (p-val < 0.001) for dating suggestions, as considered to be a subjective topic. Conversely, although the sample still showed that for financial recommendations, as considered to be objective in nature, the CTR remains higher for those generated by specialized humans rather than by algorithms (1.8% versus 1.6% respectively), this is only marginally true, since the significance level of the difference shows a p-val of 0.092 (Castelo, Bos, & Lehmann, 2019).

Finally, in the same study the researchers found that testing the same subjective topics while giving information about the benefits of relying on data based recommendations increased the CTR from 0.39% to 0.87% in those subjective advice realized by algorithms (Castelo, Bos, & Lehmann, 2019).

Indeed, a complementary stream of evidence on how individuals rely on algorithmic versus human judgments comes from Logg's research on what he defined to be "algorithm appreciation". Across three experiments that systematically varied task characteristics, the authors demonstrated that people tend to rely more heavily on algorithmic advice rather than on human advice, even when both sources provide identical inputs and about either subjective or objective topics as identified by Castelo (Logg, Minson, & Moore, 2019).

In his first experiment, participants judged a person's weight from a photograph and revised their estimate after receiving advice labelled either as human- or algorithm-generated. Participants relied significantly more on the advice when it was framed as algorithmic rather than human. As a measure of trust, they used the WOA (Weight On Advice), measured as the weighted difference between the initial respondent's estimation and the adjusted respondent's estimation after receiving the advice. The analysis showed a WOA of 0.45 for the algorithm advice and 0.30 for the human advice, with an acceptable level of statistical significance (p-val of 0.003) (Logg, Minson, & Moore, 2019).

Likewise, the second experiment still demonstrated a higher reliance on algorithmic suggestions over human about a more subjective and taste-based domain, that is predicting the popularity of

songs. Once again, participants relied more on algorithmic advice with a WOA 0.37 than on human advice with a WOA 0.21 and an acceptable significance level (p-val<0.001), despite both forms of advice being identical (Logg, Minson, & Moore, 2019).

Finally, the third experiment tested predicting romantic attraction, a pure interpersonal domain, and found that participants continued to place significantly more trust in algorithmic suggestions. In the algorithmic case the results indicated a WOA of 0.38, against a WOA for the human suggestion of 0.26, with an acceptable level of significance highlighted by a p-val of 0.001 (Logg, Minson, & Moore, 2019).

However, Dietvorst and Bartels add a moral dimension: consumers object to algorithms especially when they make morally relevant trade-offs, because they are seen as using a cold maximisation strategy that feels inappropriate in ethical domains (Dietvorst & Bartels, 2022).

They demonstrate that consumers' trust in algorithmic systems is also constrained by moral acceptability. Their research shows that people intuitively associate algorithms with maximisation-based decision processes, meaning that algorithms focus exclusively on producing the best expected outcome according to a given criterion, disregarding whether the process through which the outcome is reached violates moral rules or social norms (Dietvorst & Bartels, 2022).

Because many individuals view consequentialist, maximising decision strategies as ethically problematic in morally sensitive contexts, they become more skeptic of algorithmic involvement in such domains as the tradeoffs that a health insurance company makes as pursued in their studies. In the first study, they found that participants systematically believed the algorithm was more likely to use maximization as a decision strategy than a human in each one of the seven dimensions investigated (advertisement, agent, plan, premium, prescription, recommendation and renewal), with an overall mean 0.83 points out of 5 higher for the algorithm's likelihood to use maximisation than humans' ratings. These results support the hypothesis that people believe that decisions made by an algorithm have more of a basis in maximization rather than decisions made by humans (Dietvorst & Bartels, 2022).

At the same time, in their second study they examined whether people are less tolerant of algorithmic decision-makers when the decision domain is morally relevant. Results from the second study highlighted that consumers manifest stronger objections to algorithmic decision makers when they perceive that the decision in question is highly morally relevant. The study tested the objection rate through participants' likelihood to switch health insurance company when it declared to be going to use an algorithm to make decisions on the same seven dimensions included in the previous study. The results confirmed that participants who felt more moral conviction for their assigned

decision reported stronger switching intentions with a statistically significant (p-val<0.001) correlation rate (r) of 0.42 between participants' ratings of moral conviction and switching intention (Dietvorst & Bartels, 2022).

Taken together, these streams of evidence confirm that consumers' trust in AI-driven communications remains fundamentally ambivalent. Many individuals are willing to rely on algorithmic and personalised outputs in everyday practice and, in some cases, even place more weight on algorithmic input than on equivalent human advice across diverse tasks ranging from perceptual judgments to cultural preferences and interpersonal inferences.

Yet this reliance is clearly conditional: it is strongest when AI operates in low-stakes, instrumental and predominantly objective domains as numerical forecasting and when personalisation is perceived as clearly beneficial and supportive of users' goals. Conversely, trust is withheld or becomes fragile when algorithms are deployed in morally sensitive, intimate or identity-relevant domains, particularly where underlying decision logics resemble "cold" maximisation strategies that appear to disregard moral rules, social norms or users' sense of authenticity and privacy.

In this sense, trust in AI does not hinge solely on accuracy, performance or transparency but also on whether the decision process and trade-offs enacted by the system align with lay moral intuitions about what constitutes an acceptable way of reaching an outcome. This implies that contemporary consumers can and do rely heavily on AI-based personalisation in digital marketing contexts yet, such reliance remains contingent and fragile: it is maximised when AI is framed and experienced as a tool that augments human communication in a controllable and value-congruent way and may erode when AI appears to displace human agency, operate opaquely on personal data or cross perceived boundaries of authenticity and moral acceptability.

4.2. AI Authenticity Perceptions

Linked to the concept of trust, authenticity emerges not as an inherent property of a message but as a subjective perception constructed by audiences. In the context of digital and AI-mediated communication, authenticity refers to the extent to which content is perceived as real, credible, sincere and grounded in truthful or original expression. As outlined by Gilpin, authenticity is fundamentally relational: it depends on audiences' evaluations of attributes such as authority, fidelity, origin, credibility and sincerity, which shape whether communicators are regarded as genuine and trustworthy (Gilpin, Palazzolo, & Brody, 2010).

From a consumer-evaluation perspective, authenticity also encompasses truthfulness and reality, understood as correspondent to the real world, as well as originality and sincerity. All of these are qualities that consumers associate with human creativity, emotional depth and the use of appropriate techniques or processes in producing content. When these expectations are met, audiences tend to form more favourable evaluations and stronger positive responses toward the producer of the content (Han, 2024).

Jago further emphasise that authenticity is not a fixed attribute, but a perception shaped by contextual cues, including expectations about authorship, the nature of the task and the transparency of the production process. In automated or AI-supported work, audiences rely on signals of identity, transparency and engagement to judge whether the output still reflects a sincere and legitimate expression of the creator, thereby highlighting how authenticity is socially constructed through interaction between producers, technologies and audiences (Jago, Carroll, & Lin, 2022).

Altogether, these sources portray authenticity as a socially and perceptually constructed judgement about whether a content feels truthful, sincere and credibly rooted in an identifiable and legitimate creative or communicative process, a judgement that becomes especially salient when AI participates in content creation.

Drawing on prior authenticity findings, Jago's research argues that people often infer authenticity from biographical narratives about creation, that is who made something, why they made it and how this reflects their values and efforts, and that these logics extend to automated work as well. The researchers show that people are hesitant to attribute authenticity to autonomous systems by default, particularly in domains where they characteristically expect human authenticity, such as poetry or art (Jago, Carroll, & Lin, 2022).

In their first study the researchers wanted to test two measures of correlation between automated agents and authenticity perceptions. They hypothesised that if an automated agent is somehow seen as being connected to humans (created or cared for specific people), its work might be judged as more authentic. So as if people see the agent as having a human-like mind including emotions, intentions and some sort of consciousness, its work might also be judged as more authentic. The respondents were then asked to evaluate an automated agent on several dimensions linked to one specific out of twenty application domains (as designing logos, composing music, grading essays or helping with dispatching) to which automation is or could be applied. Therefore, the results showed significant ($p\text{-val} < 0.001$) levels of correlation between both "connection" and perceived authenticity, and "humanness" and perceived authenticity, with values of 0.44 and 0.43 respectively (Jago, Carroll, & Lin, 2022).

Among others, an insightful result comes from the second study, realised after manipulating the previous experiment's cues. The researchers found that when the AI is explicitly disclosed to be designed by a specific person, with a bit of information about that person's background and skills, the authenticity perception about that agent significantly (p-val < 0.007) increases on average (Jago, Carroll, & Lin, 2022).

A more diagnostic test comes from Study 4. Participants used an online algorithm to generate a haiku and then evaluated the poem. They hypothesized that in ordinary circumstances users may not consciously consider the human creators behind automated systems. However, when they are prompted to do so, perceptions of human connection are reinforced and the agent's work is evaluated as more authentic, even without explicitly stating the involvement of a human creator. The researchers did this by asking five questions to the respondents about their perceptions about the creators in order to indirectly drive them to think about the human developer's skills and interests behind the algorithm. Participants rated the AI-generated poem as being more authentic with an average evaluation of 4.57 in a 1 to 7 scale, when subjected to human presence clues. When no clues were given the average decreased to 4.10. The study confirmed (p-val = 0.032) that human origin clues are able to drive people's authenticity perceptions on algorithmic works (Jago, Carroll, & Lin, 2022).

Taken together, studies 1, 2 and 4 demonstrate that authenticity judgments are not inherent or fixed properties of AI-generated outputs but rather malleable evaluations that depend on how the system and its origins are framed to users. Across these studies, subtle differences in presentation, ranging from explicit origin information that identifies a human creator to minimal prompts that merely encourage users to infer human involvement, systematically alter perceptions of an automated agent's connection to people. In turn, this perceived connection operates as a psychological mechanism that elevates authenticity judgments (Jago, Carroll, & Lin, 2022).

In other words, people can see AI-generated outputs as authentic, but mainly when they can anchor them in human origins and values rather than viewing them as detached, purely mechanical products. Users evaluate the very same AI output differently when they are encouraged, either directly or indirectly, to imagine the individuals, intentions or expertise behind the system. These findings collectively indicate that authenticity is not solely a function of technical output quality but also of the social and cognitive framing surrounding the AI, highlighting how even light-touch interventions can shift evaluations in a meaningful way.

Other studies show a slightly different reality in social-media and branding contexts. In these cases, authenticity is a central evaluative lens and generative AI triggers more ambivalent reactions. Brüns

and Meißner examined how followers react when creators adopt GAI to produce social media content.

In their first experiment, they asked the respondents to think about their favourite brand followed on social media and then asked one group to think said brand to begin using GAI to create certain social media content in all their steps and another group to think said brand to refuse using GAI to create social media content in any step. The hypothesis behind this experiment was that adopting GAI to automate content creation attenuates favourable follower reactions, including credibility, brand attitudes, EWOM (Electronic Word of Mouth) intentions and brand loyalty compared to the no adoption situation. Indeed, in the automation scenario, perceived post credibility, EWOM intentions and brand loyalty means (all three mediated by the negative effect on perceived brand authenticity) were significantly (all p-val < 0.001) lower than in the no adoption scenario, therefore supporting the hypothesis (Brüns & Meißner, 2024).

In the second study, the researchers illustrated how brands can use GAI applications to create marketing content and focused directly on disclosure. Consistently, this study confirmed that participants had significantly lower perceptions on post credibility and brand attitudes (both p-val < 0.02), again mediated by a reduction on perceived brand authenticity, when it was said that the visualised post was realised using GAI compared to the no adoption disclosure case. EWOM showed no significant differences instead, possibly due to the respondents' missing relationship with the fictitious brand and thus the lower involvement. Moreover, participants could not distinguish between artificial or human content, unless the description was clear. Indeed, despite it was exactly the same AI generated content, when the content was disclosed as AI-generated, authenticity perceptions of the fictitious brand and subsequent responses were more negative rather than when the content was labelled as human-made. It follows that the sample was unable to notice a perceptual difference between the two content (Brüns & Meißner, 2024).

These findings echo broader evidence from art and communication. Experimental work on AI-generated artworks shows that when images are labelled as being AI-created rather than created by humans, participants tend to rate them to be less authentic, profound and worthy, even though they cannot always tell AI and human works apart in blind tests (Bellaiche, et al., 2023; Van Hees, Grootswagers, Quek, & Varlet, 2025).

A third study was conducted to test an intermediate case: the case of AI assistance rather than full adoption. This experiment's conditions are the same as the first one. Participants were told to imagine that their favourite influencer on social media began using GAI to assist them in content creation, meaning that for some posts, certain steps would have been performed by GAI. While for

the other two cases of complete adoption and non-absolute adoption the statements of experiment 1 remained valid. The researchers confirmed that the case of complete automation, compared to that of absolute non-adoption, significantly (p-val < 0.001) reduced all the previously investigated metrics (brand authenticity, EWOM intentions and brand loyalty) through an indirect effect via perceived brand authenticity. Similarly, the case of assistance, compared to no adoption of any AI, also obtained a significant (p-val < 0.001) attenuating effect on the aforementioned metrics through the effect of reduced perceived brand authenticity. Finally, the possibility of exclusive use of AI similarly significantly reduces (p-val < 0.001) said metrics compared to the case of its use as a mean to simply assist content creation (Brüns & Meißner, 2024).

Summarizing all these Brüns and Meißner's work's findings, the deeper and stronger is the role of GAI in content creation, from no adoption to assistance and total automation, the more negative were the subsequent reactions. Also relevant is that the effect of AI application on digital content creation has been fully mediated by the perceived feeling of authenticity in any situation on each one of the effects on credibility, EWOM intentions and brand loyalty.

In conclusion, literature therefore converges on several aspects of how audiences perceive authenticity in AI-generated content. First, human involvement works as a key authenticity signal: content that is framed as the outcome of human creativity, emotion or effort, even when it is still produced by AI tools, tends to be seen as more authentic than content framed as fully automated. Origin stories that highlight human creators, their motivation and expertise increase perceived authenticity of AI outputs which, in turn, acts as a boost individuals' willingness to engage with them (Jago, Carroll, & Lin, 2022). Second, perceptions of authenticity are highly sensitive to labeling and disclosure rather than to sensitivity to artificially created content itself. Informing consumers that a content is fully AI-generated usually reduces related authenticity perceptions, particularly in creative or identity-relevant domains, whereas presenting AI as an assisting or collaborative tool has smaller and sometimes even negligible negative effects (Brüns & Meißner, 2024). Third, authenticity concerns are very domain-specific. In functional tasks with low emotional stakes audiences care relatively less about authenticity and more about accuracy or convenience; in creative, intimate or morally loaded contexts, the authenticity of the "voice" behind the message becomes much more salient (Han, 2024; Ja□o, Carroll, & Lin, 2022).

4.3. Humanness Attribution in AI-Generated Content

Humanness attribution refers to the inference that a message (text, image, audio) was produced by a human rather than an automated system, based on perceived cues such as stylistic variability, emotional tone, narrative specificity and apparent “effort” or intentionality (Jakesch, Hancock, & Naaman, 2023). It has been already discussed how in AI-mediated marketing communications, this attribution matters because it affects downstream judgments about trust, authenticity and ultimately, persuasiveness and subsequent behaviours. As discussed, this often happens independently of whether recipients can accurately identify the true source and a declaration of belonging to a human or artificial origin is sufficient to deceive the subjects' perceptions. In other words, the feeling of a human-generated content acts as a credence attribute (Brüns & Meißner, 2024). In practice, people frequently hold robust beliefs that AI outputs are less genuine or less worthy, strongly structuring evaluations when inferred or clearly disclosed. Nevertheless, AI intervention is rarely verifiable in real time and people’s actual ability to discriminate AI from human content is frequently weak, context-dependent and driven by unreliable surface heuristics (Jakesch, Hancock, & Naaman, 2023).

This effect is visible across different modalities. The first evidence of human preference emerges in aesthetic and creative domains, where attribution labels often matter more than intrinsic detectability. Bellaiche performed an experimental work on AI-generated artworks which showed that when images are labelled as being AI-created rather than by humans, participants tend to rate them to be less authentic, profound and worthy (Bellaiche, et al., 2023).

In the study, they submitted 30 images to each participant, randomly labelling half of them as being AI-generated and the remaining part as human-generated, although all of them were in fact artificially generated. Importantly, since the stimuli were AI-generated in actuality, any consistent preference for human-labelled works cannot be explained by genuine detection skill, but rather by belief-driven evaluation. Each image was then assessed by every participant across four dimensions: liking, aesthetic appeal, perceived profundity or meaningfulness, and perceived monetary value. T-tests performed on the results showed significantly higher ratings for human-labelled over AI-labelled images for all four criteria with a p-val below 0.05 in all categories, demonstrating a robust labelling penalty in evaluation when identical artworks are randomly assigned to human or AI labels (Bellaiche, et al., 2023).

Having understood this distinction, in parallel, Jakesch, Hancock and Naaman examined whether people can accurately detect AI-generated self-presentations in different contexts and experiments. Participants were presented with short written self-presentations and asked to classify each text according to its presumed authorship (human or artificial). The initial setup established a baseline classification task under standard conditions within three communicative contexts: professional,

hospitality and dating self-presentations. Subsequent experiments systematically modified the task by introducing monetary incentives for correct classifications and by providing trial-by-trial feedback, in order to assess whether increased motivation or learning opportunities affected participants' judgments. Overall, as a result, accurate identifications across every context and treatment represented 51.7% of total guesses, confirming that participants could not detect self-presentations generated by the AI language models beyond chance level in any experiment (Jakesch, Hancock, & Naaman, 2023).

However, researchers observed that participants were not guessing randomly, yet they converged on shared heuristics, even though these heuristics were largely miscalibrated with some cues not being diagnostic at all and others pushed judgments in the wrong direction. These heuristics were mainly based on intuitive cues related to content, grammar and tone. In most cases they were references to personal experiences, family, first-person narration and a warm or emotionally expressive tone were commonly interpreted as signals of human authorship. On the other hand, impersonal, monotonous or generic language were associated with artificial texts (Jakesch, Hancock, & Naaman, 2023).

Nevertheless, subsequent computational and labelling analyses showed that most of these cues were poorly aligned with actual authorship, either human or AI, with only a limited subset, such as nonsensical or repetitive content, being genuinely more prevalent in AI-generated texts. Overall, this resulted in consistent but inaccurate authorship judgments (Jakesch, Hancock, & Naaman, 2023).

This pattern implies that in everyday "non-disclosed" settings, consumers may feel they are detecting AI, while in practice they are mostly reacting to stylistic features that models can learn to mimic. Indeed, the same work shows that AI outputs can be optimized to exploit these heuristics. The researchers deliberately modified AI-generated texts to incorporate features that were previously found to be commonly associated with human writing. Validation experiments show that participants evaluated the optimized AI-generated self-presentations as more human than the human-written and the nonoptimized AI-generated ones. Optimized presentations were rated as human in 65.7% of cases against 51.6% for regularly generated ($p\text{-val} < 0.0001$) and 51.7% for human-written ($p\text{-val} < 0.0001$) presentations (Jakesch, Hancock, & Naaman, 2023).

Together, these experimental manipulations allow the authors to examine not only whether participants can distinguish between human- and AI-generated language, but also how such judgments are formed under different informational and motivational conditions. At the same time, these findings align with a broader conclusion that is, when source is not disclosed, evaluations may converge on the content's surface appeal. For AI-assisted marketing, the implication is direct:

recipients' humanness attributions are often based on predictable authenticity cues, which can be manufactured at scale, allowing artificial content to appear convincingly human.

At the same time, evidence suggests that detectability is modality-specific rather than uniformly impossible. In a study about visual art, Van Hees et al. separated preference without labels and discrimination ability. The researchers designed two experiments to examine how people perceive and evaluate human-made versus AI artworks. The stimuli set consisted of matched pairs of images, each pair including one human-made painting and one artificially generated image produced with DALL·E 2 AI, carefully matching style and composition (Van Hees, Grootswagers, Quek, & Varlet, 2025).

In the first experiment, participants viewed the paired images without any information about their origin and were asked to indicate which artwork they preferred, allowing aesthetic judgments to be formed without authorship cues. Without any origin information, observers preferred the AI-generated over human-made artworks significantly (p-val < 0.0001) more often, above 50% chance level. In the second experiment, a separate group of participants viewed the same image pairs but were explicitly informed that one image in each pair was artificial and were asked to identify which one. Again, observers could identify which image was AI-generated significantly (p-val < 0.0001) above 50% chance level (Van Hees, Grootswagers, Quek, & Varlet, 2025).

This design allowed the authors to isolate aesthetic preference from authorship detection while holding the visual stimuli constant. Notably, the authors found a positive correlation between the paintings that were more preferred and those that were more AI-detectable, suggesting that the very features making artificial images more appealing may also create detectable regularities. This indicates that in non-disclosed contexts, people may sometimes detect "artificiality" in images better than in text, but this does not □uarantee ne□ative reactions; absent labels, preference can even tilt toward AI outputs (Van Hees, Grootswagers, Quek, & Varlet, 2025).

In addition, in text-based creativity, Köbis and Mossink provided experimental evidence that participants cannot reliably differentiate AI-generated from human-written poetry, reinforcing that "human-like" language can pass casual scrutiny even when people hold strong beliefs about what real human creativity should look like. Taken together, these findings suggest a general mechanism relevant to marketing: when AI use is not disclosed, consumers often cannot reliably detect it, especially in text, but once AI authorship becomes salient through disclosure, suspicion or labelling, evaluations may shift sharply due to beliefs about authenticity, intention and deservingness rather than changes in perceived quality (Köbis & Mossink, 2021).

Overall, this body of evidence indicates that humanness attribution operates less as a function of reliable source detection and more as a belief-driven evaluative mechanism. Across textual and visual domains, individuals consistently rely on intuitive cues to infer human versus artificial origin, yet their actual ability to discriminate AI- from human-produced content is often limited, uneven across modalities and shaped by heuristics that can be strategically reproduced by generative systems.

When source information is absent, evaluations tend to privilege surface qualities such as fluency, coherence and aesthetic appeal, allowing artificial outputs to perform on par with, or even exceed, human content. Conversely, when AI authorship becomes salient through disclosure, evaluations frequently shift towards moral criteria, including authenticity, intentionality and deservingness, producing penalties independent from objective quality.

Taken together, these findings suggest that humanness functions as a fragile and manipulable signal in AI-assisted communication. It plays a central role in structuring trust and authenticity perceptions. For AI-assisted marketing, this implies that perceived humanness can be systematically engineered at scale, underscoring the importance of balancing between disclosure and artificial design to improve effectiveness and perceived quality of AI-assisted or AI-generated content.

5. Effectiveness of AI-Driven Marketing

The diffusion of generative AI is reshaping marketing communications by altering how messages are produced, targeted and evaluated, with consequences that extend from short-term campaign performance to longer-term brand relationships. At the operational level, AI enables scalable personalization and rapid experimentation, potentially improving attention and engagement metrics such as opens and clicks while also influencing harder-to-move outcomes closer to economic value, including conversions and sales. Yet, effectiveness is not uniform: empirical evidence suggests that performance gains depend on context, product type, stage of the funnel and may be constrained by boundary conditions such as overpersonalization and privacy-related reactance. Beyond performance, AI also operates as a cue that can shift consumers' psychological inferences about intent, authenticity and human involvement, thereby affecting trust and the perceived legitimacy of persuasion attempts. These perceptual mechanisms help explain why similar content can generate different responses depending on whether its origin is salient and how it is framed. Overall, AI's impact in marketing communication is best understood as a set of contingent effects in which measurable metric shifts, consumer perceptions and governance choices interact to determine both immediate results and relational outcomes over time.

5.1. Effectiveness and Impact on Metrics

A central claim in marketing practice is that incorporating AI into campaign design and execution is associated with measurable improvements in performance. In empirical terms, effectiveness is most often operationalised through quantitative outcome metrics that capture first, attention and engagement through open rates and click-through rates (defined as the proportion of message recipients who click on a link contained in a marketing communication) and second, downstream actions closer to the generation economic value as conversion rates, purchases or sign-ups (Grewal, Satornino, Davenport, & Guha, 2024; Nesterenko & Olefirenko, 2023). The fundamental step is then understanding how the use of AI in marketing content and personalisation correspond to systematic variation in key performance metrics, especially clicks and conversions.

Direct evidence linking AI-driven personalisation to measurable behavioural outcomes comes from a Brachten's field study in the email domain that compares non-personalised generic messaging

against AI-personalised variants. In his work, compared to the generic email versions' benchmark of 13.4%, AI-personalized versions scored a 60.1% and 55.1% click-through rates (CTR) for organizational and individual levels respectively. In parallel, conversion rates (CVR) reached 1.9% for non-personalized email versions, against 26% and 15.5% for personalized versions at the organizational and individual levels respectively (Brachten, 2025).

Two considerations need to be made for this result interpretation. First, these outcomes are purely behavioural, measuring clicks and completed actions, not only self-declared intentions, which strengthens their relevance to the analysis of effectiveness as defined here. Second, the study's context is a simulation environment rather than a real commercial persuasion campaign. So, the absolute levels of CTR and CVR should not be treated as market benchmarks directly generalisable to consumer marketing. However, the results still demonstrate important evidence for this chapter's core claim, that is AI-driven tailoring can be associated with very large shifts in click and completion metrics.

Complementary quantitative evidence is reported in a study adopting a mixed-methods approach combining data from interviews with marketing professionals and business clients and complemented by case studies, industry reports and academic literature to examine the use of generative AI in B2B email marketing. The analysis covers a dataset of 50 AI-enabled B2B email campaigns across multiple industries, evaluating performance against "traditional" campaigns. The study reports higher average outcomes for AI-driven campaigns across common email KPIs. AI-personalised campaigns exceeded by 15 percentage points with an average of 35% against 20% for non-AI campaigns in open rates. Consistently, the other investigated rates were on average moderately higher for AI-assisted campaigns with 12% and 8% with respect to CTR and CVR compared to 7% and 4% registered on average in non-AI campaigns (Andrew & Blake, 2024).

Nevertheless, this last analysis is a collection of other research projects and expert opinions presented in a short-form article format with limited methodological accuracy and it should therefore be treated as suggestive rather than dispositive evidence of causal uplift. However, this evidence is directionally consistent with the field study above, confirming that AI-driven personalisation is associated with higher engagement and downstream conversion metrics and it is still useful to locate AI-related effects directly in the metrics that marketers want to optimise and highlights where campaign datasets report measurable differences.

Beyond email marketing, experimental evidence from Matz et al. demonstrates that AI-generated personalized messages significantly increase persuasive effectiveness compared to non-personalized content. Although the study does not directly measure standard marketing metrics such as click-

through or conversion rates, its outcome measures capture core antecedents of downstream behavioural responses, thereby providing relevant evidence for the effectiveness logic underlying CTR and CVR.

In their first study (Study 1a) the researchers employed a controlled experimental design to test whether personality-tailored advertising messages generated by the pre-trained LLM open-source version of GPT-3 were perceived as being more persuasive. The recruited sample was first exposed to a set of advertising messages tailored to the high and low ends of the Big Five model personality traits (Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism) promoting the same product. For each trait (excluding Neuroticism for technical reasons), two contrasting versions of the advertisement were produced by prompting the model with trait-specific descriptors derived from established personality inventories (Ten-Item-Personality-Inventory, TIPI) (Matz, et al., 2024).

Participants evaluated these paired messages using bipolar rating scales designed to compare the persuasiveness and interest of trait-aligned versus opposing messages, with effectiveness measured as the average between the two items. After completing the advertisements' evaluations, participants' individual personality profiles were measured using a Big Five measure (BFI-2S). This procedure allowed the researchers to examine whether the orientation of individual preference of AI-generated messages mirrored individual's measured personality traits (Matz, et al., 2024).

The findings provided evidence that the messages were aligned to the recipients' measured personality traits and judged to be more persuasive when they were personality-tailored by the AI. Regression analyses indicated that individual differences in Openness, Conscientiousness and Extraversion significantly ($p\text{-val} < 0.001$ in all categories) predict preferences for advertising messages designed to appeal to those respective traits, even after controlling for socio-demographic variables. However, in only one case, no significant ($p\text{-val} = 0.152$) alignment effect is observed for Agreeableness, suggesting that not all personality dimensions contribute equally to the effectiveness of AI-driven personalization (Matz, et al., 2024).

Nevertheless, overall, these findings support the researchers' hypothesis that GAI can produce psychologically differentiated content which effectiveness can indeed depend on the match between message framing and recipient's personality.

The work also documents heterogeneity since some topics show smaller or non-significant differences, indicating that AI-driven uplift is not uniform across domains. In fact, Study 2 was designed to test the generalizability of AI-driven personalized persuasion across multiple message domains and psychological targeting frameworks. The sample was exposed to a series of persuasive messages generated by ChatGPT. The messages covered three distinct domains: commercial

advertising of Nike sneakers, health-related persuasion by encouraging exercise and political communication related to appeals for climate action. For each domain, messages were systematically tailored to different psychological characteristics relevant to that context, including the Big Five traits. In this study two more psychological characteristics were included. First, regulatory focus, which differentiates individuals according to whether they are motivated by attaining positive outcomes or avoiding negative ones and prior research indicated that persuasive messages matched to this orientation are more effective, especially in health domains. Second, moral foundations, which capture systematic differences in moral reasoning across dimensions such as care, fairness, loyalty, authority and purity, and prior research on moral reframing showed that political messages are more persuasive when they are framed to align with individuals' dominant moral foundations or with those associated with their political ideology (Matz, et al., 2024).

Again, AI-generated messages were created using minimal, trait-specific prompts that instructed the model to tailor each message to a particular psychological profile. Participants evaluated these messages using design-appropriate measures: bipolar comparative scales for personality- and regulatory-focus-based messages and a point-allocation task for moral-foundation messages that required participants to distribute a fixed number of points across competing arguments. After evaluating the messages, participants completed validated self-report instruments measuring their personality traits, regulatory focus, moral foundations and political ideology. This design enabled the researchers to examine whether alignment between participants' measured psychological profiles and the orientation of AI-generated messages influenced message effectiveness across diverse persuasive contexts (Matz, et al., 2024).

For commercial advertising, the results replicated previous findings by showing significant alignment ($p\text{-val} < 0.05$) between certain personality traits, particularly Openness and Extraversion, and higher perceived message effectiveness. In this case, Conscientiousness and Agreeableness showed no systematic effects, with a p-val equal to 0.424 and 0.984 respectively. In the health domain, messages tailored to individuals' regulatory focus exhibited effects in the expected direction, but these effects were too small and did not reach conventional levels of statistical significance ($p\text{-val} = 0.125$). In contrast, in the political domain, climate-change appeals framed around specific moral foundations demonstrate clearer matching effects: messages aligned with recipients' moral foundations of Loyalty, Fairness and Authority, and political orientation significantly (all $p\text{-val} < 0.05$) anticipated their preference for the matching messages. On the other hand, no significant (all $p\text{-val} > 0.45$) effects were recorded for Purity and Care. Thus, Study 2 provides convergent evidence that AI-generated persuasive messages are more effective if aligned

to recipients' psychological profiles, although the strength of the matching effect varies across domains and characteristics (Matz, et al., 2024).

Further studies extended earlier findings by examining the effectiveness of AI-personalized communication under more conservative and ecologically realistic conditions. Specifically, the experimental design moved from comparative evaluations to single-message assessments, approximating how individuals encounter marketing content in everyday digital environments. Furthermore, message effectiveness was evaluated using willingness-to-pay (WTP) measures as a proxy for downstream behaviour, thereby providing a closer link between perceived persuasiveness and economically relevant outcomes (Matz, et al., 2024).

As said, in their fourth experiment the researchers examined the effectiveness of AI-generated personality-tailored smartphone advertisements using both attitudinal and monetary measures. Participants were exposed to a set of AI-generated iPhone advertisements, each one tailored only to the high end of a different Big Five personality trait from the first study, focusing particularly on Openness and Extraversion since the two were the only two trait which showed significant effects in the previous experiment. Participants evaluated each advertisement individually, reporting perceived persuasiveness and interest, with effectiveness again measured as the average of the two. In addition, participants were also asked to indicate their WTP for the advertised product within a price range. Participants' personality traits were then measured ex post using the same Big Five inventory (Matz, et al., 2024).

Using a more conservative design that approximates real-world exposure, translated into a single-message evaluation, the results provided evidence that AI-generated communication matching personality not only can affect persuasiveness but also translates into economically meaningful differences. In particular, messages tailored to individuals' levels of Openness and Extraversion were again evaluated as significantly (with p-val = 0.005 and p-val < 0.001 respectively) more effective when recipients' measured traits were aligned, confirming the same effects observed above under stricter conditions (Matz, et al., 2024).

Importantly, alignment between observed WTP and Extraversion emerged as especially consequential. Participants reported a significantly (p-val = 0.004) higher WTP for the advertised smartphone when the promotional message reflected an extraverted framing. Specifically, a one–standard-deviation increase in participants' extraversion was associated with an average increase of $33 in stated willingness to pay for the iPhone when the advertisement employed matching framing. Although not statistically significant at conventional thresholds (p-val = 0.085), higher Openness

was associated with a modest increase in willingness to pay, particularly, $19 for a one–standard-deviation increase (Matz, et al., 2024).

These results indicate that AI-driven psychological personalization can influence not only how persuasive a message feels, but also how much value consumers attach to the promoted product.

Compared to the previous studies, their last experiment has considerable theoretical importance since it more closely approximates real-world marketing practice by relying on pre-existing consumer profiles to generate one-to-one personalized messages at scale. This design reflects how firms can leverage previously collected data to tailor communications prior to exposure, rather than inferring preferences after message exposure. Study 4 was designed to simulate a real-world deployment of large-scale AI-driven personalized persuasion by dynamically generating individualized advertising messages. Rather than relying on predefined message sets, the study leveraged previously collected personality data to personalize content at the individual level. However, as will be further discussed in Chapter 7, this approach also foregrounds ethical concerns related to data governance, privacy and the legitimacy of using pre-existing consumer profiles for AI-driven persuasion.

Returning to the analysis of the benefits of AI implementation, in this study participants were recontacted from earlier studies in which they had completed the personality assessment, allowing the researchers to access participants' personality profiles prior to the experiment. Each participant was categorized as low or high on their salient trait (Openness, Extraversion or Conscientiousness), which served as input for one-to-one personalization of message generation. Using OpenAI's ChatGPT, the researchers generated one personalized advertisement per participant, tailored to their salient trait. An experiential product (a weekend getaway to Rome) and a material product (Nike sneakers) were examined as product domains. At the same time, direct personality labels for the travel ad and adjective-based trait descriptions for the sneaker ad were used as different prompting strategies. Generic ads for each product were also generated using non-personalized prompts to establish a comparison baseline. Each participant evaluated both a personalized and a generic ad for each product. Same way as the prior study, outcome measures included perceived ad effectiveness and WTP (Matz, et al., 2024).

As a result, personality-optimized ads for a trip consistently outperformed generic ones. Between-subjects experiment showed that participants exposed to personalized travel ads rated them as significantly more persuasive translating into monetary terms in a substantially higher WTP, both within a significance threshold of 0.05. Indeed, exposure to a personalized travel message increased stated WTP by approximately $117. Within-subjects comparison corroborated this pattern, showing

that the same participants evaluated personalized travel ads as more effective and valuable than generic alternatives (Matz, et al., 2024).

By contrast, results for the Nike sneaker product were directionally consistent but statistically weaker, within a non-significant p-val ranging from about 0.15 to 0.32. However, personalized sneaker ads tended to receive higher effectiveness ratings and willingness-to-pay estimates than generic ads, confirming the expected trend, yet within a non-conventional significance range. This asymmetry suggests that the impact of AI-driven personalization may depend on product category and consumption context, with experiential, hedonic, or aspirational offerings being more responsive to personality-based tailoring than relatively standardized consumer goods (Matz, et al., 2024).

Although all these outcomes are often measured as intentions rather than observed conversions, they are still relevant for this chapter's effectiveness logic because they provide a mechanism-level foundation for why CTR and CVR might move. If AI can generate psychologically matched messages that systematically improve persuasiveness, then this should translate into measurable differences in intermediate and downstream metrics when embedded into campaigns.

Taken together, the reviewed evidence indicates that the use of AI in marketing communications, when operationalised through scalable content personalisation, is associated with measurable variation in general effectiveness, key performance metrics and impacts on customers' behaviour and opinion. Across experimental and applied studies, AI-driven personalisation frequently enhances observed persuasiveness, providing a plausible pathway through which downstream behavioural metrics such as CTR and CVR may improve when such messages are deployed in real campaigns. In addition, GAI can produce messages that systematically increase other economically meaningful behaviours, first, the willingness to pay, strengthening the inference that AI-enabled personalisation can positively influence business performance outcomes when embedded in operational marketing systems (Andrew & Blake, 2024; Brachten, 2025; Matz, et al., 2024).

At the same time, it is important to acknowledge that the literature clearly indicates that effectiveness is heterogeneous rather than universal. Not all psychological traits, message domains or product categories benefit equally from AI-driven tailoring and some studies report small or non-significant effects depending on context (Matz, et al., 2024).

Accordingly, the most defensible conclusion is a cautious but affirmative one. The integration of AI into marketing communications through scalable and data-driven personalisation is associated with frequently positive and empirically observable changes in performance metrics. However, the size and reliability of these effects depend on contextual factors. As such, AI-driven effectiveness gains

should be understood not as automatic outcomes of adoption, but as contingent results that require careful empirical validation within each specific marketing context.

5.2. Human- and AI-Led Content Production: Comparative Effectiveness

Building on the previous Chapter 4's discussion about trust, authenticity and humanness perceptions, differences can be detected on stricter observable performance outcomes, such as the aforementioned click and conversion metrics, when marketing content is created by humans, by generative AI or through hybrid ("merged") workflows. The central question is not only whether AI can produce good and reliable copy in the abstract, but which production paradigm maximizes performance.

First, a useful way to frame the comparison is to distinguish between content quality from marketplace response. General quality is typically proxied by perceived persuasiveness or usefulness and willingness-to-pay for the content. Marketplace response is captured by behavioural metrics such as CTR, click volume, downstream conversion rate until sales.

Zhang & Gosline's experimental evidence is fundamental to investigate effects on perceived quality. In particular, their study employed a between-subjects design to examine how different content creation paradigms are evaluated under varying levels of information disclosure. Participants were randomly assigned to a 3 × 4 design, consisting of three informational conditions and four content generation paradigms (Zhang & Gosline, 2023).

The content generation paradigm involved four different possibilities for content creation. In the human-only condition, the content was produced exclusively by human experts, in the AI-only condition, the content was generated solely by the LLM ChatGPT-4, in the augmented human condition, the content was primarily created by humans but with AI assistance and last, in the augmented AI condition, the content was generated by AI but with human input or oversight. To further clarify the distinction between the two intermediate conditions, which is extremely important for understanding the possible differences in evaluation: in the augmented human condition, the content is primarily created by humans, with AI providing assistive support while humans retain full decision-makin□ authority and determine the final output; in the au□mented AI condition, the content is primarily generated by AI, with humans providing guidance and oversight validating or constraining the final output by giving feedbacks, while AI leads the creative process.

To keep in mind, each participant was exposed to content generated under only one of these paradigms throughout the whole study (Zhang & Gosline, 2023).

The informational condition manipulated their content creation process' awareness. In the baseline condition, participants evaluated the quality of the content without any information about how it was produced. No reference to humans, AI or human–AI collaboration was made, ensuring that evaluations were based solely on the textual input. In the partially informed condition, participants were informed at the beginning of the study that the content they would evaluate had been generated using one of the four possible paradigms. However, they were not told which paradigm generated each specific piece of content they were evaluating. In the last condition, the fully informed condition, participants received the same general information as in the partially informed condition and, in addition, were explicitly informed of the exact content generation paradigm responsible for each piece of content they were about to evaluate. This design is very important to analyse and interpret the effectiveness of AI under more real circumstances, where the use of AI is not always declared (Zhang & Gosline, 2023).

During the experiment, participants evaluated two types of content in sequence. First, advertising content for five different products and second, persuasive campaign content for five different social and behavioural campaigns. For each content type, participants evaluated three outcome measures capturing perceived content quality. Specifically, they reported their satisfaction with the generated content and their willingness to pay to use it as advertising or persuasive material, both assessed from the perspective of the seller. In addition, participants rated their interest in the advertised product or their level of persuasion regarding the campaign message, assessed from the perspective of the consumer (Zhang & Gosline, 2023).

Across the study's four paradigms, the clearest pattern is that outputs are evaluated as being of higher quality when AI is the sole generator or is assisted by human (augmented AI condition), compared with paradigms where humans are the sole generator or retain final authority (augmented human condition). When the content is produced under AI-led paradigms evaluations of content quality are consistently higher than under human-led paradigms. At the same time, within-cluster differences are small. AI-only results are similar to augmented AI results and human-only similar to augmented human ones. This second important interpretation is that adding the other agent does not change perceived quality within each cluster by itself. AI-dominant scenarios are very close in perceived quality (both on satisfaction and WTP), while human-dominant scenarios are also very close between each other. Interpreted together, this suggests that the hybrid form is not automatically superior. What matters is not collaboration per se, but which agent is dominant in

generating the content and thus, human oversight does not meaningfully improve AI's output quality in absolute terms (Zhang & Gosline, 2023).

These results lead to three important considerations. First, AI can produce high-quality persuasive and advertising text even without being assisted by humans. Second, human–AI collaboration is not inherently optimal, but hybridization adds value only if it places the stronger agent, which is AI, in the decisive role. Last, the "maximum efficient point" in content production, as suggested by these comparisons, is likely to be AI-led generation with selective human governance, rather than any human-dominant creation.

While this study does not measure clicks or conversions directly, their design matters for deeper performance interpretation because it highlights a key asymmetry that is highly relevant for those statistics. Thus, moving to marketplace responses captured by behavioural metrics, recent field evidence help to confirm that AI creative processes can outperform human ones on click-based outcomes.

A large study on visual generative AI in advertising by Lee et al. compared different human and AI combinations for ad generation. The researchers adopted a multi-method research design combining a controlled laboratory experiment with a real-world field experiment to examine the effects of GAI on advertising effectiveness. First, a laboratory experiment was conducted to assess consumer responses under controlled conditions. This experiment focused on participants' evaluations of advertising content and was designed to measure perceived ad effectiveness by measuring purchase intention. To complement and extend the laboratory findings, the authors then conducted a field experiment using Google Ads, measuring real-world advertising performance through click-through rates. The field study allowed to enhance robustness and external validity compared to the controlled conditions of the lab experiment. By combining these two approaches, the study leveraged the strengths of each method: the laboratory experiment provided internal validity and detailed insight into evaluative responses, while the field experiment tested whether these effects generalized to actual consumer behaviour in a naturalistic advertising environment (Lee, Todri, Adamopoulos, & Ghose, 2025).

Both studies compared two AI-based advertising conditions against a human-generated benchmark, enabling a direct comparison. In the baseline condition, ads were entirely created by human experts and served as the reference category. In the AI-generated condition, ads were produced autonomously by a GAI system and used without any human modification while, in the AI-modified condition the original human-created advertisement was subsequently revised or enhanced by a generative AI system (Lee, Todri, Adamopoulos, & Ghose, 2025).

The field experiment provided strong evidence of the effectiveness of GAI in digital advertising as measured by clicks. Overall, the results revealed a clear divergence between AI-created and modified ads. Evidence showed that AI-created advertisements achieved higher CTRs than human expert–created ads, whereas AI-modified ads performed worse than the human baseline. The results remained consistent across all model specifications, namely controlling for brand, product, day and device effects, as well as ad style and prompt effects for modified and AI-generated content respectively. The effect of AI-modified ads on CTR was negative but not statistically significant (p-val > 0.1), indicating no reliable improvement over expert–created ads. In contrast, GAI-created ads exhibited a positive and statistically significant (p-val < 0.05) effect on CTR across all model specifications compared to the human baseline. In particular, taking as a reference the most conservative specification and given an average CTR of 3.73% for human–created ads, the estimated odds ratio implied a CTR of approximately 4.44% for GAI-created ads. Which means an increase of 0.71 percentage points, or a 19.0% of relative increased effectiveness over the human baseline. Overall, the field study corroborated the previous laboratory findings and demonstrated that GAI is highly effective when used to create advertising content from scratch, translating into meaningful gains in real-world consumer engagement. By contrast, using AI to modify existing human-created advertisements does not yield comparable benefits, instead, it might even reduce effectiveness (Lee, Todri, Adamopoulos, & Ghose, 2025).

Beyond the pure content creation, it is also crucial to consider how targeted content is delivered and presented to its intended audience. Indeed, a study from Tian et al. investigated in a field study AI-powered advertising in terms of targeting and delivery optimization, rather than the generation of the creative content itself. By analysing how AI-driven targeting affects key performance metrics such as CTR and sales, the study provided valuable insights into how AI influenced advertising effectiveness through mechanisms other than content generation. Including this evidence allows for a more comprehensive assessment of AI's impact on marketing performance, distinguishing between the role of AI as a creative agent and its role as a decision-making tool in advertising delivery.

The primary objective of the study was to examine how AI-powered targeted advertising influenced consumer engagement and purchase behaviours relative to manually targeted advertising, with particular attention to the potential trade-off between click-through rates (CTR) and conversion outcomes. Specifically, the study was seeking to test the existence of a so called selective conversion effect, whereby AI-driven targeting may reduce exploratory clicking while increasing the likelihood of purchase once consumers engage with the content. A second objective was to assess whether the impact of AI-powered advertising varies across different product characteristics,

focusing on product popularity (popular against niche products) and product innovativeness (innovative against traditional) as key factors. The study aims to determine whether AI-driven targeting differentially affects CTR and sales. Overall, the experiment design intended to clarify how AI-powered advertising enhances effectiveness distinguishing between short-term engagement metrics and downstream sales outcomes. At the same time, the study tried to identify the conditions under which AI-driven targeting is most likely to generate superior commercial performance (Tian, Zhao, Sun, & Qiu, 2025).

The study implemented a large-scale field experiment between January 2023 and January 2024 in collaboration with a major Chinese home-appliance manufacturer. For each product and on each day, AI-powered and manually managed advertising campaigns were run in randomly assigned but comparable local markets, with equal budget allocations across methods. Consumers in each market were exposed to only one advertising approach. Advertising performance was tracked at the product–day level, recording clicks, costs and subsequent purchase outcomes, enabling a causal comparison of AI-powered and manual advertising inside real-world conditions (Tian, Zhao, Sun, & Qiu, 2025).

The results showed that AI-powered advertising significantly reduced click-through behaviour while simultaneously improving downstream conversion outcomes. Specifically, relative to manually managed advertising, AI-powered advertising led to a 3.81 percentage-point reduction in CTR compared to manually managed advertising. Given an average CTR for manual advertising of 6.6%, this corresponds to a 57.7% relative decrease in CTR. Since advertising expenditure is click-based, this lower CTR translates into a reduction in daily advertising costs per product, indicating higher targeting efficiency. Nevertheless, despite lower engagement at the clicking stage, AI-powered advertising significantly increases sales performance. On average, AI-powered campaigns generate 4.151 additional orders per product per day increasing daily revenue. Moreover, AI-powered advertising increased shopping-cart additions by 15.82 instances per product per day, representing a 25.37% increase relative to manual advertising. Thus, AI-powered advertising delivered a 29% higher return on investment (ROI) compared with traditional advertising approaches overall (Tian, Zhao, Sun, & Qiu, 2025).

The analysis reveals that the selective conversion effect is amplified for popular products. For products with higher baseline popularity, AI-powered advertising led to an even stronger reduction in CTR, reflecting greater consumer certainty and lower need for exploratory clicking. At the same time, the increase in sales associated with AI-powered advertising is stronger for popular products than for niche products. This pattern confirms that AI-driven targeting is particularly effective when consumers already possess substantial product knowledge, reinforcing conversion at the purchasing

stage while reducing inefficient ad spending. However, the effects for innovative products were different, where consumer knowledge is more limited. In this context, AI-powered advertising increased CTR by 2.37 percentage points relative to traditional products, indicating higher curiosity and engagement at the clicking stage. Although the overall effect on CTR remained negative when accounting for the main AI effect, the selective conversion effect clearly weakened at the engagement stage for innovative products. Nevertheless, AI-powered advertising still improved sales outcomes for innovative products, suggesting that once consumers click and reduce uncertainty, AI-driven targeting enhances conversion at the purchasing stage. This confirms that AI-powered advertising is particularly effective for promoting novel products by balancing higher engagement with stronger conversion (Tian, Zhao, Sun, & Qiu, 2025).

These findings demonstrate that AI-powered advertising does not maximize effectiveness by increasing clicks, but by improving targeting precision and conversion efficiency. Across products, AI-powered advertising lowers engagement metrics such as CTR while consistently enhancing sales, revenue and finally ROI. These results provide robust empirical support for the selective conversion effect and highlight that the effectiveness of AI-powered advertising depends critically on where in the consumer decision process performance is evaluated, clicking or purchasing.

To wrap up, from a practical perspective, the evidence suggests that the most efficient mix of human and AI involvement in marketing overall is typically not a symmetric division of creative labour, but rather a governance-oriented hybrid in which AI leads content generation and optimization, while humans retain control over boundary-setting, brand alignment and other strategic conditions. Zhang and Gosline's paradigm-level results are fully coherent with this logic. Indeed, performance is strongest when AI has final decision authority, whereas human-only and augmented-human configurations do not dominate on perceived quality. In this sense, the hybrid optimum is less about shared authorship and more about placing humans at the highest-leverage control points, while allowing AI to exploit scale, variation and search efficiency.

5.3. Effectiveness of AI-Personalization in Email Communication

Email marketing is a particularly suitable domain for assessing the behavioural effectiveness of generative AI–driven personalization because its core outcomes are directly observable. In fact, unlike other marketing channels, email allows researchers and practitioners to directly observe and track behavioural variables, most notably link clicks and downstream action completions. As a

result, it provides a particularly suitable setting for assessing whether AI-personalized textual content translates into measurable improvements in performance, beyond stated intentions or attitudinal measures. Recent research projects have therefore increasingly relied on email-based field experiments to quantify how AI-enabled personalization affects user behaviour in realistic conditions. Particular evidence is drawn from a large-scale real-world experiment by Brachten that uses a simulated email campaign as a controlled setting to measure clicks and conversions-like actions.

Although Brachten's study was conducted as a phishing simulation rather than a commercial marketing campaign, it can still be considered a useful proxy for analysing email marketing effectiveness as evidenced by Brachten himself. Both phishing and marketing emails rely on similar persuasive mechanisms and can be assessed using identical behavioural metrics, such as click-through rates and call-to-action completions. While differing in ethical intent, both domains use personalization and psychological triggers to influence perceived relevance and trust, making phishing simulations a controlled and behaviourally grounded setting for examining how AI-driven personalization affects email engagement (Brachten, 2025).

The study was implemented inside a large professional services organization. The study was designed to evaluate the behavioural impact of AI-personalized email messages by comparing them with a baseline control condition of non-personalized generic emails derived from previous simulations conducted in the same organization and year. The experiment operationalized performance with two deterministic outcomes. First, CTR measured as whether the recipient clicked the link provided in the email. Second, CVR measured as whether the recipient completed the requested action after clicking which in this case was submitting credentials on a landing page (Brachten, 2025).

The metrics calculated for the generic email baseline highlighted a CTR of 13.4% and a CVR of 1.9%. The first research question investigated how AI-personalized email messages affected link clicks compared to generic ones. As a result, AI-personalized emails produced a total CTR of 57.4% across the sample, representing an overall increase of 44 percentage points compared to the baseline within a statistically significant range of less than 1% ($p\text{-val} < 0.001$). Hence, this result confirms that within this setting, AI-personalized content is associated with a very large uplift in clicking behaviour. The second objective wanted to understand how emails personalized through AI can affect action completions compared to generic texts. Again, AI-personalized emails produced a statistically significant ($p\text{-val} < 0.001$) increase in email effectiveness over conversions. AI-personalization produced an 18.5 percentage points increase in action completions, from 1.9% in the generic benchmark to 20.4% in the AI-optimised manipulation. Overall, AI personalization is

not only associated with more clicks, but also with substantially higher completion of the requested downstream actions (Brachten, 2025).

Finally, although the study used a simulated context within a phishing experiment context, the results provide strong behavioural evidence that implementing AI systems for text creation in email content can substantially increase both clicking and downstream actions compared with generic, non-personalized copy. The setting strengthens inference by relying on observed behaviour in a real organizational environment and by benchmarking against prior campaigns from the same setting. This makes the evidence highly relevant for the core question of whether GAI-driven personalization in email can move CTR/CVR outcomes.

Building on the evidence that AI-personalised email copy can materially shift *observed* CTR and CVR outcomes, a complementary stream of research focuses on an even earlier gateway in the email engagement funnel: the subject line or email title. In large-scale email programs, particularly those driven by recommender systems, the body of the message is often already personalised through item recommendations, while the subject line remains template-based. This means the title is generated using a fixed, pre-defined structure with minimal linguistic variation, in which only a small number of elements such as the item name are dynamically inserted. This design choice may limit engagement at the entry point, as the subject line plays a critical role in setting expectations and signalling relevance before the email is even opened. To address this limitation, Jobson et al. examined whether LLMs can be used to generate more thematic, recommendation-consistent email titles that better communicate the content of the email and thereby enhance engagement in real-world conditions (Jobson, et al., 2025).

The study was implemented as a randomised controlled A/B test at platform scale on Mercari, a large online marketplace, and targeted users who had not accessed the app in the previous seven days but had remained active in the broader customer base, having used the app within the past year and completed at least one purchase in the previous six months. Eligible users were randomly assigned at the individual level to either a treatment condition, in which the email subject line was generated by the LLM GPT-4o-mini, or a control condition, in which the subject line followed a standard, template-based format. This design ensured that observed differences in engagement could be causally attributed to the subject-line generation approach rather than to differences in user characteristics or prior behaviour (Jobson, et al., 2025).

In the control condition, subject lines were constructed using a fixed structure based on the top recommended item, typically following a formula equivalent to “<item name> and other items are on sale.” In contrast, the treatment condition employed the LLM to generate titles that were less

repetitive and more thematic, while still remaining consistent with the set of items recommended within the email body. Performance was evaluated using both email-specific engagement metrics and downstream business indicators, allowing the authors to observe how subject-line optimisation affects different stages of the email funnel (Jobson, et al., 2025).

In terms of outcomes, the experiment is particularly informative for distinguishing between engagement metrics and harder-to-move conversion metrics. In fact, the use of LLM-generated subject lines produced only a small (although statistically weak) increase in relative email open rates of 0.46% but led to a substantial and statistically significant relative lift in within-email engagement. Specifically, the promoted item clicks or tap rate increased by 23.63% in the treatment condition, indicating a markedly higher propensity for users to interact with recommended items once the email is opened. However, the intervention did not generate a statistically significant change in overall buyer conversion rates which also remains slightly negative (–0.42%) (Jobson, et al., 2025).

This pattern is consistent with a funnel-based interpretation in which improvements at the entry point, represented here by clearer and more engaging subject lines, translate into higher intermediate engagement without necessarily overcoming downstream constraints such as price sensitivity or other market conditions (Jobson, et al., 2025).

Overall, the evidence reviewed in this chapter indicates that AI–driven personalization can substantially improve the behavioural effectiveness of email marketing, particularly on directly observable metrics such as click-through and intermediate engagement. Field experiments conducted in realistic organizational and platform settings show that AI-generated email content can produce large uplifts in CTR and, in some cases, in downstream action completions, compared with generic or template-based messages. At the same time, the results highlight an important funnel dynamic. Indeed, while GAI is highly effective at capturing attention and stimulating interaction, its impact on final conversion outcomes appears more limited and contingent on factors beyond message design. Taken together, these findings support the view that generative AI enhances email marketing effectiveness primarily by optimising early- and mid-funnel behaviours, thereby creating the conditions for improved performance without, however, guaranteeing immediate purchase conversion.

5.4. Generative AI and Long-Term Effectiveness in Brand–Consumer Relationships

Apart from short-term interactions, generative AI can also influence long-term brand–consumer relationships for several reasons. Its adoption can alter first, how consumers interpret the brand's intentions and humanness, second, how consistently the brand communicates across touchpoints, and third, how consumers translate brand perceptions into future-oriented behaviours as advocacy and repurchase intention. In marketing, these dynamics are usually captured through constructs such as trust, commitment and loyalty, which can empirically be operationalized through behavioural intentions (which refers to the customers' likelihood of engaging in future actions that support or involve the brand), EWOM and identity-based ties as self-brand congruity (which refers to the extent to which the customers feel represented by the brand's identity)that predict longer-run outcomes including repeated purchases and retention.

A core implication emerging across the topics covered before is that GAI rarely affects relationship outcomes directly. Instead, it functions as a more or less explicit cue about the source of brand communication, which shifts psychological evaluations such as perceived authenticity and trust which drive downstream responses relevant to relationship strength. This mediation logic is particularly explicit in Ali et al.'s discussion, which frames GAI content as a stimulus and tests how it changes internal brand perceptions such as authenticity, image and self-brand congruity and, in turn, behavioural outcomes including EWOM and behavioural intentions (Ali, Ali, & Alotaibi, 2025). In other words, GenAI's long-term relational effect depends on whether it strengthens or weakens the key psychological "bridges" that stabilize brand relationships over time.

In their study the researchers adopted a scenario-based, between-subjects experimental design aimed at isolating the effect of GAI adoption as a content source cue on consumer perceptions and intentions. The core idea was to keep the content itself constant while manipulating who (or what) is described as having produced the content. This allowed the authors to attribute any observed differences in consumer responses to the presence vs. absence of GenAI involvement, rather than to differences in message quality, tone, or informational content (Ali, Ali, & Alotaibi, 2025).

Participants were presented with a hypothetical restaurant brand scenario, framed as a realistic digital marketing context. Specifically, respondents viewed a restaurant-related social media post designed to resemble a typical brand communication they might encounter online. While the content of the post was identical, the experimental manipulation consisted solely of a source

attribution cue, which indicated either that the content was created by a human employee or that the content was generated using generative AI technology adopted by the restaurant. This design ensured that the only difference between groups was the knowledge that an AI was involved in the process. Participants were randomly assigned to one of the two experimental conditions (AI- or human-created content) which included the manipulation cue about content creation. After exposure, participants were asked to evaluate the restaurant and the communication (Ali, Ali, & Alotaibi, 2025).

However, rather than measuring immediate behavioural outcomes, the study focused on psychological and intentional constructs widely accepted as proxies for longer-term brand relationships including brand authenticity and image, self-brand congruity, behavioural intentions as the likelihood of choosing or supporting the restaurant, and electronic word-of-mouth intention. Importantly, the authors also modelled the structural relationships among these constructs, allowing them to test not only whether GAI adoption changed evaluations, but how changes in brand perceptions translated into downstream intentions (Ali, Ali, & Alotaibi, 2025).

The findings provided evidence of a relational headwind. The results indicate that disclosing the use of AI was associated with a broad deterioration in relationship-relevant brand perceptions and downstream consumer responses and it also changed which psychological drivers most strongly translate brand perceptions into future-oriented behaviours.

The authors first run a multivariate analysis of variance to test whether the two experimental conditions differed jointly across key brand perception constructs. The MANOVA showed a strong overall multivariate effect of condition on the set of dependent variables, meaning that attributing the communication to GAI, rather than a human, produced a statistically significant ($p\text{-val} < 0.001$) shift in how the brand was evaluated across the central dimensions for longer-term relationship quality (Ali, Ali, & Alotaibi, 2025).

The univariate comparisons showed consistent differences between conditions across all focal constructs. In the AI adoption condition, participants reported lower perceived brand authenticity compared with the no AI condition. The same pattern held for brand image and self-brand congruity. Crucially for the long-term relationship topic, the downstream outcome variables, although both framed as mere intentions, also followed the same direction. Behavioural intentions were lower under AI adoption likewise EWOM intention. These differences support the interpretation that labelling content as AI-produced is associated with weaker perceived authenticity and brand equity cues, alongside lower willingness to advocate for the brand and lower stated

likelihood of future supportive behaviours, indicators of long-term relationship strength (Ali, Ali, & Alotaibi, 2025).

Even though these are not observed long-run retention metrics, they can be used as proximal indicators of relationship strength since intentions are widely used as a proxy to predict future choices and positive EWOM is usually tied to advocacy and relational attachment. Hence, the key implication is that AI, at least when explicitly framed as the main content source, can create a relationship headwind by weakening the very perceptions that usually translate into loyalty over time.

The second stage of the analysis clarified how the relationship pathways change depending on whether content is AI-driven or human-created. First, in the AI adoption group, self-brand congruity became the dominant driver of both downstream outcomes, strongly predicting behavioural intentions and EWOM intention (both p-val < 0.001). Under the same condition, brand authenticity had essentially no effect on behavioural intentions (p-val = 0.853) and only a borderline effect on EWOM (p-val = 0.051), while brand image significantly predicted EWOM (p-val < .05) but has only a marginal association with behavioural intentions (p-val = 0.065). It follows that when communication is framed as AI-generated, consumers' future-oriented responses hinge primarily on identity alignment with the brand, whereas authenticity becomes less diagnostic for deciding whether to engage in further actions. These results remain consistent with the idea that AI mediation can weaken traditional relational cues (Ali, Ali, & Alotaibi, 2025).

In contrast, in the no AI adoption sample, the explanatory structure appeared to be more aligned to the classic brand relationship drivers. Brand image resulted as the strongest predictor of behavioural intentions, followed by brand authenticity and self-brand congruity (all p-val equal or lower than 0.005). For EWOM intention, brand authenticity and self-brand congruity were significant (p-val < 0.02 for both), while brand image was not (p-val = 0.457). it derives that when content is framed as being created by humans, authenticity and congruity are central to advocacy intentions, whereas image is especially important for broader behavioural intention (Ali, Ali, & Alotaibi, 2025).

This is a critical long-term insight: GAI can change what holds the brand relationships together. When consumers perceive that AI-produced communication dilutes traditional cues (authenticity, warmth, human intent), they appear to rely more on internal identity alignment as if they asked if the brand can still represent themselves as an anchor for future-oriented behaviours.

Finally, the multigroup tests confirmed that several key paths differ significantly across conditions, important for long-term relationship because it shows that AI involvement does not only shift levels of perceptions and intentions, but also the mechanisms. Indeed, the effect of brand authenticity on

behavioural intentions is significantly (permutation p-val = 0.027) stronger in the no AI group but it is not significant on EWOM. Differences are also significant for the effect of brand image on behavioural intention (permutation p-val = 0.043) and on EWOM (permutation p-val = 0.005) and for the effect of self-brand congruity on both behavioural intention (permutation p-val = 0.001) and EWOM (permutation p-val = 0.005) (Ali, Ali, & Alotaibi, 2025).

In practical terms, the study suggests a mechanism shift. This implies that AI may not only affect relationship strength but also relationship structure. Loyalty becomes more contingent on self-brand fit when the communication source is AI. In AI-labelled contexts, when authenticity and image weaken, relationship-relevant outcomes, especially advocacy and future behavioural support, become disproportionately dependent on whether the consumer feels the brand fits their identity. On the other hand, in human-labelled contexts authenticity and image retain stronger explanatory power for these same outcomes.

Recalling the previous authenticity-focused Chapter 4., it can help explain why AI-labelled content can damage relationship formation and how brands can mitigate it.

In particular, in the aforementioned chapter it has already been discussed how, among others, in the Jago et al.'s paper, the researchers demonstrate that authenticity is not merely a property of output quality. Instead, it is shaped by whether people can concretely connect an output to its, at least presumed, human origins and the human intentions behind the system. When participants are encouraged to consider the creators behind an automated system, perceived authenticity increased, even when the output itself did not change (Jago, Carroll, & Lin, 2022).

For long-term brand relationships, this concept matters because authenticity is a stable, trust-adjacent cue. When the usage of an AI breaks the perception of realness or sincerity, trust accumulation becomes harder across exposures. Conversely, framing a GAI as value-aligned and human-governed (rather than pure detached automation) can preserve authenticity perceptions that support long-run loyalty-oriented positive outcomes as advocacy.

A further long-term complication is that consumers may not simply penalize AI content, but they may reward human provenance in a way that impacts brand equity over time. If the same content can be evaluated more positively when labelled as created solely by a human expert, this implies a durable risk for relationship-building. If consumers developed a learned association that a certain brand uses AI instead of people, the brand can lose a provenance premium that normally supports emotional attachment and loyalty resilience. Conversely, brands that position GAI as assistive with human oversight and responsibility may avoid triggering that comparative disadvantage (Brüns & Meißner, 2024).

Overall, the findings suggest that generative AI influences long-term brand–consumer relationships primarily through indirect relational mechanisms rather than through immediate behavioural effects. The study by Ali et al. clearly shows that merely disclosing AI involvement in brand communication can weaken key relationship cues and, crucially, alter the pathways through which brand perceptions translate into future-oriented outcomes. When content is AI-labelled, loyalty-related intentions and advocacy become more dependent on identity alignment (self-brand congruity), while authenticity and brand image lose explanatory power. In contrast, these traditional drivers remain central when communication is framed as human-created. This makes the study a clean reference point for arguing that AI's impact on loyalty and relationship strength operates largely via changes in perceived authenticity, brand image and identity alignment. These findings finally imply that preserving human connection and value alignment in AI-mediated communication is critical to sustaining trust accumulation and relationship resilience over time.

5.5. Overpersonalization as a Boundary Condition of AI-Driven Personalization Effectiveness

While generative AI enables marketers to scale personalization efficiently, the relationship between personalization intensity and effectiveness is not necessarily monotonic. For this reason, overpersonalization can be defined as the level of personalization that exceeds recipients' privacy expectations or violates norms about what a sender should reasonably know, thereby triggering perceived intrusiveness, suspicion or psychological reactance. In marketing communications, these negative responses can offset (or even reverse) the relevance gains of personalization, producing counterproductive effects in engagement and downstream actions.

A useful way to conceptualize this risk is to combine insights from privacy-calculus and persuasion-knowledge theories: consumers trade off utility (relevance, convenience) against perceived costs (loss of control, surveillance, manipulation) and they update these judgments when personalization reveals how much the sender knows in reality. When the level of personal detail appears excessive, consumers may infer covert data collection or manipulative intent, which raises vulnerability perceptions and can reduce trust and compliance. This general logic is consistent with evidence on the personalization–privacy paradox in digital communications. Personalization can increase perceived relevance yet simultaneously activate privacy concerns and resistance, depending on

informational cues and perceived boundary violations (Aguirre, Roggeveen, Grewal, & Wetzels, 2016; Bol, et al., 2018).

Recalling Brachten's real-world AI-personalized email simulation, it provides particularly direct behavioural evidence that more personalized content is not always better. In the same study, the researcher added a further layer of investigation in which two personalization levels were compared. The organizational-level personalization referred to a form of contextual personalization at scale, tailoring the content to the organizational environment and the general role context of the recipient, without incorporating highly specific personal identifiers. At this level the email adapted its content to the recipient's broader organizational role context, enhancing relevance while remaining consistent with what an institutional sender is expected to know about a large group of recipients. By contrast, the individual level involved a higher degree of granularity and specificity. At the individual-level personalization the email incorporated more recipient-specific information, such as department or job role, increasing specificity but also the salience of data usage. Both organizational-level and individual-level personalization conditions were implemented within the original randomized controlled experiment, with recipients randomly assigned to one of the personalization levels (Brachten, 2025).

The empirical results showed that the organizational-level personalization achieved a CTR of 60.1%, while individual-level personalization achieved a CTR of 55.1%. The difference between the two AI personalization levels was statistically significant with a p-val lower than 0.001, indicating that personalization level can shift click rates. However, the practical difference is comparatively small relative to the very large gap between AI-personalized and generic emails. In a similar way, for action completion, the gap between personalization levels remained statistically significant (p-val < 0.001), although it was even more pronounced. The conversion rate for organizational-level personalization was 26% against 15.5% for individual-level personalization, with an overall discrepancy of around 10 percentage points. Similarly to CTR, higher granularity personalization did not improve performance for conversion outcomes. Indeed, organizational-level personalization outperformed the individual-level one also on action completion rate (Brachten, 2025).

This pattern is consistent with a perception- rather than technical-based explanation discussed by the author: hyper-personalization may trigger suspicion, perceived intrusiveness, or a mismatch between the sender identity and the level of personal detail used, which can reduce trust and dampen downstream compliance with the call-to-action even if recipients still click. Organizational-level personalization instead tends to preserve trust and legitimacy, helping explain its higher

performance on downstream action completions, clearly showing that the optimal level of personalization may be conditional (Brachten, 2025).

Aguirre et al. provide a clear explanation of why personalization can become counterproductive once it crosses a certain personalization threshold. The core finding is called “personalization paradox”. If increasing personalization can raise advertising relevance and, thus, response, it can also simultaneously heighten discomfort when the message suggests that data have been collected in ways that feel non-transparent, creating a negative reaction that suppresses engagement. When personalization cues make people suddenly realize that information has been gathered without their consent, the ad becomes less effective not because personalization is inherently bad, but because it triggers a perceived loss of control and a sense of vulnerability (Aguirre, Mahr, Grewal, De Ruyter, & Wetzels, 2015).

The paper substantiates this paradox with behavioural evidence gathered in a Facebook field test. They observed that for a financial services brand the CTR significantly increased (from 0.017% to 0.077% with a p-val < 0.001) initially as ads moved from non-personalized to moderately personalized, but then dropped sharply (from 0.077% to 0.032% with a p-val = 0.002) once the ad became more personalized. The same pattern repeated itself for a dog food brand. The CTR raised from 0.047% to 0.064% (at a statistically significant level of p-val = 0.084), then fell to 0.033% at a significance level of p-val =0.009 when high personalization was introduced. This inverted pattern supports the overpersonalization argument according to which personalization improves relevance up to a point, but can reduce engagement when it becomes salient enough to reveal extensive tracking or unexpectedly intimate knowledge (Aguirre, Mahr, Grewal, De Ruyter, & Wetzels, 2015).

A second piece of evidence from their second study used the enforcement of a Dutch national regulation requiring websites to explicitly inform users about the use of tracking cookies and obtain consent as external condition that suddenly increased awareness around tracking. In a real campaign dataset, the authors reported that CTR dropped by about 11% immediately after the law’s introduction, then later rebounded as users became accustomed to the new transparency. This shows that it is not personalization alone that depresses response, but the perception that tracking is occurring without an expected form of notice or consent, for example when personalization acts as a cue of covert information collection (Aguirre, Mahr, Grewal, De Ruyter, & Wetzels, 2015).

Therefore, Aguirre et al. imply that the effectiveness of personalization depends not only on relevance, but also on perceived appropriateness and data-collection legitimacy. This confirms that higher granularity can fail to improve behavioural outcomes because it increases the salience of

personal data usage and can trigger suspicion or discomfort, especially when the sender-context does not earn that level of intimacy. Even though Aguirre et al. do not specifically study AI-personalization, their evidence helps generalize the core boundary condition which can still be applied in automatic personalized content generation. Personalization can produce diminishing or negative returns when it violates privacy expectations and the mechanism is strongly tied to perceived vulnerability rather than purely cognitive privacy concern, representing a strong theoretical anchor to explain the overpersonalization risk in AI-enabled marketing communications as highlighted by Brachten.

Taken together, the evidence supports an effectiveness-oriented conclusion: overpersonalization is a boundary condition that constrains the returns to generative AI personalization. Rather than assuming that increasing granularity monotonically improves outcomes, marketers should treat personalization intensity as having a maximum efficient point that depends on channel norms, sender identity and the sensitivity of the cues used.

This also implies that evaluating overpersonalization requires tracking multi-stage metrics, not only CTR. The clearest signal of backfire risk may appear in mid-to-lower-funnel behaviours (action completion, purchase or subscription), where trust and perceived legitimacy become more consequential. More broadly, the academic literature suggests that firms can mitigate overpersonalization risks by increasing transparency and perceived control, using less sensitive cues, and aligning the level of personal detail with what recipients expect the sender to know.

5.6. Human Favouritism versus AI Aversion in Marketing Communications

Building on Chapter 4's discussion of trust, authenticity and humanness as context-dependent judgments and the effects discussed in Chapter 5 of these perceptions on performance analysed previously, it is fundamental to clarify a closely related distinction that matters for AI-driven marketing communications, that is AI aversion versus human favouritism. In much of the earlier literature, lower acceptance of algorithmic outputs, when disclosed, has often been interpreted as a generalized reluctance toward AI. However, emerging evidence suggests that, in many persuasion contexts, what looks like "AI aversion" is more precisely a positive bias toward human creators, a preference that raises evaluations when content is believed to be human-made, rather than a penalty applied when AI involvement is disclosed.

First of all, there is a conceptual distinction that needs to be done. In absolute terms, AI aversion implies that knowing AI was involved reduces perceived quality or willingness to engage relative to an otherwise identical baseline. Human favouritism, instead, implies a comparative asymmetry, that is knowing a human expert created the content increases perceived quality relative to baseline, while knowing AI was involved does not equivalently reduce evaluations. This distinction matters because the managerial implication changes. If the main effect is a premium for human content, firms can focus on human-authorship cues and human accountability signals, rather than assuming that disclosure inevitably harms performance, as demonstrated by Jago et al. in their study (Jago, Carroll, & Lin, 2022; Zhan & Gosline, 2023).

Recalling again Zhang and Gosline's research from Chapter 5.2, their experiment provides a direct test of this asymmetry in persuasive marketing content, comparing four content-generation paradigms: human-only, AI-only, augmented human (human final decision using AI draft) and augmented AI (AI final decision using human draft) (Zhang & Gosline, 2023).

Crucially, when participants evaluated persuasive texts focusing exclusively on content quality, without any information about how the content was produced, AI-generated outputs were not penalized. On the contrary, in this baseline condition, content generated solely by ChatGPT-4 achieved higher average satisfaction than content generated solely by human experts (5.29 against 4.93 with a significant p-val < 0.001) and a higher WTP (log WTP: 4.83 against 4.61 within a significant rage of p-val = 0.01). in addition, similar patterns emerged when comparing AI-final and human-final paradigms more broadly. Content in which AI made the sole or final decision was consistently evaluated more positively than content in which the human expert retained final control, although the magnitude of these differences was modest. It derives that, in the absence of source cues, AI is fully capable of producing persuasive content perceived as equal or superior in quality to human-generated content, undermining the notion of an inherent aversion to AI outputs (Zhang & Gosline, 2023).

Importantly, this pattern remained largely unchanged when participants were made aware that AI may have been involved in content generation but were not told which specific paradigm produced each piece of content. Even under this mild salience of AI, AI-final paradigms continued to outperform human-final paradigms on satisfaction, and WTP remained broadly similar across AI-only and augmented-AI conditions. For instance, average satisfaction for AI-generated content remained significantly higher than for human-generated content (5.12 vs 4.80 significant at p-val < 0.001). These results suggest that the mere possibility of AI involvement does not trigger automatic resistance or discounting, further reinforcing the idea that negative reactions to AI are not driven by quality perceptions per se (Zhang & Gosline, 2023).

However, the critical evidence in favour of human favouritism emerged when the same content was evaluated under full disclosure of its origin. When participants were explicitly informed that a piece of content was generated solely by a human expert, evaluations increased significantly relative to the baseline condition. Average satisfaction slightly, but significantly (p-val = 0.003), raised, and WTP increased even more markedly (p-val < 0.001). Crucially, no symmetric negative effect was observed when participants were fully informed that the content was generated by AI or involved to assist the creation process. For AI-only content, disclosure had no significant effect on satisfaction, being the p-val = 0.35 and WTP with the p-val = 0.87). This asymmetry clearly confirms that the disclosure bias operates through a premium assigned to human authorship, rather than through a penalty imposed on AI involvement (Zhang & Gosline, 2023).

Zhang and Gosline further ruled out the alternative explanation that this effect is driven by a generic quality prime. If simply knowing that content could come from highly skilled humans or state-of-the-art AI enhanced perceived quality, then evaluations in the partially informed condition should exceed those in the baseline condition. Empirically, this did not occur as for human-generated content, moving from baseline to partially informed had virtually no effect on WTP and slightly reduced satisfaction. These findings suggest that the uplift observed under full disclosure is not attributable to heightened expectations of competence, but rather to a pure biased favouritism toward human creators when their involvement is made explicit (Zhang & Gosline, 2023).

To move further across the effects more relevant for marketing performance, the researchers extended their study beyond perceived quality. In fact, when human authorship was disclosed, participants reported significantly (p-val < 0.001) greater interest in the advertised products and higher persuasion in campaigns' messages (p = 0.022). By contrast, the authors identified only a single instance in which persuasion decreased when AI-only campaign messages were explicitly disclosed (specifically, from 5.07 to 4.88 with a p-val = 0.03). However, this effect did not persist when comparing informed and partially informed conditions. The absence of a consistent negative gradient as AI involvement becomes more salient suggests that this result does not reflect robust AI aversion. Rather, it appears to be a limited context-specific framing effect (Zhang & Gosline, 2023).

Finally, the results remained qualitatively similar across task types, including campaign persuasion and product advertising. In both cases, AI-generated or AI-final content was evaluated as at least as effective as human-generated content in baseline conditions, and favouritism toward human creators appeared consistently when authorship was explicitly revealed (Zhang & Gosline, 2023).

Moreover, strong correlations between satisfaction and downstream outcomes such as product interest (r = 0.62) and campaign persuasion (r = 0.86), suggest that these perceptual biases are not

merely symbolic, but can translate into behaviourally relevant concrete responses (Zhang & Gosline, 2023).

In conclusion, the dominant and replicable pattern across these studies is a positive bias toward human-authored content when human involvement is made explicit, supporting the interpretation of human favouritism rather than generalized aversion to AI-generated content. Consumer resistance to AI-generated marketing content is fundamentally relative rather than absolute, emerging only when content origin is explicitly disclosed and only in direct comparison with human-generated outputs. Anyhow, this brings very important implications for how firms frame, disclose and position AI-assisted communications.

6. Experimental Examination of Subtle Humanization Cues in AI-Generated Marketing Emails

6.1. Empirical Study Overview

As discussed in Chapter 4, consumer responses to AI-mediated communication are shaped not only by message content, but by inferences about intent, sincerity and human involvement. In this context, trust, authenticity and perceived humanness operate as central evaluative dimensions that influence how marketing messages are interpreted and judged. In particular, the literature has showed that people frequently expect AI-produced outputs to be less genuine or less worthy, even when objective quality differences are absent. This creates an evaluative environment in which AI becomes a salient cue that can either support perceived efficiency or trigger scepticism, depending on how human involvement is inferred or framed (Jago, Carroll, & Lin, 2022).

Building on this, and as developed in Chapter 5, a key clarification is that what is often described as being an AI aversion may be better understood as human favouritism. Rather than consistently penalizing AI-generated content, audiences tend to reward content when it is believed to be created by human experts. Zhang and Gosline's evidence is especially instructive here, showing that evaluations shift asymmetrically when authorship is disclosed. Human provenance statistically generates a premium, while AI provenance does not always produce a penalty. This distinction matters for marketing communication because it implies that negative reactions to AI are frequently comparative and belief-driven, biased towards humans, not simply a stable rejection of AI output quality (Zhang & Gosline, 2023).

At the same time, as outlined in Chapter 4, consumers' ability to reliably discriminate AI from human content is often limited, particularly in text-based domains, so evaluations frequently rely on surface heuristics and arbitrary cues rather than accurate detection of origin. In this sense, perceived humanness functions as a psychologically meaningful but fragile attribution. In fact, recipients may feel they are identifying AI involvement, or reversed, yet in practice they often respond to stylistic signals that can be manufactured on purpose. Jago et al.'s findings reinforce this mechanism by showing that authenticity judgments toward automated work are malleable and can be increased when recipients are prompted to consider human origins or some kind of human connection behind the system, even when the output itself does not change (Jago, Carroll, & Lin, 2022).

Finally, the prior chapters have also highlighted that the perceived disadvantage of AI is attenuated when human participation remains visible or implied. In social media contexts, Brüns and Meißner have showed that full GAI adoption in content creation tends to reduce perceived brand authenticity and downstream responses, whereas human-in-the-loop configurations in which AI assists rather than replaces human creators mitigate, though may not eliminate, negative reactions. This suggests that audiences respond not only to AI use per se, but to the degree to which AI appears to replace versus support human agency (Brüns & Meißner, 2024).

Beyond explicit source disclosure and labelling effects, a complementary stream of research suggests that audiences rely heavily on micro-level cues embedded in the message itself when forming judgments about authorship, competence and sincerity. In online communication contexts, typographical and orthographical errors have been shown to trigger systematic attribution processes. Readers often interpret such errors not merely as accidental noise, but as socially meaningful signals that shape perceptions of conscientiousness, credibility and trustworthiness. In particular, evidence from research on online consumer reviews, which typically consist of texts written by ordinary users rather than professional communicators, shows that in this context typographical and orthographical errors can reduce perceived expertise and professionalism. Importantly, on the other hand, these imperfections are socially associated with pure human production rather than automated systems. As a result, such imperfections function as socially interpretable cues that may simultaneously reinforce perceptions of human origin while undermining judgments of reliability. making them potentially ambiguous cues that may simultaneously signal human presence and reduced competence (Cox, Cox, & Cox, 2017).

Research on conversational agents within chatbot interactions and service communications further reinforces this duality. Experimental evidence on typing errors in chatbot communications indicates that mistakes are cognitively processed as human-like behaviours, yet they do not automatically enhance perceived humanness or social presence. Instead, in this framework, recipients may interpret such errors as signs of system incompetence or poor design. This suggests that textual imperfections function as context-dependent social signals, capable of activating human-like inferences while also generating negative competence-related evaluations (Bührke, Brendel, Lichtenberg, Greve, & Mirbabaie, 2021).

Importantly, these studies have primarily examined online reviews and service chatbot interactions, rather than AI-generated marketing emails embedded in persuasive contexts.

In the literature on send-time optimization in email marketing timing is predominantly treated as a technical variable aimed at maximizing open rates and click-through rates through predictive

modelling and machine-learning algorithms. Research in this area conceptualizes send-time as an optimization problem, focusing on identifying statistically optimal delivery windows based on historical engagement data (Araújo, et al., 2022).

However, this stream does not examine whether specific temporal patterns, such as perfectly rounded send-times versus irregular, non-round timestamps, may operate as perceptual cues that shape recipients' inferences about automation versus human agency. In other words, while timing has been studied extensively as a performance variable, its potential role as a paratextual signal of humanness or automation remains underexplored.

Taken together, these strands of research suggest that imperfections and delivery characteristics may carry socially meaningful information beyond their functional role. In addition, prior studies have largely investigated either explicit disclosure effects or error-related credibility penalties in isolation. Limited attention has been devoted to how subtle, non-declared micro-cues, such as minor typographical mistakes or non-standardized send-times, may implicitly shape perceptions in AI-generated email marketing communication. Addressing this gap is particularly relevant in contexts such as email texts where recipients mostly rely on heuristic signals to interpret message origin and intent. Overall, these findings point to a cohesive premise for the present empirical study: subtle "humanizing" cues in AI-generated marketing emails may shift perceived humanness and authenticity, thereby shaping trust-related evaluations and behavioural intentions.

6.2. Methodology

6.2.1. Research Design

The empirical study adopts a quantitative, between-subjects experimental design implemented through an online questionnaire administered via Qualtrics. The aim is to assess whether subtle "humanizing" cues embedded in a promotional email, namely minor typographical errors and the precision of the displayed sending time, systematically influence recipients' perceptions and behavioural intentions, in the absence of any explicit disclosure about the origin of the content.

The design was developed with two manipulated independent variables: (1) typos (absent vs present) and (2) send-time precision (exact round time vs inexact non-round time). This yields four experimental conditions: (V1) exact time (09:00) / no typos; (V2) inexact time (09:37) / no typos;

(V3) exact time (09:00) / typos; and (V4) inexact time (09:37) / typos. Each respondent was exposed to one version only, consistent with a between-subjects logic.

Random assignment to conditions was conducted automatically by Qualtrics. The randomizer was configured to maintain an approximately balanced distribution across conditions, ensuring that each version was presented in comparable proportions across the overall sample. The experimental condition viewed by each participant was stored as embedded data, enabling subsequent group comparisons in the statistical analysis.

Following exposure to the email stimulus, participants completed multi-item measures capturing key perceptual evaluations (trust-related and authenticity-related perceptions) as well as intention-based outcomes (engagement/interaction intention and purchase intention). All items were measured on seven-point Likert scales (1 = Strongly disagree, 7 = Strongly agree).

To ensure comparability across respondents and across languages, the questionnaire was administered in two languages (Italian and English). The two versions were designed to preserve semantic equivalence across all items, the email body and the manipulation features (typos and send-time display), maintaining parallel meaning and consistent experimental conditions across languages.

The final dataset was analysed in STATA.

6.2.2. Stimuli Development

The experimental stimuli consisted of four variants of the same promotional email. The starting point was a real commercial email originally sent by Deliveroo. To ensure that the stimulus could be used ethically and consistently across respondents, the email text was adapted with the support of ChatGPT (GPT-5.2) by removing any personally identifiable references and by standardising the content while preserving a realistic promotional structure. The resulting email offered a time-limited discount and encouraged recipients to interact with the message to redeem the offer, while also suggesting a set of example restaurant options (labelled A, B and C) and the possibility of exploring through the email additional restaurants in the app or website.

The four experimental versions were created by systematically manipulating two micro-level cues while keeping all other elements constant: (1) the presence of typographical errors in the email copy and (2) the displayed sending time, shown either as an exact round time (09:00) or as an inexact

non-round time (09:37). This procedure generated four stimuli that were identical in substantive content (offer structure, call-to-action and overall message layout) but differed only in the targeted manipulation features, allowing a focused test of how these cues affect perceived humanness-related evaluations and downstream intentions.

6.2.3. Participants

Participants were recruited through digital word-of-mouth distribution (online snowballing) using the anonymous Qualtrics survey link. A total of 137 responses were collected. Data screening identified 2 duplicate submissions and 29 incomplete responses (one of which was also duplicated). These cases were excluded prior to analysis, resulting in a final sample of 107 valid responses retained for statistical analyses.

Before exposure to the experimental stimulus, respondents completed a set of socio-demographic questions (age group, geographical origin, education level, occupation and gender). The survey also included behavioural background questions administered after the main dependent variables, capturing participants' self-reported frequency of opening promotional emails and their familiarity with AI technologies.

The sample was relatively young and highly educated. The majority of respondents were between 18 and 27 years old (53.27%), followed by participants over 37 (31.78%) and between 28 and 37 (13.08%), while a small proportion was under 18 (1.87%). In terms of education, 42.99% held or were pursuing a Master's degree, 22.43% a Bachelor's degree, and 34.58% reported a high school qualification. Regarding employment status, 53.27% were employed, 28.97% were students, 10.28% reported being both students and employed, while 7.48% were unemployed. The gender distribution was relatively balanced, with 54.21% identifying as female, 44.86% as male, and 0.93% as third gender.

With respect to email-related behaviour, most respondents reported incidental exposure to promotional emails (63.55%), while 26.17% indicated no active engagement, 5.61% reported checking emails prior to purchase decisions, while 4.67% declared actively searching for offers and purchase ideas via email. In terms of familiarity with artificial intelligence technologies, 44.86% identified as AI-experienced and frequent users, 26.17% as AI-aware but not active users, and 28.97% as unfamiliar with AI tools.

6.2.4. Procedure

Participants accessed the questionnaire through an anonymous Qualtrics link. Upon entering the survey, respondents first completed a set of socio-demographic questions including age range, country of origin, education level, occupation and gender.

After completing these preliminary questions, participants were automatically assigned to one of the four experimental conditions through Qualtrics' randomization function. Each participant viewed only one version of the promotional email stimulus. The email was presented in a realistic format and included the displayed send-time as part of the visual layout.

Following exposure to the stimulus, participants were asked to evaluate the email on several perceptual and intention-based measures. All evaluations were collected immediately after stimulus exposure in order to capture spontaneous impressions.

Subsequently, respondents answered three behavioural background questions concerning their proficiency in the language in which the questionnaire was administered, their frequency of opening promotional emails and their familiarity with AI technologies.

The questionnaire was available in both Italian and English. The two language versions were designed to maintain semantic equivalence across all items and manipulation features. Participation was voluntary and anonymous, and responses were collected without any personally identifiable information.

6.2.5. Measures

The study assessed four primary outcome constructs: perceived trust, perceived authenticity, intention to interact with the email (click intention) and purchase intention.

Perceived trust toward the email was measured using three items reflecting credibility and reliability of the communicated offer. Participants indicated their agreement with the following statements:

- "The promises or information presented in this email seem credible to me."
- "This email gives me the impression that the company would keep what it promises."
- "Overall, this email conveys a sense of trust toward the offer being presented."

The three observed items have been respectively renamed as “Credible”, “Reliable” and “Trustworthy”.

Perceived authenticity was assessed using three items capturing genuineness, naturalness and absence of artificiality as follows:

- “This email gives me a genuine impression (not ‘forced’ or ‘fabricated’).”
- “The content of this email does not feel artificial.”
- “The tone and style of this email feel natural.”

These items have been subsequently renamed as “Genuine”, “Non Artificial” and “Natural” respectively.

The measurement items were developed based on established scales and theoretical conceptualizations of trust, credibility and authenticity in prior research (Bruhn, Schoenmüller, Schäfer, & Heinrich, 2012; Erdem & Swait, 2004; Yan & Battocchio, 2021), and were contextually adapted to reflect evaluations of promotional email communication.

Two additional single-item measures captured downstream behavioural tendencies as:

- Click intention: “I would be willing to click on the links in this email to explore the offer or the company’s app/website.”
- Purchase intention: “This email makes me more likely to purchase the promoted product or service.”

It is important to point out that no functional hyperlinks or real purchase options were embedded in the stimulus. The study measured self-reported intentions rather than observed behavioural actions.

All above items were measured using a seven-point Likert scale ranging from 1 (Strongly disagree) to 7 (Strongly agree).

6.3. Research Expectations

Building on the theoretical framework outlined, the present study examines how subtle variations in promotional email communication may shape consumer evaluations and behavioural intentions. In particular, the experimental manipulations (grammatical errors and send-time punctuality) were designed to introduce cues that could plausibly trigger divergent interpretative mechanisms.

From a signalling and credibility perspective, grammatical errors may function as negative quality signals. Prior research suggests that inconsistencies or deviations from professional standards can activate scepticism and scam-related heuristics, especially in digital environments where consumers are frequently exposed to phishing attempts and low-quality communications. Under this interpretation, the presence of typographical or grammatical errors would be expected to undermine perceived trust and, consequently, reduce engagement behaviours such as clicking on embedded links or expressing purchase intention.

However, an alternative theoretical mechanism should also be considered. As discussed, in highly automated and AI-driven communication contexts, minor imperfections may serve as cues of human involvement. Given increasing awareness of algorithmically generated content, perfectly polished messages in timing and text might be interpreted as overly artificial or machine-produced. In contrast, irregularities could enhance perceived humanness and reduce perceptions of artificiality, potentially increasing authenticity perceptions and mitigating resistance. From this perspective, errors might not uniformly reduce behavioural intentions and could, under certain conditions, even generate a “humanization bonus”.

A similar pattern may apply to the manipulation of send-time punctuality. Strictly punctual communication may also signal automation, while slightly irregular timing could reduce perceptions of artificial scheduling but might be interpreted as less professional organization.

Given these competing mechanisms, the study does not assume strictly unidirectional effects of the experimental manipulations on click intention or purchase intention. Rather, it examines whether and how such cues influence downstream behavioural tendencies. Nevertheless, based on the broader literature on credibility and persuasion, perceived trust and authenticity are expected to play a central role in driving engagement responses. These are considered as evaluative mechanisms through which message characteristics may translate into behavioural intentions.

In summary, the study adopts a theoretically grounded but partially open approach. While trust and authenticity are expected to function as key explanatory mechanisms, the net effect of grammatical errors and punctuality cues on behavioural intentions remains an empirical question shaped by the interplay between credibility-based scepticism and authenticity-based humanization processes.

6.4. Results

6.4.1. Measurement Model and Construct Validation

Prior to analysing the effects of the experimental manipulations, the psychometric properties of the multi-item constructs were examined to ensure adequate construct validity and internal consistency. In line with the theoretical conceptualization presented in the Measures section, perceived trust and perceived authenticity were operationalized as latent constructs.

As an exploratory step, pairwise correlations among the six items were inspected (Table 1). Items designed to capture perceived trust showed significant and strong positive correlation values (all above 0.82 with a significance level of p-val<0.001) among themselves and weaker with the remaining items. The three authenticity items were also significantly and strongly correlated among themselves (with correlation values higher than 0.6 with a significance level of p-val<0.001). Cross-construct correlations were positive but comparatively lower than within-construct correlations, with "Genuine" as being the only one showing a more ambiguous result. These results provided preliminary evidence that the items grouped in line with their intended theoretical dimensions.

Table 1
Pairwise Correlations Among Items

Variable	1	2	3	4	5	6
1. Credible	—					
2. Reliable	.82***	—				
3. Trustworthy	.84***	.85***	—			
4. Genuine	.59***	.61***	.67***	—		
5. Non-Artificial	.44***	.36***	.50***	.61***	—	
6. Natural	.52***	.52***	.58***	.76***	.73***	—

Note. $N = 107$. Correlations are Pearson's r. ***$p < .001$.

To formally assess the measurement structure, a confirmatory factor analysis (CFA) was conducted using structural equation modelling (SEM). A two-factor model was specified, with the three trust items loading on a latent trust factor and the three authenticity items loading on a latent authenticity factor. The two latent constructs were allowed to correlate, reflecting theoretical expectations that trust and authenticity are related yet distinct evaluations of marketing communication (Figure 1).

After constraining equal to 1 the links between the new "Trust" item and "Credible", and the new "Authenticity" item and "Genuine", the standardized solution (Table 2) indicated strong and statistically significant (with a p-val level lower than 0.001) factor loadings for all items on their respective latent constructs, with loadings ranging approximately from 0.77 to above 0.94. These magnitudes exceed conventional threshold for acceptable indicator reliability of 0.05 and support validity of both constructs. In addition, the correlation between Trust and Authenticity was

substantial (approximately 0.70) at a si□nificant value (p-val < 0.001), indicatin□that while related, the two concepts remain distinct and are not empirically redundant. The two-factor model thus provided a theoretically coherent and statistically supported representation of the measurement structure.

Consequentially, based on the satisfactory factor loadin□s and overall model consistency, composite scores were computed as the arithmetic mean between the correspondin□items to □enerate the two new latent variables, namely "Trust" and "Authenticity". These mean indices were used in the analyses reported from now on.

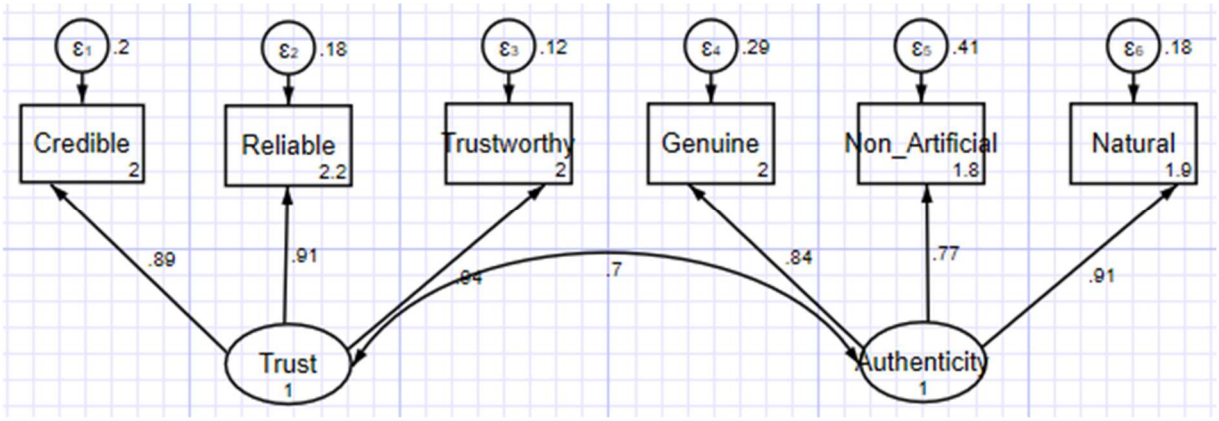
Table 2

Confirmatory Factor Analysis Standardized Loadings for Trust and Authenticity Constructs

Item	Factor	Loading	SE	z	95% CI
Credible	Trust	.89***	.02	37.03	[.85, .94]
Reliable	Trust	.91***	.02	40.87	[.86, .95]
Trustworthy	Trust	.94***	.02	50.61	[.90, .98]
Genuine	Authenticity	.84***	.04	21.13	[.77, .92]
Non-Artificial	Authenticity	.77***	.05	16.46	[.68, .86]
Natural	Authenticity	.91***	.03	27.38	[.84, .97]
Factor	**Factor**	**Covariance**	**SE**	**z**	**95% CI**
Trust	Authenticity	.70***	.06	11.26	[.58, .82]

Note. Standardized coefficients reported. SE = standard error. CI = confidence interval. ***p < .001. *N* = 107.

Figure 1 - Confirmatory Factor Analysis Standardized Model



6.4.2. Effects of Experimental Manipulations on Behavioural Intentions

To examine whether the four email versions differed in their overall impact on the primary outcome variables, a multivariate analysis of variance (MANOVA) was initially conducted with Email

Version (V1–V4) as the between-subjects factor, and “Trust”, “Authenticity”, “Click Intention” and “Purchase Intention” as dependent variables. The multivariate test did not reveal a statistically significant global effect across the combined dependent variables (Pillai’s Trace $p > .05$), suggesting that, when considered simultaneously, the four versions did not produce substantially different multivariate response profiles.

Given the conceptual relevance of engagement-related behaviours, subsequent univariate analyses were conducted to inspect potential differences at the individual outcome level. One-way ANOVAs were performed for each dependent variable separately. No statistically significant differences emerged across versions for “Trust” (p-val ≈ 0.35), “Authenticity” (p-val ≈ 0.46) and “Purchase Intention” (p-val ≈ 0.57). However, “Click Intention” displayed a comparatively stronger pattern of variation across conditions (Table 3), approaching statistical significance in the initial four-group comparison (p-val ≈ 0.115), thereby warranting closer inspection of the mean differences.

Table 3
One-Way ANOVA for “Click Intention” Across “Email Versions”

Source	SS	df	MS	F	p
Between Groups (Email Version)	22.09	3	7.36	2.03	.115
Residual	374.15	103	3.63		
Total	396.24	106	3.74		

Note. $N = 107$.

Descriptive statistics revealed a systematic pattern across the four experimental versions (Table 4). Version 1 (no typos and punctual timing) produced the highest mean for “Click Intention” (M ≈ 3.58), followed by Version 2 (no typos and non-punctual timin□; M ≈ 3.34). Version 3 (typos and punctual timing) yielded a lower mean (M = 3.04), while Version 4 (typos and non-punctual timing) generated the lowest results for “Click Intention” (M ≈ 2.37). Although the ANOVA did not reach conventional significance levels, the mean structure suggested a consistent downward shift in engagement when grammatical errors were present.

Table 4
Descriptive Statistics for “Click Intention” by “Email Version”

Email Version	n	M	SD
1	26	3.58	2.21
2	29	3.34	1.76
3	25	3.04	1.99
4	27	2.37	1.64
Total	107	3.08	1.93

Note. M = mean; SD = standard deviation.

This pattern suggests that the presence of errors may systematically reduce click-related engagement, even if the four-group comparison does not capture the underlying factorial structure efficiently. The progressive decline in average click propensity from error-free to error-containing versions suggests that grammatical imperfections function as a negative cue affecting immediate interaction tendencies. Notably, this deterioration is observable both in absolute terms and within comparable timing conditions, as error-containing versions consistently display lower means than their error-free counterparts regardless of send-time punctuality. In contrast, the punctuality manipulation does not appear to produce a clear or consistent mean separation across versions at this stage of analysis.

Taken together, the MANOVA and subsequent univariate comparisons suggest that the primary behavioural variation across email versions concerns click intention rather than purchase intention or perceptual evaluations considered in isolation. These findings motivated a more theoretically aligned factorial analysis (2×2 design) to directly examine the independent effects of grammatical errors and send-time punctuality, as well as their potential interaction.

6.4.3. Factorial Analysis of Grammatical Errors and Send-Time Cues

To more precisely examine the structure underlying the four experimental conditions, a series of 2×2 factorial ANOVAs were conducted. Rather than treating the four email versions as unrelated categories, the analysis leveraged the factorial design of the experiment, distinguishing between the presence versus absence of grammatical errors and send-time punctuality. To operationalize the factorial design, two dichotomous variables were created to reflect the factorial structure of the experimental design. The variable “Typos” was coded as 0 for email versions without grammatical errors and 1 for versions containing grammatical errors. Similarly, the variable “Time” was coded as 0 for punctual send-time conditions and 1 for non-punctual timing conditions. This approach allowed for the estimation of independent main effects as well as their interaction, thereby providing a theoretically aligned test of the experimental manipulations.

The factorial ANOVA revealed a statistically significant (p-val ≈ 0.043) main effect of grammatical errors on click intention (Table 5). As previously evidenced by the comparison of mean values, error-containing emails yielded lower click intention than their error-free counterparts. This mean difference was evident within both punctual and non-punctual timing conditions, indicating that the observed reduction in engagement is robust to punctuality cues.

In contrast, the main effect of send-time punctuality was not statistically significant (p-val $p \approx 0.225$), nor was the interaction between "Typos" and "Time", being the p-val ≈ 0.555. These findings suggest that while grammatical imperfections function as a meaningful engagement-reducing cue, variations in send-time punctuality do not independently influence click intention within the present design.

Table 5

Two-Way ANOVA for "Click Intention" ("Typos" × "Time")

Source	SS	df	MS	F	p
Typos	15.23	1	15.23	4.19	.043
Time	5.42	1	5.42	1.49	.225
Typos × Time	1.28	1	1.28	0.35	.555
Residual	374.15	103	3.63		
Total	396.24	106	3.74		

Note. $N = 107$. ηp^2 = partial eta squared.

For perceived trust, the analysis indicated a marginal main effect of grammatical errors, being the p-val ≈ 0.08. Although this effect did not reach conventional significance levels, this value suggests that emails containing errors tend to generate a negative pattern over trust perceptions, confirmed by the error-free versions' means being comparatively lower than the errors-filled versions' ones. No significant main effect of "Time" (p-val ≈ 0.66) or interaction effect ($p \approx 0.912$) was observed. This pattern implies that the trust impact is not meaningfully moderated by punctuality cues.

For perceived authenticity, neither grammatical errors nor send-time punctuality produced statistically significant effects and no interaction effect was detected (all p-val > 0.11 with "Typos" being the closest variable to this value). Similarly, "Purchase Intention" was not significantly influenced by either manipulation or their interaction (all p-val > 0.17). These results indicate that the experimental cues primarily affected immediate engagement tendencies rather than broader evaluative perceptions or downstream purchase intentions.

Taken together, the factorial analyses clarified that the primary behavioural impact of the manipulations concerns click intention, with grammatical errors emerging as the central experimental driver. While punctuality cues did not produce effects, the consistent reduction in click intention in error-containing conditions suggests that imperfections may act as strong negative quality signals in promotional email communication.

Furthermore, the marginal reduction in trust further hints at a potential explanatory mechanism linking message quality cues to behavioural engagement. This possibility is explored more directly in the subsequent regression and mediation analyses.

6.4.4. The Mediation Role of Trust in the Relationship Between Grammatical Errors and Click Intention

Building upon the factorial analyses, which identified grammatical errors as the primary experimental driver of click intention, subsequent analyses aimed to examine the psychological mechanism underlying this behavioural effect. Specifically, regression and SEM were employed to assess whether perceived trust mediates the relationship between grammatical errors and click intention.

A first regression model examined the direct association between “Trust”, “Authenticity” and “Click Intention”. The model was statistically significant (p-val < 0.001), explaining approximately 52% of the variance in click intention (R^2 adjusted ≈ 0.52). “Trust” significantly emerged as a strong and positive predictor of “Click Intention” (β ≈ 0.72 at a p-val level lower than 0.001), indicating that higher levels of perceived trust were associated with substantially greater willingness to click on the email content. “Authenticity”, in contrast, did not significantly (p-val ≈ 0.15) predict “Click Intention” when included alongside “Trust”, suggesting that trust plays a more central behavioural role in this context.

A second regression model incorporated the experimental manipulation of typographical errors alongside perceived trust as predictors of click intention. “Trust” remained a strong and statistically significant predictor (β ≈ 0.79 with p-val < 0.001), while the direct effect of grammatical errors on click intention was negative (β ≈ -0.3), although not statistically significant (p-val > 0.25) once “Trust” was included in the model. This pattern suggests that the behavioural impact of grammatical errors may operate primarily through their influence on trust rather than through a direct pathway. Taken together, these findings provide preliminary evidence for a mediation mechanism.

To formally test this mechanism, a structural equation model (SEM) with bootstrap estimation was conducted (Table 6). Bootstrap estimation with 12345 as fixed random seed and 5,000 replications was employed to obtain robust standard errors and confidence intervals for the indirect effects. The model specified “Typos” as the independent variable, “Trust” as the mediator, and “Click Intention” as the outcome variable.

The direct path from “Trust” to “Click Intention” was positive (β ≈ 0.79) and highly significant (p-val < 0.001), confirming the central role of trust as a driver of engagement. The direct path from grammatical errors to click intention was negative (β ≈ -0.3) but not statistically significant (p-val <

0.001) in the presence of "Trust". Importantly, the indirect effect of grammatical errors on click intention through the effect of trust was negative (β ≈ -0.46) but not significant (p-val ≈ 0.073), although approached conventional significance levels, while the total effect of grammatical errors was statistically significant (p ≈ 0.04) and negatively (β ≈ -0.76) related to click propensity.

Table 6

Mediation Analysis: Effects of "Typos" on "Click Intention" via "Trust"

Effect	Path	b	SE (boot)	z	p	95% CI
Direct Effects						
	Trust → Click Intention	.79	.07	10.98	< .001	[.65, .94]
	Typos → Click Intention	−.30	.26	−1.13	.257	[−.82, .22]
Indirect Effect	Typos → Trust → Click Intention	−.46	.26	−1.79	.073	[−.97, .04]
Total Effect	Typos → Click Intention	−.76	.37	−2.06	.040	[−1.49, −.04]

Note. N = 107. Standardized coefficients reported. SE = bootstrap standard error. CI = normal-based 95% confidence interval.

This pattern is consistent with a mediation structure in which grammatical errors reduce click intention primarily by lowering perceived trust. Although the indirect effect was marginal rather than conventionally significant, the combined evidence from the regression and SEM analyses supports the interpretation of trust as a central psychological mechanism linking message quality cues to behavioural engagement.

Overall, the analyses indicate that grammatical imperfections function as a negative quality signal that undermines engagement largely through trust-related processes. While the direct behavioural effect of errors diminishes once trust is accounted for, the total effect remains significant, suggesting that trust plays a substantive mediating role in shaping click propensity.

6.5. Discussion

The present study examined whether subtle "humanizing" micro-cues embedded in promotional email communication shape consumer evaluations and behavioural intentions in the absence of explicit disclosure regarding content origin. Drawing on prior literature on AI-mediated communication, signalling processes and human favouritism, the study explored whether such cues would activate authenticity-based humanization mechanisms or credibility-based scepticism mechanisms. The empirical results provide a coherent answer: in this context, credibility considerations dominate.

Across analyses, grammatical errors emerged as the only experimental manipulation producing a statistically meaningful behavioural effect. Specifically, error-containing emails generated lower click intention compared to error-free versions. This reduction was observable both in absolute comparisons across the four versions and within matched timing conditions, indicating that the negative association between errors and engagement is robust to send-time precision. In contrast, send-time punctuality did not exert any detectable influence on trust, authenticity, click intention, or purchase intention. The absence of interaction effects further suggests that these cues operated independently rather than synergistically.

The findings are particularly informative in light of the theoretical ambiguity outlined in the Research Expectations. On the one hand, typographical imperfections may function as cues of human origin, potentially increasing perceived humanness and authenticity in a digital environment where AI-generated communication is increasingly prevalent. On the other hand, such imperfections may signal reduced professionalism and competence, activating scepticism. The results indicate that, within persuasive email communication, the latter mechanism prevails. Error-containing emails did not increase authenticity perceptions and did not generate any compensatory behavioural benefit. Instead, they appear to have functioned as negative quality signals that reduce engagement propensity. In this sense, deliberate or “forced humanization” through imperfection does not appear to be an effective strategy in promotional email contexts.

The central explanatory role of trust reinforces this interpretation. Regression analyses revealed that trust strongly predicts click intention, accounting for a substantial proportion of its variance. Moreover, when trust was introduced into the model, the direct effect of grammatical errors on click intention became non-significant, and the indirect pathway through trust exhibited a negative sign consistent with mediation. Although the bootstrap estimate of the indirect effect did not reach conventional significance levels, the overall pattern aligns with a trust-transmission mechanism: grammatical errors reduce perceived trust and lower trust translates into lower willingness to click.

Equally instructive is the asymmetry observed between click intention and purchase intention. The manipulations influenced click intention but did not produce significant differences in purchase intention. Rather than representing an inconsistency, this divergence is theoretically coherent when interpreted through a funnel-based lens. Clicking constitutes a low-cost, exploratory action requiring minimal commitment, whereas purchase intention reflects a higher-commitment, economically consequential decision that depends on additional factors beyond message-level cues. In the context of a realistic Deliveroo promotional email, baseline predispositions toward the category may vary substantially across participants. Individuals who are not actively interested in ordering food may nonetheless be willing to click out of curiosity, yet remain unlikely to report

increased purchase intention. In this sense, the email operates as an initial engagement trigger: it can stimulate exploratory engagement through trust-related cues, but it does not independently overcome broader determinants of purchase decisions.

The null findings for send-time precision further refine the theoretical contribution of the study. While timing has been extensively investigated in email marketing research as an optimization variable aimed at maximizing open rates, the present results suggest that minor variations in displayed timestamp precision do not operate as salient paratextual cues of automation or human agency. Rounded versus non-rounded send-times appear insufficient to meaningfully alter recipients' inferences about intent, authenticity or competence. This implies that not all micro-cues embedded in digital communication are psychologically consequential. Some may function purely at a technical level without triggering attribution processes.

Taken together, the results contribute to a more nuanced understanding of humanization strategies in AI-mediated marketing communication. They suggest that human-likeness cues are not inherently advantageous and that the effectiveness of such cues depends on the balance between perceived human presence and perceived competence. In the present setting, typographical errors, although socially associated with human production, simultaneously signalled reduced reliability, thereby undermining trust and engagement. The findings therefore challenge the assumption that introducing imperfections can systematically mitigate scepticism toward automated communication. Instead, they indicate that preserving professionalism and trustworthiness remains critical in persuasive email contexts, even when brands seek to appear more "human".

More broadly, the study underscores the centrality of trust as the key evaluative mechanism linking message characteristics to behavioural intentions. In environments where digital scams and automated outreach are common, recipients rely heavily on cues that reduce perceived risk before engaging with promotional content. Subtle signals that erode trust can meaningfully reduce engagement. Consequently, strategies aimed at enhancing AI-generated marketing communication should prioritize trust-preserving forms of real human involvement rather than artificial imperfections designed to simulate humanness.

Overall, the evidence suggests that micro-level textual imperfections do not generate a "humanization bonus" in promotional email marketing. Instead, they function primarily as competence-related signals that shape engagement through trust-based evaluations. Send-time precision, by contrast, appears perceptually neutral. These findings refine the broader debate on AI aversion and human favouritism by demonstrating that in the absence of explicit authorship

disclosure, recipients rely on credibility-sensitive cues rather than on subtle stylistic irregularities when deciding whether to engage with marketing emails.

6.6. Limitations

Several limitations should be considered when interpreting the findings. First, the final sample size (N = 107) limits statistical power, particularly for detecting smaller effects and indirect effects in mediation models. This may partly explain why some relationships exhibited only marginal significance despite being directionally consistent with the proposed mechanism. Relatedly, the between-subjects 2×2 design implies relatively small cell sizes, which can increase standard errors and reduce the robustness of estimates, especially for interaction effects.

Second, the study relied on self-reported intentions rather than observed behaviours. Click intention and purchase intention capture meaningful downstream tendencies, but they do not perfectly map onto real-world clicking and conversion, especially in the absence of functional hyperlinks or an actual purchase pathway. As a consequence, the results should be interpreted as evidence about perceived willingness to engage rather than verified behavioural outcomes.

Finally, the external validity of the results may be constrained by the use of a single stimulus based on a specific brand (Deliveroo) and product category (food delivery). Participants' baseline involvement with the category, prior attitudes toward the brand, or situational relevance of the offer may have shaped purchase-related responses independently of the manipulations, potentially attenuating effects on purchase intention. Future research could test whether the same cue effects generalize across brands, industries and categories with different baseline purchase propensities.

7. Future Perspectives

As AI becomes increasingly embedded across the firms' communications ecosystem, its trajectory moves beyond incremental efficiency gains toward a deeper restructuring of how firms design and govern interactions with consumers. The technologies discussed in the following sections signal the emergence of a communication environment that is more predictive and contextually adaptive than any prior paradigm.

Against this backdrop, this section explores the future directions that are likely to shape the evolution of AI-driven relationships. It examines the technological innovations that are poised to mature, the behavioural dynamics they will influence as AI systems assume a more active role in crafting and delivering personalised messages. The purpose is to outline the pathways that will guide practical implementations in the next years.

7.1. Next-Gen Personalisation Tools

As seen, personalisation has already moved beyond rule-based targeting and static segmentation toward adaptive and context-aware generation systems not only capable to efficiently identify who gets the message but also what the message is, how and when it is delivered for every individual consumer.

This shift is being driven by a stack of already existing technologies, yet under unstoppable active development and in many cases already undergoing early commercial experimentation as LLMs fine-tuned for brand voice, multimodal generation systems, real-time adaptive agents and autonomous offer-personalisation engines. Together, these tools are continuously optimised through feedback loops (Büyüksomer & Tümbek Tekeoğlu, 2024; Patil, 2024).

This section will outline the core technological pillars, describing their emerging applications in marketing communication and highlighting early empirical tests and deployments. The emphasis is on technologies that are either already being piloted inside digital marketing workflows or sufficiently mature in research prototypes to be credible short-term candidates to be used in brand–consumer interaction.

Talking about LLMs, these resources are no longer used only to generate generic copy. They are increasingly prompt-steered to reproduce a specific brand voice and persuasion style for different micro-audiences, in real time. In practice, this means that the model writes many aligned text variants, each one calibrated to a different audience cluster or even to the inferred traits of a single recipient (Büyüksomer & Tümbek Tekeoğlu, 2024). Patil describes this as a move to psychographic and situational micro-targeting that is not hand-authored but generated on demand, with LLMs able to reframe the same offer using different value propositions, from safety to savings or exclusivity (Patil, 2024).

Early applications can be identified in AI-assisted email drafting systems which already generate tailored subject lines, preheaders and body copy conditioned on past open/click history, cart content and behavioural intent signals (Islam, et al., 2024; Patil, 2024), and “smart reply” and “assistive response suggestion” systems in customer service already deployed in email and chat support to draft answers that reflect customer sentiment, prior purchase context and cross-sell opportunities. Their purpose is twofold: reducing response latency (increasing efficiency) and maintaining tone consistency (guaranteeing brand safety) (Kannan, et al., 2016; Adam, Wessel, & Benlian, 2021).

What is new here is the emergence of LLMs for personalised persuasion. Recent experiments show notably improvements on persuasion effectiveness, implying that generative models are capable of delivering not only relevance but attitudinal influence, representing an early, research-validated pathway toward optimized one-to-one messaging (Matz, et al., 2024).

However, as seen earlier, next-gen personalisation is not confined to text. Diffusion models and multimodal generators now allow marketers to adapt every creative asset from text, image, audio and video to match segments or individuals (Ho, Jain, & Abbeel, 2020; Rombach, Blattmann, Lorenz, Esser, & Ommer, 2022).

In this direction, an emerging practice described by Büyüksomer is the creative modularisation. Brands define core narrative elements and value promises and generative systems then localise, restyle or recompose those elements for specific audiences or channels. AI could, for example, adapt the main banner image in a campaign to reflect the recipient’s recently browsed product category, lifestyle cues or geographic context (Büyüksomer & Tümbek Tekeoğlu, 2024).

Patil argues that multimodal generation is able to collapse the traditional bottleneck in campaign workflows since instead of briefing a designer for each market, AI can create channel-ready ad creatives and email hero banners in less time. In early commercial pilots, these systems are also being used for influencer-style content synthesis, in which AI generates short-form visual assets that

mimic an influencer's tone or aesthetic to maintain continuity across sponsored posts, targeted ads and promotional emails, while reducing turnaround time and cost (Patil, 2024).

Multimodal generators and optimisation mechanisms require campaign managers for personalising content and integrate and coordinate different communication channels for small groups of individuals. Conventional approaches to communication orchestration rely extensively on manual marketer input, thereby constraining the ability to personalise messages in terms of content, timing, frequency and tone. This is the reason why emerging deployments are trying to manage orchestration through AI systems that can act as always-on personal campaigns managers for each consumer. The objective is to be able to infer the next-best moment to contact a specific user, through the best channel, generate the creative and finally enforce fatigue and compliance constraints (Abboud, Hanna, Jeunen, Raheja, & Wheeler, 2025).

This underscores the critical role of shaping next-generation personalisation strategies around both temporal and contextual dimensions. Predictive send-time optimisation to automatise when to speak and frequency capping to determine the level of persistence have existed for years, but the new generation couples these predictive layers with generative layers so that timing, channel selection and message framing are co-optimised dynamically for each target (Araújo, et al., 2022). (Araújo)

Such real time adaptation processes go beyond static marketing automation workflows. The system continuously re-evaluates the state of the individual through recent browsing, prior opens or last purchases and proactively creates and deploys individualised contact messages that are both personalised and situationally aware (Abboud, Hanna, Jeunen, Raheja, & Wheeler, 2025).

As an example, early tests in AI-augmented customer service chatbots in sales and post-purchase support are not simply reactive FAQ bots: they integrate the recommendation logic to propose personalised upsells or recover abandoned carts in real time, using phrasing adapted to the customer's sentiment and purchase history. This strategy demonstrates a fusion of service communication and promotional communication (Adam, Wessel, & Benlian, 2021). The same orchestrating logic is being adapted to outbound marketing communications, where proactive outreach is generated and timed by the system itself rather than manually scheduled.

Another important emerging frontier regards Autonomous Offer Personalisation and Generative Offer Design. This development allows personalised offer construction, which means identifying what tailored offer should exist for one specific individual in a particular moment and not only which one among the existing offers should be more or less frequently shown. The key is using generative models to map a consumer profile and behavioural context and link it to a specific incentive or bundle that maximises acceptance probability (Challapalli, et al., 2025).

In Challapalli's research, empirical tests report double-digit improvements (from 80% to 94%) in predicted offer acceptance rate, that is the percentage of offers generated by the model that evaluators predicted to be likely accepted by customers. The research compared Supervised Fine-Tuning (SFT) versus Contrastive Fine-Tuning (CFT), associated to 80% and 94% acceptance rate respectively (Challapalli, et al., 2025).

SFT is the traditional method of adapting a pretrained language model to a specific task using labelled examples. In this case the model learns by minimising cross-entropy loss that is, it tries to predict the correct output (the offer) given an input (the customer profile). SFT can produce a model that performs well on known cases, matching known patterns, but struggles with new customer profiles (Howard & Ruder, 2018; Zie□ler, et al., 2019).

CFT instead adds another layer of learning by teaching the model not only what is correct, but also what is different or opposite using a contrastive loss function which pulls together the representations of accepted offers and the corresponding customer profiles, while pushing apart rejected offers and the same customer profiles. This way, the model learns a semantic understanding of what makes an offer appealing or not to different customers not just based on labels but based on similarity in meaning and motivation (Oord, Li, & Vinyals, 2018).

These findings suggest that future marketing communications may increasingly contain dynamically composed value propositions that did not exist in a static playbook (Challapalli, et al., 2025). Whereas historically, personalisation in communications meant choosing one out of others pre-written offers, this new direction involves generating offer variants that fit individual's past signals, price sensitivity and other motivational frames and then delivering it at the optimal time, through the optimal channel with appropriate tone. This is especially relevant for channels as email and in-app messages, where space is limited and the offer must be extremely context-relevant to justify an interruption.

A final class of near-future tools is simulation-driven. Firms are beginning to experiment with the idea of consumer digital twin. This is a behavioural model that approximate how a given profile (or small cluster) is likely to respond to different stimuli including tone, argument, incentive and timing. These predictive twins allow marketers to pre-screen which variants are most likely to succeed, reducing the number of weak or risky communications that reach actual consumers. The great advantage is that a consumer digital twin differs fundamentally from a static customer's model because it updates dynamically as new behavioural, attitudinal and contextual information becomes available, effectively becoming a "living" representation of how a customer thinks, reacts and decides (Shedlock, 2025).

In marketing contexts, consumer digital twins integrate diverse data sources, including behavioural histories, digital-channel interactions, qualitative feedback and broader market or environmental signals. The digital twin emerges from the integration of various technologies into a single loop that includes recommendation systems (ranking what to show), LLMs (generating how to say it) and response models (estimating whether it will work). This integration and simulation capability allows to simulate how a consumer is likely to interpret and respond to marketing stimuli, enabling firms to forecast message performance and refine communications before any real exposure occurs (Hornik & Rachamim, 2025).

Recent research shows that digital twins are already being explored in the domain of personalised advertising. Cui describes how digital media technologies combined with digital-twin models can create adaptive content environments in which marketing stimuli (visuals, messages and narrative cues) are aligned with predicted preferences, leading to stronger engagement and more favourable attitudinal responses (Cui, 2025). At the same time, Polimetla's review similarly emphasises that digital twins are transforming traditional segmentation, shifting from broad audience categories to fluid, real-time digital identities that affect the tone, timing and nature of a communication alongside consumers behaviour evolution (Polimetla, 2025).

Digital twins offer marketers a form of strategic foresight, enabling them not just to react to consumer behaviour, but to anticipate it in advance through predictive modelling and scenario testing. Although still in the early stages of adoption, the trajectory across marketing research suggests that consumer digital twins may become central to future communication strategies by supporting more relevant, accurate and context-sensitive personalisation at scale (Hornik & Rachamim, 2025; Shedlock, 2025).

Summarising, next-generation personalisation tools in marketing communications are characterised by four converging capabilities: generative production of copy and visuals conditioned on individual si□nals; real-time orchestration of timing, channel and tone at the level of the single recipient; autonomous construction of personalised offers and incentives; and simulation-based pre-testing of communication variants using behavioural twins. These technologies are not purely speculative. They are already under active experimentation in email marketing, social and influencer-style content creation, customer service chat and targeted promotional outreach, and early empirical studies indicate measurable gains in engagement and conversion outcomes.

7.2. AI in Multichannel and Omnichannel Strategies

A multichannel strategy refers to a model in which a retailer is present across several different channels, but each channel is managed separately. The goal is primarily to increase customer convenience by enhancing the performance of each individual channel independently. In this configuration, channels coexist but remain compartmentalised, without integration of information, processes, or customer interactions. Multichannel retailing emerged when increasing internet use encouraged consumers to browse and purchase through multiple options. In this scenario, AI has helped marketers to separately enhance the efficiency of their communication strategies (Faria & Carvalho, 2025).

However, AI power is not only limited to manage and strengthen communication channels individually, but it can also be implemented to develop coordinated communication strategies which involve different channels simultaneously, allowing the improvement of the general communicative performance. Taking a step forward, an omnichannel strategy represents an evolution of the previous approach. Omnichannel retailing refers to the deliberate use of several communication channels ranging from emails to websites, social media, mobile apps, search engines and physical stores, to reach and engage customers, with each channel managed as an integrated part of a broad portfolio of touchpoints. Digital marketing research typically classifies email, social media, SEO/SEM, display and text messaging as core elements of such omnichannel architectures, emphasizing their joint contribution to reach engagement and sales rather than the performance of any single channel in isolation, making consumers to encounter a seamless, consistent and unified experience (Faria & Carvalho, 2025).

Compared with the more tightly integrated omnichannel systems, multichannel strategies often suffer from fragmented data, inconsistent messaging and limited visibility of the end-to-end customer journey. For this reason, while multichannel approach improves access, omnichannel strategies ensure that information, services and customer interactions remain coherent and synchronised across the entire customer journey. This step increasingly positions AI as a coordinating layer that sits above individual channels and improves both integration and effectiveness. This approach allows customers to switch effortlessly between channels at any stage (search, purchase, delivery, post-purchase) while retailers guarantee transparency, visibility and real-time consistency of information throughout the whole system (Faria & Carvalho, 2025).

As discussed, Gen-AI allows marketers to address a long-standing problem highly important especially in multi and omni channel campaigns: the need to produce and adapt content for many

formats while maintaining consistency. Building on this, the MARK-GEN framework proposed by Islam provides a structured foundation for using generative AI as a central creative engine capable of supporting these architectures. Since MARK-GEN proposes a detailed planning and develops a single generative model from unified data, brand guidelines and behavioural insights, it allows firms to generate channel-specific variations while preserving a coherent narrative and tone of voice. This directly addresses first, the fragmentation that often undermines multi-channel campaigns, where content must be repeatedly rewritten and reformatted for each platform and second, the framework's need for iterative operations mirroring the real-time adaptation required in omnichannel orchestration. In this sense, MARK-GEN creates the operational infrastructure through which generative AI can maintain scalability and cross-channel integration within multimodal strategies (Islam, et al., 2024).

Support for the superiority of integrated AI in omnichannel strategies comes from Sujeet Dutta's empirical analysis, which demonstrates how AI enhances both customer experience and operational coherence when deployed across interconnected touchpoints rather than isolated channels. The study shows that GAI strengthens omnichannel performance by maintaining context as customers move between online, mobile and physical environments, enabling real-time personalisation, unified content delivery and synchronised service interactions (Dutta, 2024).

The analysis of performance on Sephora's AI-powered chatbot, called "Sephora's Virtual Artist", had meaningful results. This chatbot integrates language models to understand and fulfill requests in a conversational way, while image recognition technologies can analyse and provide tailored recommendations based on pictures uploaded and Sephora's inventory systems allows customers to check products availability all into a single flow. The results have evidenced an 11% increase in bot-assisted conversion rates, a 32% rise in average order value and a significant reduction in return rates because customers receive more accurate and consistent advice across platforms (Dutta, 2024).

Again, IKEA's Augmented Reality-based app provides further evidence. The app, called "IKEA place", allows customers to place pieces of furniture in a virtual representation of theirs houses by using GAI. The AI is able to use 3D Rendering and spatial recognition systems to copy their products and place them in a virtual to scale representation of the customer's physical space. Also, a Generative Design system can suggest complementary products and generate room layout based on customers' preferences and existing furniture. The app use achieves a 35% decrease in product returns due to size issues and a 22% higher conversion rate compared to non-app users. These findings confirm that customers benefit when AI systems maintain continuity rather than resetting cognitive and informational context at each channel boundary (Dutta, 2024).

Other studies show that, although not specifically focused on comparing performance differences in omnichannel AI marketing, the integration of multiple AI technologies and the analysis of aggregated data coming from each one of them provide a more complete and detailed picture, as well as more accurate estimates and predictions of consumer behaviour.

Evidence from analytic research confirms the hypothesis that AI systems that integrate data across multiple channels outperform approaches that treat each stream in isolation. Yin's research proposes an AI-integrated decision support system that fuses heterogeneous marketing data coming from social media interactions (Twitter or TikTok), advertising records (Meta Ads and Google Ads), consumer behaviour logs, sentiment dynamics and market indicators, into a single hybrid AI model that combines a Graph Neural Network (learns relationships across different channels or data sources) and a Temporal Transformer (learns how these signals change over time). This multi-source integration allows the model to jointly understand cross-channel connections and temporal patterns, making predictions more accurate than single-channel models in content diffusion. The study shows that this system combining information from multiple channels performs better than traditional single-channel models that only analyse one stream of data at a time and cannot capture how signals interact across platforms over time. For example, it has been demonstrated that the integrated system achieves the lowest Root Mean Squared Error (0.063) and highest R^2 (0.911). The authors explicitly attribute these gains to the ability to combine signals from multiple platforms and data types, arguing that multi-source integration enhances robustness and decision relevance in complex digital ecosystems (Yin, Chen, & Zhang, 2025).

A complementary perspective comes from Liu, who develops a unified AI analysis system for multi-channel marketing data using advanced anomaly-detection algorithms. The system collects and standardises data from social media, email campaigns, web analytics, CRM systems, mobile apps and traditional media (TV, radio and print advertising) into a unified analytics environment and then applies machine-learning and ensemble methods to detect abnormal performance patterns across channels. Experimental results show anomaly detection accuracy above 94% with false positives below 3.8% and participating organisations report on average a 23.7% improvement in ROI due to enhanced marketing campaigns effectiveness and reduction in operational waste after deploying the integrated framework. These improvements arise precisely because the AI monitors cross-channel patterns that would be invisible or delayed if each channel were analysed separately (Liu, 2025).

Overall, the findings clearly show that generative AI's value in retail emerges most strongly when it is embedded as an integrative layer that synchronises information, personalisation and operations

across channels, thereby solving fragmentation and delivering measurable improvements in both marketing performance and customer satisfaction.

To conclude, recent research projects on AI-driven personalised journeys confirm how these systems now shape the customer experience across entire pathways rather than within isolated channels. They explain that AI-generated recommendations, curated options and automated prompts accompany consumers as they transition between platforms, progressively influencing how they perceive relevance, convenience, empowerment, trust and support throughout the journey (Hardcastle, Vorster, & Brown, 2025).

Their analysis suggests that AI-enabled multi-channel strategies can enhance engagement and strengthen brand relationships by simplifying search and expanding perceived choice when coordination is transparent and respectful of autonomy. Yet, it can also create feelings of intrusiveness or reduced autonomy when the underlying mechanisms are opaque or overly prescriptive. In these cases, consumers may feel monitored or steered, which weakens trust and undermines the perceived value of the relationship (Hardcastle, Vorster, & Brown, 2025).

Taken together, this emerging evidence indicates that AI enhances multi-channel strategies along three interdependent dimensions: integrating data to form a coherent view of the customer across channels; □eneratin□ and adaptin□ content that is both channel-specific and strate□ically consistent; and orchestrating sequencing so that touchpoints reinforce rather than cannibalise each other. When these capabilities are combined, multi and omnichannel communication shifts to a more unified learning structure in which each interaction informs the next, increasing the overall effectiveness of the firm’s marketing communications portfolio.

8. Ethical Dilemmas of Using Generative AI in Marketing

While generative AI offers unprecedented opportunities in marketing communications, its adoption also introduces complex ethical challenges. As discussed in Chapter 4, AI-generated and human-generated content are often perceptually indistinguishable, and consumer evaluations are strongly shaped by disclosure, labelling and inferred authorship rather than by objective detection. This perceptual ambiguity amplifies the ethical stakes of AI-driven marketing: when users cannot reliably identify AI involvement, issues such as transparency, data privacy, consent, analysis biases and the risk of over-automation raise questions about fairness and accountability. Addressing these dilemmas is essential to ensure that the benefits of AI-driven marketing do not come at the expense of ethical standards. The examination of the major concerns related to AI application will be the topic of this section because being conscious of the risks besides the advantages of any technology is the key to maintain control over it and foster broader social other than economic growth.

8.1. Ethical Dilemmas in Detail

To understand the reason why AI assisted processes gave way to the raise of ethical concerns it is necessary to understand which these concerns are in detail.

First, there are privacy and surveillance concerns. AI-enabled personalization typically depends on fine-grained tracking and profiling with invisible tracking pixels, tagged links and cross-device identifiers features which expand the scope of data capture and inference beyond what most recipients can reasonably anticipate. Large-scale measurements show email tracking is widespread and technically sophisticated, with pixels and tracking links prevalent across marketing emails. Detection studies document both the scale and stealth of these practices (Haupt, Bender, Fabian, & Lessmann, 2018). Even when notice is provided, the notice-and-choice model struggles in practice: reading and understanding disclosures is prohibitively costly, privacy judgments are context-dependent and malleable and interface dark patterns can steer users away from privacy-protective choices, together undermining the substance of consent in programs powered by AI (McDonald & Cranor, 2008). Normatively, these problems are well captured by contextual integrity. In fact, privacy depends on whether data flows conform to the norms of a given context while cross-context

repurposing risks violating those norms even if some form of consent was obtained for example, the practice of inferring sensitive traits from click/open histories and enriching profiles for persuasion (Nissenbaum, 2009).

Another important risk is the risk that involves manipulation and loss of consumers' autonomy. Generative systems make message personalization cheap and scalable, increasing the persuasive power of messages. Evidence from psychological targeting shows that tailoring persuasive appeals to psychometric attributes changes behaviours, raising autonomy concerns when such tailoring exploits decision-making vulnerabilities with scarce recipients' awareness (Matz, Kosinski, Nave, & Stillwell, 2017). As discussed in Chapter 4, these concerns are further intensified by evidence that consumers frequently rely on perceived humanness and authenticity cues when evaluating personalised messages, even when such cues are artificially engineered, limiting their awareness of how and why specific persuasive strategies are applied. Philosophical and legal analyses distinguish legitimate persuasion from online manipulation, defined as "hidden influence" that covertly subverts decision power. Moreover, big-data personalization enables "hypernudges" (dynamic, adaptive choice architectures) that can shape preferences and behaviours continuously. These concepts frame AI-crafted content as potential sites of autonomy harm when individuals cannot discern how the message was tailored or why they were targeted (Susser, Roessler, & Nissenbaum, 2019).

About fair and non-discriminatory rules in targeting, AI-driven selection can inadvertently skew exposure to offers, discounts, or service messages across demographic groups, even when advertisers specify inclusive targets. Audit and field experiments on advertising delivery show algorithmic bias arising from optimization itself, with gender- and race-linked disparities in who sees opportunity-related content (Ali, et al., 2019). Relatedly, opacity in profiling pipelines can hide pathways by which sensitive traits are inferred and acted upon, complicating internal accountability and external scrutiny (Datta, Tschantz, & Datta, 2014).

Given that this is an operating field situated within the broader context of global digital interconnectivity, other risks come from a security perspective with informatic scams and frauds. The same generative capabilities that improve marketing copy can lower the cost of crafting convincing, personalized phishing or business-compromise communications, eroding the broader trust environment around electronic interactions. Experimental and systematization studies find that LLMs can generate high-quality spear-phishing content and accelerate attacker workflows, raising in fact a dual-use risk (Hazell, 2023; Heidin□, Schneier, Vishwanath, Bernstein, & Park, 2024).

Finally, content creation is a major ground of marketing work including tasks like copywriting, design and experimentation. Analyses indicate that LLMs expose a large share of communications and marketing tasks under substitution pressure for junior creative roles. At the same time, field evidence shows augmentation effects which ensure productivity gains concentrated among less-experienced workers and reducing the amount of specialised and skilled workers. It is clear that involving AI in marketing processes has heavy impacts on professional dynamics that raise ethical questions about fair transition, reskilling and the potential for substitution and intensified algorithmic management of creative labour (Eloundou, Manning, Mishkin, & Rock, 2023).

8.2. Data Privacy and Consumer Consent (Theoretical Exercisability of the "Right to Object")

In the previous section it has been investigated how one of the major concerns involves consumers' data security and privacy. But the central question is to what extent they are free of exerting the possibility to object to have their data stored and used indeed. In addition, when AI-generated communications are perceived as human-authored, users may not even realise that a data-driven system is operating, undermining the practical exercisability of the right to object before any interface-level control is encountered.

In marketing contexts, a right to object can be understood, in purely normative terms, as an expression of individual autonomy: the power to refuse or withdraw participation in data-driven personalization when continued processing conflicts with one's preferences, interests, or identity. Exercising such a right presupposes two ethical preconditions: explicability, which means that individuals are able to understand what a system is doing and why, and agency, which means that individuals have a real ability to act on that understanding (Floridi, et al., 2018).

Theoretically, a right to object becomes actionable when several conditions hold. First, intelligibility: people receive contextually appropriate and decision useful explanations. Second, low-friction choice architecture: mechanisms to object or withdraw are accessible, immediate and reversible, without imposing cognitive and procedural burdens that negate the choice. Third, effectiveness beyond the interface: objections propagate across the full socio-technical system from datasets to downstream activations, so that opting out actually changes future inferences. Last, accountability artifacts: organizations maintain documentation among which change logs and decision summaries, and operational routines that make user control auditable and reliable in

practice. Managerial guidance on ethical AI stresses translating values into infrastructures, processes and practice, so that these conditions can be concretely realized (Blackman, 2020).

Although notice-and-choice frameworks aim to preserve autonomy by requiring disclosure and user decisions, empirical research shows that they rarely deliver meaningful control. Cognitive overload, bounded rationality and the complexity of privacy policies prevent users from making informed choices, while organizational practices often add friction that undermines the theoretical right to object (McDonald & Cranor, 2008; Acquisti, Brandimarte, & Loewenstein, 2015). People's privacy judgements are highly context-dependent and malleable, so choices made under one framing may not reflect stable, autonomous preferences. For example, default settings have a high impact on consumers' decisions and sticking to default settings usually appears convenient other than being considered as implicit recommendations (Acquisti, Brandimarte, & Loewenstein, 2015). Moreover, repeated consent prompts produce privacy/consent fatigue and exhaustion, weakening the very intentions that a right to object is meant to preserve (Choi, Park, & Jung, 2018). Interface dark patterns further erode agency by nudging or obstructing users away from privacy-protective choices, thereby converting a nominal right into a practically hollow one (Mathur, et al., 2019). Finally, privacy is often contextual and relational: data flows that violate contextual norms, as repurposing signals across settings, or infer attributes from seemingly innocuous data can circumvent individual control, limiting the practical scope of an objection even when an interface-level choice exists (Nissenbaum, Privacy as contextual integrity, 2004).

From a theoretical standpoint, the right to object is exercisable when intelligible explanations, low-friction and reversible controls, system-wide propagation and auditability are jointly present. However, empirical findings on cognitive load and manipulative choice architectures imply that without careful design and governance, the right risks becoming performative, superficial or symbolic rather than substantial. Therefore, realizing this right requires aligning explanation, choice architecture and organizational operations with human cognitive limits and contextual privacy norms, not merely placing a control in an interface.

8.3. Algorithmic Biases and Fairness

Within the guidelines proposed by Floridi in his paper, justice (fairness) and explicability (intelligibility and accountability) are core ethical principles for a "Good AI Society". However, since algorithmic outcomes are always conducted by humans which set principals, directions and

main functioning mechanisms, human biases also have a cascading effect on these automations. Hence, fairness problems are not merely technical defects but value-laden choices that must be governed (Floridi, et al., 2018).

Human biases shape algorithms through decisions made at each stage of the machine-learning lifecycle: how the problem is framed, how data are collected and labelled, which objectives are optimized, how models are evaluated and how systems are deployed and updated. Within this process, taxonomies in the scholarly literature identify recurring failure modes: historical bias, representation (sampling) bias, measurement/label bias, aggregation bias, evaluation bias and deployment/feedback bias. Each one of these biases represents a distinct pathway by which sociocultural inequities can be encoded into automated decisions (Suresh & Guttag, 2021). Among the most common biases, the following can certainly be identified.

First, there is what can be considered as an "objective selection bias": seemingly neutral goals ("maximize click-through at lowest cost" for example) embed business values and constraints that can systematically disadvantage certain groups when costs, reach, or response differ by socio-demographic features. Empirical audits of online delivery systems show that optimization itself can yield skewed exposure even without explicit use of protected attributes (Lambrecht & Tucker, 2019).

Second, training data often reflect prior human judgments or structural inequities, generating an "historical bias". Models that "learn from history" reproduce those patterns, so that improving accuracy alone can preserve injustice rather than correct it (Barocas & Selbst, 2016).

Moreover, "sampling biases" can arise. In fact, if some populations are under-represented or missing in the training sets, error rates will be uneven. Landmark audits of commercial systems showed large accuracy gaps by gender and skin tone, illustrating how imbalanced data propagate into disparate performance (Buolamwini & Gebru, 2018).

When constructs of interest (quality, engagement or relevance) are proxied by noisy or partial signals, labels inherit the prejudices and contexts of their creation including for example complaint counts and click proxies, leading models to optimize toward biased or incomplete targets, originating what can be defined as a "label bias" (Mehrabi, Morstatter, Saxena, Lerman, & Galstyan, 2021).

"Proxy biases" can arise when protected attributes are removed. In this case, other variables (location, device, time-of-day or interests) can act as proxies, re-introducing sensitive information

and enabling indirect discrimination through correlated attributes when the main ones that the model exploits are missing or omitted (Barocas & Selbst, 2016).

A single global model that assumes homogeneity can misfit subpopulations whose data-generating processes differ in language use or access patterns for example, thus yielding systematically unequal errors across groups and creating an “aggregation bias” (Suresh & Guttag, 2021).

Benchmarks and validation disruptions often mirror training skews and “evaluation biases” arise. If a subgroup’s performance is not separately measured, disparities remain invisible and “good overall metrics” can mask harms to minorities (Mehrabi, Morstatter, Saxena, Lerman, & Galstyan, 2021).

Finally, when an AI system is deployed in the real world, its decisions influence the new data generated. That means the future data isn’t independent of the system. Instead, it is partly produced by the system’s own outputs. Therefore, once deployed, models influence behaviour and future data by shaping their own data and creating “feedback biases” into loops that entrench initial skews unless actively monitored and corrected (Suresh & Guttag, 2021).

However, formal results show that widely used group fairness criteria such as equal error rates (which focuses on estimating rates of mistakes) and predictive parity (which focuses on estimating the correctness of the model’s predictions) cannot, in general, be satisfied simultaneously. This means that practitioners must choose which notion of fairness to prioritize for each specific context and justify that choice (Corbett-Davies, Gaebler, Nilforoshan, Shroff, & Goel, 2023).

As Floridi emphasizes, fairness is anchored in justice and autonomy. Biases arise not only from flawed data or models but also from human decisions about objectives, acceptable errors and oversight. It follows that decisions must be made transparent and contestable to align with societal values (Floridi, et al., 2018).

8.4. Regulation and Policies

To remedy and limit these ethical problems and risks, many governments have set rules to the use of automations in business processes when operating on consumers. Public governance of AI relevant to marketing communications is anchored in a small set of governmental and inter- and supra-national instruments through binding laws or public standards that define principles, duties and controls for trustworthy AI use. Below, there will be summarised the most salient frameworks and

specific obligations they introduce for AI-mediated content and personalization such as transparency of AI interaction, management of risks arising from AI, documentation and oversight.

The AI Act is the first comprehensive horizontal AI law issued by the European Union in 2024. It classifies AI systems by risk evaluation and imposes obligations on providers and deployers including documentation, risk management, transparency and oversight duties. For marketing-adjacent uses, two points are especially salient. First, transparency duties for AI systems that interact with people which imposes that users must be informed and for synthetic content (deepfakes) mandatory disclosure through clear advice and watermarks. Second, governance requirements for general-purpose models that may be embedded in content workflows. These provisions set the baseline for disclosure and accountability in AI-authored communications (REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, 2024).

For any email, SMS or any other push personalization, the EU privacy duo remains foundational. The “General Data Protection Regulation” (GDPR) issued in 2016 establishes lawful-basis requirements. These strict requirements focus, among others, on the concepts of consent and legitimate interests. The law aims to minimise the amount of data collected and processed and limit the purposes for which specific categories of sensitive data can be managed. The law also strongly enforced data-subject rights including principles that directly constrain data collection and profiling used to tailor messages giving, on the other hand, right to the consumers to object and withdraw from consent at any time (REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, 2016). The second piece of regulation comes from the “ePrivacy Directive” of 2002, which provides channel-specific rules for electronic communications with opt-in and opt-out mechanics for direct marketing, thus complementing GDPR’s general principles (DIRECTIVE 2002/58/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, 2002). Together, they shape the data inputs and control interfaces for AI-assisted personalization in direct channels.

For platforms and very large online intermediaries, the “Digital Services Act” (DSA) adds advertising and recommendation-system transparency by informing users why content is suggested and the main parameters involved, offering meaningful control. While not being an AI-specific statute, these explainability obligations affect AI-mediated content ranking and targeting in platform environments and create a disclosure baseline that influences cross-channel marketing practices (REGULATION (EU) 2022/2065 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL, 2022).

The Council of Europe Framework Convention on AI is the first international, legally binding treaty specifically on AI, opened for signature in September 2024. It requires parties to ensure AI activities are consistent with human rights, democracy and any other established and publicly known laws to be equally and fairly applied, and to adopt proportionate safeguards across the AI lifecycle. Although high-level, it will guide national legislation and supervisory expectations in jurisdictions that ratify it, thereby shaping the normative environment for AI-generated communications (Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, 2024).

Beyond regional borders, OECD members and partners adopted in 2019 intergovernmental principles which are now the de facto global soft-law baseline for a trustworthy AI management. These principles focus on human-centred values, fairness, transparency and explainability, robustness and safety, accountability and they also include recommendations for international cooperation and policies on AI. They are frequently incorporated by reference in national strategies and sectoral guidance, and many organizations align their internal AI governance and disclosure practices to them (OECD.AI Policy Observatory, 2025).

Moreover, a universal normative instrument was issued by UNESCO and endorsed by its member States in 2021. This “Recommendation on the Ethics of AI” articulates human-rights-based ethical principles and policy actions which include transparency, fairness, human oversight, data governance and also education on the general adoption of AI. For marketing and communication, the Recommendation strengthens the expectation that AI-mediated content must be identifiable, explainable and respectful of autonomy, influencing policy debates and professional standards (Unesco, 2022).

Issued by the International Organization for Standardization and International Electrotechnical Commission, the ISO/IEC 23894 of 2023 provides process-level guidance for AI risk identification, analysis, treatment and monitoring across the lifecycle. It is increasingly used by public authorities and industry as a compliance-by-design reference, complementing statutory duties as AI Act risk management and privacy by design, giving concrete structure to model documentation, testing and monitoring in content pipelines (Simonetta & Paoletti, 2024).

All these pieces of regulations and recommendation elevate transparency and accountability as normative expectations on an international level.

8.5. Private and Voluntary Responsible Innovation Frameworks

Beyond binding regulation, many organizations adopt voluntary frameworks to structure the safe use of AI in content creation, personalization and analytics. These range from principles of good practices to AI management and AI risk frameworks about how to run processes and operational artifacts which rules how to document and audit specific models and disclosure. Together they translate ethical ideals of explicability and justice into ordinary, daily governance for marketing teams. These practices help companies to improve transparency and inspire trust among the audience of users and targets of their communication and personalization actions (Floridi, et al., 2018).

Some researchers distil beneficence, non-maleficence, autonomy, justice and explicability as five principles for a good AI-based society, which many firms use as a north star for disclosure, oversight and fairness in AI-mediated communications. As a voluntary foundation, these principles frame transparency as both intelligibility and accountability, shaping internal policies for personalized content pipelines (Floridi, et al., 2018).

As discussed earlier, OECD's AI principles function as widely adopted soft law, ensuring human-centred values, fairness, transparency/explainability, robustness/safety and accountability. Even when not legally required, companies align disclosure practices, model monitoring and incident response to these intergovernmental principles to signal trustworthy use of AI in consumer engagement (OECD.AI Policy Observatory, 2025).

Operational frameworks for risk management in AI use come from NIST AI Risk Management Framework (AI RMF 1.0) which provides a voluntary, lifecycle-oriented method to identify, measure and mitigate AI risks across govern, map, measure and manage functions. Marketing teams adapt it to register personalization models, assess risk levels for privacy, biases or manipulative targeting, define controls including guardrails or human-in-the-loop systems and monitor incidents (NIST AI, Artificial intelligence risk management framework (AI RMF 1.0), 2023). In addition, NIST's 2024 Generative AI Profile extends the Risk Management Framework with concrete practices for content generation and disclosure (NIST AI, 2024).

Same as for OECD's principles, even when they are not mandatory, ISO/IEC 23894 of 2023, focused on AI risk management, guides organizations to integrate AI-specific risk analysis and treatment into existing governance and ISO/IEC 42001of 2023, focused on AI management systems, specifies a certifiable management-system approach for AI governance, following the

steps of Plan-Do-Check-Act. Firms adopt these to move from ad-hoc best efforts to auditable routines, particularly useful when personalizing large volumes of digital content (Simonetta & Paoletti, 2024).

Moreover, to strengthen and operationalize exercisability of the right to object without invoking specific legal regimes firms also could minimize interaction costs for objection or withdrawal right, eliminate coercive or misleading patterns in interfaces, propagate objections through downstream activations, maintain documentation and provide purpose-specific explanations tied to concrete actions (Floridi, et al., 2018; Blackman, 2020).

In addition, firms increasingly complement ethical principles and internal risk frameworks with detailed, model-level documentation that enhances transparency throughout the AI lifecycle. One of the most widespread tools in this regard is the Model Card, a concise document that outlines a model's purpose, data sources, evaluation metrics, limitations and recommended uses. Within the context of marketing personalization, model cards clarify which variables influence targeting decisions, identify content domains that should be avoided and specify the points at which human review becomes necessary before automated outputs are released at scale (Mitchell, et al., 2019).

Another dominant documentation practice is the use of datasheets for datasets, which record the provenance, composition, collection methods and consent assumptions underlying the data used to train AI systems. These datasheets help prevent the silent repurposing of behavioural data and encourage privacy-respecting personalization, ensuring that datasets remain aligned with the context for which they were originally collected (Gebru, et al., 2021).

Empirical studies in human–computer interaction show why these documentation artifacts matter: they make ethical and fairness considerations concrete, improving a team's capacity to identify risks early and to translate abstract ethical values into reviewable design decisions. This effect is particularly significant in large, cross-functional product teams, where operationalizing shared ethical standards can otherwise be difficult (Madaio, Stark, Wortman Vaughan, & Wallach, 2020).

A growing literature describes internal algorithmic auditing as a voluntary practice that spans the development lifecycle from scoping to data curation, modelling, evaluation and deployment, and produces an auditable trail as policies or tests (Raji, et al., 2020).

In practice, adopters combine: principles (OECD) to set expectations for autonomy, justice and explicability; a risk/mana□ement system (NIST RMF; ISO/IEC 42001 and 23894) to assign roles, controls and monitorin□; documentation artifacts (Model Cards, Datasheets and FactSheets) to make models/datasets le□ible to reviewers and business owners; audits and checklists to ensure new

campaigns pass fairness, privacy and disclosure gates before scaling. Field studies indicate that checklists and templates must be co-designed with practitioners to be used consistently, and that no single template suffices (Madaio, Stark, Wortman Vau□han, & Wallach, 2020; Raji, et al., 2020).

In conclusion, focusing specifically on the implications for personalised content, the voluntary toolset above enables firms to state when and how AI is used in copy creation, document data sources and limits. Beyond that, test targeting and content for bias and manipulation risks, and finally, govern continuous experimentation under explicit ethical guardrails, thereby operationalizing explicability and justice from the theory in day-to-day marketing operations (Floridi, et al., 2018).

Such governance mechanisms become particularly important in light of evidence that AI-generated content can appear convincingly human, making transparency and disclosure not merely formal requirements but central conditions for ethical and trustworthy marketing communication.

9. Conclusion

9.1. Recap on Findings

This work set out to clarify how generative AI is reshaping consumer decision-making and brand-related outcomes in marketing communication, with particular attention to personalisation and the behavioural metrics that organisations use to evaluate campaign effectiveness. Across the discussion, the central conclusion that emerged was that generative AI can meaningfully improve performance-relevant outcomes in marketing communication, especially at early stages of the customer journey, but its effectiveness is neither uniform nor purely technical. Rather, it is mediated by consumer inferences about authorship, humanness and intent, and it is bounded by perceptual constraints that can offset short-term gains.

The conceptual discussion established the background necessary to interpret this tension. The analysis positioned generative AI within a longer evolution path, from rule-based automation and predictive modelling, toward systems capable of producing persuasive language at scale, thereby shifting marketing practice from optimisation of targeting parameters to the scalable generation and adaptation of content. This shift does not merely automate delivery, but it intervenes in a layer of marketing communication where meaning and perceptions are constructed. In this perspective, the consumer response to AI-driven communication is not only a function of relevance or quality. It also reflects how the communication is interpreted as a social signal: who, or what, is perceived to be speaking and with which motives.

Building on this foundation, the thesis synthesised evidence on effectiveness in a way that connects outcomes to mechanisms. The review of AI applications in marketing communication highlighted that generative systems are increasingly used to support copywriting, creative variation and scaling personalisation, enabling rapid experimentation and message adaptation. However, the effectiveness findings stressed that improved content generation does not automatically translate into improved business outcomes. The literature reviewed points to a consistent “funnel” pattern. AI can be highly effective in improving attention and engagement metrics such as clicks, yet downstream outcomes closer to economic value, primarily conversions, are typically harder to move and more contingent on factors beyond message design, as offer attractiveness, product–market fit and situational relevance. This is reflected in the thesis’ framing of performance as a perception-dependent process

where trust and authenticity function as psychological gates that determine whether certain cues translate into action.

Indeed, trust, authenticity and humanness attributions became central in AI-mediated persuasion contexts. The results evidenced that when recipients suspect automation, the same communication can be evaluated through a different interpretative lens: what previously looked like brand competence may become perceived as impersonality or manipulation, and what previously signalled professionalism may be reinterpreted as machine-like detachment. Accordingly, the thesis argues that consumer judgments about authenticity are not reducible to output quality alone. They also depend on whether the message can be psychologically anchored in human origins, values or oversight and this insight becomes especially important when explicit disclosure or other authorship cues make AI involvement salient.

This synthesis culminates in the empirical study, which tests a specific and practically relevant question: whether “humanising” cues in promotional emails can shift perceptions and behavioural intentions. Using a between-subjects design, the study manipulated the presence of grammatical errors and the precision of the displayed send-time (rounded or non-rounded), measuring perceived trust, authenticity, click and purchase intention in a realistic email stimulus.

The results provided a clear empirical message: in persuasive email contexts, credibility considerations dominate over any presumed “humanisation bonus.” Grammatical errors reduce click intention, and the pattern of results indicates that this effect operates primarily through trust-related psychological paths. In contrast, send-time precision does not exert a statistically detectable influence on trust, authenticity, click intention, nor purchase intention, suggesting that not all paratextual cues are psychologically salient signals of automation or humanness.

Two additional findings refine the overall conclusion. First, the data showed an asymmetry between click intention and purchase intention: the manipulations affect the low-cost, exploratory action of clicking, but do not meaningfully shift stated purchase propensity. This divergence is theoretically coherent with a funnel-based interpretation of effectiveness, whereby early-stage engagement is more responsive to communicative cues, while higher-commitment decisions remain constrained by broader determinants. Second, the study demonstrated that “forced humanisation” through imperfections is not an effective strategy in this context. Imperfections socially associated with original human production assumptions can be overcome and simultaneously function as negative competence signals, thereby undermining trust and reducing engagement rather than enhancing authenticity.

Overall, the thesis delivers an integrated conclusion: generative AI's influence on consumer decision-making is best understood as a dual phenomenon, combining scalable performance optimisation with meaning creation. AI can improve measurable engagement by enhancing relevance and facilitating rapid content variation, but its ability to support deeper brand outcomes is conditional on the interpretative environment in which communication is received. In short, generative AI can strengthen marketing performance, but it does so within a perception-driven system in which trust and authenticity remain pivotal, and where some "human-like" design choices can backfire when automation erodes credibility.

9.2. Research Implications

Building on the theoretical synthesis and empirical findings presented in this thesis, several implications emerge for ongoing academic debates on artificial intelligence in marketing. Rather than treating AI as a purely technological optimisation tool, the overall evidence contributes to clarifying how generative systems reshape persuasion, consumer judgment and relational outcomes in digitally mediated contexts.

One key theoretical contribution lies in clarifying the distinction between predictive and generative AI in marketing practice. Much of the earlier literature on AI in marketing has focused on predictive applications as segmentation, targeting, recommendation engines and optimisation algorithms operating in the background of decision systems. In these contexts, AI improves accuracy and efficiency but does not directly intervene in brand communication. By contrast, generative AI actively produces persuasive content, entering domains traditionally associated with human creativity and judgment. The thesis demonstrates that this shift has profound implications: once AI is involved in creative processes it can no longer be evaluated solely in terms of relevance or performance but also in terms of authorship. The contribution, therefore, is to reposition generative AI as a socially interpreted communicative agent rather than merely a backend optimisation tool. This moves the discussion beyond efficiency gains and situates GAI within broader theories of relational marketing.

Research on algorithm aversion and appreciation has traditionally focused on judgments in forecasting or artistic contexts. This thesis extends those dynamics into persuasive marketing communication, particularly email marketing. The findings suggest that consumer reactions to AI-mediated communication are not reducible to simple aversion or appreciation. Instead, responses

are shaped by interpretive frames. When cues undermine perceived competence or professionalism, engagement declines. When credibility is preserved, AI-generated communication can perform effectively. Importantly, negative reactions in the empirical study were triggered not exclusively by explicit AI disclosure but by signals that weakened trust. This refines the aversion/appreciation debate by demonstrating that in persuasion contexts, consumer responses are mediated by trust and authenticity inferences rather than by automation awareness alone.

Another research implication concerns the so called “human premium”. The literature reviewed indicates that consumers often attribute greater value to content perceived as human-generated, or at least not perceived as artificially generated, particularly in subjective or identity-relevant domains. However, the presented empirical evidence complicates simplistic interpretations of this premium. The attempt to simulate humanness through typically human imperfections did not increase authenticity feelings. Instead, it deteriorated engagement propensity. This suggests that the human premium does not stem from mere surface-level markers of human fallibility or artificial mimicry. Rather, it appears to operate through attributional framing, perceived expertise and genuine intention. Theoretically, this reframes the human premium as a cognitive attribution process where authenticity benefits arise when consumers can anchor communication in human governance or meaningful intent. This distinction contributes to a more nuanced understanding of how authenticity is constructed in AI-mediated marketing.

A further theoretical clarification concerns the asymmetry between click intention and purchase intention observed in the empirical study. The manipulations significantly influenced click intention but did not alter purchase intention. Clicking represents a low-cost exploratory action, highly sensitive to surface-level cues. Purchase intention, in contrast, reflects a higher-commitment evaluation likely shaped by product value and broader brand perceptions beyond the immediate stimulus. This finding refines theoretical models of AI effectiveness by demonstrating that engagement and conversion or loyalty cannot be treated as interchangeable outcomes. AI-generated communication may be particularly effective in exploratory or informational phases, while its impact in identity-relevant or high-stakes decisions is opaquer. Future research should therefore model persuasion as a staged process, examining how attribution and trust mechanisms influence transitions between attention, engagement and purchase.

The synthesis of this thesis opens several parallel hints for future inquiry. First, future studies should examine the long-term relational impact of repeated exposure to AI-generated communication, particularly in terms of brand loyalty. Second, experimental research manipulating disclosure framing could clarify how transparency interacts with authenticity perceptions. Third, cross-category comparisons could test whether the click–conversion asymmetry varies for hedonic

versus utilitarian products, or for identity-relevant versus standardized goods. Fourth, behavioural field data measuring actual clicks and purchases would complement self-reported intentions and allow stronger causal inference. Finally, future research should integrate AI-mediated persuasion into broader customer-brand relationship models, exploring how perceived human oversight, governance structures and ethical positioning shape AI legitimacy in marketing communication.

Overall, this thesis contributes to research by demonstrating that generative AI effectiveness is not merely a function of technological sophistication but of social interpretation. By linking persuasion theory and performance metrics, it offers a framework in which AI's impact on consumer decision-making is understood as contingent on authorship framing and credibility signals within the decision journey.

9.3. Managerial Implications

The findings of this work carry clear managerial implications for organisations adopting AI for their operations, in particular when related to content creation and personalisation. Overall, the evidence supports a balanced conclusion: generative AI can deliver substantial operational and strategic advantages, yet its impact on consumer-facing outcomes is not inherently positive or uniform. As a result, effective adoption requires a deliberate trade-off logic, in which the efficiency and scalability benefits of AI are weighed against potential risks to credibility and authenticity of brand imagine.

A first implication concerns productivity and capability access. Generative AI meaningfully reduces the time and cost required to produce marketing communication assets, enabling rapid drafting, iterative testing and large-scale adaptation of tone and message variants across segments. In practical terms, AI adoption allows to compress production cycles and lower the marginal cost of experimentation, making activities such as copy variation, localisation and format tailoring feasible even for small teams. Moreover, generative systems can increase access to competences that would otherwise require specialised personnel, thereby functioning as an enabling technology that broadens the range of firms able to execute sophisticated marketing programs. Crucially, these advantages also underpin the operational feasibility of mass personalisation: the capacity to generate large volumes of tailored messaging that can be aligned with consumer profiles, contexts and stages of the customer journey.

However, the second and equally important implication is that performance effects cannot be assumed a priori. The thesis highlights that improvements in communication efficiency and output

volume do not mechanically translate into consistently higher effectiveness throughout any marketing context. More broadly, the results reinforce the idea that generative AI may more reliably improve upper-funnel behaviours, such as attention and clicks, rather than deeper commitment-related outcomes until purchase, which are shaped by broader determinants about consumer, product and market fit. From a managerial perspective, this implies that AI adoption should be evaluated by which outcome the organisation seeks to optimise and at which stage of the consumer decision process.

Consequently, the central managerial recommendation is to treat generative AI as a contingent tool rather than a universal performance lever. Decision-makers should adopt an explicit trade-off framework that considers multiple variables: the type of model and tool employed, the degree of automation, the product and industry context, the channel, the sensitivity of the communication task and the intended behavioural endpoint. The same AI-enabled personalisation strategy can plausibly largely succeed in one setting and underperform in another depending on how recipients interpret the message. This means that implementation decisions should be made at the level of use case design, not through broad organisational statements such as “adopt GenAI in marketing”. In practice, firms should investigate where AI can add value (speed, scale, ideation or variation generation) and where it creates risk (relational communications, high-stakes decisions and identity-relevant positioning), and then align governance accordingly before implementation.

A further implication is that the “safest” and often most effective pathway is hybridisation, rather than full automation. The thesis consistently points to the importance of human oversight as a credibility-preserving mechanism. In operational terms, this supports an AI-assisted model in which AI is used to generate drafts, alternatives and variants, while humans remain responsible for strategic intent and vision, factual correctness and brand-consistency. Hybrid governance can mitigate the risk that AI outputs inadvertently introduce wrong cues that reduce trust or appear manipulative, while still capturing efficiency gains. It also provides a defensible organisational posture: when marketing communication is understood as a relational and reputational asset, accountability cannot be fully delegated to automated systems without exposing the brand to preventable reputation erosion.

In line with this, firms should operationalise hybrid adoption through concrete processes rather than informal expectations. This includes establishing clear quality thresholds (linguistic accuracy standards, brand voice checks, compliance reviews, …), building feedback loops from performance metrics back into prompt design and content guidelines, and implementing channel-specific playbooks. For email marketing specifically, where recipients often process messages quickly and

rely on heuristic cues, the study implies that maintaining clarity, coherence and professionalism as competence signals should take precedence.

Finally, managerial evaluation of generative AI should extend beyond efficiency measures to include strategic and relational indicators. While cost reduction and speed are immediate benefits, firms should also monitor whether AI-enabled communication changes credibility-related perceptions over time. This requires treating measurement as multidimensional: clicks and engagement should be analysed alongside brand outcomes, complaint signals and indicators of consumer discomfort. In other words, the appropriate managerial question is not simply whether AI works, but under which conditions it improves performance without compromising the brand's relational foundations.

To conclude, the managerial implications can be summarised as follows: generative AI enables unprecedented efficiency, capability access and scalability, but its effectiveness remains context-dependent and can involve non-trivial reputational trade-offs. The most robust strategy is therefore not maximal automation, but purposeful integration, by deploying AI where it amplifies productivity and relevance, while retaining human oversight to preserve credibility, align communication with brand identity and protect credibility as a long-term asset for building loyalty.

Statement on the Use of AI Tools

In accordance with the University's guidelines on the responsible use of generative artificial intelligence tools, ChatGPT was used during the preparation of this thesis exclusively as a support instrument for writing assistance, research facilitation, methodological clarification and coding support.

All theoretical arguments, analytical decisions, interpretations and conclusions presented in this thesis are the result of the author's independent academic work. ChatGPT was employed as an auxiliary tool to improve clarity of expression, refine sentence structure, enhance academic tone and proofread written sections. Every output generated by the system was carefully reviewed, critically evaluated and substantially revised by the author before inclusion. No AI-generated content was incorporated without thorough human verification and intellectual oversight.

ChatGPT was also used to accelerate the identification of relevant academic literature and to obtain suggestions for potentially pertinent scholarly sources. All suggested references were independently verified, accessed and critically evaluated by the author before being included in the discussion. In addition, the tool facilitated the localisation of specific information within complex documents, such as regulatory texts, helping to navigate and interpret technical language more efficiently. However, the full reading, interpretation and contextual integration of these materials were conducted autonomously.

Furthermore, ChatGPT supported the reformulation and linguistic refinement of the experimental email stimuli, ensuring clarity and consistency while preserving the author-designed manipulations. The tool was also used to assist in drafting and debugging statistical code in STATA. Although the analytical framework, model selection and interpretation of empirical results were determined independently, ChatGPT contributed to the technical implementation of commands and scripts, significantly accelerating the learning and execution process of the software.

In addition, ChatGPT functioned as a conceptual help to structure arguments, clarify methodological questions and enhance coherence across chapters. In these cases, it served as an interactive support mechanism rather than a source of original academic content.

Finally, the use of generative AI tools in this thesis enhanced efficiency and supported skill development, particularly in academic writing and statistical programming.

References

Abboud, S., Hanna, E., Jeunen, O., Raheja, V., & Wheeler, S. (2025, September). Agentic Personalisation of Cross-Channel Marketing Experiences. *Proceedings of the Nineteenth ACM Conference on Recommender Systems*, 907-910.

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science, 347*(6221), 509-514.

Adam, M., Wessel, M., & Benlian, A. (2021). AI-based chatbots in customer service and their effects on user compliance. *Electronic markets, 31*(2), 427-445.

Adamopoulou, E., & Moussiades, L. (2020, May). An Overview of Chatbot Technology. *IFIP international conference on artificial intelligence applications and innovations*, 373-383.

Afroogh, S., Akbari, A., Malone, E., Kargar, M., & Alambeigi, H. (2024). Trust in AI: progress, challenges, and future directions. *Humanities and Social Sciences Communications, 11*(1).

Agrawal, A., Gans, J., & Goldfarb, A. (2022). *Prediction machines, updated and expanded: The simple economics of artificial intelligence.* Harvard Business Press.

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.

Aguirre, E., Roggeveen, A., Grewal, D., & Wetzels, M. (2016). The personalization-privacy paradox: implications for new media. *Journal of consumer marketing, 33*(2), 98-110.

Ali, L., Ali, F., & Alotaibi, S. (2025). Beyond the hype: Evaluating the impact of generative AI on brand authenticity, image, and consumer behavior in the restaurant industry. *International Journal of Hospitality Management, 131*.

Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., & Rieke, A. (2019). Discrimination through optimization: How Facebook's Ad delivery can lead to biased outcomes. *Proceedings of the ACM on human-computer interaction, 3*(CSCW), 1-30.

Andrew, J., & Blake, H. (2024). Personalized Email Campaigns Using Generative AI in B2B Marketing. *Research Gate*.

Araújo, C., Soares, C., Pereira, I., Coelho, D., Rebelo, M., & Madureira, A. (2022). A novel approach for send time prediction on email marketing. *Applied Sciences, 12*(16).

Balasubramaniam, N., Kauppinen, M., Rannisto, A., Hiekkanen, K., & Kujala, S. (2023). Transparency and explainability of AI systems: From ethical guidelines to requirements. *Information and Software Technology, 159*.

Barocas, S., & Selbst, A. (2016). Big Data's Disparate Impact. *Calif. L. Rev., 104*(3), 671-732.

Bellaiche, L., Shahi, R., Turpin, M., Ragnhildstveit, A., Sprockett, S., Barr, N., . . . Seli, P. (2023). Humans versus AI: whether and why we prefer human-created compared to AI-created artwork. *Cognitive research: principles and implications, 8*(1).

Blackman, R. (2020). A Practical Guide to Building Ethical AI. *Harvard Business Review, 15*.

Bol, N., Dienlin, T., Kruikemeier, S., Sax, M., Boerman, S., Strycharz, J., . . . De Vreese, C. (2018). Understanding the effects of personalization as a privacy calculus: Analyzing self-disclosure across health, news, and commerce contexts. *Journal of Computer-Mediated Communication, 23*(6), 370-388.

Brachten, C. (2025). Impact of AI Personalization on Email Clicks and Conversions. *Utrecht University*.

Brown, T. e. (2020). Language models are few-shot learners. *Advances in neural information processing systems, 33*, 1877-1901.

Bruhn, M., Schoenmüller, V., Schäfer, D., & Heinrich, D. (2012). Brand authenticity: Towards a deeper understanding of its conceptualization and measurement. *Advances in consumer research, 40*.

Brüns, J., & 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*.

Bührke, J., Brendel, A., Lichtenberg, S., Greve, M., & Mirbabaie, M. (2021). Is making mistakes human? On the perception of typing errors in chatbot communication. *University of Bamberg*.

Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Conference on fairness, accountability and transparency*, 77-91.

Büyüksomer, A., & Tümbek Tekeoğlu, A. (2024). Enhancing Marketing Communications with Generative AI: A Systematic Literature Review. *Journal of Industrial Policy and Technology Management, 7*(2), 109-122.

Castelo, N., Bos, M., & Lehmann, D. (2019). Task-dependent algorithm aversion. *Journal of marketing research, 56*(5), 809-825.

Challapalli, V., Sai, K., Singh, P., Prasad, R., Maurya, A., & Singh, A. (2025). SLM4Offer: Personalized Marketing Offer Generation Using Contrastive Learning Based Fine-Tuning. *Cornell University*.

Chintalapati, S., & Pandey, S. (2022). Artificial intelligence in marketing: A systematic literature review. *International Journal of Market Research, 64*(1), 38-68.

Choi, H., Park, J., & Jung, Y. (2018). The role of privacy fatigue in online privacy behavior. *Computers in Human Behavior, 81*, 42-51.

Corbett-Davies, S., Gaebler, J., Nilforoshan, H., Shroff, R., & Goel, S. (2023). The measure and mismeasure of fairness. *Journal of Machine Learning Research, 24*(312), 1-117.

Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (September 5, 2024).

Cox, D., Cox, J., & Cox, A. (2017). To Err is human? How typographical and orthographical errors affect perceptions of online reviewers. *Computers in Human Behavior, 75*, 245-253.

Cui, J. (2025). The Impact of Digital Media Technology on Digital Twins: The Moderating Role of Personalized Advertising. *Accounting, Marketing and Organization, 1*(1).

Datta, A., Tschantz, M., & Datta, A. (2014). Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. *Cornell University*.

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2019, October 10). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science, 48*(1), 24-42.

Dietvorst, B., & Bartels, D. (2022). Consumers Object to Algorithms Making Morally Relevant Tradeoffs Because of Algorithms' Consequentialist Decision Strategies. *Journal of Consumer Psychology, 32*(3), 406-424.

DIRECTIVE 2002/58/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL (July 12, 2002).

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM, 55*(10), 78-87.

Dutta, S. (2024). Revolutionizing retail: An empirical study on the impact of generative AI in omnichannel strategies. *International Journal for Multidisciplinary Research, 6*(5).

Dwivedi, Y., & al. (2023, August). Opinion Paper: "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*.

Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models. *Science, 384*(6702), 1306-1308.

Erdem, T., & Swait, J. (2004). Brand credibility, brand consideration, and choice. *Journal of consumer research, 31*(1), 191-198.

European Union. (2025). *EU AI Act: first regulation on artificial intelligence*. Retrieved from https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

European Union. (2025). *Shaping Europe's digital future*. Retrieved from https://digital-strategy.ec.europa.eu/policies/regulatory-framework-ai

Faria, S., & Carvalho, J. (2025). From a Multichannel to an Optichannel Strategy in Retail. *Journal of Theoretical and Applied Electronic Commerce Research, 20*(1).

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., . . . Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines, 28*(4), 689-707.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J., Wallach, H., Iii, H., & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM, 64*(12), 86-92.

Gilpin, D., Palazzolo, E., & Brody, N. (2010). Socially mediated authenticity. *Journal of communication management, 14*(3), 258-278.

Grewal, D., Satornino, C., Davenport, T., & Guha, A. (2024, December 14). How generative AI Is shaping the future of marketing. *Journal of the Academy of Marketing Science, 53*(3), 702-722.

Han, S. (2024, June). Consumer Perceptions of AI-Generated Content and Disclaimer in Terms of Authenticity, Deception, and Content Attribute. *24th Biennial Conference of the International Telecommunications Society (ITS)*.

Hardcastle, K., Vorster, L., & Brown, D. (2025). Understanding customer responses to AI-Driven personalized journeys: impacts on the customer experience. *Journal of Advertising, 54*(2), 176-195.

Haupt, J., Bender, B., Fabian, B., & Lessmann, S. (2018). Robust identification of email tracking: A machine learning approach. *European Journal of Operational Research, 271*(1), 341-356.

Hazell, J. (2023). Spear Phishing With Large Language Models. *Cornell university*.

Heiding, F., Schneier, B., Vishwanath, A., Bernstein, J., & Park, P. (2024). Devising and detecting phishing emails using large language models. *IEEE Access, 12*, 42131-42146.

Hermann, E., & Puntoni, S. (2024). Artificial intelligence and consumer behavior: From predictive to generative AI. *Journal of Business Research, 180*.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems, 33*, 6840-6851.

Hornik, J., & Rachamim, M. (2025). AI-enabled consumer digital twins as a platform for research aimed at enhancing customer experience. *Management Review Quarterly*, 1-34.

Howard, J., & Ruder, S. (2018, July). Universal language model fine-tuning for text classification. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 1*(Long Papers), 328-339.

Huang, M.-H., & Rust, R. (2020). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science, 49*(1), 30-50.

Islam, T., Miron, A., Nandy, M., Choudrie, J., Liu, X., & Li, Y. (2024). Transforming Digital Marketing with Generative AI. *Computers, 13*(7).

Jago, A., Carroll, G., & Lin, M. (2022). Generating Authenticity in Automated Work. *Journal of Experimental Psychology: Applied, 28*(1), 52-70.

Jakesch, M., Hancock, J., & Naaman, M. (2023). Human heuristics for AI-generated language are flawed. *Proceedings of the National Academy of Sciences, 120*(11).

Jobson, D., Shukla, M., Dinh, P., Young, J., Pittoni, N., Chen, N., & Ginstrom, R. (2025). Using item recommendations and LLMs in marketing email titles. *Cornell University*.

Kannan, A., Kurach, K., Ravi, S., Kaufmann, T., Tomkins, A., Miklos, B., . . . Ramavajjala, V. (2016, August). Smart Reply: Automated Response Suggestion for Email. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 955-964.

Kietzmann, J., Paschen, J., & Treen, E. (2018). Artificial intelligence in advertising: How marketers can leverage artificial intelligence along the consumer journey. *Journal of Advertising Research, 58*(3), 263-267.

Kirk, C., & Givi, J. (2025). The AI-authorship effect: Understanding authenticity, moral disgust, and consumer responses to AI-generated marketing communications. *Journal of Business Research, 186*.

Kirkby, A., Baumgarth, C., & Henseler, J. (2023). To disclose or not disclose, is no longer the question – effect of AI-disclosed brand voice on brand authenticity and attitude. *Journal of Product & Brand Management, 32*(7), 1108-1122.

Köbis, N., & Mossink, L. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in human behavior, 114*.

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer, 42*(8), 30-37.

Krizhevsky, A., Sutskever, I., & Hinton, G. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84-90.

Kshetri, N., Dwivedi, Y., Davenport, T., & Panteli, N. (2024). Generative artificial intelligence in marketing: Applications, opportunities, challenges, and research agenda. *International journal of information management, 75*.

Lambrecht, A., & Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. *Management science, 65*(7), 2966-2981.

Lee, H., Todri, V., Adamopoulos, P., & Ghose, A. (2025). The Impact of Visual Generative AI on Advertising Effectiveness. *Available at SSRN 5638311*.

Liu, Y. (2025, July). Intelligent Analysis Methods for Multi-Channel Marketing Data Based on Anomaly Detection Algorithms. *Proceedings of the 2nd International Conference on Image Processing, Machine Learning, and Pattern Recognition*, 198-206.

Logg, J., Minson, J., & Moore, D. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes, 151*, 90-103.

Madaio, M., Stark, L., Wortman Vaughan, J., & Wallach, H. (2020, April). Co-designing checklists to understand organizational challenges and opportunities around fairness in AI. *Proceedings of the 2020 CHI conference on human factors in computing systems*, 1-14.

Mathur, A., Acar, G., Friedman, M., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on human-computer interaction, 3*(CSCW), 1-32.

Matz, S., Kosinski, M., Nave, G., & Stillwell, D. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the national academy of sciences, 114*(48), 12714-12719.

Matz, S., Teeny, J., Vaid, S., Peters, H., Harari, G., & Cerf, M. (2024). The potential of generative AI for personalized persuasion at scale. *Scientific Reports, 14*(1).

McDonald, A., & Cranor, L. (2008). The cost of reading privacy policies. *Journal of Law and Policy for the Information Society, 4*(3), 543-568.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM computing surveys, 54*(6), 1-35.

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., . . . Gebru, T. (2019, January). Model cards for model reporting. *Proceedings of the conference on fairness, accountability, and transparency*, 220-229.

Nesterenko, V., & Olefirenko, O. (2023). The impact of AI development on the development of marketing communications. *Marketing i menedžment innovacij, 14*(1), 169-181.

Nilsson, N. (2009). The quest for artificial intelligence.

Nissenbaum, H. (2004). Privacy as contextual integrity. *Wash. L. Rev., 79*, 119-158.

Nissenbaum, H. (2009). Privacy in context: Technology, policy, and the integrity of social life. *Privacy in context*.

NIST AI. (2023). Artificial intelligence risk management framework (AI RMF 1.0). Retrieved from https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

NIST AI. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile. *NIST Trustworthy and Responsible AI Gaithersburg, MD, USA*.

OECD.AI Policy Observatory. (2025). *OECD AI Principles overview*. Retrieved September 24, 2025, from OECD.AI Policy Observatory Website: https://oecd.ai/en/ai-principles

Oord, A., Li, Y., & Vinyals, O. (2018). Representation learning with contrastive predictive coding. *Cornell University*.

Patil, D. (2024). Generative artificial intelligence in marketing and advertising: Advancing personalization and optimizing consumer engagement strategies. *Available at SSRN 5057404*.

Polimetla, J. (2025). Literature Review on Digital Twins in Marketing. *Available at SSRN 5588290*.

Raji, I., Smart, A., White, R., Mitchell, M., Gebru, T., Hutchinson, B., . . . Barnes, P. (2020, January). Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 33-44.

REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL (April 27, 2016).

REGULATION (EU) 2022/2065 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL (October 19, 2022).

REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL (June 13, 2024).

Ricci, F., Rokach, L., & Shapira, B. (2021). Recommender systems: Techniques, applications, and challenges. *Recommender systems handbook*, 1-35.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10684-10695.

Sahni, N., Wheeler, S., & Chintagunta, P. (2016, October 23). Personalization in email marketing: The role of noninformative advertising content. *Marketing Science, 37*(2), 236-258.

Shedlock, A. (2025, September 18). What is a Digital Twin in AI Marketing Research? *Greenbook*.

Simonetta, A., & Paoletti, M. (2024). ISO/IEC Standards and Design of an Artificial Intelligence System. *IWESQ@ APSEC*, 39-43.

Suresh, H., & Guttag, J. (2021, October). A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, 1-9.

Susser, D., Roessler, B., & Nissenbaum, H. (2019). Online manipulation: Hidden influences in a digital world. *Geo. L. Tech. Rev., 4*(1).

Tian, J., Zhao, V., Sun, Y., & Qiu, L. (2025, August 12). AI-Powered vs. Manual Advertising: A Field Experiment in E-Commerce. *Manual Advertising: A Field Experiment in E-Commerce*.

Unesco. (2022). Recommendation on the ethics of artificial intelligence. *United Nations Educational, Scientific and Cultural Organization*.

Van Hees, J., Grootswagers, T., Quek, G., & Varlet, M. (2025). Human perception of art in the age of artificial intelligence. *Frontiers in Psychology, 15*.

Vaswani, A. e. (2017). Attention Is All You Need. *Advances in neural information processing systems, 30*.

Wedel, M., & Kannan, P. (2016). Marketing Analytics for Data-Rich Environments. *Journal of marketing, 80*(6), 97-121.

Yang, J., & Battocchio, A. (2021). Effects of transparent brand communication on perceived brand authenticity and consumer responses. *Journal of Product & Brand Management, 30*(8), 1176-1193.

Yin, Z., Chen, X., & Zhang, X. (2025). AI-Integrated Decision Support System for Real-Time Market Growth Forecasting and Multi-Source Content Diffusion Analytics. *Cornell University*.

Zhang, R., & Tetreault, J. (2019, July). This email could save your life: Introducing the task of email subject line generation. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 446-456.

Zhang, Y., & Gosline, R. (2023). Human favoritism, not AI aversion: People's perceptions (and bias) toward generative AI, human experts, and human–GAI collaboration in persuasive content generation. *Judgment and Decision Making, 18*.

Ziegler, D., Stiennon, N., Wu, J., Brown, T., Radford, A., Amodei, D., . . . Irving, G. (2019). Fine-Tuning Language Models from Human Preferences. *Cornell University*.

Acknowledgements

As these final pages bring this work to a close, I find myself reflecting not only on the research questions explored, the data analysed, or the arguments developed, but on the people and experiences that quietly shaped this journey behind the scenes. What appears here as a structured academic contribution is, in reality, the outcome of encounters, challenges, guidance, courage and support that extended far beyond these pages.

First and foremost, I would like to express my sincere gratitude to my professor and supervisor for having been a constant intellectual reference throughout the development of this thesis. His guidance played a fundamental role in shaping the theoretical foundation upon which this work is built. Through his teaching during the course, I progressively developed a strong interest in the statistical and computational dimensions of marketing data analysis, an interest that ultimately inspired the empirical and operational component of this research. I am equally grateful for the opportunity he, together with the other professors involved and the entire department, gave me to participate in the 2024 Summer School on responsible innovation for businesses in Ingolstadt. That experience enriched me academically beyond the ordinary coursework and significantly influenced the ethical reflections on artificial intelligence that are discussed in this thesis.

I am deeply thankful to my family for never limiting my choices, but instead constantly encouraging me to aim higher, even during moments of uncertainty and self-doubt. Their unwavering support has been the foundation that allowed me to pursue opportunities across countries and disciplines with confidence and determination.

I would also like to thank the University of Bologna for allowing me to be part of such a prestigious and intellectually vibrant community. Being part of this institution also enabled me to broaden my horizons and diversify my academic path, offering me the possibility to engage in meaningful international experiences.

In this regard, I owe a special acknowledgment to two cities that, for a significant period of my life, became home: Sydney and Valencia. They confronted me with my fears, insecurities and personal limits. Only to show me that they could be overcome. They introduced me to new dreams, new dawns, new lights and new perspectives. Both cities enriched me academically, professionally and, above all, humanly, through the people I had the privilege to meet.

Finally, I would like to thank all those who shared parts of this journey with me. Those I met only for a brief walk towards the Opera House or a bachata class between an exam and a chapter written, and those who became a lasting presence in my life, whether permanently or temporarily. From

each of them, through both light and shadow, I have always learned something valuable about who I was, who I am and who I aspire to become. Thank you to everyone who contributed, knowingly or unknowingly, to shaping the person I am today and who helped me step out of those unexpected whirlwinds that life sometimes places in our path, moments in which nothing can be seen but the darkness of the deep.

In many ways, as the final step of this journey, this thesis represents more than an academic achievement. It is the result of years of uncertainty, disappointments and fears, but finally and most important growth, courage and discovery. It stands as a reminder that every challenge faced, every horizon crossed has been part of a larger journey, one that began long before these pages and will continue far beyond them.