

**Stated Click-Through Intention Across Two FocusForm Advertisement Conditions:
An Online Experiment Comparing AI-Generated and Human-Written Copy**

This is to certify that the work I am submitting is my own. All external references and sources are clearly acknowledged and identified within the contents. I am aware of the University of Warwick regulation concerning plagiarism and collusion.

No substantial part(s) of the work submitted here has also been submitted by me in other assessments for accredited courses of study, and I acknowledge that if this has been done an appropriate reduction in the mark I might otherwise have received will be made.

Abstract

This dissertation compared average stated click-through intention between adults randomly assigned to one of two FocusForm advertisement conditions: AI-generated or human-written copy. Each participant viewed one advertisement, and recruitment occurred through digital and social-media channels. The study also compared perceived authenticity and examined whether perceived relevance and trust were associated with stated click-through intention.

A quantitative, between-subjects online experiment was administered through Qualtrics. Participants were randomly assigned to view one of two static advertisements for FocusForm, a fictional productivity app. One advertisement contained AI-generated copy and the other contained human-written copy. The advertisements used the same broad visual template, and participants were not told who created the copy. After the exclusion criteria fixed during survey design were applied, the final analytical sample consisted of 60 participants, with 30 in each condition. Click-through intention, perceived authenticity, perceived relevance and trust were measured using three-item, seven-point scales. Welch independent-samples t-tests compared the conditions, and multiple linear regression with HC3 robust standard errors modelled stated click-through intention using perceived relevance, trust and advertisement condition as predictors.

The study did not identify a statistically significant difference in stated click-through intention between the AI-generated and human-written copy conditions. Perceived authenticity also did not differ significantly between conditions. By contrast, perceived relevance and trust were both positively associated with stated click-through intention after accounting for each other and advertisement condition. The condition coefficient was not statistically significant; its point estimate was near zero, but the confidence interval allowed differences in either direction.

The findings do not show that AI-generated copy is generally better or worse than human-written copy. They provide context-specific evidence that, for the two selected FocusForm advertisements, relevance and trust had significant positive conditional associations with stated click-through intention in the fitted regression model, whereas advertisement condition was not significantly associated with the outcome. The experiment did not directly test AI-assisted advertising workflows. However, when considered alongside the literature, the findings support the broader managerial recommendation that AI should assist copy development while humans retain responsibility for strategy, review, testing and final approval. The conclusions remain limited to stated intention, one fictional product, two selected advertisement executions and a convenience sample of 60 participants.

Table of Contents

Abstract.....	ii
List of Tables	vi
List of Figures	vii
1. Introduction	1
1.1 Background and Research Context	1
1.2 Research Problem	1
1.3 Research Aim, Question and Objectives.....	2
1.4 Contribution of the Study	2
1.5 Dissertation Structure	3
2. Literature Review	4
2.1 Digital Advertising Effectiveness and Click-Through Intention.....	4
2.2 Drivers of Click-Through Intention.....	5
2.3 Human-Written versus AI-Generated Advertising Copy	6
2.4 Theoretical Basis for Human–AI Differences.....	8
2.5 Hypothesis Development	9
3. Methodology	11
3.1 Research Design	11
3.2 Experimental Treatments and Stimuli	11
3.3 Copy Creation Procedure	12
3.4 Survey Design and Measures	13
3.5 Participants, Sampling and Data Collection	15
3.6 Data Preparation and Ethics	15
3.7 Data Analysis.....	16
4. Results	18
4.1 Sample Profile and Data Cleaning	18
4.2 Reliability and Descriptive Statistics.....	19
4.3 Randomisation Check.....	19
4.4 Hypothesis Testing	20
4.5 Regression Analysis	20
4.6 Summary of Findings.....	22
5. Discussion	23
5.1 Overview of Main Findings.....	23
5.2 Interpretation in Relation to Existing Literature.....	23
5.3 Theoretical Implications	25

5.4 Managerial Implications	26
5.5 Implications for AI-Assisted Advertising	27
6. Limitations and Conclusion	28
6.1 Limitations	28
6.2 Future Research	29
6.3 Conclusion	30
Reference List	32
Appendices	36
Appendix A: Experimental Advertisements	37
A1. Final Advertisement Stimuli	37
Appendix B: Creative Brief and Copy-Creation Materials	38
B1. Shared Creative Brief	38
B2. Human-Written Copy	38
B3. Exact AI-Generation Prompt	39
B4. Full ChatGPT Output	40
B5. Final AI-Generated Copy	40
Appendix C: Full Survey Instrument and Qualtrics Flow	41
C1. Participant Information Sheet	41
C2. Eligibility and Consent	42
C3. Click-Through Intention Measure	43
C4. Perceived Authenticity Measure	44
C5. Perceived Relevance Measure	45
C6. Trust Measure	46
C7. Attention Check and Demographic Questions	47
C8. Prior Productivity-App Use	48
C9. Debriefing Statement	49
C10. Qualtrics Survey Flow and Randomisation	50
Appendix D: Recruitment Materials and Sensitivity Power Calculation	51
D1. Recruitment Materials	51
D2. Sensitivity Power Calculation	51
Appendix E: Ethics Documentation	53
E1. Ethics Approval Confirmation	53
Appendix F: Data Preparation, Variable Coding and Scoring	54
F1. Data-Cleaning Summary	54
F2. Variable Dictionary	54
F3. Composite Scoring	58

F4. Data-Cleaning and Analysis Code	58
Appendix G: Additional Descriptive and Reliability Results	78
G1. Item-Level Descriptive Statistics	78
G2. Detailed Reliability Diagnostics	78
G3. Pearson Correlations Between Composite Measures	79
Appendix H: Detailed Statistical Outputs and Regression Diagnostics.....	80
H1. Welch Independent-Samples t-Tests.....	80
H2. Visual Distribution Checks for H1 and H2.....	81
H3. Conventional OLS Regression Output.....	81
H4. HC3 Robust Regression Results.....	82
H5. Multicollinearity Diagnostics	82
H6. Regression Residual Diagnostics.....	83
H7. Influence Diagnostics	84
H8. Leave-One-Out Sensitivity Analysis	85

List of Tables

Table 3.1. Item wording for the four study constructs.	14
Table 4.1. Data-cleaning and analytical-sample flow.....	18
Table 4.2. Sample characteristics overall and by advertisement condition.	19
Table 4.3. Reliability and descriptive statistics for the study constructs.	19
Table 4.4. Welch independent-samples t-test results for H1 and H2.	20
Table 4.5. Multiple regression predicting click-through intention.....	21
Table 4.6. Summary of hypothesis outcomes.....	22

List of Figures

Figure 3.1. Final FocusForm advertisements used in the AI-generated and human-written copy conditions.	12
---	----

1. Introduction

1.1 Background and Research Context

Generative artificial intelligence (GenAI) refers to artificial intelligence systems that create new content, including text, images and video, from patterns learned from existing data (Grewal et al., 2024). In marketing, these systems can support tasks such as producing advertisements, product descriptions, social-media content and sales material (Grewal et al., 2024; Kshetri et al., 2024). GenAI is therefore increasingly relevant to how firms create and adapt digital marketing communications.

In digital advertising, GenAI can be used to produce advertising copy, generate alternative message versions and adapt content for different audiences or platforms. This may increase the speed, scale and variety of creative production. However, greater production efficiency does not remove the need for human involvement. Firms still need to review whether generated content is accurate, suitable and consistent with the brand before it is shown to consumers (Grewal et al., 2024).

The growing use of GenAI in marketing makes consumer responses to AI-generated content an important research topic. Existing studies have examined AI-generated advertising in specific settings, including social-media content and emotional marketing communication (Brüns and Meißner, 2024; Kirk and Givi, 2025). However, the ability to create advertising content does not show whether consumers respond similarly to AI-generated and human-written copy.

This dissertation addresses this issue through a controlled online experiment. Participants viewed one of two visually matched static advertisements for FocusForm, a fictional productivity application. One advertisement contained AI-generated copy and the other contained human-written copy. Participants were not told who created the copy. The study therefore examined responses to the final advertisement executions rather than reactions to an AI label.

1.2 Research Problem

Existing research provides different expectations about how consumers may respond to AI-generated advertising. AI-generated copy may be accepted more readily when the message is practical, rational or focused on functional benefits. Bakpayev et al. (2020) found that AI- and human-created rational advertisements received similar evaluations, while AI-created emotional advertisements were evaluated less favourably. They also found that AI-created advertising was more acceptable for utilitarian than hedonic products. Castelo et al. (2019) similarly found that consumers relied less on algorithms for subjective than objective tasks.

However, human-written communication may be valued more when consumers expect emotional sincerity or authenticity. Kirk and Givi (2025) found that emotional messages believed to be AI-authored were viewed as less authentic. Brüns and Meißner (2024) also found that GenAI use in social-media content creation could reduce perceived brand authenticity. These studies mainly concern situations in which AI involvement was known or believed. They do not establish whether undisclosed AI-generated advertising copy will be evaluated differently from human-written copy.

The research problem is therefore whether average stated click-through intention and perceived authenticity differ between participants exposed to the AI-generated and human-written FocusForm advertisement conditions in a controlled digital-advertising setting. The study focuses on stated click-through intention rather than actual clicking behaviour.

1.3 Research Aim, Question and Objectives

The aim of this dissertation is to examine whether average stated click-through intention differs between the AI-generated and human-written FocusForm advertisement conditions among adult digital consumers. In this study, “adult digital consumers” refers broadly to adults recruited through digital and social-media channels rather than to a separately screened consumer category.

The dissertation addresses the following research question:

Does average stated click-through intention differ between participants exposed to the AI-generated and human-written FocusForm advertisement conditions?

The study has four objectives:

1. To compare average stated click-through intention between the two advertisement conditions.
2. To examine whether perceived authenticity differs between the conditions.
3. To assess the association between perceived relevance and stated click-through intention.
4. To assess the association between trust and stated click-through intention.

1.4 Contribution of the Study

This dissertation adds context-specific evidence to research on consumer responses to AI-generated marketing communication. It compares selected AI-generated and human-written advertising copy within two static advertisements for a fictional productivity application. Because authorship was not disclosed, the study focuses on responses to the final advertisement executions rather than reactions to an AI label.

The study also contributes by examining stated click-through intention together with perceived authenticity, relevance and trust. It compares click-through intention and authenticity across the two conditions and assesses whether relevance and trust are associated with stated click-through intention.

The contribution therefore lies in combining an undisclosed comparison of AI-generated and human-written advertising copy with measures of stated click-through intention, perceived authenticity, relevance and trust within one controlled experiment.

The study also has practical relevance for marketers considering GenAI for advertising-copy development. It provides evidence on how two selected copy executions were received in a functional productivity-app setting. However, it does not show that AI-generated copy is generally more or less effective than human-written copy. It also does not measure actual advertising-platform performance.

1.5 Dissertation Structure

This dissertation consists of six chapters. Chapter 2 reviews the literature on digital advertising effectiveness, click-through intention, consumer perceptions and AI-generated advertising. It also develops the hypotheses. Chapter 3 explains the methodology, including the experimental design, advertisement conditions, measures and data-collection process. Chapter 4 presents the results. Chapter 5 discusses the findings and their theoretical and managerial implications. Chapter 6 outlines the limitations, identifies directions for future research and presents the conclusion.

2. Literature Review

2.1 Digital Advertising Effectiveness and Click-Through Intention

Digital advertising effectiveness can be assessed through both behavioural and brand-related outcomes. Behavioural outcomes include clicks, website visits and purchases, while brand-related outcomes include awareness, recall, recognition, image and purchase intention (Lohtia, Donthu and Hershberger, 2003; Drèze and Hussherr, 2003). This distinction matters because digital advertisements can influence consumers even when they do not click immediately. Hoban and Bucklin (2015), for example, show that display advertising can affect website visitation, while Lewis and Reiley (2014) demonstrate that online advertising can influence both online and offline purchases. Clicks are therefore useful indicators of response, but they do not capture the full effect of digital advertising.

Click-through rate (CTR) is one of the most common behavioural metrics in online advertising. It is usually calculated as the number of clicks divided by the number of impressions, where a click refers to a user action that redirects them from the advertisement to another web location (Lohtia et al., 2003). Within the hierarchy-of-effects model, clicking can be understood as a behavioural response following advertising exposure (Yoo, Kim and Stout, 2004). CTR is therefore relevant when the aim of an advertisement is to encourage immediate action.

However, low CTR does not necessarily mean that an advertisement is ineffective. Drèze and Hussherr (2003) argue that consumers may avoid looking directly at banner advertisements while still being influenced through awareness or memory. Lewis and Reiley (2014) similarly found that advertising increased purchases even though much of the effect came from consumers who had not clicked on the advertisement. CTR should therefore be treated as one measure of effectiveness rather than the only relevant outcome.

Advertising execution can also affect click-related responses. Lohtia et al. (2003) found that content and design elements influenced banner-ad CTR, although the effects differed between B2B and B2C contexts. Yoo et al. (2004) found that animated advertisements produced higher click-through intention than static advertisements, while Yoo (2011) showed that message framing, keyword insertion and product involvement jointly influenced click behaviour in search advertising. These findings support the view that differences in message content can affect whether consumers intend to click.

The present dissertation uses a survey-based experiment rather than a live advertising campaign. It therefore measures stated click-through intention instead of actual CTR. Click-through intention captures a participant's reported willingness or likelihood to click after

viewing an advertisement (Yoo et al., 2004). The Theory of Planned Behaviour also treats behavioural intention as an important predictor of behaviour. Armitage and Conner (2001) found that the theory explained substantial variance in both intention and behaviour.

Nevertheless, intention is not the same as action. Chandon, Morwitz and Reinartz (2005) show that self-reported intentions do not perfectly predict later behaviour and that measuring intentions can itself affect the intention–behaviour relationship. Click-through intention was therefore an appropriate and theoretically supported outcome for this study. Because participants viewed a static, non-clickable advertisement within the survey, actual click behaviour could not be measured. Instead, stated click-through intention was used as a consistent measure of intended engagement across participants.

2.2 Drivers of Click-Through Intention

Advertising value provides a useful framework for explaining why consumers respond positively or negatively to advertisements. Ducoffe (1995) defines advertising value as the consumer's subjective evaluation of the worth or utility of advertising. Informativeness and entertainment can increase perceived value, while irritation and deceptiveness can reduce it. Brackett and Carr (2001) extend this framework by identifying credibility as another important influence on advertising value and attitude. Later research supports this general view. Maseeh et al. (2021) find that personalisation, entertainment, credibility and informativeness are positively associated with attitudes towards mobile advertising, while irritation is negatively associated with them. Chen et al. (2022) also identify credibility and emotional appeal as important predictors of advertising attitude in social-media advertising.

Advertising attitude is also associated with behavioural responses. MacKenzie and Lutz (1989) define attitude towards the ad as a favourable or unfavourable response during exposure. Wang and Sun (2010) find that online-advertising beliefs predict attitudes, which then predict ad clicking and shopping frequency. Chen et al. (2022) similarly find that advertising attitude predicts advertising clicks. Consumers who evaluate an advertisement positively may therefore be more likely to engage with it.

Relevance is particularly important for this dissertation. Kim and Huh (2016) define perceived relevance as the extent to which consumers view an advertising message as self-related or useful for achieving their goals. They find that relevance strongly predicts attention, attitude towards the advertisement and clicking. De Keyser et al. (2021) also show that perceived personalisation increases personal relevance, which then improves brand attitude and click intention. Bleier and Eisenbeiss (2015) identify perceived usefulness as an important mechanism linking personalised advertising to click-related outcomes. These

findings support the expectation that consumers will report stronger click-through intention when the message appears useful and relevant to their needs.

Trust in this dissertation refers to the extent to which participants perceive FocusForm and the claims made in its advertisement as believable and dependable. This construct is related to advertising credibility, which concerns whether advertising claims appear truthful and believable (MacKenzie and Lutz, 1989). Prior research also links credibility to advertising value and attitudes towards advertising (Brackett and Carr, 2001; Chen et al., 2022). Evidence from personalised advertising further suggests that existing trust in a retailer can shape how consumers respond to advertising. Bleier and Eisenbeiss (2015), for example, find that deeper personalisation increases click-through rates for a trusted retailer but reduces them for a less trusted retailer. Aguirre et al. (2015) similarly show that trust-building strategies can reduce negative reactions to data collection. Although retailer trust and advertising credibility are distinct from trust in a specific fictional app, these findings jointly suggest that perceived believability and dependability may increase willingness to engage with an advertisement.

Positive responses are not guaranteed. Irritation can reduce advertising value when messages are viewed as annoying, offensive or manipulative (Ducoffe, 1995). Maseeh et al. (2021) confirm that irritation is negatively associated with attitudes towards mobile advertising. Intrusiveness, privacy concern and reactance may also weaken engagement. De Keyser et al. (2021) find that perceived intrusiveness reduces brand attitude and click intention, while Kim and Huh (2016) show that privacy concern negatively affects attitudes towards online behavioural advertising. Bleier and Eisenbeiss (2015) also find that low trust can increase reactance towards personalised advertising.

Overall, the cited literature links consumer evaluations of advertisements to click-related and attitudinal responses. Across the different settings reviewed, relevance, trust, credibility, informativeness and entertainment are generally associated with more favourable responses, while irritation, intrusiveness and privacy concern are associated with less favourable attitudes or click-related intentions. AI-generated and human-written copy may therefore receive different responses if the two executions are evaluated differently on these dimensions.

2.3 Human-Written versus AI-Generated Advertising Copy

AI can support advertising production by increasing speed, scale and variation. Bakpayev et al. (2020) describe programmatic creative as an algorithm- and data-driven process that adds automation, dynamic content production and optimisation to creative design and testing. GenAI can automate large parts of the process, but it can also support individual

stages while humans retain oversight and refine the output (Brüns and Meißner, 2024). AI-generated copy is therefore commercially attractive because it can support rapid content creation, alternative versions and adaptation for different audiences.

Existing research does not show that either AI-generated or human-written copy is always superior. Instead, outcomes appear to depend on the task and message. Bakpayev et al. (2020) find that consumers evaluate AI- and human-created rational appeals similarly, but rate human-created emotional appeals more favourably. They also find that AI-created advertising is more acceptable for utilitarian products than for hedonic products. This is consistent with Castelo et al. (2019), who show that consumers rely less on algorithms for subjective tasks than for objective tasks. Algorithms are often associated with analytical and data-processing ability, while humans are associated with emotional understanding and contextual judgement (Desveaud and Pavone, 2025).

This distinction is relevant because FocusForm is a productivity app with a functional value proposition. Its benefits concern planning, organisation, efficiency and goal management. Bakpayev et al. (2020) therefore provide reason to expect that AI-generated copy could be evaluated similarly to human-written copy in this setting. However, productivity advertising can also use motivational language, which consumers may evaluate partly through perceived sincerity and emotional understanding. If consumers value those qualities, human-written copy could still have an advantage.

Human-written communication may be valued more when consumers expect sincerity, warmth or genuine emotion. Kirk and Givi (2025) find that emotional marketing communications believed to be AI-authored are perceived as less authentic. Lower authenticity then increases moral disgust and reduces positive word of mouth and customer loyalty. Brüns and Meißner (2024) also find that GenAI use in social-media content creation can reduce favourable consumer responses through lower perceived brand authenticity. These findings provide a basis for examining whether perceived authenticity differs between the two copy conditions.

Consumer response may also depend on whether AI involvement is disclosed. Baek et al. (2024) find that disclosing AI-generated content in prosocial advertising reduces attitudes through lower perceived credibility. Sands et al. (2025) similarly show that awareness of AI-generated advertising can reduce brand credibility and attitude, although responsible brand signals may weaken these effects. Brüns and Meißner (2024) also find stronger negative reactions when consumers are aware that GenAI was used.

The present experiment does not disclose the source of either advertisement. It therefore examines responses to the final copy executions rather than reactions to an AI label. This

distinction is important because a negative reaction to AI authorship is not necessarily a negative reaction to the content itself.

Algorithm aversion offers one explanation for why consumers may resist AI-generated advertising in subjective contexts. Castelo et al. (2019) find that consumers trust algorithms less for subjective than objective tasks. In their field study, human-generated dating advice produced a higher CTR than algorithm-generated dating advice, while the difference was weaker for financial advice. Dietvorst, Simmons and Massey (2015) also show that consumers can avoid algorithms after seeing them make mistakes, even when the algorithm performs better overall. However, Logg, Minson and Moore (2019) provide a counterpoint by showing that people may rely more on algorithms in numerical estimation and forecasting tasks. Consumer reactions to AI are therefore not consistently negative.

Advertising production can also involve AI–human collaboration rather than full automation or fully human production. AI may contribute information processing, speed and variation generation, while humans contribute contextual understanding, emotional judgement and creative refinement. Ryoo et al. (2025) find that consumers expect AI–human collaboration to improve advertising by combining information management, innovative ideas, creative research and efficiency. Brüns and Meißner (2024) also find that negative reactions are weaker when AI assists humans rather than replaces them. This literature is included to inform the later managerial discussion; the present study did not test collaborative production workflows.

2.4 Theoretical Basis for Human–AI Differences

Differences between AI-generated and human-written copy are likely to arise through consumer perceptions of the message rather than authorship alone. Advertising value theory suggests that consumers respond more positively when an advertisement appears useful, informative, credible or entertaining, and less positively when it appears irritating or deceptive (Ducoffe, 1995). The copy condition may therefore affect click-through intention if one execution is viewed as more useful, believable or valuable than the other.

Perceived relevance is one theoretically relevant perception examined in this study. Consumers are more likely to attend to and engage with advertisements that match their needs or goals (Kim and Huh, 2016; De Keyzer et al., 2021). In the present study, relevance concerns whether the FocusForm message appears useful and connected to participants' needs for organisation, planning and goal management. Greater perceived relevance was therefore expected to be positively associated with stronger stated click-through intention.

Trust is a second theoretically relevant perception. Consumers are more willing to engage when advertising claims appear believable and the advertiser appears trustworthy

(MacKenzie and Lutz, 1989; Bleier and Eisenbeiss, 2015). In this study, trust concerns FocusForm and the believability of the claims made in its advertisement. Greater trust in FocusForm was therefore expected to be positively associated with stronger stated click-through intention.

In this dissertation, perceived authenticity refers to whether the advertisement and its message appear authentic, genuine and sincere. The three items therefore capture an overall judgement of the execution's genuineness rather than brand authenticity or perceived human authorship. Authenticity is particularly relevant when a message is emotional or personally expressive. Kirk and Givi (2025) show that known or believed AI authorship can reduce authenticity for emotional marketing messages. Friestad and Wright's (1994) Persuasion Knowledge Model provides theoretical context for how perceived AI authorship could influence authenticity judgements because consumers use their knowledge of persuasion agents and tactics when interpreting advertising attempts. However, authorship was not disclosed, and perceived authorship and persuasion knowledge were not measured. The model therefore provides theoretical context rather than a tested explanation. The study instead tests whether the two final copy outputs differ in perceived authenticity.

The literature therefore produces two competing expectations for the condition comparisons. Research on functional and objective tasks suggests that the AI-generated advertisement may be evaluated similarly to the human-written advertisement. Research on authenticity and emotional understanding suggests that the human-written advertisement could be evaluated more favourably when those qualities matter. Because FocusForm contains both practical and motivational elements, the literature does not establish a clear directional advantage. This supports non-directional comparisons for stated click-through intention and authenticity, while relevance and trust are examined for their associations with stated click-through intention.

2.5 Hypothesis Development

Existing research provides competing expectations regarding AI-generated and human-written advertising copy. AI-generated rational appeals can receive similar evaluations to human-created appeals, while human-created emotional appeals may be preferred (Bakpayev et al., 2020). Consumers also tend to rely more on algorithms for objective than subjective tasks (Castelo et al., 2019). Because FocusForm is primarily functional but may also involve motivational language, the literature does not establish a clear advantage for either condition. Therefore:

H1: Average stated click-through intention differs between the AI-generated and human-written copy conditions.

Human-written copy may be perceived as more authentic when consumers expect emotional sincerity or a human perspective. However, much of the evidence for an authenticity disadvantage concerns disclosed or believed AI authorship and emotional messages. In the present undisclosed, functional setting, the direction of any difference was uncertain, and differences in the final wording could favour either execution. H2 was therefore specified as non-directional. Accordingly:

H2: Perceived authenticity differs between the AI-generated and human-written copy conditions.

Perceived relevance was expected to be positively associated with click-through intention because relevant advertisements appear more useful, self-related and aligned with consumer goals. Kim and Huh (2016) and De Keyser et al. (2021) both report positive relationships between relevance and click-related responses. The present study tests this relationship as a conditional association rather than a causal effect. Therefore:

H3: Perceived relevance is positively associated with stated click-through intention.

Trust is expected to be positively associated with stated click-through intention. Consumers who view an advertisement and its claims as believable, rather than manipulative or suspicious, may also report stronger willingness to engage. Bleier and Eisenbeiss (2015) and Aguirre et al. (2015) link trust to fewer negative reactions and more favourable advertising responses. Therefore:

H4: Trust is positively associated with stated click-through intention.

3. Methodology

3.1 Research Design

This dissertation used a quantitative, between-subjects, survey-based online experiment to compare consumer responses to two static advertisements for FocusForm, a fictional productivity and task-management app. The A/B-style design contained two randomly assigned conditions and one advertisement exposure per participant. The eligibility and consent process, stimulus instructions, post-exposure measures and debriefing were the same across conditions. It was not a live advertising-platform test. The main purpose was to test whether stated click-through intention differed between conditions. Perceived authenticity was also compared, while perceived relevance and trust were examined as predictors of stated click-through intention.

The survey was administered through Qualtrics. Participants aged 18 or over who provided consent were randomly assigned to one condition and viewed only one advertisement. They were not told whether the copy had been generated by AI or written by a human. The advertisements used the same broad visual template, including the logo, colours, layout, phone mock-up, icons and “Learn More” call-to-action. The experimental treatment was exposure to one of two complete advertisement-copy executions, not a pure manipulation of authorship with identical wording.

The advertisements were static and the call-to-action was not clickable. The study therefore measured stated click-through intention rather than actual clicks or click-through rate. Random assignment supported a causal comparison between exposure to these two advertisements. However, the design did not establish the effect of all AI-generated copy relative to all human-written copy. Relevance and trust were measured rather than manipulated, so H3 and H4 concerned associations rather than causal effects.

3.2 Experimental Treatments and Stimuli

The stimuli were two 1080 × 1080 pixel social-media-style advertisements for FocusForm. The app was presented as helping users organise tasks, plan their day, prioritise work, set objectives, track progress and stay focused.

Both advertisements used the same fixed non-copy elements: the brand name and logo, phone mock-up, screen content, colour palette, background, icons, layout and call-to-action. The headline, body sentence and three feature statements differed between conditions, and minor spacing differences resulted from the different copy lengths. These visible copy differences were part of the treatment.

One version contained AI-generated copy. The other contained human-written copy supplied by an external freelance graphic designer with commercial advertising experience. The experiment therefore compared responses to two complete executions that differed in wording, framing, emphasis and length. These differences were allowed because each creator produced a complete copy execution from the shared brief, but they prevent the condition effect from being attributed to authorship alone.

The final advertisement stimuli are shown in Figure 3.1 and reproduced in Appendix A.

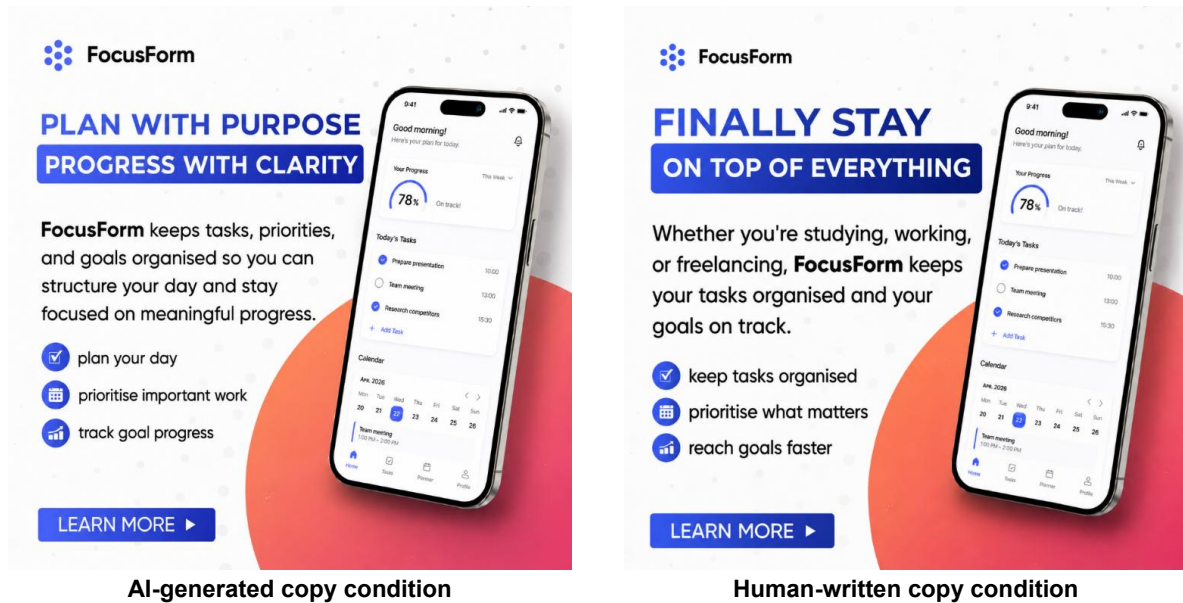


Figure 3.1. Final FocusForm advertisements used in the AI-generated and human-written copy conditions.

Note. Full-resolution versions of both advertisement stimuli are provided in Appendix A.

3.3 Copy Creation Procedure

The researcher developed one shared creative brief for both conditions. It described FocusForm, its main benefits, the intended adult target audience, the brand tone, the static social-media setting and the required copy structure. The external designer received the shared brief reproduced in Appendix B1 and used it to create the visual template and the human-written copy. No AI was used to produce the human-written version.

The AI-generated copy was created using OpenAI's ChatGPT GPT-5.5 High Intelligence model on 24 June 2026. The AI prompt reproduced the product description, target audience, brand tone, advertisement context, output structure and content requirements contained in the shared brief. The requirements to produce exactly one version, omit a separate call-to-action, avoid mentioning artificial intelligence and avoid exaggerated claims were already part of the shared brief. The AI prompt additionally stated that the copy would be used in an academic survey experiment and instructed the model not to refer to human-written copy,

the experiment or the comparison condition. The exact wording is reproduced in Appendix B3.

This approach followed guidance on prompt engineering and GenAI stimulus development. Prompt engineering uses task-specific instructions and context to guide model output without changing the model itself (Sahoo et al., 2024). Van Berlo, Campbell and Voorveld (2024) recommend specifying the manipulation, context, execution and output requirements when GenAI is used to create experimental stimuli. The prompt therefore defined the product, audience, tone, format, structure and content constraints.

The model produced one final copy version, which was used as generated and inserted into the same visual template as the human-written version. Both outputs addressed FocusForm, the adult audience, the functional benefits, the clean and motivating tone, and the required headline, body and feature structure. The fixed “Learn More” call-to-action remained part of the visual template and was not generated as part of either copy output. The human-written body sentence contained 16 words, slightly below the brief’s approximate 18–25-word range, and was retained as the designer’s final output. The final wording, framing, emphasis and length therefore differed as part of the two complete executions.

The shared creative brief, exact AI-generation prompt, full model output and final AI-generated copy are provided in Appendix B.

3.4 Survey Design and Measures

The Qualtrics survey was designed to take approximately four to six minutes. Participants first viewed the Participant Information Sheet, confirmed that they were aged 18 or over and provided consent. Eligible participants were then randomly assigned to one condition and shown one FocusForm advertisement. They then completed the four shared construct measures, followed by the attention check, demographic questions and debriefing statement. The procedure therefore followed the sequence: participant information; age eligibility and consent; random assignment; advertisement exposure; construct measures; attention check and demographics; and debriefing.

After exposure, all participants completed the same measures. Stated click-through intention was the primary dependent variable. Perceived authenticity, perceived relevance and trust in FocusForm were also measured. Click-through intention was used as a practical proxy where actual clicks were unavailable (Yoo, Kim and Stout, 2004; Armitage and Conner, 2001; Chandon, Morwitz and Reinartz, 2005). Relevance was included because it predicts attention, advertising attitude and click intention (Kim and Huh, 2016; De Keyzer et al., 2021). Trust was included because credibility and trust affect responses to digital advertising (MacKenzie and Lutz, 1989; Bleier and Eisenbeiss, 2015; Aguirre et al., 2015). Authenticity

was included because AI involvement may influence authenticity perceptions in creative communication (Brüns and Meißner, 2024; Kirk and Givi, 2025).

Each construct was measured using three positively worded items developed for this study rather than copied unchanged from a single validated scale. The click-through-intention items were informed by Yoo et al. (2004) and the behavioural-intention literature discussed in Section 2.1, then adapted to the non-clickable “Learn More” button. The relevance items reflected the self-relatedness and usefulness concepts discussed by Kim and Huh (2016) and De Keyzer et al. (2021). The trust items reflected the trustworthiness and claim-believability concepts discussed by MacKenzie and Lutz (1989), Bleier and Eisenbeiss (2015) and Aguirre et al. (2015). The authenticity items reflected the authenticity, genuineness and sincerity concepts discussed by Brüns and Meißner (2024) and Kirk and Givi (2025). Responses used a seven-point Likert scale from “Strongly disagree” to “Strongly agree”. The item sets and arithmetic-mean scoring rules were fixed during survey design and retained unchanged when the final data were analysed. Cronbach’s alpha assessed internal consistency rather than construct validity, and the sample was not intended to provide formal scale validation. No item was removed. The three items for each construct were averaged to create separate scores for click-through intention, authenticity, relevance and trust. No reverse coding was required. The exact item wording is shown in Table 3.1.

Construct	Item wording
Stated click-through intention	1. I would be likely to select the “Learn More” button in this advertisement. 2. I would consider selecting “Learn More” to find out more about FocusForm. 3. If I saw this advertisement online, I would intend to select “Learn More”.
Perceived authenticity	1. The FocusForm advertisement feels authentic. 2. The FocusForm advertisement comes across as genuine. 3. The message in this advertisement feels sincere.
Perceived relevance	1. The FocusForm advertisement is relevant to me. 2. The FocusForm app appears relevant to my needs. 3. The message in this advertisement relates to things that matter to me.
Trust	1. Based on this advertisement, FocusForm seems trustworthy. 2. Based on this advertisement, I would trust FocusForm to deliver what it promises. 3. The claims made about FocusForm in this advertisement seem believable.

Table 3.1. Item wording for the four study constructs.

A separate attention check stated: “To show that you are paying attention, please select ‘Disagree’ for this statement.” No question asked participants whether they believed the copy was AI- or human-written. The study was designed to compare responses under non-disclosure rather than to test whether participants could detect authorship. Although an authorship-perception question could have been placed after the outcome measures,

perceived authorship was outside the pre-specified hypotheses and was not collected. Demographic questions covered age, gender and prior productivity-app use.

A single informal pretest was conducted by a friend before data collection. It checked advertisement display, question wording, randomisation, survey flow and the CSV export. The pretest confirmed that the required variables were present and had the expected data types. The pretest response was excluded from the study data. Participants later received a debrief explaining the fictional brand, the two copy conditions, the non-clickable button and the measurement of stated intention.

The full survey instrument, screenshots, attention-check wording, debriefing statement and Qualtrics flow are provided in Appendix C.

3.5 Participants, Sampling and Data Collection

The eligible population was adults aged 18 or over. Participants were recruited through digital and social-media channels, so the term adult digital consumers is used as a broad descriptive label rather than a screened eligibility characteristic. A non-probability convenience sample was used. Participation was voluntary and no incentive was offered. Recruitment took place through Instagram Stories and WhatsApp, with further sharing by friends and relatives. Data collection ran from 30 June to 13 July 2026.

The study aimed for at least 60 usable responses with equal allocation between conditions. With 30 participants per condition, $\alpha = .05$ and 80% power, the design could detect an effect of approximately Cohen's $d = 0.74$ or larger. The sample was therefore suitable for detecting moderate-to-large differences, but smaller effects may not have been identified. Non-significant findings were not treated as proof that no difference existed.

After exclusions, the final analytical sample contained 60 participants, with 30 in each condition. Random assignment supported the internal comparison, but convenience sampling limited generalisability beyond adults reached through the researcher's social-media networks.

The recruitment materials and sensitivity power calculation are provided in Appendix D.

3.6 Data Preparation and Ethics

Qualtrics responses were exported to CSV and prepared in Python. Two metadata rows were removed first, leaving 93 participant-response records. Qualtrics export fields not required for eligibility, completion checks, scoring, demographics or the statistical analyses were removed. The retained completion and eligibility indicators, experimental condition, construct items, attention check and demographic variables were converted into the numerical or categorical formats required for analysis. Condition labels were standardised

and one written age response was converted to a numeric value. Composite scores were calculated as the mean of the three items for each construct. The retained variables and coding are listed in Appendix F2.

Exclusions were applied sequentially. One under-18 response and one non-consenting response were removed, followed by 20 incomplete responses and 11 attention-check failures. This produced the final sample of 60. No imputation was required because the retained responses were complete for all substantive items. Survey duration was inspected, but no duration-based exclusion was applied because no threshold had been specified in advance. No duplicate Qualtrics Response IDs were identified.

Ethical approval was granted by the Warwick Business School Programme Quality Team, University of Warwick (reference R_236XsTp5oUFzJAJ). Participants provided informed consent and could withdraw before submission by closing the survey. Post-submission withdrawal was not possible because participants did not provide an identifier that could be used to locate an individual response. The survey was configured to collect anonymous responses. No names, email addresses, other directly identifying information or IP addresses were collected. Qualtrics export metadata not required for eligibility, data-quality checks, scoring or analysis were removed and were not analysed. The anonymised analysis files were retained intact in an appropriate format on the researcher's private laptop and will be kept for at least 10 years from the date of any publication based on the study, in line with University policy, before disposal under University guidelines.

The ethics approval is provided in Appendix E. The data-cleaning summary, variable dictionary, composite-scoring details and selected Python code are provided in Appendix F.

3.7 Data Analysis

The final dataset was analysed in Python. Pandas and NumPy were used for data preparation, scoring and descriptive tables. SciPy and statsmodels were used for inferential analysis and diagnostics, while Matplotlib was used for plots.

Cronbach's alpha assessed the internal consistency of each three-item construct (Cronbach, 1951). Descriptive statistics and Pearson correlations between the four composite measures were calculated before hypothesis testing. The correlations were treated as descriptive analyses rather than additional hypothesis tests. Composite means were treated as approximately continuous variables (Norman, 2010). All primary tests were two-sided with $\alpha = .05$.

H1 and H2 were tested using Welch independent-samples t-tests. H1 compared mean stated click-through intention between conditions, while H2 compared mean perceived

authenticity. Welch's test was chosen because it does not assume equal population variances (Welch, 1947). Means, standard deviations, mean differences, degrees of freedom, *p*-values and 95% confidence intervals were reported. Hedges' *g* was used as the preferred standardised mean difference because it corrects for small-sample bias (Hedges, 1981). Differences were calculated as AI minus human.

H3 and H4 were tested using multiple linear regression. Stated click-through intention was regressed on perceived relevance, trust and condition, coded 0 for human-written and 1 for AI-generated. H3 was evaluated through the relevance coefficient and H4 through the trust coefficient. The model was specified to test H3 and H4; perceived authenticity was treated as the outcome for H2 rather than as an additional predictor. The regression coefficients represented conditional associations rather than causal effects.

Regression diagnostics included variance inflation factors, residual-versus-fitted and Q-Q plots, Breusch–Pagan (Breusch and Pagan, 1979) and White tests (White, 1980), the Jarque–Bera test and Ramsey RESET (Ramsey, 1969). Because the formal tests indicated heteroscedasticity, HC3 robust standard errors were used for final inference. Cook's distance, leverage, studentised residuals and leave-one-out analysis assessed whether individual observations materially affected the conclusions. Observations crossing the screening thresholds were examined using all four diagnostics. An observation would not be removed solely because it was influential; exclusion would require evidence that the response was invalid. No such evidence was identified.

Additional descriptive and reliability results are provided in Appendix G. Detailed t-test outputs, regression results, diagnostics and sensitivity analyses are provided in Appendix H.

4. Results

This chapter presents the final analytical sample, the reliability and descriptive results for the four study constructs, the descriptive randomisation check, the hypothesis tests, and the multiple regression findings. The results are reported in the same order as the hypotheses introduced in Chapter 2.

4.1 Sample Profile and Data Cleaning

The raw Qualtrics export contained 95 rows. Two rows contained Qualtrics question-text and ImportId metadata and were removed before participant-level cleaning, leaving 93 participant-response records. One age-ineligible response and one non-consenting response were excluded, followed by 20 incomplete responses. A further 11 responses were excluded for failing the instructional attention check. The final analytical sample therefore comprised 60 participants. No response was excluded based on survey duration or substantive scores, and no duplicated Qualtrics Response IDs were identified.

Stage	Records removed	Records remaining
Raw CSV rows	—	95
Qualtrics metadata rows removed	2	93
Age-ineligible response removed	1	92
Non-consenting response removed	1	91
Incomplete responses removed	20	71
Attention-check failures removed	11	60
Final analytical sample	—	60

Table 4.1. Data-cleaning and analytical-sample flow.

The final sample was evenly allocated between conditions, with 30 participants viewing the AI-generated copy advertisement and 30 viewing the human-written copy advertisement. Participants had a mean age of 30.65 years ($SD = 11.83$), with ages ranging from 22 to 64 years. The sample comprised 33 male participants (55.0%) and 27 female participants (45.0%); no participant selected the non-binary/third-gender or prefer-not-to-say options. Regarding prior productivity-app use, 20 participants (33.3%) reported using such an app during the previous 12 months, 36 (60.0%) reported no prior use, and four (6.7%) were unsure.

Characteristic	Overall (<i>n</i> = 60)	AI (<i>n</i> = 30)	Human (<i>n</i> = 30)
Age, <i>M</i> (<i>SD</i>)	30.65 (11.83)	30.90 (12.76)	30.40 (11.02)
Age, median (range)	26 (22–64)	26 (22–64)	26 (23–61)
Male, <i>n</i> (%)	33 (55.0%)	18 (60.0%)	15 (50.0%)
Female, <i>n</i> (%)	27 (45.0%)	12 (40.0%)	15 (50.0%)
Prior app use: No, <i>n</i> (%)	36 (60.0%)	15 (50.0%)	21 (70.0%)
Prior app use: Yes, <i>n</i> (%)	20 (33.3%)	12 (40.0%)	8 (26.7%)
Prior app use: Not sure, <i>n</i> (%)	4 (6.7%)	3 (10.0%)	1 (3.3%)

Table 4.2. Sample characteristics overall and by advertisement condition.

4.2 Reliability and Descriptive Statistics

Cronbach's alpha was .935 for click-through intention, .897 for perceived authenticity, .915 for perceived relevance, and .873 for trust. The highest overall mean was observed for trust ($M = 4.99$, $SD = 1.18$), followed by authenticity ($M = 4.89$, $SD = 1.30$), relevance ($M = 4.36$, $SD = 1.53$), and click-through intention ($M = 4.16$, $SD = 1.79$). Compared with the human condition, the AI condition mean was 0.19 points higher for click-through intention, 0.28 points higher for relevance, 0.04 points higher for trust and 0.16 points lower for authenticity.

Construct	Items	α	Overall <i>M</i>	Overall <i>SD</i>	AI <i>M</i>	AI <i>SD</i>	Human <i>M</i>	Human <i>SD</i>	Minimum	Maximum
Click-through intention	3	.935	4.16	1.79	4.26	1.94	4.07	1.65	1.00	7.00
Perceived authenticity	3	.897	4.89	1.30	4.81	1.19	4.97	1.42	1.00	7.00
Perceived relevance	3	.915	4.36	1.53	4.50	1.58	4.22	1.49	1.67	7.00
Trust	3	.873	4.99	1.18	5.01	1.13	4.97	1.25	1.00	7.00

Table 4.3. Reliability and descriptive statistics for the study constructs.

Pearson correlations showed that click-through intention was positively correlated with authenticity ($r = .531$), relevance ($r = .465$), and trust ($r = .449$). Authenticity and trust were also positively correlated ($r = .764$), while relevance had little correlation with either authenticity ($r = .049$) or trust ($r = .003$). These correlations are presented descriptively and were not used as separate hypothesis tests.

Detailed item-level descriptive statistics, reliability diagnostics and the correlation matrix are provided in Appendix G.

4.3 Randomisation Check

The measured background characteristics were compared descriptively across conditions. Mean age differed by 0.50 years between the AI condition ($M = 30.90$, $SD = 12.76$) and the

human condition ($M = 30.40$, $SD = 11.02$). The AI condition contained 18 male and 12 female participants, while the human condition contained 15 participants in each gender category. Prior productivity-app use was reported by 40.0% of the AI condition and 26.7% of the human condition. These values describe the observed sample composition only. No formal tests of balance were conducted and random assignment provides the basis for the experimental comparison.

4.4 Hypothesis Testing

H1 proposed that average click-through intention would differ between the advertisement conditions. Mean click-through intention was slightly higher in the AI-generated copy condition ($M = 4.26$, $SD = 1.94$) than in the human-written copy condition ($M = 4.07$, $SD = 1.65$). However, the mean difference of 0.19 was not statistically significant, $t(56.52) = 0.41$, $p = .686$, 95% CI $[-0.74, 1.12]$. The confidence interval included zero, and the standardised mean difference was small (Hedges' $g = 0.103$). There was therefore insufficient evidence to reject the null hypothesis, and H1 was not supported.

H2 proposed that perceived authenticity would differ between the advertisement conditions. Mean authenticity was slightly lower in the AI-generated copy condition ($M = 4.81$, $SD = 1.19$) than in the human-written copy condition ($M = 4.97$, $SD = 1.42$). However, the mean difference of -0.16 was not statistically significant, $t(56.23) = -0.46$, $p = .648$, 95% CI $[-0.83, 0.52]$. The confidence interval included zero, and the standardised mean difference was small (Hedges' $g = -0.117$). There was therefore insufficient evidence to reject the null hypothesis, and H2 was not supported. The results did not show a statistically significant difference, but smaller differences may still exist and may not have been detected by the study.

Hypothesis	Outcome	AI M (SD)	Human M (SD)	Difference	95% CI	t	df	p	Hedges' g	Decision
H1	Click-through intention	4.26 (1.94)	4.07 (1.65)	0.189	$[-0.743, 1.121]$	0.406	56.519	.686	0.103	Not supported
H2	Perceived authenticity	4.81 (1.19)	4.97 (1.42)	-0.156	$[-0.834, 0.523]$	-0.459	56.230	.648	-0.117	Not supported

Table 4.4. Welch independent-samples t -test results for H1 and H2.

4.5 Regression Analysis

A multiple linear regression examined whether perceived relevance and trust were associated with click-through intention after accounting for experimental condition. The model included relevance, trust and condition, with the human-written copy condition coded

as 0 and the AI-generated copy condition coded as 1. The model explained 41.6% of the variance in click-through intention ($R^2 = .416$; adjusted $R^2 = .385$). The HC3 robust overall model test was statistically significant, $F(3, 56) = 12.57$, $p < .001$.

The Breusch–Pagan ($p = .020$) and White ($p = .006$) tests indicated non-constant residual variance. Final coefficient inference was therefore based on HC3 robust standard errors. Perceived relevance was positively associated with click-through intention, $B = 0.543$, HC3 SE = 0.156, $z = 3.49$, $p < .001$, 95% CI [0.238, 0.848]. Holding trust and condition constant, a one-point increase in relevance was associated with an estimated 0.54-point increase in click-through intention. H3 was therefore supported.

Trust was also positively associated with click-through intention, $B = 0.679$, HC3 SE = 0.156, $z = 4.36$, $p < .001$, 95% CI [0.374, 0.984]. Holding relevance and condition constant, a one-point increase in trust was associated with an estimated 0.68-point increase in click-through intention. H4 was therefore supported. These coefficients describe conditional associations rather than causal effects. After accounting for perceived relevance and trust, the condition coefficient was close to zero and was not statistically significant, $B = 0.008$, HC3 SE = 0.396, $p = .984$, 95% CI [-0.767, 0.783].

Predictor	<i>B</i>	HC3 SE	<i>z</i>	<i>p</i>	Lower 95% CI	Upper 95% CI
Intercept	-1.598	1.056	-1.514	.130	-3.667	0.471
Perceived relevance	0.543	0.156	3.487	< .001	0.238	0.848
Trust	0.679	0.156	4.363	< .001	0.374	0.984
Condition	0.008	0.396	0.020	.984	-0.767	0.783

Table 4.5. Multiple regression predicting click-through intention.

The predictor VIF values were approximately 1.00, indicating no material multicollinearity concern. The Q-Q plot and Jarque–Bera test did not show a clear departure from residual normality ($p = .252$). Ramsey RESET did not provide strong evidence of functional-form misspecification ($p = .096$), although these diagnostics do not prove that the model was correctly specified. Although the residual plot did not show a severe variance pattern, the Breusch–Pagan and White tests indicated heteroscedasticity, supporting the use of HC3 inference. Potentially influential observations were inspected using Cook’s distance, leverage and studentised residuals. In the leave-one-out analysis, the relevance and trust coefficients remained positive and statistically significant in every model. The condition coefficient changed direction across models but remained statistically non-significant in every case. No observations were removed on this basis.

Residual-versus-fitted and Q-Q plots and the formal diagnostic-test outputs are provided in Appendix H6. Influence diagnostics and the leave-one-out sensitivity analysis are provided in Appendices H7 and H8.

4.6 Summary of Findings

The condition comparisons did not identify statistically significant differences in either click-through intention or perceived authenticity, meaning that H1 and H2 were not supported. In the regression analysis, both relevance and trust were positively associated with click-through intention after accounting for each other and experimental condition. H3 and H4 were therefore supported.

Hypothesis	Analysis	Concise finding	Decision
H1	Welch independent-samples t-test	No statistically significant click-through-intention difference between conditions	Not supported
H2	Welch independent-samples t-test	No statistically significant authenticity difference between conditions	Not supported
H3	Multiple regression with HC3 inference	Relevance was positively associated with stated click-through intention	Supported
H4	Multiple regression with HC3 inference	Trust was positively associated with stated click-through intention	Supported

Table 4.6. Summary of hypothesis outcomes.

5. Discussion

This chapter interprets the main findings in relation to the research question and the literature reviewed in Chapter 2. It then considers the theoretical and managerial implications. The final section discusses how AI can support advertising-copy development without replacing human judgement.

5.1 Overview of Main Findings

The study did not find a statistically significant difference in stated click-through intention between the AI-generated and human-written copy conditions. The mean was slightly higher in the AI-generated condition, but the difference was small. H1 was therefore not supported. The result did not show that either advertisement produced higher stated click-through intention. However, it also did not prove that the two advertisements were equivalent. Smaller differences may still exist.

Perceived authenticity also did not differ significantly between conditions. The human-written condition had a slightly higher mean, but the observed AI-minus-human difference was only -0.16 points, its confidence interval included zero and the test was not statistically significant. H2 was therefore not supported. This does not mean that AI-generated advertising will always be viewed as equally authentic.

Perceived relevance and trust were both positively associated with stated click-through intention after accounting for each other and advertisement condition. H3 and H4 were therefore supported. The relevance and trust coefficients were positive and statistically significant, while the condition coefficient was close to zero and was not statistically significant. The detailed findings are shown in Tables 4.4 and 4.5.

Overall, the condition comparison did not identify a statistically significant difference in stated click-through intention. In the regression model, relevance and trust showed positive conditional associations with the outcome.

5.2 Interpretation in Relation to Existing Literature

The small observed difference in click-through intention is compatible with research showing that responses to AI-generated advertising can depend on the message and product. Bakpayev et al. (2020) found that AI- and human-written rational advertisements received similar evaluations, while AI-created emotional advertisements were evaluated less favourably. They also found that AI-created advertising was more acceptable for utilitarian than hedonic products. Castelo et al. (2019) similarly found that consumers trusted algorithms more for objective tasks than for subjective tasks. However, the confidence

interval in the present study allowed differences in either direction, so the non-significant result should not be interpreted as evidence that the two advertisements were equivalent.

FocusForm was presented as a functional productivity app. Its benefits focused on organisation, planning, prioritisation and progress tracking. One possible interpretation is that this functional context reduced any disadvantage of the AI-generated copy. This explanation is speculative because the study did not compare rational and emotional appeals, utilitarian and hedonic products, or participants' perceptions of the advertisement's functional character. It also tested only two advertisements. The result therefore applies to this setting and should not be generalised to AI-generated copy as a whole.

The result for relevance was consistent with the literature reviewed in Chapter 2. Kim and Huh (2016) found that relevance predicted attention, attitude towards the advertisement and clicking. De Keyzer et al. (2021) also found that personal relevance improved brand attitude and click intention. Bleier and Eisenbeiss (2015) identified perceived usefulness as an important part of personalised advertising.

Participants who viewed the FocusForm message as more relevant also reported stronger stated click-through intention. This positive association is consistent with the view that advertising perceived as useful is linked to stronger intended engagement. It also fits Ducoffe's (1995) advertising-value framework, which links advertising value to the utility an advertisement provides. However, advertising value was not measured directly, and relevance was not manipulated. The result therefore shows an association rather than a causal effect.

Trust was also positively associated with stated click-through intention. This fits research linking trust and credibility to favourable advertising responses. MacKenzie and Lutz (1989) define advertising credibility in terms of whether claims appear truthful and believable. Brackets and Carr (2001) also identify credibility as part of advertising value and attitude. Bleier and Eisenbeiss (2015) found better click-related outcomes for a trusted retailer, while Aguirre et al. (2015) found that trust-building strategies could reduce negative reactions to personalised advertising. Chen et al. (2022) also linked credibility to advertising attitude.

FocusForm was fictional, so participants had no previous experience with the app. The trust scale combined judgements of the app's trustworthiness, whether it would deliver what it promised and whether its claims appeared believable. Perceived believability and dependability are shared elements across this measure, advertising credibility and retailer trust, but the constructs are not identical. The positive coefficient therefore provides context-specific evidence that stronger trust in FocusForm and its claims was associated with stronger stated click-through intention.

The authenticity result was less straightforward. Brüns and Meißner (2024) found that GenAI use in social-media content could reduce perceived brand authenticity. Kirk and Givi (2025) also found that emotional marketing messages believed to be AI-authored were seen as less authentic. However, these studies involved situations in which consumers knew or believed that AI had been used.

Participants in the present study were not given an authorship label. However, the study did not measure whether they inferred that either advertisement was AI-generated. Non-disclosure may help explain why the authenticity difference was small and non-significant, but this remains speculative. The study also did not measure perceived emotionality or compare different levels of emotional content.

The study also did not directly test algorithm aversion or algorithm appreciation. Dietvorst et al. (2015) found that consumers can avoid algorithms after seeing them make mistakes, while Logg et al. (2019) found that people may prefer algorithmic judgement in some numerical tasks. Castelo et al. (2019) showed that responses can depend on whether a task is viewed as objective or subjective. The present study did not measure attitudes towards AI or disclose which advertisement was AI-generated. It therefore cannot show how consumers would have responded if AI involvement had been made clear.

Relevance and trust were almost uncorrelated with each other despite both having positive conditional associations with stated click-through intention. Within this sample, they may therefore reflect different aspects of advertisement evaluation. However, this pattern does not establish separate causal mechanisms. The advertisement-condition coefficient remained close to zero and non-significant. This does not mean that authorship never matters. Its effect may depend on the product, the message and whether AI involvement is disclosed.

Taken together, the small and non-significant condition differences may reflect the functional product context, the non-disclosure of AI authorship, the similar functional requirements given to both copy creators and the study's limited ability to detect small effects. These explanations remain tentative because product type, disclosure and message appeal were not independently manipulated.

5.3 Theoretical Implications

This study adds context-specific empirical evidence to research on AI-generated marketing communication by comparing responses to AI-generated and human-written advertising copy when authorship was not disclosed. Participants were not told whether the copy was AI-generated or human-written. The results therefore reflect responses to the two final advertisement executions rather than reactions to an AI label. This differs from research

examining situations in which consumers know or believe that AI was involved, such as Brüns and Meißner (2024) and Kirk and Givi (2025).

This matters because a negative response to AI authorship is not necessarily a negative response to the content itself. Under non-disclosure, the tests did not detect statistically significant differences in stated click-through intention or perceived authenticity between the two advertisement conditions. The confidence intervals allowed differences in either direction, so the results do not establish that the AI-generated condition was no lower or equivalent. The study also did not isolate authorship from differences in wording, framing and copy length or test how disclosure would have changed the results.

The findings are context-specific and are compatible with prior evidence that responses to AI-generated advertising may vary across settings. Earlier research suggests that AI may be more accepted for practical and objective tasks than for emotional or subjective tasks (Bakpayev et al., 2020; Castelo et al., 2019). FocusForm was a functional productivity app, but the study did not compare different contexts or estimate a context interaction. It therefore cannot establish that context caused the observed result or show a general advantage or disadvantage for AI-generated advertising.

Finally, the results provide additional context-specific evidence that relevance and trust are positively associated with stated click-through intention after accounting for each other and advertisement condition. This is consistent with the relevance, credibility and advertising-value literature reviewed in Chapter 2. The study does not establish that these perceptions are causal, universally important or more important than variables that were not included in the model.

5.4 Managerial Implications

Marketers should not assume that AI-generated copy will automatically receive weaker consumer responses. The tests did not detect statistically significant differences in stated click-through intention or perceived authenticity between conditions. They did not establish that the AI-generated advertisement was equivalent to or no lower than the human-written advertisement.

The final copy should be judged on its quality rather than only on how it was created. Marketers should assess whether the message is relevant to the intended audience and whether the product appears trustworthy. They should also check its clarity, tone, brand fit and factual accuracy.

A message can be fluent and well written without being useful to the consumer. Marketers should therefore make sure that the copy connects product benefits to problems or goals

that matter to the target audience. Trust should also be considered, especially for unfamiliar products. The study did not test specific ways of building trust, but it showed that trust in the app was associated with stronger stated click-through intention.

AI-generated copy should be reviewed and tested before use. Internal review can assess clarity, factual accuracy and brand fit, while consumer pretesting can assess intended reactions before launch. Live A/B testing can then measure behavioural outcomes such as clicks, conversions or another campaign objective. This is especially important because the present study measured stated intention rather than actual campaign performance.

The results are based on two advertisements, one fictional product and 60 participants. Firms should therefore avoid treating one AI-generated output as evidence of what AI can produce in every context. Each output should be assessed on its own merits.

5.5 Implications for AI-Assisted Advertising

The experiment did not compare AI-assisted, fully automated and fully human production workflows. The recommendation in this section is therefore based on the literature and managerial judgement rather than a workflow effect tested in the study.

The literature suggests that AI can support advertising development by generating ideas, early drafts and alternative copy versions, while humans retain responsibility for strategy and final decisions (Bakpayev et al., 2020; Brüns and Meißner, 2024). Human marketers should define the intended audience, communication goals, brand tone and constraints before using AI.

AI can then generate alternatives, which the marketer should review for accuracy, clarity, tone and brand fit before consumer pretesting and behavioural A/B testing. The final decision should remain with the human marketer, who is responsible for the strategic direction and suitability of the advertisement.

This division of responsibilities is consistent with research on human–AI collaboration. Ryoo et al. (2025) found that consumers expected collaboration to combine efficiency and information management with human creativity. Brüns and Meißner (2024) also found weaker negative reactions when AI supported human creators rather than replacing them. AI-generated copy can therefore be treated as one input into advertising development rather than a replacement for human marketing judgement.

6. Limitations and Conclusion

This final chapter considers the limitations of the study, outlines directions for future research, and provides a final answer to the research question.

6.1 Limitations

The findings should be interpreted in consideration of several limitations. The first concerns the use of stated click-through intention rather than actual clicking behaviour. Participants reported how likely they would be to click the “Learn More” button, but the advertisement was static and the button was not interactive. Stated intentions do not always translate into behaviour (Chandon et al., 2005). The study therefore cannot show whether either advertisement would have produced a higher actual click-through rate in a live campaign.

Stated click-through intention, perceived authenticity, perceived relevance and trust were all measured through self-report scales. Participants may have interpreted the items differently or used the response scale in different ways. They also knew that they were taking part in a study and may have paid more attention to the advertisement than they would during normal social-media use. The survey therefore provided a controlled comparison, but it did not fully reproduce real digital-advertising conditions. Because all four constructs were self-reported in the same survey, common-method influence cannot be ruled out. The strong correlation between authenticity and trust may also reflect some conceptual overlap.

The sampling approach also limits the extent to which the findings can be generalised. Participants were recruited mainly through the researcher’s social-media networks, particularly Instagram and WhatsApp. The final sample may therefore differ from adults beyond those networks. Only age, gender and prior productivity-app use were included in the reported sample profile, so possible differences in education, social-media use, exposure to digital advertising or broader technology experience were not assessed.

The final analytical sample consisted of 60 participants, with 30 in each condition. This was sufficient for detecting moderate-to-large differences, but smaller effects may not have been identified. The non-significant findings for click-through intention and authenticity should therefore not be interpreted as strong evidence that no differences exist.

The use of the fictional FocusForm brand created both an advantage and a limitation. It reduced the influence of prior brand experience because participants had no existing relationship with the app. However, it also reduced realism. Participants could not base their trust or click-through intention on previous experience, reputation, reviews, or established brand knowledge. For this reason, responses to a real or familiar brand may differ.

The study also examined only one product category. FocusForm was presented as a practical productivity app. Responses to AI-generated copy may be different for products and services involving emotion, personal identity, luxury, financial risk, healthcare, or charitable giving. The findings should therefore not be generalised across all advertising categories.

Participants viewed only one static advertisement on one occasion. Real consumers may encounter advertisements repeatedly and across several platforms and formats. Repeated exposure may affect familiarity, trust, recall, irritation, and clicking behaviour. A single survey exposure cannot show how responses might change over time.

A further limitation concerns the specific advertisements used. The study compared one AI-generated copy execution with one human-written execution. The findings therefore depend partly on the particular wording, framing, message appeal and fit of those versions within the visual template. A different AI model, prompt, human writer, brief or editing process could have produced different copy and different results. The study does not compare the general ability of AI and humans to create advertising copy. It compares responses to these two advertisements.

Although the advertisements followed the same broad visual template, authorship was not the only feature that differed. The wording, framing, emphasis, spacing and copy length varied between the final advertisement executions. The condition therefore combined authorship with content differences. Comparing complete executions increased realism but reduced experimental control because the condition effect cannot be attributed to authorship alone.

Finally, participants were not told whether the copy was AI-generated or human-written. This allowed the study to examine responses to the final advertisements without an AI label. However, the findings cannot show whether disclosure would have changed authenticity, trust or stated click-through intention.

These limitations do not remove the value of the study. They define the boundaries within which the findings should be interpreted. The results apply to one controlled survey experiment involving two advertisements, one fictional product and a convenience sample of 60 participants.

6.2 Future Research

Future research should test AI-generated and human-written advertising copy in more realistic settings. Live advertising-platform experiments could use interactive advertisements and measure actual clicks, click-through rates, engagement and conversions where

appropriate. This would provide behavioural evidence rather than relying only on stated intention.

Larger and more diverse samples would also strengthen future studies. A larger sample would improve the ability to detect smaller differences between conditions. Recruiting participants beyond personal networks through research panels, targeted recruitment, or stratified sampling would also improve representativeness. Future research could examine whether responses differ by age, education, occupation, digital-advertising experience, or familiarity with AI.

Future research should test separate sources of contextual variation. One set of studies could compare utilitarian and hedonic products. Other studies could vary practical and emotional appeals, familiar and unfamiliar brands, or low-risk and high-risk decisions. Treating these dimensions separately would show which feature, if any, changes responses to AI-generated and human-written advertising copy.

Using several advertisements in each condition would reduce reliance on one AI-generated and one human-written execution. Future studies could vary AI models, prompts, human writers, copy styles, emotional and rational appeals, and levels of human editing. This would provide a broader comparison of AI-generated and human-written advertising copy.

The effect of disclosure should also be tested directly. A future study could compare disclosed AI-generated copy, undisclosed AI-generated copy, disclosed human-written copy and undisclosed human-written copy. This would help separate reactions to the content from reactions to an AI or human label.

Future studies could also examine repeated exposure and different advertising formats, including static social-media advertisements, video, search advertising, and email marketing. Finally, relevance and trust could be manipulated experimentally. This would help determine whether increasing relevance or trust causes stronger click-through intention rather than only being associated with it.

6.3 Conclusion

This dissertation examined whether average stated click-through intention differed between participants exposed to the AI-generated and human-written FocusForm advertisement conditions.

Within this experiment, the study did not find evidence that the AI-generated FocusForm advertisement produced different stated click-through intention from the human-written FocusForm advertisement. The mean was slightly higher in the AI-generated condition, but the difference was small and not statistically significant. This does not prove that the two

advertisements were equivalent. The conclusion applies only to these two FocusForm advertisements and to stated intention rather than actual clicking behaviour.

The study also did not identify a statistically significant difference in perceived authenticity. By contrast, perceived relevance and trust had positive and statistically significant conditional coefficients, while the advertisement-condition coefficient was close to zero and non-significant. Relevance and trust were measured rather than manipulated, so these findings do not establish causal effects or rank the predictors by importance.

The dissertation adds context-specific empirical evidence to research on AI-generated marketing communication. It compares responses to AI-generated and human-written advertising copy when participants were not told who created it. The study therefore focuses on responses to the two final advertisement executions rather than reactions to an AI label.

Overall, the findings do not show that AI-generated copy is generally better or worse than human-written copy. As a cautious recommendation informed by the literature rather than directly tested by the experiment, AI-generated copy can be used as an input into advertising development while human marketers retain responsibility for product and audience understanding, strategic direction, review, testing and final approval.

Reference List

- Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K. and 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), pp. 34–49. doi:10.1016/j.jretai.2014.09.005.
- Armitage, C.J. and Conner, M. (2001) 'Efficacy of the theory of planned behaviour: A meta-analytic review', *British Journal of Social Psychology*, 40(4), pp. 471–499. doi:10.1348/014466601164939.
- Baek, T.H., Kim, J. and Kim, J.H. (2024) 'Effect of disclosing AI-generated content on prosocial advertising evaluation', *International Journal of Advertising*, 45(1), pp. 171–192. doi:10.1080/02650487.2024.2401319.
- Bakpayev, M., Baek, T., van Esch, P. and Yoon, S. (2020) 'Programmatic creative: AI can think but it cannot feel', *Australasian Marketing Journal*, 30(1), pp. 90–95. doi:10.1016/j.ausmj.2020.04.002.
- Bleier, A. and Eisenbeiss, M. (2015) 'The Importance of Trust for personalized online advertising', *Journal of Retailing*, 91(3), pp. 390–409. doi:10.1016/j.jretai.2015.04.001.
- Brackett, L.K. and Carr, B.N. (2001) 'Cyberspace advertising vs. other media: Consumer vs. mature student attitudes', *Journal of Advertising Research*, 41(5), pp. 23–33. doi:10.2501/jar-41-5-23-32.
- Breusch, T.S. and Pagan, A.R. (1979) 'A simple test for heteroscedasticity and random coefficient variation', *Econometrica*, 47(5), p. 1287. doi:10.2307/1911963.
- Brüns, J.D. and 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, p. 103790. doi:10.1016/j.jretconser.2024.103790.
- Castelo, N., Bos, M.W. and Lehmann, D.R. (2019) 'Task-dependent algorithm aversion', *Journal of Marketing Research*, 56(5), pp. 809–825. doi:10.1177/0022243719851788.
- Chandon, P., Morwitz, V.G. and Reinartz, W.J. (2005) 'Do intentions really predict behavior? self-generated validity effects in survey research', *Journal of Marketing*, 69(2), pp. 1–14. doi:10.1509/jmkg.69.2.1.60755.
- Chen, W.-K., Ling, C.-J. and Chen, C.-W. (2022) 'What affects users to click social media ads and purchase intention? the roles of advertising value, emotional appeal and

- credibility', *Asia Pacific Journal of Marketing and Logistics*, 35(8), pp. 1900–1916. doi:10.1108/apjml-01-2022-0084.
- Cronbach, L.J. (1951) 'Coefficient alpha and the internal structure of tests', *Psychometrika*, 16(3), pp. 297–334. doi:10.1007/bf02310555.
- De Keyzer, F., Dens, N. and De Pelsmacker, P. (2021) 'How and when personalized advertising leads to brand attitude, click, and wom intention', *Journal of Advertising*, 51(1), pp. 39–56. doi:10.1080/00913367.2021.1888339.
- Desveaud, K. and Pavone, G. (2025) 'Consumer perceptions of AI-generated marketing content', *Encyclopedia of Artificial Intelligence in Marketing*, pp. 1–7. doi:10.1007/978-3-031-75316-9_94-1.
- Dietvorst, B.J., Simmons, J.P. and Massey, C. (2015) 'Algorithm aversion: People erroneously avoid algorithms after seeing them err.', *Journal of Experimental Psychology: General*, 144(1), pp. 114–126. doi:10.1037/xge0000033.
- Drèze, X. and Hussherr, F.-X. (2003) 'Internet advertising: Is anybody watching?', *Journal of Interactive Marketing*, 17(4), pp. 8–23. doi:10.1002/dir.10063.
- Ducoffe, R.H. (1995) 'How consumers assess the value of advertising', *Journal of Current Issues & Research in Advertising*, 17(1), pp. 1–18. doi:10.1080/10641734.1995.10505022.
- Friestad, M. and Wright, P. (1994) 'The persuasion knowledge model: How people cope with persuasion attempts', *Journal of Consumer Research*, 21(1), pp. 1–31. doi:10.1086/209380.
- Grewal, D., Saturnino, C., Davenport, T. and Guha, A. (2024) 'How generative AI is shaping the future of marketing', *Journal of the Academy of Marketing Science*, 53(3), pp. 702–722. doi:10.1007/s11747-024-01064-3.
- Hedges, L.V. (1981) 'Distribution theory for Glass's estimator of effect size and related estimators', *Journal of Educational Statistics*, 6(2), p. 107. doi:10.2307/1164588.
- Hoban, P.R. and Bucklin, R.E. (2015) 'Effects of internet display advertising in the purchase funnel: Model-based insights from a randomized field experiment', *Journal of Marketing Research*, 52(3), pp. 375–393. doi:10.1509/jmr.13.0277.
- Kim, H. and Huh, J. (2016) 'Perceived relevance and privacy concern regarding online behavioral advertising (OBA) and their role in consumer responses', *Journal of Current Issues & Research in Advertising*, 38(1), pp. 92–105. doi:10.1080/10641734.2016.1233157.

- Kirk, C.P. and Givi, J. (2025) 'The ai-authorship effect: Understanding authenticity, moral disgust, and consumer responses to ai-generated marketing communications', *Journal of Business Research*, 186, p. 114984. doi:10.1016/j.jbusres.2024.114984.
- Kshetri, N., Dwivedi, Y., Davenport, T. and Panteli, N. (2024) 'Generative Artificial Intelligence in Marketing: Applications, opportunities, challenges, and Research Agenda', *International Journal of Information Management*, 75, p. 102716. doi:10.1016/j.ijinfomgt.2023.102716.
- Lewis, R.A. and Reiley, D.H. (2014) 'Online ads and offline sales: Measuring the effect of retail advertising via a controlled experiment on Yahoo!', *Quantitative Marketing and Economics*, 12(3), pp. 235–266. doi:10.1007/s11129-014-9146-6.
- Logg, J.M., Minson, J.A. and Moore, D.A. (2019) 'Algorithm appreciation: People prefer algorithmic to human judgment', *Organizational Behavior and Human Decision Processes*, 151, pp. 90–103. doi:10.1016/j.obhdp.2018.12.005.
- Lohtia, R., Donthu, N. and Hershberger, E.K. (2003) 'The impact of content and design elements on banner advertising click-through rates', *Journal of Advertising Research*, 43(4), pp. 410–418. doi:10.2501/jar-43-4-410-418.
- MacKenzie, S.B. and Lutz, R.J. (1989) 'An empirical examination of the structural antecedents of attitude toward the ad in an advertising pretesting context', *Journal of Marketing*, 53(2), pp. 48–65. doi:10.1177/002224298905300204.
- Maseeh, H.I., Jebarajakirthy, C., Pentecost, R., Ashaduzzaman, M., Arli, D. and Weaven, S. (2021) 'A meta-analytic review of Mobile Advertising Research', *Journal of Business Research*, 136, pp. 33–51. doi:10.1016/j.jbusres.2021.06.022.
- Norman, G. (2010) 'Likert scales, levels of measurement and the "laws" of statistics', *Advances in Health Sciences Education*, 15(5), pp. 625–632. doi:10.1007/s10459-010-9222-y.
- Ramsey, J.B. (1969) 'Tests for specification errors in classical linear least-squares regression analysis', *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 31(2), pp. 350–371. doi:10.1111/j.2517-6161.1969.tb00796.x.
- Ryoo, Y., Bakpayev, M., Jeon, Y., Kim, K. and Yoon, S. (2025) 'High hopes, hard falls: Consumer expectations and reactions to ai-human collaboration in advertising', *International Journal of Advertising*, 45(1), pp. 48–80. doi:10.1080/02650487.2025.2458996.

- Sahoo, P., Singh, A., Saha, S., Jain, V., Mondal, S. and Chadha, A. (2024) *A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications* [Preprint]. Available at: <https://doi.org/10.48550/arXiv.2402.07927> (Accessed: 22 July 2026).
- Sands, S., Demsar, V., Ferraro, C., Wilson, S., Wheeler, M. and Campbell, C. (2025) 'Easing AI-advertising aversion: How leadership for the greater good buffers negative response to ai-generated ads', *International Journal of Advertising*, 45(1), pp. 27–47. doi:10.1080/02650487.2025.2457080.
- van Berlo, Z.M., Campbell, C. and Voorveld, H.A. (2024) 'The made framework: Best practices for creating effective experimental stimuli using generative AI', *Journal of Advertising*, 53(5), pp. 732–753. doi:10.1080/00913367.2024.2397777.
- Wang, Y. and Sun, S. (2010) 'Assessing beliefs, attitudes, and behavioral responses toward online advertising in three countries', *International Business Review*, 19(4), pp. 333–344. doi:10.1016/j.ibusrev.2010.01.004.
- Welch, B.L. (1947) 'The generalization of "student's" problem when several different population variances are involved', *Biometrika*, 34(1–2), pp. 28–35. doi:10.1093/biomet/34.1-2.28.
- White, H. (1980) 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *Econometrica*, 48(4), p. 817. doi:10.2307/1912934.
- Yoo, C.Y., Kim, K. and Stout, P.A. (2004) 'Assessing the effects of animation in online banner advertising', *Journal of Interactive Advertising*, 4(2), pp. 49–60. doi:10.1080/15252019.2004.10722087.
- Yoo, C.Y. (2011) 'Interplay of message framing, keyword insertion and levels of product involvement in click-through of Keyword Search ads', *International Journal of Advertising*, 30(3), pp. 399–424. doi:10.2501/ija-30-3-399-424.

Appendices

Stated Click-Through Intention Across Two FocusForm Advertisement Conditions: An Online Experiment Comparing AI-Generated and Human-Written Copy

The following pages consolidate the supporting experimental materials, survey documentation, data-preparation records and detailed statistical outputs referenced in the dissertation.

Overview. Appendix Contents

Appendix	Title
A	Experimental Advertisements
B	Creative Brief and Copy-Creation Materials
C	Full Survey Instrument and Qualtrics Flow
D	Recruitment Materials and Sensitivity Power Calculation
E	Ethics Documentation
F	Data Preparation, Variable Coding and Scoring
G	Additional Descriptive and Reliability Results
H	Detailed Statistical Outputs and Regression Diagnostics

Appendix A: Experimental Advertisements

A1. Final Advertisement Stimuli

Participants were randomly assigned to view one of the two static FocusForm advertisements shown below. The visual template and fixed “Learn More” call-to-action were held constant, while the headline, body copy and feature statements differed between conditions.

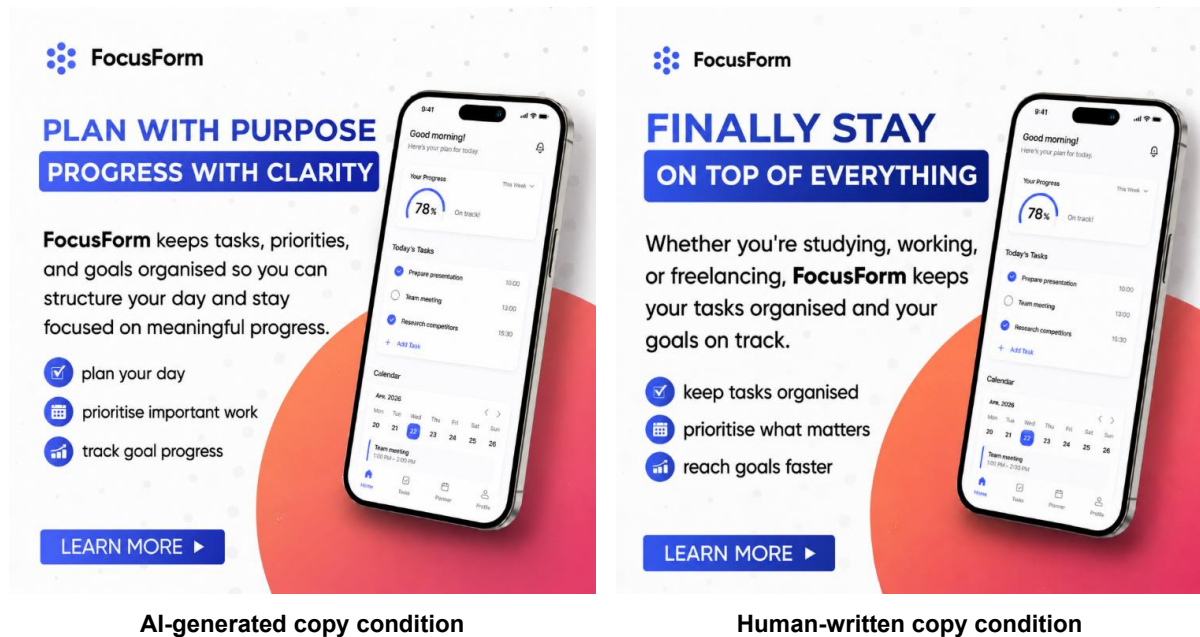


Figure A1. Final FocusForm advertisements used in the two experimental conditions.

Note. Both images were 1080 × 1080 pixels. Minor differences in text spacing resulted from the different copy lengths. The call-to-action was embedded in the image and was not clickable in the survey.

Appendix B: Creative Brief and Copy-Creation Materials

B1. Shared Creative Brief

Product: FocusForm is a fictional productivity app that helps users create and manage tasks, plan their day, prioritise important work, set clear objectives, track progress, and stay focused on what matters.

Target audience: Adult digital consumers, including students, professionals, freelancers, and people who want to be more organised and productive.

Brand tone: Clean, modern, minimal, professional, trustworthy, motivating, and easy to understand.

Advertisement context: The copy will appear in a square social-media-style advertisement for a productivity app. The advertisement already includes the FocusForm brand name, a phone mock-up, and a “Learn More” call-to-action button.

Required output:

1. A headline consisting of two short lines.
2. One concise body sentence of approximately 18–25 words.
3. Three short feature or benefit statements, each between 3 and 5 words.

Requirements:

- Focus on organisation, daily planning, priorities, goals, progress, or productivity.
- Keep the language concise and suitable for a modern digital advertisement.
- Do not mention artificial intelligence.
- Do not include emojis, hashtags, exclamation marks, or exaggerated claims.
- Do not include a call-to-action, because “Learn More” is already part of the advertisement.
- Provide exactly one final copy version and no explanation.

Required output format: Headline line 1; headline line 2; body copy; feature 1; feature 2; feature 3.

B2. Human-Written Copy

The external freelance designer supplied the following copy for the human-written condition:

Headline line 1: FINALLY STAY

Headline line 2: ON TOP OF EVERYTHING

Body copy: Whether you're studying, working, or freelancing, FocusForm keeps your tasks organised and your goals on track.

Feature 1: keep tasks organised

Feature 2: prioritise what matters

Feature 3: reach goals faster

Note. The fixed "Learn More" call-to-action was part of the advertisement template and was not treated as part of the human-written copy.

B3. Exact AI-Generation Prompt

Tool used: ChatGPT GPT-5.5 High Intelligence

Date of generation: 24 June 2026

Clarification. The phrase "independently from the product brief below" is reproduced exactly as used in the original prompt. It was intended to instruct the model to create its own copy version from the brief, without reference to the human-written copy; it was not intended to tell the model to ignore the brief.

Create one version of advertising copy for a fictional productivity app called FocusForm.

This copy will be used in an academic survey experiment. Generate the advertising copy independently from the product brief below. Do not refer to AI, human-written copy, the experiment, or any comparison condition.

Creative brief:

Product:

FocusForm is a fictional productivity app that helps users create and manage tasks, plan their day, prioritise important work, set clear objectives, track progress, and stay focused on what matters.

Target audience:

Adult digital consumers, including students, professionals, freelancers, and people who want to be more organised and productive.

Brand tone:

Clean, modern, minimal, professional, trustworthy, motivating, and easy to understand.

Ad context:

The copy will appear in a square social-media-style advertisement for a productivity app. The ad already includes the FocusForm brand name, a phone mock-up, and a "Learn More" call-to-action button.

Required output:

1. A headline consisting of two short lines.
2. One concise body sentence of approximately 18-25 words.
3. Three short feature or benefit statements, each between 3 and 5 words.

Requirements:

- * Focus on organisation, daily planning, priorities, goals, progress, or productivity.
- * Keep the language concise and suitable for a modern digital advertisement.
- * Do not mention artificial intelligence.
- * Do not include emojis, hashtags, exclamation marks, or exaggerated claims.
- * Do not include a call-to-action, because "Learn More" is already part of the advert.
- * Provide exactly one final copy version and no explanation.

Output format:

Headline line 1:

Headline line 2:

Body copy:

Feature 1:

Feature 2:

Feature 3:

B4. Full ChatGPT Output

Headline line 1: Plan with purpose

Headline line 2: Progress with clarity

Body copy: FocusForm keeps tasks, priorities, and goals organised so you can structure your day and stay focused on meaningful progress.

Feature 1: Plan your day

Feature 2: Prioritise important work

Feature 3: Track goal progress

B5. Final AI-Generated Copy

Headline line 1: PLAN WITH PURPOSE

Headline line 2: PROGRESS WITH CLARITY

Body copy: FocusForm keeps tasks, priorities, and goals organised so you can structure your day and stay focused on meaningful progress.

Feature 1: plan your day

Feature 2: prioritise important work

Feature 3: track goal progress

Note. The generated wording was used unchanged. Capitalisation reflects the visual styling applied within the advertisement template.

Appendix C: Full Survey Instrument and Qualtrics Flow

The following screenshots reproduce the participant-facing survey materials and the Qualtrics flow used to administer the experiment. Variable names shown in the survey builder are retained to support reproducibility.

C1. Participant Information Sheet

Consumer Responses to Digital Advertising

Start of Block: Eligibility_and_Consent

information **Participant Information** **Study title** Responses to Digital Advertising You are invited to take part in a research study conducted by Alexander Bade as part of an MSc dissertation at Warwick Business School. **What is the purpose of the study?** This study examines how people respond to digital advertising. You will be shown one mock digital advertisement for a fictional productivity app and asked to answer a short set of questions about your response to it. The survey should take approximately 4–6 minutes to complete. **What will taking part involve?** If you decide to take part, you will: - view one static digital advertisement; - answer questions about the advertisement and the advertised app; and - answer a small number of general background questions. The “Learn More” button shown in the advertisement is part of the image and is not interactive. No actual clicking behaviour will be recorded. **Do I have to take part?** No. Taking part is entirely voluntary. You may stop completing the survey at any time before submitting your responses by closing the survey. You do not need to give a reason. Because the survey is anonymous, it will not be possible to withdraw your individual response after you have submitted it. **How will my information be used?** The information collected will be used for an MSc dissertation at Warwick Business School. No names, email addresses, or other directly identifying information will be requested in this survey. Responses will be analysed anonymously and reported only in aggregate form. Survey data will be stored and handled in accordance with the approved study data-handling procedures. **Why is some information not explained in full before the survey?** To avoid influencing participants’ responses, some aspects of the study’s precise purpose cannot be fully explained before participation. A full explanation will be provided in the debrief at the end of the survey. **Who can I contact?** For questions about the study, please contact: Alexander Bade MSc Business Analytics Warwick Business School Alexander.Bade@warwick.ac.uk This study has received ethical approval from the Warwick Business School Programme Quality Team, University of Warwick. Ethics reference: R_236XsTp5oUFzJAJ If you have concerns about the conduct of the research, please contact Seyoung Seol. Email: Seyoung.Seol@warwick.ac.uk Please read this information carefully before deciding whether to take part.

Figure C1. Participant Information Sheet shown before eligibility and consent.

C2. Eligibility and Consent

age_eligibility Are you aged 18 years or over?

☐ Yes (1)

☐ No (0)



consent Having read the Participant Information above, do you consent to take part in this study?

☐ Yes, I consent to take part (1)

☐ No, I do not consent (0)

End of Block: Eligibility_and_Consent

Start of Block: Stimulus_Human

Image_Human Please take a moment to view the advertisement below. The “Learn More” button is part of the advertisement image and is not clickable in this survey.

End of Block: Stimulus_Human

Start of Block: Stimulus_AI

Image_AI Please take a moment to view the advertisement below. The “Learn More” button is part of the advertisement image and is not clickable in this survey.

End of Block: Stimulus_AI

Start of Block: Shared_Measures

Figure C2. Age-eligibility and informed-consent questions.

C3. Click-Through Intention Measure

cti Please indicate how much you agree or disagree with the following statements about the FocusForm advertisement you just viewed.

	Strongly disagree (1)	Disagree (2)	Somehwat disagree (3)	Neither agree nor disagree (4)	Somehwat agree (5)	Agree (6)	Strongly agree (7)
I would be likely to select the "Learn More" button in this advertisement. (1)	☐	☐	☐	☐	☐	☐	☐
I would consider selecting "Learn More" to find out more about FocusForm. (2)	☐	☐	☐	☐	☐	☐	☐
If I saw this advertisement online, I would intend to select "Learn More". (3)	☐	☐	☐	☐	☐	☐	☐

Figure C3. Three-item stated click-through intention measure.

C4. Perceived Authenticity Measure

auth Please indicate how much you agree or disagree with the following statements about the FocusForm advertisement you just viewed.

	Strongly disagree (1)	Disagree (2)	Somehwat disagree (3)	Neither agree nor disagree (4)	Somehwat agree (5)	Agree (6)	Strongly agree (7)
The FocusForm advertisement feels authentic. (auth_1)	☐	☐	☐	☐	☐	☐	☐
The FocusForm advertisement comes across as genuine. (auth_2)	☐	☐	☐	☐	☐	☐	☐
The message in this advertisement feels sincere. (auth_3)	☐	☐	☐	☐	☐	☐	☐

Figure C4. Three-item perceived authenticity measure.

C5. Perceived Relevance Measure

rel Please indicate how much you agree or disagree with the following statements about the FocusForm advertisement you just viewed.

	Strongly disagree (1)	Disagree (2)	Somehwat disagree (3)	Neither agree nor disagree (4)	Somehwat agree (5)	Agree (6)	Strongly agree (7)
The FocusForm advertisement is relevant to me. (1)	☐	☐	☐	☐	☐	☐	☐
The FocusForm app appears relevant to my needs. (2)	☐	☐	☐	☐	☐	☐	☐
The message in this advertisement relates to things that matter to me. (3)	☐	☐	☐	☐	☐	☐	☐

Figure C5. Three-item perceived relevance measure.

C6. Trust Measure

trust Please indicate how much you agree or disagree with the following statements about the FocusForm advertisement you just viewed.

	Strongly disagree (1)	Disagree (2)	Somehwat disagree (3)	Neither agree nor disagree (4)	Somehwat agree (5)	Agree (6)	Strongly agree (7)
Based on this advertisement, FocusForm seems trustworthy. (1)	☐	☐	☐	☐	☐	☐	☐
Based on this advertisement, I would trust FocusForm to deliver what it promises. (2)	☐	☐	☐	☐	☐	☐	☐
The claims made about FocusForm in this advertisement seem believable. (3)	☐	☐	☐	☐	☐	☐	☐

Figure C6. Three-item trust measure.

C7. Attention Check and Demographic Questions

attn_check Please indicate how much you agree or disagree with the following statement.

	Strongly disagree (1)	Disagree (2)	Somehwat disagree (3)	Neither agree nor disagree (4)	Somehwat agree (5)	Agree (6)	Strongly agree (7)
To show that you are paying attention, please select “Disagree” for this statement. (attn_check)	☐	☐	☐	☐	☐	☐	☐

age What is your age in years?

gender Which of the following best describes your gender?

- ☐ Male (1)
- ☐ Female (2)
- ☐ Non-binary / third gender (3)
- ☐ Prefer not to say (4)

Figure C7. Instructional attention check and demographic questions.

C8. Prior Productivity-App Use

prior_app_use Have you used a task-management or productivity app in the past 12 months?

- ☐ Yes (1)
- ☐ No (0)
- ☐ Not sure (2)
- ☐ Prefer not to say (3)

End of Block: Shared_Measures

Figure C8. Prior productivity-app-use question.

C9. Debriefing Statement

Thank you for taking part.

This study compares responses to two versions of a static digital advertisement for the fictional productivity app FocusForm.

Participants were randomly assigned to view one of two advertisements that used the same brand, visual template, phone mock-up, colours, icons, layout and “Learn More” button. The headline, body copy and feature statements differed between conditions, and minor spacing differences resulted from the different copy lengths. One version used human-written copy and the other used AI-generated copy. This information was not provided before the survey because knowing the source of the copy may have influenced participants’ responses.

The “Learn More” button was not interactive, and no actual clicking behaviour was recorded. The study measured participants’ reported intention to select the button, rather than actual clicking behaviour.

If you have any questions about this study, please contact:

Alexander Bade

MSc Business Analytics Student

Warwick Business School, University of Warwick

Email: Alexander.Bade@warwick.ac.uk

You may withdraw from the study at any time before submitting your responses by closing the survey. As the survey is anonymous and no identifying information will be collected, it will not be possible to identify and remove your individual response after submission.

This study has received ethical approval from the Warwick Business School Programme Quality Team, University of Warwick.

Ethics reference: R_236XsTp5oUFzJAJ

If you have concerns about the conduct of the research, please contact Seoyung Seol.

Email: Seoyung.Seol@warwick.ac.uk

Thank you again for your participation.

Figure C9. Debriefing statement shown after survey completion.

Note. The debriefing statement is reproduced as text rather than as a screenshot to improve readability. The wording is unchanged. The survey was configured to collect anonymous responses, and the exported dataset did not contain names, email addresses, IP addresses or location fields.

C10. Qualtrics Survey Flow and Randomisation

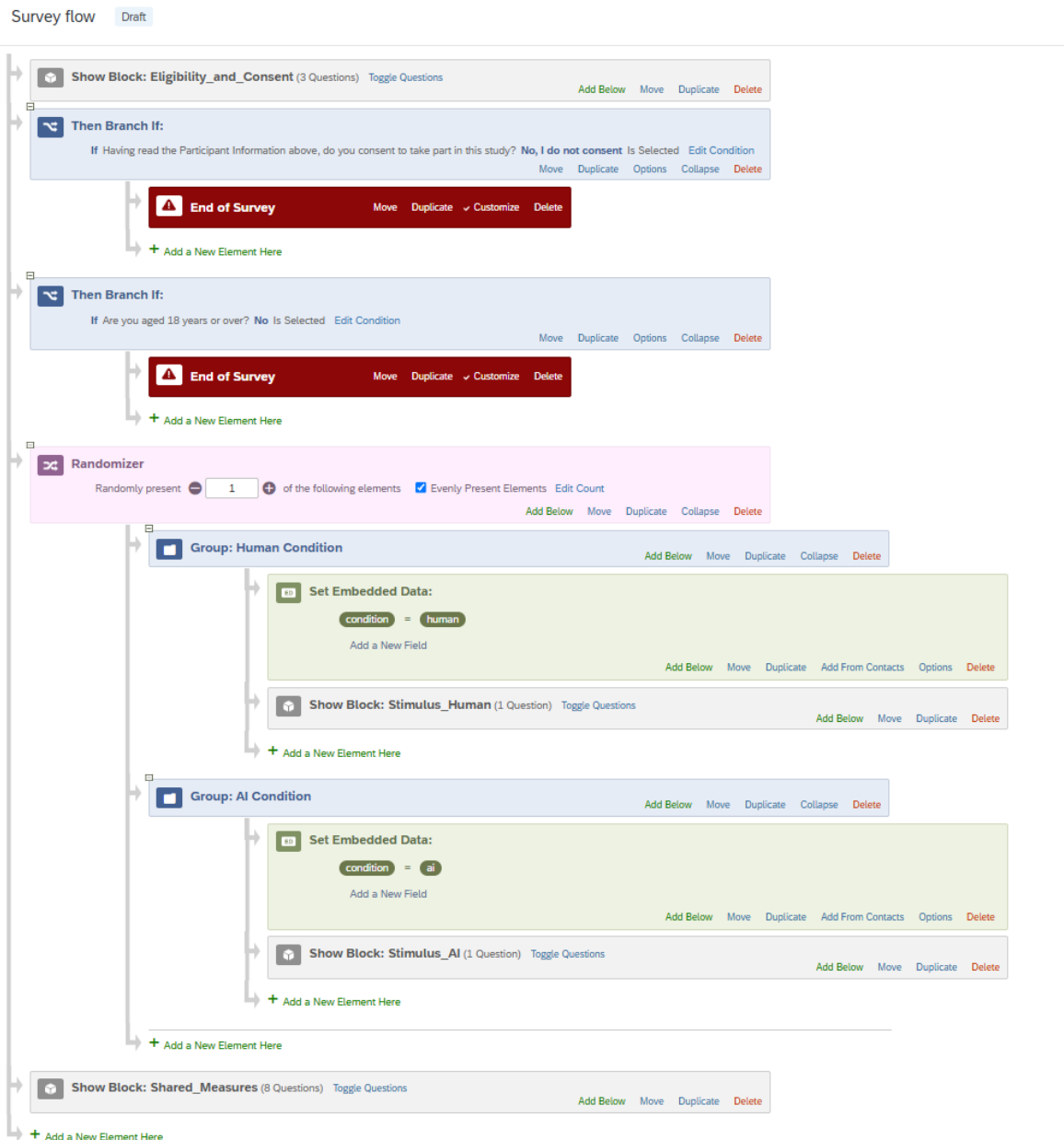


Figure C10. Qualtrics survey flow, condition assignment and embedded-data settings.

Note. The “Learn More” button shown in each stimulus was part of the static image and was not interactive. No AI-authorship attribution question was included because the study focused on responses under non-disclosure and perceived authorship was outside the pre-specified hypotheses. An attribution question placed before the outcome measures could also have revealed the purpose of the study and influenced those responses.

Appendix D: Recruitment Materials and Sensitivity Power Calculation

D1. Recruitment Materials

Participants were recruited through Instagram Stories and WhatsApp. The examples below show the main recruitment messages used during data collection.

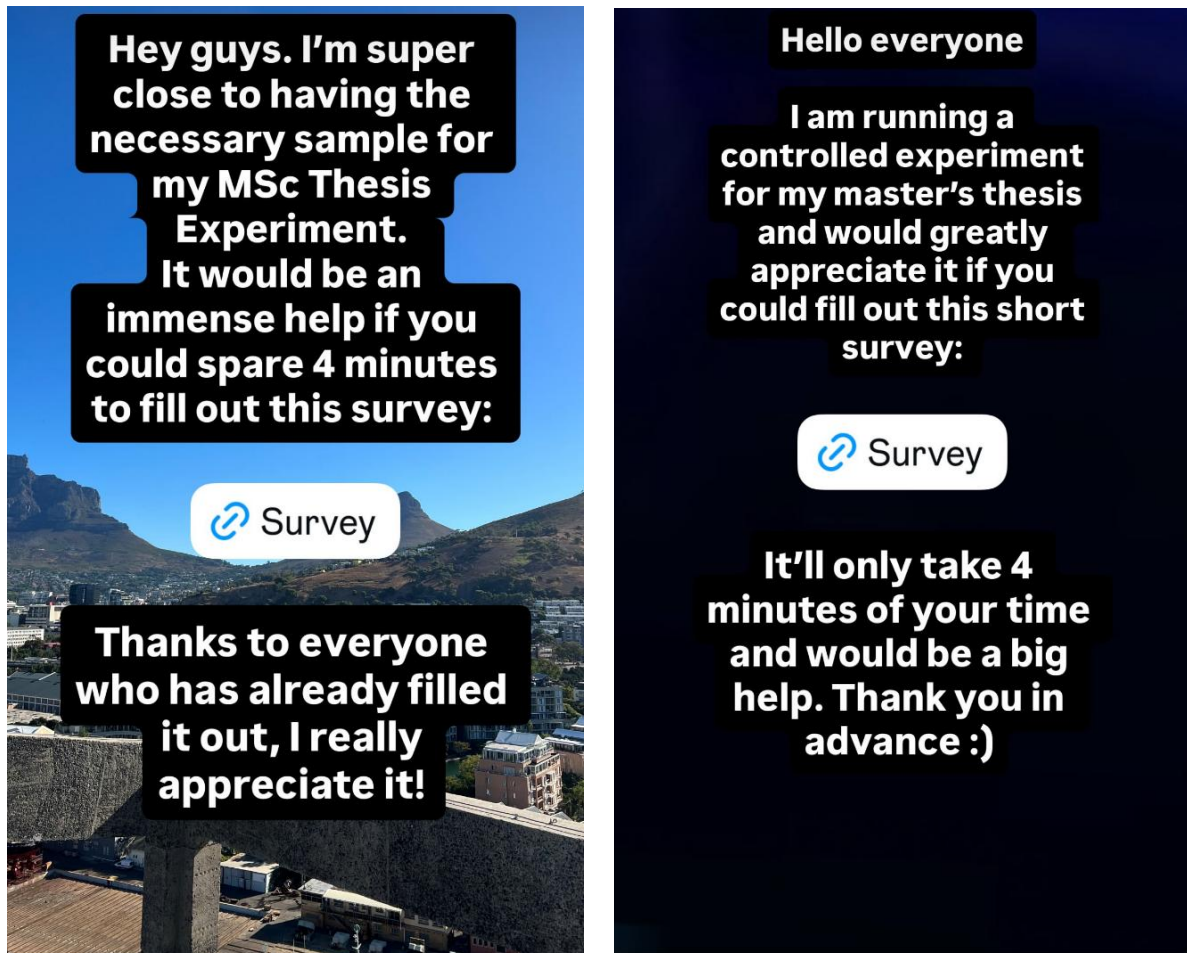


Figure D1. Instagram Story recruitment materials used during data collection.

Note. Participation was voluntary and no incentive was offered. Private contact names and individual WhatsApp conversations are not reproduced.

D2. Sensitivity Power Calculation

The calculation below shows that a balanced two-group comparison with 30 participants per condition, $\alpha = .05$ and 80% power could detect an effect of approximately Cohen's $d = 0.74$ or larger.

```
from statsmodels.stats.power import TTestIndPower

# Create power-analysis object
analysis = TTestIndPower()

# Assumptions
n_per_group = 30      # participants in each condition
alpha = 0.05          # significance level
```

```

power = 0.80          # desired statistical power
ratio = 1             # equal group sizes
alternative = "two-sided"

# Calculate the minimum detectable standardised effect size
detectable_effect_size = analysis.solve_power(
    effect_size=None,
    nobs1=n_per_group,
    alpha=alpha,
    power=power,
    ratio=ratio,
    alternative=alternative
)

print(f"Sample per group: {n_per_group}")
print(f"Total sample: {n_per_group * 2}")
print(f"Minimum detectable effect size: Cohen's d = {detectable_effect_size:.3f}")

```

Output	Value
Sample per group	30
Total sample	60
Minimum detectable effect size	Cohen's d = 0.736

Table D1. Output from the sensitivity power calculation.

Note. With 30 participants per condition, the achieved sample had 80% power at $\alpha = .05$ to detect a standardised mean difference of approximately Cohen's $d = 0.74$ or larger. Smaller effects may not have been detectable.

Appendix E: Ethics Documentation

E1. Ethics Approval Confirmation

The study received ethical approval from the Warwick Business School Programme Quality Team before data collection began. The approval reference was R_236XsTp5oUFzJAJ.

Dear Alexander (and cc Seyoung Seol)

Course: MSc Business Analytics (MSBA)

Your Research Ethics Form has been reviewed by your supervisor Seyoung Seol (included on this email for information). Your Programme Team has also received this email for information.

The outcome of this review is: Research Ethics Approval granted with conditions

Conditions for approval provided by supervisor: Your study includes an intervention in a lab experiment (this was not ticked correctly earlier). Approval is granted as it does not change the risk of this study.

Your Research Ethics Form is approved, subject to the implementation of the conditions and/or revisions stated by your supervisor. If you are unsure about these conditions, please consult your supervisor. You do not need to resubmit your Research Ethics Form.

Summary of responses: [Click to download PDF of responses](#)

You must now upload this PDF to my.wbs as proof of your Research Ethics Approval. By uploading this confirmation of Research Ethics Approval with conditions to my.wbs, you are accepting **responsibility for the implementation of the stated conditions**, with support from your supervisor.

Remember, you must also include this confirmation of approval as an appendix as part of your final submission. If your research methodology changes since receiving this approval, you will need to submit a new Research Ethics Form to your supervisor. [Click here to revise and resubmit your original REF](#). This link will only work for minor revisions to questions you originally answered. If you need to start over and submit a new REF, please contact the Programme Quality Team at QualityPG@wbs.ac.uk

If you have any questions at all, please contact your Programme Team, or your supervisor.

Wishing you all the best with your research,

Programme Quality Team

Research Ethics Approval reference number: R_236XsTp5oUFzJAJ

Figure E1. Ethics approval confirmation.

Appendix F: Data Preparation, Variable Coding and Scoring

F1. Data-Cleaning Summary

Cleaning step	Decision
Qualtrics metadata	Two non-participant metadata rows removed
Eligibility	One under-18 and one non-consenting response removed
Completion	Twenty incomplete responses removed
Attention check	Eleven failures removed
Missing data	No imputation performed
Survey duration	Inspected, but no duration-based exclusions applied
Duplicate responses	No duplicate Response IDs identified
Final sample	60 participants, 30 per condition

Table F1. Data-cleaning decisions and final analytical sample.

Note. Exclusions were applied sequentially. No response was excluded solely because of survey duration or substantive response values.

F2. Variable Dictionary

Variable name	Variable type	Description	Coding or scale	Use in data preparation or analysis
ResponseId	Administrative identifier	Qualtrics-generated identifier assigned to each survey response.	Unique text value	Used only to check for duplicated Response IDs. It was not included as a predictor, outcome or reported diagnostic identifier.
Progress	Data-quality variable	Percentage of the survey completed by the respondent.	Integer from 0 to 100	A response was treated as complete only when Progress = 100 and Finished = 1.
Duration (in seconds)	Data-quality variable	Total recorded survey-completion time.	Integer measured in seconds	Inspected as a data-quality indicator. No duration-based exclusion was applied because no threshold had been specified in advance.
Finished	Data-quality variable	Qualtrics indicator of whether the survey response was completed.	0 = not finished; 1 = finished	A response was treated as complete only when Finished = 1 and Progress = 100.

Variable name	Variable type	Description	Coding or scale	Use in data preparation or analysis
age_eligibility	Eligibility variable	Indicates whether the participant confirmed that they were aged 18 years or over.	1 = aged 18 or over; responses not coded 1 were not retained	Used as the first participant-level exclusion criterion. One age-ineligible response was removed.
consent	Eligibility variable	Indicates whether the participant provided informed consent.	1 = consent provided; responses not coded 1 were not retained	Used as the second participant-level exclusion criterion. One non-consenting response was removed.
condition	Experimental variable	Advertisement-copy condition assigned through the Qualtrics randomiser.	human = human-written copy; ai = AI-generated copy	Used as the grouping variable for the condition-level descriptive statistics and Welch t-tests. Also used to create condition_code.
cti_1	Construct item	"I would be likely to select the 'Learn More' button in this advertisement."	Seven-point Likert scale: 1 = Strongly disagree to 7 = Strongly agree	First item included in the cti composite.
cti_2	Construct item	"I would consider selecting 'Learn More' to find out more about FocusForm."	Seven-point Likert scale: 1 = Strongly disagree to 7 = Strongly agree	Second item included in the cti composite.
cti_3	Construct item	"If I saw this advertisement online, I would intend to select 'Learn More'."	Seven-point Likert scale: 1 = Strongly disagree to 7 = Strongly agree	Third item included in the cti composite.
auth_1	Construct item	"The FocusForm advertisement feels authentic."	Seven-point Likert scale: 1 = Strongly disagree to 7 = Strongly agree	First item included in the auth composite.
auth_2	Construct item	"The FocusForm advertisement comes across as genuine."	Seven-point Likert scale: 1 = Strongly disagree to 7 = Strongly agree	Second item included in the auth composite.
auth_3	Construct item	"The message in this advertisement feels sincere."	Seven-point Likert scale: 1 = Strongly disagree to 7 = Strongly agree	Third item included in the auth composite.
rel_1	Construct item	"The FocusForm advertisement is relevant to me."	Seven-point Likert scale: 1 = Strongly disagree to 7 = Strongly agree	First item included in the rel composite.

Variable name	Variable type	Description	Coding or scale	Use in data preparation or analysis
rel_2	Construct item	"The FocusForm app appears relevant to my needs."	Seven-point Likert scale: 1 = Strongly disagree to 7 = Strongly agree	Second item included in the rel composite.
rel_3	Construct item	"The message in this advertisement relates to things that matter to me."	Seven-point Likert scale: 1 = Strongly disagree to 7 = Strongly agree	Third item included in the rel composite.
trust_1	Construct item	"Based on this advertisement, FocusForm seems trustworthy."	Seven-point Likert scale: 1 = Strongly disagree to 7 = Strongly agree	First item included in the trust composite.
trust_2	Construct item	"Based on this advertisement, I would trust FocusForm to deliver what it promises."	Seven-point Likert scale: 1 = Strongly disagree to 7 = Strongly agree	Second item included in the trust composite.
trust_3	Construct item	"The claims made about FocusForm in this advertisement seem believable."	Seven-point Likert scale: 1 = Strongly disagree to 7 = Strongly agree	Third item included in the trust composite.
attn_check	Data-quality variable	Instructional attention check stating: "To show that you are paying attention, please select 'Disagree' for this statement."	Seven-point Likert scale; correct response: 2 = Disagree	Responses were excluded when attn_check was not equal to 2. Eleven attention-check failures were removed.
age	Demographic variable	Participant's age in completed years after data cleaning.	Integer measured in years	Used to describe the sample and assess descriptive comparability between conditions. One written response, "Twenty four years", was converted to 24.
gender	Demographic variable	Numeric version of the participant's self-reported gender.	1 = Male; 2 = Female; 3 = Non-binary / third gender; 4 = Prefer not to say	Used to create gender_label for the demographic summaries and descriptive randomisation check.
prior_app_use	Background variable	Indicates whether the participant had used a productivity or task-management app during the previous 12 months.	0 = No; 1 = Yes; 2 = Not sure; 3 = Prefer not to say	Used to create prior_app_use_label for the demographic summaries and descriptive randomisation check.
age_raw	Data-cleaning helper variable	Copy of the original age response before text standardisation and numeric conversion.	Original text value from the survey export	Retained as an audit variable to document the correction of the written age response. Not used in the statistical analysis.

Variable name	Variable type	Description	Coding or scale	Use in data preparation or analysis
prior_app_use_label	Descriptive helper variable	Text-labelled version of prior_app_use.	No / Yes / Not sure / Prefer not to say	Used to produce readable descriptive tables for prior productivity-app use.
gender_label	Descriptive helper variable	Text-labelled version of gender.	Male / Female / Non-binary or third gender / Prefer not to say	Used to produce readable gender-frequency tables.
cti	Derived composite	Participant-level mean of cti_1, cti_2 and cti_3.	Continuous composite score from 1 to 7	Primary dependent variable. Used for descriptive statistics, H1, correlations and the multiple regression analysis.
auth	Derived composite	Participant-level mean of auth_1, auth_2 and auth_3.	Continuous composite score from 1 to 7	Used for descriptive statistics, correlations and H2.
rel	Derived composite	Participant-level mean of rel_1, rel_2 and rel_3.	Continuous composite score from 1 to 7	Used for descriptive statistics, correlations and as the relevance predictor for H3.
trust	Derived composite	Participant-level mean of trust_1, trust_2 and trust_3.	Continuous composite score from 1 to 7	Used for descriptive statistics, correlations and as the trust predictor for H4.
condition_code	Derived experimental variable	Numeric version of the experimental condition created for the regression analysis.	0 = human-written copy; 1 = AI-generated copy	Included as the experimental-condition indicator in the multiple regression model predicting cti. This allows the relevance and trust coefficients to be estimated conditional on condition.

Table F2. Variable dictionary for the analysis-ready dataset.

Note. All substantive construct items were positively worded and measured on a seven-point Likert scale ranging from 1 (“Strongly disagree”) to 7 (“Strongly agree”). No reverse coding was required. Each composite was calculated as the arithmetic mean of its three corresponding items. The final analytical sample contained no missing values for the 12 substantive items, and no imputation was performed. The text variable condition retained the categories human and ai, while condition_code was created separately for the regression analysis.

F3. Composite Scoring

The four study constructs were each measured using three positively worded items. Composite scores were calculated only where responses were available for all three intended items. Higher values represented higher stated click-through intention, perceived authenticity, perceived relevance or trust. Each composite remained on the original 1–7 scale.

Construct	Constituent items	Composite variable	Calculation	Interpretation
Stated click-through intention	cti_1, cti_2, cti_3	cti	$(cti_1 + cti_2 + cti_3) / 3$	Higher values indicate stronger stated intention to select the “Learn More” button.
Perceived authenticity	auth_1, auth_2, auth_3	auth	$(auth_1 + auth_2 + auth_3) / 3$	Higher values indicate that the advertisement was perceived as more authentic, genuine and sincere.
Perceived relevance	rel_1, rel_2, rel_3	rel	$(rel_1 + rel_2 + rel_3) / 3$	Higher values indicate that the advertisement and FocusForm appeared more relevant to the participant.
Trust	trust_1, trust_2, trust_3	trust	$(trust_1 + trust_2 + trust_3) / 3$	Higher values indicate greater trust in FocusForm and the claims presented in the advertisement.

Table F3. Composite-score calculations and interpretation.

Note. No reverse coding or missing-value imputation was required for the final analytical sample.

F4. Data-Cleaning and Analysis Code

The extracts below reproduce the principal Python code used to import and clean the Qualtrics data, calculate the composite measures, test the hypotheses and assess model assumptions. Display-only and file-export commands that do not affect the analysis are omitted. The full notebook is retained by the researcher as the complete working record; it is not reproduced in full in this dissertation.

F4.1. Import the Qualtrics Export and Remove Metadata Rows

```
# Load the anonymous Qualtrics re-export

csv_path = Path(
    "Consumer Responses to Digital Advertising_July 24, 2026_14.50.csv"
)

if not csv_path.exists():
    raise FileNotFoundError(
```

```

        f"The Qualtrics CSV was not found: {csv_path}"
    )

df_reexport = pd.read_csv(
    csv_path,
    dtype="string",
    encoding="utf-8-sig"
)

# Confirm that identifying, IP-address and location columns are absent

identifier_columns = {
    "IPAddress", "LocationLatitude", "LocationLongitude",
    "RecipientLastName", "RecipientFirstName",
    "RecipientEmail", "ExternalReference"
}

assert identifier_columns.isdisjoint(df_reexport.columns)

# Retain the responses present in the original analysis export

analysis_export_cutoff = pd.Timestamp("2026-07-15 10:14:00")
recorded_at = pd.to_datetime(
    df_reexport["RecordedDate"],
    format="%Y-%m-%d %H:%M:%S",
    errors="coerce"
)
later_sessions = recorded_at > analysis_export_cutoff

assert df_reexport.loc[
    later_sessions,
    "Finished"
].eq("0").all()

df_raw = (
    df_reexport.loc[~later_sessions]
    .copy()
    .reset_index(drop=True)
)

print(f"Source file: {csv_path}")
print(f"Full re-export shape: {df_reexport.shape}")
print(f"Analysis-export shape: {df_raw.shape}")
print(
    "Later-recorded incomplete sessions excluded:",
    int(later_sessions.sum())
)

```

```

display(
    df_raw[
        ["Progress", "Finished", "condition"]
    ].head(3)
)

# Identify Qualtrics metadata rows
question_text_row = df_raw["StartDate"].eq("Start Date")

import_id_row = df_raw.apply(
    lambda column: column.str.contains(
        '"ImportId":',
        regex=False,
        na=False
    )
).any(axis=1)

metadata_mask = question_text_row | import_id_row

print("Metadata rows identified:", int(metadata_mask.sum()))
display(
    df_raw.loc[
        metadata_mask,
        ["StartDate", "Finished", "cti_1", "attn_check", "condition"]
    ]
)

assert metadata_mask.sum() == 2, (
    "Expected exactly two Qualtrics metadata rows. Inspect the export."
)

df = (
    df_raw.loc[~metadata_mask]
    .copy()
    .reset_index(drop=True)
)

# Retain only variables required for cleaning, descriptives and analysis
columns_to_keep = [
    "ResponseId",
    "Progress",
    "Duration (in seconds)",
    "Finished",
    "age_eligibility",

```

```

    "consent",
    "condition",
    "cti_1", "cti_2", "cti_3",
    "auth_1", "auth_2", "auth_3",
    "rel_1", "rel_2", "rel_3",
    "trust_1", "trust_2", "trust_3",
    "attn_check",
    "age",
    "gender",
    "prior_app_use"
]

missing_columns = sorted(set(columns_to_keep) - set(df.columns))

if missing_columns:
    raise KeyError(f"Required columns are missing: {missing_columns}")

df = df[columns_to_keep].copy()

# Remove leading/trailing whitespace from text values
for column in df.select_dtypes(include=["string", "object"]).columns:
    df[column] = df[column].str.strip()

```

F4.2. Convert Variables and Create Demographic Labels

```

# Convert integer-coded variables

numeric_columns = [
    "Progress",
    "Duration (in seconds)",
    "Finished",
    "age_eligibility",
    "consent",
    "cti_1", "cti_2", "cti_3",
    "auth_1", "auth_2", "auth_3",
    "rel_1", "rel_2", "rel_3",
    "trust_1", "trust_2", "trust_3",
    "attn_check",
    "gender",
    "prior_app_use"
]

for column in numeric_columns:
    df[column] = pd.to_numeric(
        df[column],
        errors="coerce"
    ).astype("Int64")

```

```

# Correct and convert age
df["age_raw"] = df["age"].copy()

df["age"] = (
    df["age"]
    .str.lower()
    .str.replace(r"\s+", " ", regex=True)
    .str.strip()
    .replace({
        "twenty four years": "24",
        "twenty-four years": "24",
        "twenty four": "24",
        "twenty-four": "24"
    })
)

df["age"] = pd.to_numeric(
    df["age"],
    errors="coerce"
).astype("Int64")

# Standardise the experimental-condition labels
df["condition"] = (
    df["condition"]
    .str.lower()
    .str.strip()
)

unexpected_conditions = set(df["condition"].dropna().unique()) - {"ai", "human"}

if unexpected_conditions:
    raise ValueError(
        f"Unexpected condition labels found: {unexpected_conditions}"
    )

df["condition"] = pd.Categorical(
    df["condition"],
    categories=["human", "ai"]
)

PRIOR_APP_USE_MAP = {
    0: "No",
    1: "Yes",
    2: "Not sure",
    3: "Prefer not to say"
}

```

```
# Gender labels confirmed from the Qualtrics response coding
GENDER_MAP = {
    1: 'Male',
    2: 'Female',
    3: 'Non-binary / third gender',
    4: 'Prefer not to say'
}

df["prior_app_use_label"] = (
    df["prior_app_use"]
    .map(PRIOR_APP_USE_MAP)
    .astype("string")
)

df["gender_label"] = (
    df["gender"]
    .map(GENDER_MAP)
    .fillna(
        df["gender"].map(
            lambda value: f"Code {int(value)}" if pd.notna(value) else pd.NA
        )
    )
    .astype("string")
)
```

F4.3. Apply Exclusion Rules and Data-Quality Checks

```
# Sequential cleaning stages

stage_metadata_removed = df.copy()

stage_age_eligible = stage_metadata_removed.loc[
    stage_metadata_removed["age_eligibility"].eq(1)
].copy()

stage_consented = stage_age_eligible.loc[
    stage_age_eligible["consent"].eq(1)
].copy()

stage_complete = stage_consented.loc[
    stage_consented["Finished"].eq(1)
    & stage_consented["Progress"].eq(100)
].copy()

df_analysis = stage_complete.loc[
    stage_complete["attn_check"].eq(2)
]
```

```

].copy().reset_index(drop=True)

cleaning_flow = pd.DataFrame({
    "Stage": [
        "Raw CSV rows",
        "Qualtrics metadata rows removed",
        "Under-18 / age-ineligible responses removed",
        "Non-consenting responses removed",
        "Incomplete responses removed",
        "Attention-check failures removed",
        "Final analytical sample"
    ],
    "Excluded at stage": [
        0,
        len(df_raw) - len(stage_metadata_removed),
        len(stage_metadata_removed) - len(stage_age_eligible),
        len(stage_age_eligible) - len(stage_consented),
        len(stage_consented) - len(stage_complete),
        len(stage_complete) - len(df_analysis),
        0
    ],
    "Remaining": [
        len(df_raw),
        len(stage_metadata_removed),
        len(stage_age_eligible),
        len(stage_consented),
        len(stage_complete),
        len(df_analysis),
        len(df_analysis)
    ]
})

display(cleaning_flow)

assert len(df_analysis) == 60, (
    f"Expected 60 final observations, found {len(df_analysis)}."
)

assert df_analysis["condition"].value_counts().to_dict() == {
    "human": 30,
    "ai": 30
}, "Expected a 30/30 condition split."

# Additional data-quality checks

quality_summary = pd.DataFrame({
    "Check": [

```

```

        "Final observations",
        "AI condition",
        "Human condition",
        "Duplicate ResponseIds",
        "Missing construct-item values",
        "Minimum duration (seconds)",
        "Median duration (seconds)",
        "Maximum duration (seconds)"
    ],
    "Value": [
        len(df_analysis),
        int((df_analysis["condition"] == "ai").sum()),
        int((df_analysis["condition"] == "human").sum()),
        int(df_analysis["ResponseId"].duplicated().sum()),
        int(
            df_analysis[
                [
                    "cti_1", "cti_2", "cti_3",
                    "auth_1", "auth_2", "auth_3",
                    "rel_1", "rel_2", "rel_3",
                    "trust_1", "trust_2", "trust_3"
                ]
            ].isna().sum().sum()
        ),
        int(df_analysis["Duration (in seconds)"].min()),
        float(df_analysis["Duration (in seconds)"].median()),
        int(df_analysis["Duration (in seconds)"].max())
    ]
})

display(quality_summary)

print(
    "No duration exclusion was applied. The maximum duration is retained "
    "because a long duration alone does not demonstrate an invalid response."
)

```

F4.4. Construct Composite Measures and Assess Reliability

```

construct_items = {
    "cti": ["cti_1", "cti_2", "cti_3"],
    "auth": ["auth_1", "auth_2", "auth_3"],
    "rel": ["rel_1", "rel_2", "rel_3"],
    "trust": ["trust_1", "trust_2", "trust_3"]
}

all_items = [

```

```

    item
    for items in construct_items.values()
    for item in items
]

# Confirm all item responses use the intended 1-7 scale
item_range_check = pd.DataFrame({
    "Minimum": df_analysis[all_items].min(),
    "Maximum": df_analysis[all_items].max(),
    "Missing": df_analysis[all_items].isna().sum()
})

display(item_range_check)

assert (
    df_analysis[all_items]
    .apply(lambda column: column.dropna().between(1, 7).all())
    .all()
), "At least one construct item falls outside the expected 1-7 scale."

for construct, items in construct_items.items():
    df_analysis[construct] = (
        df_analysis[items]
        .mean(axis=1, skipna=False)
    )

assert (
    df_analysis[list(construct_items)]
    .apply(lambda column: column.dropna().between(1, 7).all())
    .all()
), "At least one composite score falls outside the 1-7 scale."

display(
    df_analysis[
        ["cti_1", "cti_2", "cti_3", "cti"]
    ].head()
)

def cronbach_alpha(items: pd.DataFrame) -> float:
    """Calculate Cronbach's alpha using complete rows and sample variances."""
    complete_items = items.dropna(axis=0, how="any")
    k = complete_items.shape[1]

    if k < 2:
        raise ValueError("Cronbach's alpha requires at least two items.")

    item_variances = complete_items.var(axis=0, ddof=1)

```

```

total_score = complete_items.sum(axis=1)
total_variance = total_score.var(ddof=1)

if total_variance == 0:
    return np.nan

return (k / (k - 1)) * (
    1 - item_variances.sum() / total_variance
)

construct_labels = {
    "cti": "Click-through intention",
    "auth": "Perceived authenticity",
    "rel": "Perceived relevance",
    "trust": "Trust"
}

reliability_rows = []

for construct, items in construct_items.items():
    reliability_rows.append({
        "Construct": construct_labels[construct],
        "Items": len(items),
        "n": df_analysis[items].dropna().shape[0],
        "Cronbach's alpha": cronbach_alpha(df_analysis[items])
    })

reliability_table = pd.DataFrame(reliability_rows)

display(
    reliability_table.style.format({
        "Cronbach's alpha": "{:.3f}"
    })
)

# Item-level descriptive statistics

item_descriptives = (
    df_analysis[all_items]
    .agg(["count", "mean", "std", "min", "max"])
    .T
    .rename(columns={
        "count": "n",
        "mean": "Mean",
        "std": "SD",
        "min": "Minimum",
    })

```

```

        "max": "Maximum"
    })
)

display(
    item_descriptives.style.format({
        "n": "{:.0f}",
        "Mean": "{:.2f}",
        "SD": "{:.2f}",
        "Minimum": "{:.0f}",
        "Maximum": "{:.0f}"
    })
)

```

F4.5. Descriptive Statistics and Correlations

```

measures = ["cti", "auth", "rel", "trust"]

overall_descriptives = (
    df_analysis[measures]
    .agg(["count", "mean", "std", "min", "max"])
    .T
    .reset_index()
    .rename(columns={
        "index": "Measure",
        "count": "n",
        "mean": "Mean",
        "std": "SD",
        "min": "Minimum",
        "max": "Maximum"
    })
)

overall_descriptives["Measure"] = (
    overall_descriptives["Measure"]
    .map(construct_labels)
)

display(
    overall_descriptives.style.format({
        "n": "{:.0f}",
        "Mean": "{:.2f}",
        "SD": "{:.2f}",
        "Minimum": "{:.2f}",
        "Maximum": "{:.2f}"
    })
)

```

```
# Pearson correlation matrix for the four composite measures

correlation_matrix = df_analysis[
    ["cti", "auth", "rel", "trust"]
].corr(method="pearson")

correlation_matrix.index = [
    construct_labels[index]
    for index in correlation_matrix.index
]

correlation_matrix.columns = [
    construct_labels[column]
    for column in correlation_matrix.columns
]

display(correlation_matrix.style.format("{:.3f}"))
```

F4.6. Welch Independent-Samples t-Tests

```
def welch_group_comparison(
    data: pd.DataFrame,
    outcome: str,
    outcome_label: str
) -> dict:
    ai = data.loc[
        data["condition"] == "ai",
        outcome
    ].dropna()

    human = data.loc[
        data["condition"] == "human",
        outcome
    ].dropna()

    n_ai = ai.count()
    n_human = human.count()

    mean_ai = ai.mean()
    mean_human = human.mean()

    sd_ai = ai.std(ddof=1)
    sd_human = human.std(ddof=1)

    mean_difference = mean_ai - mean_human
```

```

t_statistic, p_value = stats.ttest_ind(
    ai,
    human,
    equal_var=False,
    alternative="two-sided"
)

standard_error = np.sqrt(
    sd_ai**2 / n_ai
    + sd_human**2 / n_human
)

welch_df = (
    (sd_ai**2 / n_ai + sd_human**2 / n_human) ** 2
    /
    (
        (sd_ai**2 / n_ai) ** 2 / (n_ai - 1)
        + (sd_human**2 / n_human) ** 2 / (n_human - 1)
    )
)

critical_t = stats.t.ppf(
    0.975,
    df=welch_df
)

ci_lower = mean_difference - critical_t * standard_error
ci_upper = mean_difference + critical_t * standard_error

pooled_sd = np.sqrt(
    (
        (n_ai - 1) * sd_ai**2
        + (n_human - 1) * sd_human**2
    )
    / (n_ai + n_human - 2)
)

cohens_d = mean_difference / pooled_sd

correction_factor = 1 - (
    3 / (4 * (n_ai + n_human) - 9)
)

hedges_g = correction_factor * cohens_d

return {
    "Outcome": outcome_label,

```

```

        "AI n": n_ai,
        "AI Mean": mean_ai,
        "AI SD": sd_ai,
        "Human n": n_human,
        "Human Mean": mean_human,
        "Human SD": sd_human,
        "Mean difference (AI - human)": mean_difference,
        "CI lower": ci_lower,
        "CI upper": ci_upper,
        "Welch t": t_statistic,
        "df": welch_df,
        "p": p_value,
        "Cohen's d": cohens_d,
        "Hedges' g": hedges_g
    }

h1_result = welch_group_comparison(
    df_analysis,
    "cti",
    "Click-through intention"
)

h2_result = welch_group_comparison(
    df_analysis,
    "auth",
    "Perceived authenticity"
)

ttest_results = pd.DataFrame([
    h1_result,
    h2_result
])

display(
    ttest_results.style.format({
        "AI n": "{:.0f}",
        "AI Mean": "{:.2f}",
        "AI SD": "{:.2f}",
        "Human n": "{:.0f}",
        "Human Mean": "{:.2f}",
        "Human SD": "{:.2f}",
        "Mean difference (AI - human)": "{:.2f}",
        "CI lower": "{:.2f}",
        "CI upper": "{:.2f}",
        "Welch t": "{:.2f}",
        "df": "{:.2f}",
    })

```

```

        "p": "{:.3f}",
        "Cohen's d": "{:.3f}",
        "Hedges' g": "{:.3f}"
    })
)

```

F4.7. Multiple Regression, Diagnostics and HC3 Inference

```

# Code condition for regression
df_analysis["condition_code"] = (
    df_analysis["condition"]
    .map({
        "human": 0,
        "ai": 1
    })
    .astype("Int64")
)

display(
    df_analysis[
        ["condition", "condition_code"]
    ].drop_duplicates()
)

# Fit conventional OLS model for model diagnostics
ols_model = smf.ols(
    "cti ~ rel + trust + condition_code",
    data=df_analysis
).fit()

print(ols_model.summary())

X = df_analysis[
    ["rel", "trust", "condition_code"]
].astype(float)

X_with_constant = sm.add_constant(X)

vif_results = pd.DataFrame({
    "Predictor": X_with_constant.columns,
    "VIF": [
        variance_inflation_factor(
            X_with_constant.values,
            index
        )
    ]
})

```

```

        for index in range(X_with_constant.shape[1])
    ]
})

display(vif_results.style.format({"VIF": "{:.3f}"}))

print(
    "Interpret the predictor VIFs only; the intercept VIF is not "
    "substantively meaningful."
)

# Formal diagnostic tests

bp_lm, bp_lm_p, bp_f, bp_f_p = het_breuschpagan(
    ols_model.resid,
    ols_model.model.exog
)

white_lm, white_lm_p, white_f, white_f_p = het_white(
    ols_model.resid,
    ols_model.model.exog
)

jb_stat, jb_p, residual_skew, residual_kurtosis = jarque_bera(
    ols_model.resid
)

reset_result = linear_reset(
    ols_model,
    power=2,
    use_f=True
)

diagnostic_tests = pd.DataFrame({
    "Diagnostic": [
        "Breusch-Pagan LM",
        "Breusch-Pagan F",
        "White LM",
        "White F",
        "Jarque-Bera residual normality",
        "Ramsey RESET"
    ],
    "Statistic": [
        bp_lm,
        bp_f,
        white_lm,
        white_f,

```

```

        jb_stat,
        float(reset_result.fvalue)
    ],
    "p": [
        bp_lm_p,
        bp_f_p,
        white_lm_p,
        white_f_p,
        jb_p,
        float(reset_result.pvalue)
    ]
})

display(
    diagnostic_tests.style.format({
        "Statistic": "{:.3f}",
        "p": "{:.3f}"
    })
)

print(
    "The heteroscedasticity tests indicate non-constant residual "
    "variance. HC3 robust standard errors are therefore used for "
    "the final coefficient inference."
)

hc3_model = smf.ols(
    "cti ~ rel + trust + condition_code",
    data=df_analysis
).fit(cov_type="HC3")

hc3_ci = hc3_model.conf_int()

hc3_coefficients = pd.DataFrame({
    "B": hc3_model.params,
    "HC3 SE": hc3_model.bse,
    # Statsmodels uses asymptotic normal inference here because use_t=False
    "z": hc3_model.tvalues,
    "p": hc3_model.pvalues,
    "CI lower": hc3_ci[0],
    "CI upper": hc3_ci[1]
})

hc3_model_fit = pd.DataFrame({
    "n": [int(hc3_model.nobs)],
    "R-squared": [hc3_model.rsquared],
    "Adjusted R-squared": [hc3_model.rsquared_adj],

```

```

    "Robust F": [hc3_model.fvalue],
    "df model": [int(hc3_model.df_model)],
    "df residual": [int(hc3_model.df_resid)],
    "Robust model p": [hc3_model.f_pvalue]
})

display(hc3_coefficients.style.format("{:.3f}"))
display(hc3_model_fit.style.format("{:.3f}"))

```

F4.8. Influence and Leave-One-Out Sensitivity Checks

```

influence = ols_model.get_influence()
influence_frame = influence.summary_frame()

influence_results = pd.DataFrame({
    "Analysis row": np.arange(1, len(df_analysis) + 1),
    "Cook's distance": influence_frame["cooks_d"],
    "Leverage": influence_frame["hat_diag"],
    "Externally studentised residual": influence_frame["student_resid"]
})

cook_threshold = 4 / len(df_analysis)
p_parameters = len(ols_model.params)
leverage_threshold = 2 * p_parameters / len(df_analysis)

influence_summary = pd.DataFrame({
    "Screening rule": [
        "Cook's distance > 4/n",
        "Leverage > 2p/n",
        "|Externally studentised residual| > 3"
    ],
    "Threshold": [
        cook_threshold,
        leverage_threshold,
        3
    ],
    "Flagged observations": [
        int(
            (
                influence_results["Cook's distance"]
                > cook_threshold
            ).sum()
        ),
        int(
            (
                influence_results["Leverage"]
                > leverage_threshold
            )
        )
    ]
})

```

```

        ).sum()
    ),
    int(
        (
            influence_results[
                "Externally studentised residual"
            ].abs() > 3
        ).sum()
    )
]
})

display(
    influence_summary.style.format({
        "Threshold": "{:.3f}",
        "Flagged observations": "{:.0f}"
    })
)

display(
    influence_results
    .sort_values(
        "Cook's distance",
        ascending=False
    )
    .head(10)
    .style.format({
        "Cook's distance": "{:.3f}",
        "Leverage": "{:.3f}",
        "Externally studentised residual": "{:.3f}"
    })
)

# Leave-one-out HC3 sensitivity check

loo_rows = []

for excluded_index in df_analysis.index:
    reduced_data = df_analysis.drop(
        index=excluded_index
    )

    reduced_model = smf.ols(
        "cti ~ rel + trust + condition_code",
        data=reduced_data
    ).fit(cov_type="HC3")

```

```

for predictor in [
    "rel",
    "trust",
    "condition_code"
]:
    loo_rows.append({
        "Excluded row": excluded_index,
        "Predictor": predictor,
        "B": reduced_model.params[predictor],
        "p": reduced_model.pvalues[predictor]
    })

loo_results = pd.DataFrame(loo_rows)

loo_summary = (
    loo_results
    .groupby("Predictor")
    .agg(
        Minimum_B=("B", "min"),
        Maximum_B=("B", "max"),
        Minimum_p=("p", "min"),
        Maximum_p=("p", "max")
    )
    .reset_index()
)

display(
    loo_summary.style.format({
        "Minimum_B": "{:.3f}",
        "Maximum_B": "{:.3f}",
        "Minimum_p": "{:.3f}",
        "Maximum_p": "{:.3f}"
    })
)

print(
    "The leave-one-out results show whether any single observation "
    "changes the direction or statistical conclusion for a predictor. "
    "No observation is removed solely because it is influential."
)

```

Appendix G: Additional Descriptive and Reliability Results

This appendix presents item-level descriptive statistics, detailed reliability diagnostics and Pearson correlations for the four composite measures. All results are based on the final analytical sample of 60 participants.

G1. Item-Level Descriptive Statistics

Item	Construct	Item wording	n	M	SD	Min.	Max.
cti_1	Click-through intention	I would be likely to select the “Learn More” button in this advertisement.	60	4.13	1.87	1	7
cti_2	Click-through intention	I would consider selecting “Learn More” to find out more about FocusForm.	60	4.48	1.91	1	7
cti_3	Click-through intention	If I saw this advertisement online, I would intend to select “Learn More”.	60	3.87	1.93	1	7
auth_1	Perceived authenticity	The FocusForm advertisement feels authentic.	60	4.82	1.49	1	7
auth_2	Perceived authenticity	The FocusForm advertisement comes across as genuine.	60	4.85	1.40	1	7
auth_3	Perceived authenticity	The message in this advertisement feels sincere.	60	5.00	1.40	1	7
rel_1	Perceived relevance	The FocusForm advertisement is relevant to me.	60	4.25	1.70	1	7
rel_2	Perceived relevance	The FocusForm app appears relevant to my needs.	60	4.20	1.75	1	7
rel_3	Perceived relevance	The message in this advertisement relates to things that matter to me.	60	4.63	1.48	2	7
trust_1	Trust	Based on this advertisement, FocusForm seems trustworthy.	60	5.05	1.33	1	7
trust_2	Trust	Based on this advertisement, I would trust FocusForm to deliver what it promises.	60	4.85	1.39	1	7
trust_3	Trust	The claims made about FocusForm in this advertisement seem believable.	60	5.07	1.23	1	7

Table G1. Item-level descriptive statistics for the 12 construct items.

Note. All items were measured on a seven-point Likert scale ranging from 1 (“Strongly disagree”) to 7 (“Strongly agree”).

G2. Detailed Reliability Diagnostics

Construct	Item	Corrected item-total correlation	Cronbach's alpha if item deleted	Overall Cronbach's alpha
Click-through intention	cti_1	0.904	0.875	0.935
Click-through intention	cti_2	0.887	0.888	0.935

Construct	Item	Corrected item-total correlation	Cronbach's alpha if item deleted	Overall Cronbach's alpha
Click-through intention	cti_3	0.808	0.951	0.935
Perceived authenticity	auth_1	0.804	0.847	0.897
Perceived authenticity	auth_2	0.866	0.792	0.897
Perceived authenticity	auth_3	0.724	0.912	0.897
Perceived relevance	rel_1	0.861	0.851	0.915
Perceived relevance	rel_2	0.908	0.810	0.915
Perceived relevance	rel_3	0.737	0.951	0.915
Trust	trust_1	0.715	0.859	0.873
Trust	trust_2	0.805	0.777	0.873
Trust	trust_3	0.758	0.823	0.873

Table G2. Corrected item-total correlations and Cronbach's alpha if item deleted.

Note. The corrected item-total correlation is the correlation between an item and the sum of the other two items in the same construct. No items were removed; the three-item measures were retained as specified before analysis.

G3. Pearson Correlations Between Composite Measures

Measure	Click-through intention	Perceived authenticity	Perceived relevance	Trust
Click-through intention	1.000	0.531***	0.465***	0.449***
Perceived authenticity	0.531***	1.000	0.049	0.764***
Perceived relevance	0.465***	0.049	1.000	0.003
Trust	0.449***	0.764***	0.003	1.000

Table G3. Pearson correlations between the four composite measures.

Note. N = 60. Values are Pearson correlation coefficients. Two-tailed significance tests are shown as * $p < .05$, ** $p < .01$ and *** $p < .001$. These correlations are descriptive and were not used as separate hypothesis tests.

Appendix H: Detailed Statistical Outputs and Regression Diagnostics

This appendix presents the detailed statistical outputs supporting the hypothesis tests and regression analysis reported in Chapter 4. All results were reproduced using the cleaning, scoring, testing and diagnostic procedures contained in the final Python analysis notebook. The analyses are based on the final sample of 60 participants.

H1. Welch Independent-Samples t-Tests

Outcome	AI n	AI M	AI SD	Human n	Human M	Human SD	Difference	95% CI lower	95% CI upper	t	df	p	Cohen's d	Hedges' g
Click-through intention	30	4.26	1.94	30	4.07	1.65	0.189	-0.743	1.121	0.406	56.519	0.686	0.105	0.103
Perceived authenticity	30	4.81	1.19	30	4.97	1.42	-0.156	-0.834	0.523	-0.459	56.230	0.648	-0.119	-0.117

Table H1. Detailed Welch independent-samples t-test results for H1 and H2.

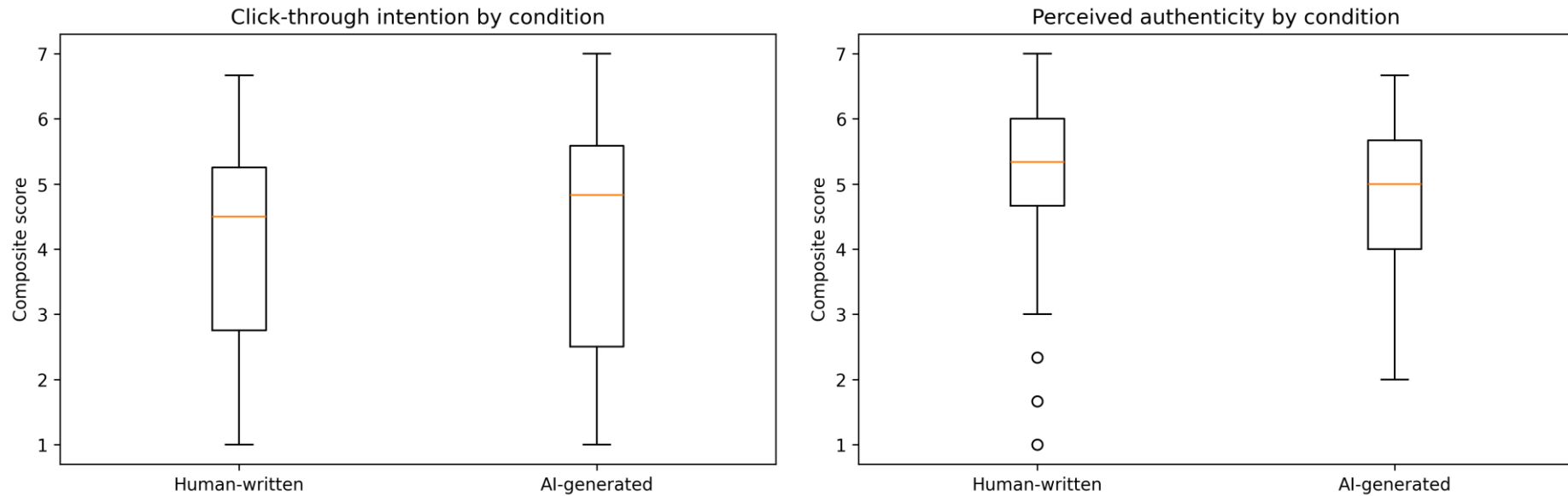
Note. Mean differences and standardised effect sizes are calculated as AI-generated minus human-written. All tests were two-sided.

Outcome	Condition	n	Skewness	Excess kurtosis	Variance
Click-through intention	Human-written copy	30	-0.442	-0.789	2.723
Click-through intention	AI-generated copy	30	-0.310	-1.161	3.775
Perceived authenticity	Human-written copy	30	-1.282	1.593	2.026
Perceived authenticity	AI-generated copy	30	-0.811	-0.053	1.415

Table H2. Distribution summaries for the outcomes used in the Welch t-tests.

Note. Welch's test was the primary comparison and does not require equal group variances. Each condition contained 30 observations.

H2. Visual Distribution Checks for H1 and H2



Figures H1 and H2. Boxplots of click-through intention and perceived authenticity by condition.

Note. The figures provide visual distribution checks only. The formal condition comparisons were based on the two-sided Welch tests shown in Table H1.

H3. Conventional OLS Regression Output

Predictor	B	SE	t	p	95% CI lower	95% CI upper
Intercept	-1.598	0.956	-1.671	0.100	-3.514	0.318
Perceived relevance	0.543	0.120	4.516	< .001	0.302	0.783
Trust	0.679	0.155	4.380	< .001	0.369	0.990
Condition (1 = AI; 0 = human-written)	0.008	0.364	0.022	0.983	-0.721	0.737

Table H3. Conventional OLS coefficient estimates.

n	R-squared	Adjusted R-squared	F	df model	df residual	Model p
---	-----------	--------------------	---	----------	-------------	---------

n	R-squared	Adjusted R-squared	F	df model	df residual	Model p
60	0.416	0.385	13.306	3	56	< .001

Table H4. Conventional OLS model-fit statistics.

Note. The conventional OLS model was used for model fit and regression diagnostics. Because the heteroscedasticity tests were significant, the final coefficient inference was based on HC3 robust standard errors.

H4. HC3 Robust Regression Results

Predictor	B	HC3 SE	z	p	95% CI lower	95% CI upper
Intercept	-1.598	1.056	-1.514	0.130	-3.667	0.471
Perceived relevance	0.543	0.156	3.487	< .001	0.238	0.848
Trust	0.679	0.156	4.363	< .001	0.374	0.984
Condition (1 = AI; 0 = human-written)	0.008	0.396	0.020	0.984	-0.767	0.783

Table H5. Multiple regression coefficients using HC3 robust standard errors.

n	R-squared	Adjusted R-squared	Robust F	df model	df residual	Robust model p
60	0.416	0.385	12.567	3	56	< .001

Table H6. HC3 robust model-fit statistics.

Note. The dependent variable was stated click-through intention. The condition coefficient compares the AI-generated condition with the human-written reference condition. The coefficients for relevance and trust describe conditional associations rather than causal effects.

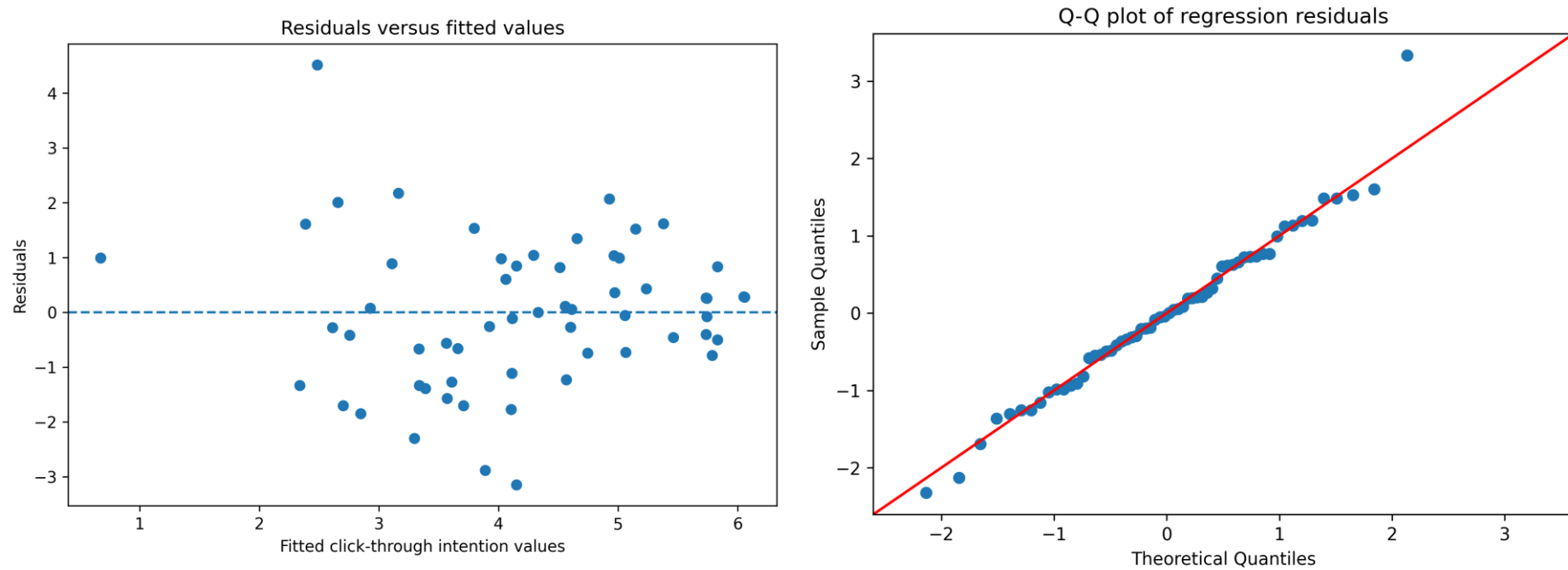
H5. Multicollinearity Diagnostics

Predictor	VIF
Intercept	27.859
Perceived relevance	1.008
Trust	1.000
Condition (1 = AI; 0 = human-written)	1.009

Table H7. Variance inflation factors for the regression model.

Note. Only the predictor VIF values are interpreted. The intercept VIF is not substantively meaningful. Predictor VIF values were approximately 1.00, indicating no material multicollinearity concern.

H6. Regression Residual Diagnostics



Figures H3 and H4. Residual-versus-fitted plot and normal Q-Q plot.

Diagnostic	Statistic	p	Interpretation
Breusch-Pagan LM	9.578	0.023	Evidence of heteroscedasticity
Breusch-Pagan F	3.546	0.020	Evidence of heteroscedasticity
White LM	19.717	0.011	Evidence of heteroscedasticity
White F	3.120	0.006	Evidence of heteroscedasticity
Jarque-Bera residual normality	2.758	0.252	No clear departure from residual normality

Diagnostic	Statistic	p	Interpretation
Ramsey RESET	2.864	0.096	No strong evidence of functional-form misspecification

Table H8. Formal regression diagnostic tests.

Note. The Breusch-Pagan and White tests indicated non-constant residual variance, supporting the use of HC3 robust standard errors. The Jarque-Bera and Ramsey RESET results did not provide clear evidence of residual non-normality or functional-form misspecification.

H7. Influence Diagnostics

Screening rule	Threshold	Flagged observations
Cook's distance > 4/n	0.067	3
Leverage > 2p/n	0.133	2
Externally studentised residual > 3	3.000	1

Table H9. Influence-screening thresholds and numbers of flagged observations.

Analysis row	Cook's distance	Leverage	Externally studentised residual	Cook flag	Leverage flag	Residual flag
7	0.295	0.094	3.753	Yes	No	Yes
13	0.114	0.077	-2.435	Yes	No	No
11	0.100	0.249	-1.101	Yes	Yes	No
55	0.065	0.055	-2.186	No	No	No
53	0.060	0.085	1.642	No	No	No
21	0.042	0.071	1.502	No	No	No
23	0.042	0.071	1.502	No	No	No
32	0.040	0.203	0.790	No	Yes	No
16	0.035	0.048	-1.705	No	No	No
38	0.035	0.091	-1.178	No	No	No

Table H10. Ten observations with the largest Cook's distances.

Note. Analysis-row numbers refer to the ordering of observations in the final cleaned dataset. Qualtrics Response IDs are not reproduced. In the screening thresholds, n is the sample size and p is the number of estimated model parameters, including the intercept. The screening rules were used to identify observations for inspection and were not automatic exclusion criteria.

H8. Leave-One-Out Sensitivity Analysis

Predictor	Minimum B	Maximum B	Minimum p	Maximum p	Direction across all models	Statistical conclusion
Condition (1 = AI; 0 = human-written)	-0.188	0.118	0.592	1.000	Varied	Non-significant in every model
Perceived relevance	0.509	0.646	< .001	0.002	Positive	Significant in every model
Trust	0.593	0.737	< .001	< .001	Positive	Significant in every model

Table H11. Leave-one-out HC3 sensitivity results.

Note. Each model was re-estimated after removing one observation. No single observation changed the positive and statistically significant conclusions for relevance or trust, or the non-significant conclusion for condition. No observations were removed on the basis of the influence checks.