International Journal of Applied Artificial Intelligence in Business (ISSN: 3056-1892)
Volume: 01 Issue: 01 Year: 2026
https://doi.org/10.51137/wrp.ijaaib.1062

Generative AI in Fashion E-Commerce Product Imagery: An Experimental Test of Image Origin and Transparency Labelling on Consumer Trust, Perceived Product Quality and Purchase Intention

Authors:

Daniel Bensch (daniel.bensch99@web.de)
University of Applied Sciences Wedel, Wedel, Germany

Abstract:

Retailers are increasingly replacing photographic model shoots with synthetic product imagery generated by text-to-image systems, a deployment decision that couples substantial production savings with an unresolved question about consumer response and about the transparency obligations attached to synthetic content. This study examines how AI-generated and photographic product images are evaluated in fashion e-commerce and how a visible origin label affects that evaluation. A 2 (image origin: AI-generated vs. photograph) × 2 (labelling: absent vs. present) between-subjects online experiment was conducted with 150 participants. Consumer trust, perceived product quality and purchase intention were measured with established scales, and general attitude toward artificial intelligence was tested as a moderator. Without labelling, no statistically significant differences between the two image types emerged on any dependent variable. The visible label reduced perceived product quality significantly but modestly, while effects on trust and purchase intention, the image origin by label interaction and moderation were not statistically significant.

Keywords:

Generative Artificial Intelligence, AI Disclosure, Fashion E-Commerce, Consumer Trust.

Submitted: 2026-07-30. Revised: 2026-09-12. Accepted: 2026-10-01.

WOHLLEBE & ROSS

Discover publication opportunities:
www.wr-publishing.org

International Journal of Applied Artificial Intelligence in Business (ISSN: 3056-1892)
Volume: 01 Issue: 01 Year: 2026
https://doi.org/10.51137/wrp.ijaaib.1062

Introduction

Generative artificial intelligence has moved from experimentation to operational use in retail product imagery. In March 2025 the fashion retailer H&M announced that it would create digital twins of 30 real models and deploy them in marketing and online channels instead of conventional photography (H&M Group, 2025), and comparable approaches are pursued by retailers such as Levi Strauss and Mango (CNN, 2025). In September 2025 the Otto Group announced that it would replace conventional model photography with its in-house system MOVEX, which generates photorealistic model images from a single cut-out product shot (Otto, 2025). Text-to-image models now produce imagery of high photorealism, and retailers expect considerable savings in production time and cost.

The consequences are particularly far reaching in fashion e-commerce, because the product image forms the central basis for the purchase decision (Jiang & Benbasat, 2004). Online, the garment can neither be touched nor tried on, so judgements about quality and fit rest largely on the visual representation. This is precisely where tension arises. Nightingale and Farid (2022) show that people can no longer reliably distinguish AI-generated faces from real ones and even tend to rate them as more trustworthy. Once the artificial origin of an image is disclosed, however, that evaluation may shift. Whether a product image marked as AI-generated is perceived differently from a photograph is therefore not a question of visual detail but one that reaches directly into the purchase decision.

The regulatory frame adds urgency. Regulation (EU) 2024/1689, the EU AI Act, establishes a uniform European legal framework for artificial intelligence, and Article 50 provides transparency obligations for certain AI-generated content that apply from 2 August 2026. A distinction has to be drawn between machine-readable marking by providers of generative systems and visible disclosure by deployers. Visible labelling is not prescribed for every AI-generated product image, but it can become relevant in particular for content classified as a deepfake within the meaning of the regulation. Retailers therefore face a twofold decision. They must weigh whether to use AI-generated product images at all, and they must anticipate how a legally required or voluntary origin notice affects the perception of those images.

Existing research does not answer this adequately. Evidence of a trust penalty following AI disclosure comes from professional performance settings and from advertising rather than from online retail with product images. How AI-generated and photographically produced product images are perceived in fashion e-commerce, and how labelling acts on that perception, has so far been examined experimentally only to a limited extent.

The aim of this study is therefore to examine empirically how AI-generated product images are perceived in comparison with conventional model photographs in fashion e-commerce, and what influence a transparency label exerts on that perception. Three variables central to the purchase decision are at the

WOHLLEBE & ROSS
Discover publication opportunities:
www.wr-publishing.org

International Journal of Applied Artificial Intelligence in Business (ISSN: 3056-1892)
Volume: 01 Issue: 01 Year: 2026
https://doi.org/10.51137/wrp.ijaaib.1062

center of the investigation: consumer trust, perceived product quality and purchase intention. The central research question is to what extent AI-generated and photographically produced product images in fashion e-commerce differ with respect to consumer trust, perceived product quality and purchase intention, and what influence a transparency label has on these variables. The question comprises four aspects: whether the two image types differ without labelling, how a notice of AI origin acts, whether this effect differs between the two image types, and what role the general attitude toward AI plays for the labelling effect.

The remainder of the paper is structured as follows. The literature review develops the theoretical frame and derives the hypotheses. The methodology section describes the experimental design, the construction of the image stimuli and the measurement of the three outcome variables. The results section reports the sample, reliability and manipulation checks and the hypothesis tests. The discussion interprets the findings, relates them to prior research and states the limitations, before the conclusion summarizes the central insights and outlines further research needs.

Literature Review

Generative AI in Image Production

Generative artificial intelligence denotes systems that create new content such as images or text from input data and have become established as a tool of visual communication. Studies indicate that even AI-generated human faces are barely distinguishable from real photographs (Nightingale & Farid, 2022). In fashion e-commerce this opens a new production logic in which a complete model image can be generated from a single cut-out shot of a garment. Real model shoots with model, studio and post-production become partially substitutable, which lowers the cost and time required for product image production (Grigsby et al., 2025).

The generation of such a model image follows a multi-stage process. A real garment serves as a visual template that is fed into an image model together with a textually described persona. On this basis the model produces a new, fully synthetic image comprising both the wearing figure and the depiction of the garment. The real product remains the point of reference but appears in the image not as a photograph but as an AI-generated reconstruction. This distinction between the real object of purchase and its synthetic depiction becomes relevant for the hypothesis development.

Whether consumers recognize an AI-generated image without a notice depends strongly on the image type. For human faces, the element central to model imagery, recognition is especially difficult. In a classification experiment with 315 participants, the hit rate between synthetic and real faces was 48.2 per cent and thus at chance level (Nightingale & Farid, 2022). In more complex scenes, visible artefacts occur more frequently and can serve as recognition cues, with the type of artefact and the complexity

International Journal of Applied Artificial Intelligence in Business (ISSN: 3056-1892)
Volume: 01 Issue: 01 Year: 2026
https://doi.org/10.51137/wrp.ijaaib.1062

of the depicted scene decisively influencing how reliably recognition succeeds (Kamali et al., 2025). Since image quality has advanced since these studies, decreasing rather than increasing detectability is to be expected. In the fleeting viewing context of online shopping, reliable identification of image origin therefore appears doubtful.

Consumer Trust, Perceived Product Quality and Purchase Intention in Online Retail

Online retail differs fundamentally from stationary retail in that consumers cannot touch, try on or closely inspect a product before purchase. Jiang and Benbasat (2004) describe this missing sensory experience as a central limitation of electronic commerce, since product attributes are conveyed solely through the user interface and the representation shown there. An information asymmetry follows, because the seller knows the actual condition of the product while the consumer cannot verify it before purchase (Akerlof, 1970). In this situation the product representation takes on a central function. With the concept of perceived diagnostic, Jiang and Benbasat (2004) describe how well a representation supports consumers in evaluating the product. The product image thus assumes the task that physical inspection performs in a store and serves as the central cue for assessing the product.

Trust plays a particular role because online purchases are bound up with uncertainty about both seller and product. Mayer et al. (1995) define trust as the willingness of a person to make herself vulnerable to another party in the expectation that this party will not act to her disadvantage, and people infer trustworthiness from three perceived attributes: ability, benevolence and integrity. Gefen et al. (2003) show that trust in the seller is as consequential for purchase intention as the perceived usefulness and ease of use of an online shop, and they understand trust as the belief that the seller has nothing to gain from deception. McKnight et al. (2002) direct attention to initial trust toward a still unknown seller and show that it helps consumers overcome perceived uncertainty and transaction risk. This perspective is particularly relevant here, because participants evaluate a product without knowledge of a specific seller and their judgement therefore rests substantially on the representation shown.

Perceived product quality constitutes an independent evaluative judgement that consumers likewise derive from the representation. Zeithaml (1988) defines it as the subjective judgement of a consumer about the excellence or superiority of a product and distinguishes it explicitly from objectively measurable condition. Because actual quality cannot be verified before purchase, consumers base their judgement on available cues. Dodds et al. (1991) show that extrinsic cues, which is information not intrinsic to the product itself such as price, brand or retail environment, decisively shape the quality judgement. In online retail the visual product representation emerges as one such cue, from which consumers infer material, workmanship and value without being able to inspect them.

WOHLLEBE & ROSS

Discover publication opportunities:
www.wr-publishing.org

International Journal of Applied Artificial Intelligence in Business (ISSN: 3056-1892)
Volume: 01 Issue: 01 Year: 2026
https://doi.org/10.51137/wrp.ijaaib.1062

Purchase intention is added as a third variable. In the model of Dodds et al. (1991) it forms the central outcome and describes a consumer's willingness to acquire a product, arising there from the interplay of perceived quality and perceived value. Purchase intention is regarded as an established predictor of actual behaviour, since according to Ajzen (1991) intentions are the immediate antecedents of action. Consumer trust, perceived product quality and purchase intention are thus independent judgements that consumers derive from the product image in online retail, so that changes in the representation can plausibly act on each of them.

The EU AI Act as a Regulatory Frame

With Regulation (EU) 2024/1689 the European Union creates a uniform legal framework for artificial intelligence. It entered into force on 1 August 2024 and becomes applicable in stages, with the transparency obligations relevant for AI-generated content under Article 50 applying from 2 August 2026.

Article 50 distinguishes between the obligations of providers and of deployers of AI systems. Providers of systems that generate synthetic image, audio, video or text content must in principle ensure that their outputs are marked in a machine-readable format and are detectable as artificially generated or manipulated. This technical marking serves the reliable detection of artificial origin and is not equivalent to a notice visible to consumers (Regulation (EU) 2024/1689, Art. 50(2)).

Visible disclosure by deployers is to be distinguished from this. A deployer is a party that uses an AI system under its own responsibility in a professional capacity, so that a fashion retailer using a generative system under its own responsibility to create product images may qualify as a deployer. For image, audio and video content, however, Article 50(4) does not require visible labelling of every AI-generated item across the board. The disclosure obligation instead attaches to whether the content qualifies as a deepfake within the meaning of the regulation. According to Article 3(60), a deepfake is AI-generated or manipulated image, audio or video content resembling existing persons, objects, places, entities or events that would falsely appear to a person to be authentic or truthful. The guidelines of the European Commission specify three cumulative conditions: the content must display a high degree of resemblance to the simulated referent, must relate to something that exists, could plausibly exist or could plausibly have existed, and must be potentially deceptive as to its authenticity or truthfulness, whereby the context of use and the composition and expectations of the audience are among the factors to be taken into account (European Commission, 2026).

A photorealistic AI model image can in principle fall within the scope of this definition, since it may depict a plausibly existing person and a real garment in the form of an apparent photograph although no such photograph was ever taken. No automatic classification of every AI-generated product image as a deepfake follows from this. Decisive are the concrete design, the degree of resemblance and the question of whether the image can be falsely understood as an authentic photograph in the respective

WOHLLEBE & ROSS
Discover publication opportunities:
www.wr-publishing.org

context of use, so that the interpretive latitude of the deepfake concept discussed in the scholarly literature remains relevant for assessing individual product images (Łabuz, 2025). Where visible disclosure is required, it must be provided clearly and distinguishably at the latest at the time of the first interaction or exposure and must meet the applicable accessibility requirements (Regulation (EU) 2024/1689, Art. 50(5)). An origin notice placed directly on the product image can in principle serve this disclosure, although whether its concrete design meets the legal requirements depends among other things on its intelligibility and perceptibility. The regulatory frame determines under which conditions AI-generated content must be marked or disclosed, but it says nothing about how consumers react to a visible origin notice.

Hypothesis Development

The starting point is the question of how AI-generated images act when their origin is unknown. Nightingale and Farid (2022) show that synthetic faces are not reliably distinguished from real ones and are on average judged somewhat more trustworthy. A comparable pattern appears in Grigsby et al. (2025): without disclosure, responses to AI-generated and conventionally produced advertising do not differ significantly, and only the notice of artificial origin changes the evaluation. Both findings argue against inferring an evaluative disadvantage from actual AI origin alone. For the fashion context, Kim (2026) reports more negative responses to social media advertising with AI models, yet that study concerns brand advertising and captures responses to the perceived model type, so it does not directly answer how a photorealistic product image acts whose origin remains unlabeled and unrecognizable. The expectation is therefore deliberately not directed in favor of either image type.

- H1a: Without labelling, AI-generated and photographically produced product images do not differ with respect to consumer trust.
- H1b: Without labelling, AI-generated and photographically produced product images do not differ with respect to perceived product quality.
- H1c: Without labelling, AI-generated and photographically produced product images do not differ with respect to purchase intention.

What happens when AI origin is disclosed is examined by Schilke and Reimann (2025) in 13 preregistered experiments. Across different tasks and actors, a consistent pattern emerges those who disclose the use of AI are judged less trustworthy. Schilke and Reimann describe this as a transparency dilemma, since the ethically desirable behaviour of disclosure is accompanied by a loss of trust, and they attribute it to lower perceived legitimacy, as disclosure signals a deviation from the expected human provision of performance. Negative disclosure effects also appear in the advertising context. Bui (2026) shows that a notice of AI involvement reduces perceived advertising value and purchase intention, with lower advertising credibility explaining part of these relationships. The evidence is nevertheless not

uniformly negative. Shi and Jiang (2026) identify two opposing paths: an AI notice increases perceived novelty and can thereby foster positive responses, while at the same time reducing the authenticity of the advertisement, which burdens attitudes and purchase intention.

For product images in fashion e-commerce the negative path is particularly plausible, because the model image serves as a carrier of information about the depicted product. A visible notice identifying the representation as AI-generated can raise doubts about its credibility and about the faithful rendering of the garment, and such doubts can impair perceived quality, trust and purchase intention. Castelo et al. (2019) additionally show that consumers judge algorithms more skeptically for subjective tasks than for objective tasks, which supports the expectation of a negative reaction given that fashion evaluations are strongly shaped by aesthetic and personal judgements.

- H2a: Labelling a product image as AI-generated lowers consumer trust.
- H2b: Labelling a product image as AI-generated lowers perceived product quality.
- H2c: Labelling a product image as AI-generated lowers purchase intention.

Grigsby et al. (2025) show that the effect of disclosure depends on which image element is disclosed as AI-generated. In three experiments on service advertising, the decline in trust and ad evaluation was stronger when disclosure concerned an intangible element such as the depicted person than when only tangible components such as equipment were AI-generated. The artificial creation of the depicted person thus constitutes a particularly sensitive case. This transfers to AI-generated model images, where both the wearing figure and the product depiction are synthetic, whereas the photographic condition rests on a real capture. An AI notice on the synthetic model image therefore concerns exactly the human element identified as critical and at the same time calls the authenticity of the entire representation into question, so that the negative authenticity path described by Shi and Jiang (2026) should carry greater weight there. In both labelled conditions the notice can trigger negative expectations, but for the AI-generated image it coincides with the actual artificial origin of the depicted person and the product depiction, and the label can act as an interpretive frame through which previously ambiguous image features are perceived as indications of artificial creation.

- H3a: Labelling affects consumer trust more strongly for AI-generated product images than for photographically produced product images.
- H3b: Labelling affects perceived product quality more strongly for AI-generated product images than for photographically produced product images.
- H3c: Labelling affects purchase intention more strongly for AI-generated product images than for photographically produced product images.

Reactions to AI vary between individuals. Dietvorst et al. (2015) show that trust in algorithmic recommendations can fall particularly quickly after observed errors, whereas Logg et al. (2019)

International Journal of Applied Artificial Intelligence in Business (ISSN: 3056-1892)
Volume: 01 Issue: 01 Year: 2026
https://doi.org/10.51137/wrp.ijaaib.1062

demonstrate algorithm appreciation, with lay people partly following algorithmic advice more than human advice. This appreciation is weaker in subjective and strongly human domains (Castelo et al., 2019). Direct moderation findings exist for the effect of AI disclosure. Schilke and Reimann (2025) show that the trust discount associated with disclosure is weaker among people with a more positive attitude toward technology, whereas mere familiarity with AI or its use showed no comparable influence. Bui (2026) likewise finds that a more positive attitude toward AI attenuates the negative effect of disclosure on advertising value and purchase intention, and Balaban and Phua (2026) report a similar moderating relationship for brand credibility.

- H4a: General attitude toward AI moderates the effect of labelling on consumer trust, such that the negative effect is weaker among people with a more positive attitude toward AI.
- H4b: General attitude toward AI moderates the effect of labelling on perceived product quality, such that the negative effect is weaker among people with a more positive attitude toward AI.
- H4c: General attitude toward AI moderates the effect of labelling on purchase intention, such that the negative effect is weaker among people with a more positive attitude toward AI.

The literature thus yields a differentiated picture and at the same time a clear gap. AI-generated images are not necessarily evaluated less favorably than photographs when their origin is unknown, whereas disclosure can reduce trust, perceived authenticity and purchase intention, with its effect depending on the image element concerned and on viewers' attitudes toward AI. Prior findings stem predominantly from professional performance contexts, service advertising and promotional social media content, and in the fashion context the focus has been on model type in brand advertising (Kim, 2026). A systematic review of consumer research on AI in fashion further shows that chatbots and general AI services predominate (Zhang & Liu, 2026). A controlled study that jointly varies image origin and labelling in fashion e-commerce and tests their effects on consumer trust, perceived product quality and purchase intention has been missing. Figure 1 summarizes the resulting research model, in which image origin, labelling and interaction act on the three dependent variables, with general attitude toward AI moderating the labelling path.

WOHLLEBE & ROSS

Discover publication opportunities:
www.wr-publishing.org

International Journal of Applied Artificial Intelligence in Business (ISSN: 3056-1892)
Volume: 01 Issue: 01 Year: 2026
https://doi.org/10.51137/wrp.ijaaib.1062

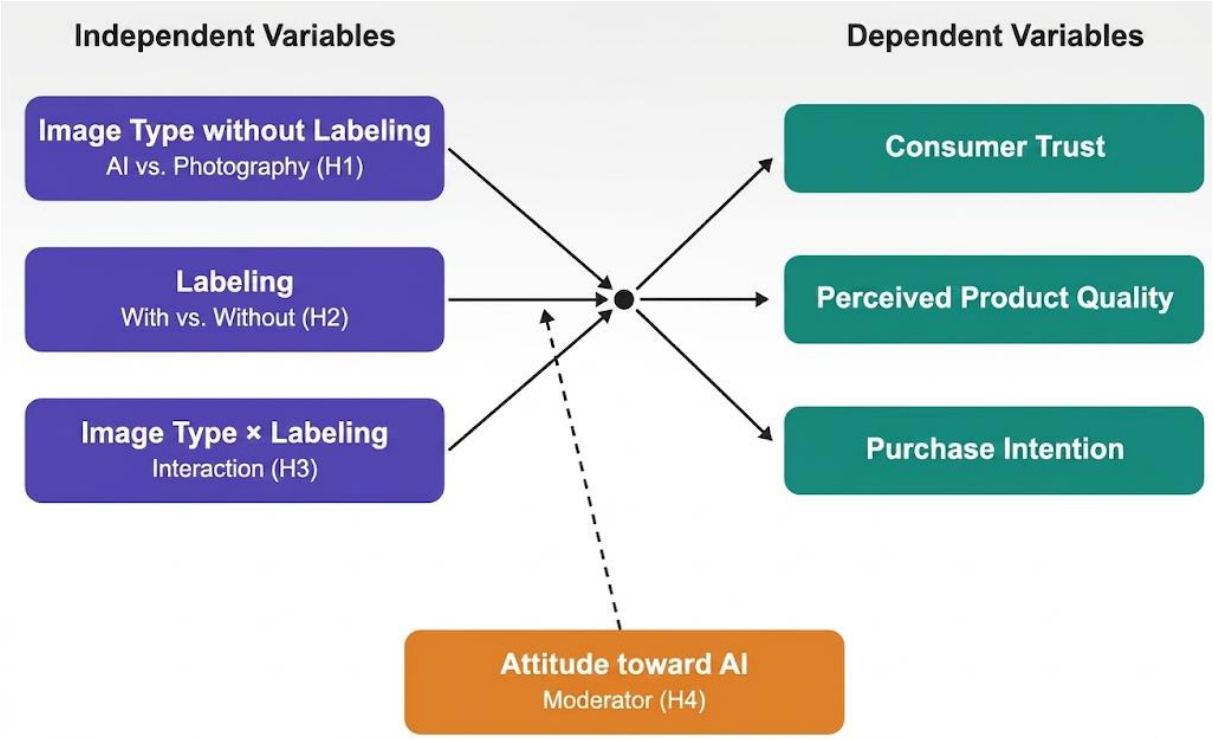

Figure 1: Research model with image origin, labelling and interaction as predictors of consumer trust, perceived product quality and purchase intention, and general AI attitude as moderator of the labelling path.

Methodology

The research question was addressed with a quantitative online experiment combining a 2 (image origin: AI-generated vs. photograph) × 2 (labelling: absent vs. present) fully crossed between-subjects design with a standardized questionnaire. Full crossing was necessary because the interaction postulated in H3 can only be tested if both factors are varied jointly. An experiment was chosen because the question targets cause and effect for the two manipulated factors, and random assignment distributes personal characteristics such as prior experience, product interest or momentary mood across the groups on average, supporting internal validity (Döring & Bortz, 2016, S. 94). A between-subjects design was preferred over a within-subjects design because the latter makes the comparison between conditions visible to participants (Charness et al., 2012) and thereby invites the behavioral changes that Zizzo (2010) describes as experimenter demand effects, in addition to order effects. Viewing exactly one product representation also matches the realistic online shopping situation. Data were collected online, which allows a broader and, with respect to AI attitude, more heterogeneous participant pool, a precondition for testing the moderator in H4, and which approximates the screen-based context of actual online shopping. The recruited convenience sample limits external validity, since it does not represent the population of online shoppers (Döring & Bortz, 2016, S. 195, 305–307).

WOHLLEBE & ROSS
Discover publication opportunities:
www.wr-publishing.org

International Journal of Applied Artificial Intelligence in Business (ISSN: 3056-1892)
Volume: 01 Issue: 01 Year: 2026
https://doi.org/10.51137/wrp.ijaaib.1062

Assignment to conditions was handled automatically by the survey software SoSci Survey. In the labelled conditions both image types carried the notice "*Mit KI generiert" (generated with AI), including the photograph. This counterfactual cell was intended, since only a condition in which the notice is present while the origin is genuine allows the effect of the notice to be separated from that of actual image origin. Because AI attitude was measured rather than manipulated, the design is an experiment with a measured moderator, so that causal claims hold for image origin and labelling, while the moderation tested in H4 is correlational in nature.

The three dependent variables were measured with established scales developed under the paradigm of Churchill (1979) using a uniform seven-point agreement format from 1 (do not agree at all) to 7 (fully agree). Consumer trust was measured with the seven-item trust scale of Gefen et al. (2003), which captures integrity, benevolence and ability as derived from Mayer et al. (1995) as well as predictability. The introductory experience stem "Based on my experience with the online vendor in the past" was dropped and the referent shifted from the vendor to the product representation shown, since participants evaluated a single unfamiliar representation without prior interaction. Perceived product quality was measured with the five items of the quality scale of Dodds et al. (1991), which operationalizes the subjective quality judgement in the sense of Zeithaml (1988) and directs evaluation toward reliability, workmanship, durability and overall excellence; adaptation was limited to orienting the items toward the representation shown. Purchase intention was measured with the five items of the willingness-to-buy scale of Dodds et al. (1991). Because the study displayed no price and did not measure perceived value, the scale was deliberately detached from its original price and value context and the price references in the affected items were removed. Items of Dodds et al. (1991) originally formulated as semantic differentials were converted into seven-point agreement statements to keep the response format uniform. All English items were translated into German and independently back translated into English by a second bilingual person, and the comparison of the backtranslation with the original served to check the content equivalence of both versions (Brislin, 1970).

General attitude toward artificial intelligence served as the moderator and was measured with the four-item AI Attitude Scale (AIAS-4) of Grassini (2023), which captures the general stance toward AI with a focus on its perceived usefulness and its expected influence on personal life, work and society. Its unidimensional structure is advantageous for the moderation test, and the items were retained in the originally validated ten-point response format, with the attitude score formed as the meaning of the four items. Age, gender, frequency of buying clothing online and prior exposure to AI-generated images were recorded as control variables, the latter two as single items. Prior exposure was included because participants with high exposure might recognize an AI-generated image even without labelling and thereby undermine the unlabeled condition. Internal consistency of the multi-item scales was assessed

WOHLLEBE & ROSS

Discover publication opportunities:
www.wr-publishing.org

International Journal of Applied Artificial Intelligence in Business (ISSN: 3056-1892)
Volume: 01 Issue: 01 Year: 2026
https://doi.org/10.51137/wrp.ijaaib.1062

with Cronbach's alpha (Cronbach, 1951) against a criterion of above 0.80, understood as orientation rather than a rigid threshold (Döring & Bortz, 2016, S. 443).

The stimulus was a real model image of a fashion product that served both as the photographic condition and as the visual template for the AI image. The product was a plain, single-colored T-shirt without print, logo or conspicuous pattern, since a restrained product draws less attention to itself, and no brand marks were included because brand information can influence perceived quality and purchase intention (Dodds et al., 1991). For the AI image, the cut-out shot of the garment without a person was the only technical reference passed to the generative model, while the real model image served solely as a template for manual alignment of pose, lighting and appearance, so that a fully synthetic image was created into which no photographic content of the original entered. The depicted person is fictitious and only modelled on the real model, whose personality rights precluded a recognizable reproduction. The model Nano Banana Pro (Gemini 3 Pro Image) by Google DeepMind was used for generation. Gender, approximate age, skin tone, hair color, hair length and style, facial hair, build and proportions, facial expression, gaze direction and posture were held constant, as were the garment, its fit and the way it was worn, and, at the level of the capture, camera perspective and eye level, framing, studio background, lighting direction, brightness, color character, image size and aspect ratio. Facial identity remained an unavoidable difference between the two images, which matters because the depicted person can influence trust judgements (Grigsby et al., 2025). The label was implemented as a small notice mark in one image corner, placed identically on both image types in the labelled conditions.

The questionnaire opened with a short greeting that described the study purpose neutrally as an investigation of the perception of product representations in online retail. This partial concealment addressed the demand characteristics described by Orne (1962), since participants who know that image origin is at issue direct their attention to it and adjust their judgement, which would deprive the unlabeled condition of its function. Such concealment is ethically defensible where no harm arises, the research question requires it and full debriefing follows (Döring & Bortz, 2016, S. 126). A data protection notice in line with the General Data Protection Regulation preceded the survey and consent was given by active confirmation. After random assignment, the stimulus of the assigned condition appeared, with the continue button released only after a short minimum display time, and the image remained visible during the evaluation of the dependent variables so that memory performance did not enter the measurement as an additional source of error. The measurement section began with the three dependent variables in the order consumer trust, perceived product quality and purchase intention, with item order randomized within each scale. AI attitude was measured only afterwards, because items on AI attitude presented before the image evaluation would have supplied precisely the interpretive cues that bias responding (Orne, 1962). The control variables followed. An attention check item requiring a specified response was embedded in the trust block, since inattentive participants increase noise and

WOHLLEBE & ROSS
Discover publication opportunities:
www.wr-publishing.org

reduce power, so that their exclusion improves the reliability of results (Oppenheimer et al., 2009). A two-part manipulation check closed the measurement section, capturing on the seven-point scale how strongly the image was perceived as AI-generated and, through a recall question, whether participants had noticed an origin notice. Without the second part, an absent labelling effect could not be distinguished from a simply overlooked notice. A debriefing disclosed the actual study purpose and stated explicitly that the label shown did not correspond to the actual origin of the image in every condition.

The questionnaire underwent a two-stage pretest. In the first stage five people completed the full questionnaire and reported on comprehensibility, instructions and technical functioning, which led to four items being rephrased and the instruction text before the stimulus being made more precise. In the second stage a separate group of ten people rated the overall similarity of the photograph and the unlabeled AI image on 4 July 2026 with regard to posture, the depicted person and the overall impression, yielding a mean similarity rating of 4.6 on a five-point scale with no rating below four, so that revision of the AI image was not required. All pretest participants were excluded from the main study. The required sample size was determined beforehand by an a priori power analysis with G*Power (Faul et al., 2007) assuming a medium effect of f = 0.25, a significance level of 0.05 and a power of 0.80 following the conventions of Cohen (1992), with a numerator degree of freedom of one, which yielded a required total sample of 128 persons, which is 32 per cell, and a targeted gross sample of about 150 to absorb exclusions. The design is therefore not powered to detect substantially smaller effects, so that a non-significant result cannot be read as evidence for the absence of an effect.

Analyses were conducted with R version 4.6.1 at $\alpha = 0.05$. Incomplete questionnaires and cases failing the attention check were excluded, and scale scores were formed as the mean of the respective items, with the attention check item not entering the trust scale. Differences between the four groups in the control variables were tested with one-way analyses of variance for the continuous variables and with a chi-square test for gender. For the first part of the manipulation check, Welch t tests compared the labelled and unlabeled conditions and, additionally, the two image types, since the Welch variant controls the type I error rate reliably even under unequal group variances (Delacre et al., 2017); the second part was analyzed with a chi-square test over the response categories, reported alongside Cramér's V. Hypotheses H2 and H3 were tested with three two-way analyses of variance, one per dependent variable, with H2 corresponding to the main effect of labelling and H3 to the interaction, and partial eta squared reported as the effect size measure (Richardson, 2011). H1 refers exclusively to the two unlabeled conditions and was tested with Welch t tests, reporting 95 per cent confidence intervals and Hedges g. Equality of error variances was examined with Levene tests (Levene, 1960), while moderate departures from normality were treated as uncritical given approximately equal group sizes, since the F test largely maintains its error rate under various departures from normality (Blanca et al., 2017). H4

International Journal of Applied Artificial Intelligence in Business (ISSN: 3056-1892)
Volume: 01 Issue: 01 Year: 2026
https://doi.org/10.51137/wrp.ijaaib.1062

was tested with three moderated regressions including labelling and image origin as coded factors, mean-centered AI attitude and the product term of labelling and AI attitude, with image origin serving as a control; R^2 and adjusted R^2 were reported for the models and the unstandardized coefficient, its standard error and the 95 per cent confidence interval for the product term.

Results

Sample and Data Preparation

Data were collected from 6 to 10 July 2026 via SoSci Survey. The study link was distributed through the study platforms SurveyCircle and SurveySwap, the social networks Facebook and Instagram, the discussion platform Reddit and an internal Teams channel for students. Of 520 recorded cases, which include all openings of the questionnaire, 161 participants reached the final page, while the remaining 359 cases were not included. Among the complete questionnaires, eight participants failed the attention check and were removed. Three further cases reported ages of 93, 93 and 95 years and thus exceeded the plausibility limit of 90 years applied in the analysis script; two of them additionally showed processing times of 38 seconds each and completely uniform response patterns across the central scale items, and careless responding can be indicated by control items, conspicuous entries and unusually short processing times (Meade & Craig, 2012). After data cleaning, 150 participants remained, exceeding the minimum of 128 established by the power analysis by 22 cases. Cell sizes were 36 for the unlabeled photograph, 36 for the unlabeled AI image, 38 for the labelled photograph and 40 for the labelled AI image, so that every condition exceeded the planned minimum of 32.

Mean age was 29.5 years (SD = 10.2) with a median of 26 years and a range from 16 to 71 years; 117 participants were younger than 35 years, 21 belonged to the 35 to 49 age group and 12 were at least 50 years old. Of the 150 participants, 76 identified as female and 71 as male, while three gave no information on gender and the category divers was not selected. The frequency of buying clothing online, measured on a seven-point scale from 1 (never) to 7 (very frequently), reached M = 4.38 (SD = 1.54), and prior exposure to AI-generated images, measured on a seven-point scale, reached M = 5.00 (SD = 1.70). The sample therefore combines regular online clothing purchasing with comparatively high prior exposure to AI-generated imagery.

Reliability and Descriptive Statistics

Internal consistency was satisfactory for all four multi-item scales. The trust scale reached $\alpha = 0.878$, perceived product quality $\alpha = 0.951$, purchase intention $\alpha = 0.949$ and AI attitude $\alpha = 0.880$, so that all scales exceeded the criterion of 0.80 and no items had to be removed. For the AIAS-4, the original study reports $\alpha = 0.902$ (Grassini, 2023), placing the present value in a comparable range. Because the items on trust, perceived quality and purchase intention were translated and partly adapted to the evaluation

WOHLLEBE & ROSS
Discover publication opportunities:
www.wr-publishing.org

International Journal of Applied Artificial Intelligence in Business (ISSN: 3056-1892)
Volume: 01 Issue: 01 Year: 2026
https://doi.org/10.51137/wrp.ijaaib.1062

of a single product representation, direct numerical comparisons with the original studies are only of limited value, and the coefficients indicate close association among the pooled items within this sample rather than content validity of the adapted measures. No responses were missing on the scale items, so that scale scores could be computed for all 150 participants.

Across all conditions, trust in the product representation reached M = 4.46 (SD = 1.21) and perceived product quality M = 4.44 (SD = 1.34), both slightly above the scale midpoint of four, while purchase intention was lower at M = 3.75 (SD = 1.64), just below the midpoint. General AI attitude, measured on a scale from one to ten, reached M = 6.87 (SD = 2.01), indicating a rather positive attitude on average with sufficient dispersion to examine the moderation effects formulated in H4. Table 1 reports means and standard deviations of the three dependent variables by condition. Descriptively, the unlabeled AI image attains the highest mean on all three variables, while the lowest values occur predominantly for the labelled AI image. For trust and perceived quality, the gap between the labelled and unlabeled AI conditions is larger than that between the two photograph conditions.

	Photo without label (*n* = 36)	AI without label (*n* = 36)	Photo with label (*n* = 38)	AI with label (*n* = 40)
Trust	4.50 (1.24)	4.71 (0.97)	4.53 (1.15)	4.13 (1.40)
Quality	4.57 (1.30)	4.79 (1.08)	4.35 (1.36)	4.10 (1.53)
Purchase Intention	3.89 (1.61)	4.01 (1.70)	3.59 (1.53)	3.54 (1.74)

Table 1: Means and standard deviations of consumer trust, perceived product quality and purchase intention by experimental condition.

Randomization and Manipulation Checks

The four groups did not differ significantly in age, $F(3, 146) = 0.60$, $p = 0.614$, in the frequency of buying clothing online, $F(3, 146) = 0.75$, $p = 0.524$, or in prior exposure to AI-generated images, $F(3, 146) = 0.61$, $p = 0.610$. For gender, the 147 participants identifying as female or male were considered, and no significant association between condition and gender emerged, $\chi^2(3) = 3.05$, $p = 0.385$. The analyses thus provide no indication of systematic differences between the groups on the characteristics examined.

In the first part of the manipulation check, the perception that the image had been generated by artificial intelligence reached M = 4.23 (SD = 2.06) in the labelled conditions and M = 3.65 (SD = 1.85) without

International Journal of Applied Artificial Intelligence in Business (ISSN: 3056-1892)
Volume: 01 Issue: 01 Year: 2026
https://doi.org/10.51137/wrp.ijaaib.1062

labelling. The difference of 0.58 points in the expected direction but does not reach the significance level, t(147.89) = 1.81, p = 0.072, 95 per cent confidence interval [-0.05, 1.21]. In the second part, 66 of 78 participants in the labelled conditions reported having noticed an origin notice, corresponding to 84.6 per cent, while nine reported no notice and three were unsure. In unlabeled conditions, five of 72 participants reported such a notice, corresponding to 6.9 per cent, with 60 answering no and seven unsure. The distribution of response categories differed significantly between labelled and unlabeled conditions, $\chi^2(2) = 91.61$, $p < 0.001$, Cramér's V = 0.78. Whether the actual image type influenced the attribution of AI origin was examined additionally: the photograph reached M = 4.04 (SD = 2.00) and the AI-generated image M = 3.87 (SD = 1.96) across both labelling conditions, a difference of 0.17 scale points that is not significant, t(147.62) = 0.53, p = 0.595, 95 per cent confidence interval [-0.47, 0.81]. The manipulation check thus yields a mixed picture: the visible notice was perceived by a large majority, while the label did not significantly change the assessment of image origin.

Hypothesis Tests

For H1a to H1c, mean differences and effect sizes are reported in the direction AI image minus photograph. Consumer trust reached M = 4.71 (SD = 0.97) for the AI image and M = 4.50 (SD = 1.24) for the photograph, a descriptive advantage of 0.21 scale points that is not significant, t(66.10) = 0.82, p = 0.418, 95 per cent confidence interval [-0.31, 0.74], Hedges g = 0.19. Perceived product quality reached M = 4.79 (SD = 1.08) for the AI image and M = 4.57 (SD = 1.30) for the photograph, a difference of 0.22 scale points, t(67.63) = 0.79, p = 0.433, 95 per cent confidence interval [-0.34, 0.78], Hedges g = 0.18. Purchase intention reached M = 4.01 (SD = 1.70) and M = 3.89 (SD = 1.61) respectively, a difference of 0.12 scale points, t(69.79) = 0.30, p = 0.765, 95 per cent confidence interval [-0.66, 0.89], Hedges g = 0.07. No significant difference between the two image types emerged on any dependent variable as long as image origin was not labelled. The descriptive differences are small and consistently favor the AI image, but the confidence intervals also include differences in the opposite direction. H1a, H1b and H1c are therefore not treated as confirmed, since a non-significant difference test does not establish equivalence of the image types, which would have required predefined equivalence bounds and a corresponding equivalence testing procedure (Lakens, 2017).

Levene tests indicated no violation of variance homogeneity for consumer trust, F(3, 146) = 1.00, p = 0.394, perceived product quality, F(3, 146) = 0.96, p = 0.413, or purchase intention, F(3, 146) = 0.42, p = 0.737. For consumer trust, the estimated marginal mean was 4.61 without labelling and 4.33 with labelling, a descriptive reduction of 0.28 scale points, with a main effect that is not significant, F(1, 146) = 2.04, p = 0.155, $\eta p^2 = 0.014$, 95 per cent confidence interval of the difference [-0.67, 0.11], so that H2a is not confirmed. For perceived product quality, the estimated marginal mean fell from 4.68 without labelling to 4.22 with labelling, a reduction of 0.46 scale points, and the main effect is significant, F(1, 146) = 4.50, p = 0.036, $\eta p^2 = 0.030$, 95 per cent confidence interval [-0.89, -0.03], confirming H2b while

WOHLLEBE & ROSS
Discover publication opportunities:
www.wr-publishing.org

International Journal of Applied Artificial Intelligence in Business (ISSN: 3056-1892)
Volume: 01 Issue: 01 Year: 2026
https://doi.org/10.51137/wrp.ijaaib.1062

indicating a limited effect size. For purchase intention, the estimated marginal mean fell from 3.95 to 3.57, a reduction of 0.38 scale points, with a main effect that does not reach significance, $F(1, 146) = 1.99$, $p = 0.160$, $\eta p^2 = 0.013$, 95 per cent confidence interval [-0.91, 0.15], so that H2c is not confirmed. The main effect of image origin was not significant for consumer trust, $F(1, 146) = 0.23$, $p = 0.636$, $\eta p^2 = 0.002$, perceived product quality, $F(1, 146) < 0.01$, $p = 0.945$, $\eta p^2 < 0.001$, or purchase intention, $F(1, 146) = 0.01$, $p = 0.909$, $\eta p^2 < 0.001$.

For H3a to H3c the interaction between image origin and labelling is decisive. For consumer trust, the photograph mean remained almost unchanged at $M = 4.50$ without and $M = 4.53$ with labelling, whereas trust in the AI image fell from $M = 4.71$ to $M = 4.13$, a difference between the two labelling effects of about 0.62 scale points; the interaction is not significant, $F(1, 146) = 2.44$, $p = 0.120$, $\eta p^2 = 0.016$. For perceived product quality, the photograph fell from $M = 4.57$ to $M = 4.35$ and the AI image from $M = 4.79$ to $M = 4.10$, corresponding to descriptive reductions of 0.22 and 0.69 scale points and a difference of about 0.48 points, with the interaction again not significant, $F(1, 146) = 1.19$, $p = 0.278$, $\eta p^2 = 0.008$. For purchase intention, the photograph fell from $M = 3.89$ to $M = 3.59$ and the AI image from $M = 4.01$ to $M = 3.54$, a difference of about 0.17 scale points, with a very small and non-significant interaction, $F(1, 146) = 0.10$, $p = 0.751$, $\eta p^2 = 0.001$. H3a, H3b and H3c are therefore not confirmed.

For H4a to H4c, the product term of labelling and AI attitude is decisive. The model for consumer trust is significant overall, $F(4, 145) = 3.80$, $p = 0.006$, $R^2 = 0.095$, adjusted $R^2 = 0.070$, while the product term is close to zero and not significant, $b = 0.001$, $SE = 0.096$, $t = 0.01$, $p = 0.989$, 95 per cent confidence interval [-0.188, 0.190], so that H4a is not confirmed. The model for perceived product quality is likewise significant overall, $F(4, 145) = 5.22$, $p < 0.001$, $R^2 = 0.126$, adjusted $R^2 = 0.102$, and the product term yields $b = 0.193$, $SE = 0.104$, $t = 1.85$, $p = 0.066$, 95 per cent confidence interval [-0.013, 0.399]; the positive sign points in the direction expected by H4b, but the p value exceeds the significance level and the confidence interval includes zero, so that H4b is not confirmed. The model for purchase intention is not significant overall, $F(4, 145) = 1.17$, $p = 0.327$, $R^2 = 0.031$, adjusted $R^2 = 0.004$, and the product term is likewise not significant, $b = 0.164$, $SE = 0.134$, $t = 1.22$, $p = 0.224$, 95 per cent confidence interval [-0.101, 0.429], so that H4c is not confirmed. Of the twelve sub-hypotheses, only H2b is confirmed.

Discussion

Interpretation of the Findings

The findings on H1 argue against a clearly recognizable evaluative disadvantage of the AI-generated image as long as its origin is not disclosed. The AI image even attained slightly higher means on all three variables, but the differences were small and not significant, and the confidence intervals include differences in both directions, so that neither an advantage of the AI image nor full equivalence of the

WOHLLEBE & ROSS
Discover publication opportunities:
www.wr-publishing.org

image types can be inferred. The manipulation check complements this: participants did not attribute artificial origin to the AI-generated image more frequently than to the photograph, which suggests that the AI image contained no unambiguous visible features by which its origin could be reliably identified and that the two stimuli were sufficiently similar in their perceived design. The check does not establish that the images were perceived as fully identical, only that no reliable recognition was possible at group level. The finding is bounded by what was tested, namely one concrete image pair, one garment and one model.

A clearer finding emerges for labelling. Perceived product quality was on average 0.46 scale points lower in the labelled conditions, a statistically detectable but limited effect at $\eta p^2 = 0.030$. A plausible explanation lies in the close link between the image shown and the quality judgement. No further product information was available in the experiment, with details on material, brand, price or customer reviews absent, so that the visual representation formed the central basis for assessing product quality. If this representation is marked as AI-generated, its reliability as a source of information can be called into question, and participants may have assumed that the image renders the actual product less faithfully. Such a mechanism is compatible with the results but was not measured directly.

For consumer trust and purchase intention the estimated marginal means were likewise lower under labelling, but the differences were not significant, so that a general negative labelling effect on all outcome variables cannot be established. It is conceivable that the quality judgement responds more immediately to the product representation than trust or purchase intention. Trust also relates to the retailer, about whom participants received hardly any information, while purchase intention additionally incorporates personal preferences, need, price expectations and expected fit, none of which were contained in the stimulus, and purchase intention showed greater dispersion, making a difference harder to detect statistically.

Interpretation of the labelling effect requires a distinction between noticing the notice and being convinced of AI origin. A large majority in the labelled conditions recalled the visible notice, yet the label did not significantly change the assessment of whether the image was in fact AI-generated. Because of random assignment, the decline in quality ratings can be attributed within the experiment to the display of the label, whereas the reason for the decline remains open: the effect need not rest on a firm belief in AI origin, since the label may also have acted as a warning signal, triggered uncertainty or reduced the perceived credibility of the representation. The main effect of labelling was moreover computed across both image types and therefore describes the effect of a visible AI notice within the chosen design rather than exclusively the effect of a truthful disclosure of an actually AI-generated image.

The interaction between image origin and labelling was not significant for any dependent variable. For trust and perceived quality, the descriptive decline after labelling was larger for the AI image than for the

International Journal of Applied Artificial Intelligence in Business (ISSN: 3056-1892)
Volume: 01 Issue: 01 Year: 2026
https://doi.org/10.51137/wrp.ijaaib.1062

photograph, so that the mean patterns point in the direction expected by H3 without supporting the hypotheses. Conversely, the non-significant interaction tests do not permit the conclusion that the effect is exactly equal for both image types, since smaller image-type-specific differences cannot be ruled out with the available sample. Nor could a moderating role of general AI attitude be established: only perceived product quality showed an interaction term in the expected direction at $p = 0.066$, which is not treated as confirmation or partial confirmation, and any such relationship would be correlational because AI attitude was measured rather than manipulated.

Relations to Prior Research

The absence of a difference between the unlabeled image types is broadly consistent with Nightingale and Farid (2022), whose participants distinguished synthetic and real faces at an average hit rate of 48.2 per cent and rated AI-generated faces as somewhat more trustworthy. In the present experiment, too, the AI-generated image was not more frequently assigned artificial origin than the photograph. A direct comparison nevertheless remains limited, since Nightingale and Farid examine isolated faces whereas the present study uses a complete model image as the basis for a product evaluation, so that the present work transfers the question of detectability to fashion e-commerce without establishing general indistinguishability of photographic and AI-generated product images.

At first sight the results diverge from Kim (2026), where AI models in social media advertising increase advertising skepticism, which in turn impairs brand support, recommendation intention and purchase intention. That study centers on a recognizable brand and on societal ideas about the use of artificial models, such as ethical concerns or possible job losses, whereas the present study shows a neutral product representation without brand reference and without information on the consequences of AI use. The divergent findings suggest that negative reactions do not depend on the technical mode of production alone but also on whether AI origin and the associated ideas become salient in the evaluation situation.

The negative labelling effect on perceived product quality fits into research on the drawbacks of AI disclosure. Grigsby et al. (2025), Schilke and Reimann (2025) and Bui (2026) show in different contexts that disclosing AI use can impair trust, advertising value or purchase intention, though their explanations differ. Schilke and Reimann attribute the loss of trust to lower perceived legitimacy of the disclosing actor, Bui identifies lower advertising credibility as a partial mediator, and Grigsby et al. emphasize the function of the image element concerned, with disclosure acting particularly unfavorably when AI has generated an intangible element such as the depicted person. The quality decline observed here is compatible with these accounts, since the model image forms the central informational basis for judging the garment and a visible AI notice can raise doubts about whether material, fit or workmanship are

WOHLLEBE & ROSS
Discover publication opportunities:
www.wr-publishing.org

rendered faithfully. Which psychological process actually carries the effect cannot be determined, as neither credibility nor authenticity nor legitimacy was measured as a mediator.

That no significant labelling effect on consumer trust emerged diverges in particular from Schilke and Reimann (2025), where evaluation is directed at persons and organizations that deploy AI to complete a task. In the present experiment no concrete retailer is recognizable as an acting party, information on brand, service or prior experience is absent and the product is immediately visible, and this different referent may explain why the detectable effect appears for product quality rather than for trust. Bui (2026) additionally reports a declining purchase intention alongside lower advertising value; the means here likewise point in a negative direction without reaching the significance level, which may reflect the reduced purchase situation in which price, brand, material information, customer reviews and further product shots are absent. Shi and Jiang (2026) show that AI notices need not have exclusively negative consequences, since disclosure can raise perceived novelty while reducing authenticity. The quality decline found here is compatible with the negative authenticity path, but as neither authenticity nor novelty was measured it remains open whether either mechanism is responsible and opposing reactions on trust and purchase intention may in part have offset one another.

The non-confirmed interaction hypotheses agree only partly with Grigsby et al. (2025), from whose results the expectation was derived that labelling harms more when it concerns an artificially generated person. A substantive difference lies in the attribution of disclosure: Grigsby et al. relate AI use to specific elements of the advertisement, whereas the notice used here marked the entire image, leaving open for participants whether the information referred to the model, the garment or the overall design. Whether an element-specific label acts differently cannot be derived from the present data. For the moderation, Schilke and Reimann (2025) and Bui (2026) report that a more positive attitude toward technology or AI can attenuate negative disclosure effects, and Balaban and Phua (2026) show a moderating influence of AI attitude, although for the relationship between recognized AI origin and brand credibility and thus not for the same interaction. The present finding is compatible with this moderation logic without confirming it, and the divergence may stem from differences in the application context, in the outcome variables or in the strength of the underlying labelling effect.

The labelling question gains additional practical weight through Article 50 of Regulation (EU) 2024/1689, which distinguishes technical marking obligations for providers of generative systems from visible disclosure obligations for deployers and does not prescribe visible labelling for every AI-generated image, focusing instead on content classified as a deepfake, with the European Commission specifying these criteria in its guidelines (European Commission, 2026). The notice mark used in the experiment therefore does not fully replicate a specific statutory labelling duty. It examines the effect of a visible origin notice in an environment in which transparent communication of synthetic content is gaining

International Journal of Applied Artificial Intelligence in Business (ISSN: 3056-1892)
Volume: 01 Issue: 01 Year: 2026
https://doi.org/10.51137/wrp.ijaaib.1062

regulatory importance, and neither a general labelling duty for AI-generated product images nor a recommendation to forgo legally required disclosure follows from the observed quality decline.

Implications

For retailers in fashion e-commerce, two conclusions follow. The first concerns the deployment decision itself. For the model image examined, there is no statistically detectable disadvantage relative to the photographic representation emerged without labelling, so that high-quality AI-generated model images are in principle a viable alternative to conventional photo production and the production advantages described above can be realized. Because the finding rests on a single image pair, it cannot be extended without qualification to all products and forms of representation. The second conclusion concerns labelling. Where a visible origin notice is legally required or used voluntarily, the observed evaluative discount is practically relevant, since average quality ratings were 0.46 scale points lower under labelling, even though the effect is limited at $\eta p^2 = 0.030$. The notice should therefore not be regarded as an effect-neutral piece of additional information. Since no statistically detectable difference in labelling effects between the AI image and the photograph emerged, attention shifts to the design of the notice. Wording, placement and depth of explanation were not the subject of this study, so no specific labelling format can be recommended, and retailers should systematically test legally reviewed variants before broad rollout. Correct attribution of the notice also requires care, since quality ratings were descriptively lower even for the incorrectly labelled photograph.

Three theoretical contributions follow. First, the study transfers prior findings on AI disclosure to product images in fashion e-commerce, an application in which the image constitutes a central source of information about the offered product and shows that image origin without labelling had no statistically detectable effect while visible labelling reduced perceived product quality. Second, the absent interaction between image origin and labelling suggests that the stronger effect for intangible image elements described by Grigsby et al. (2025) may depend on whether viewers can attribute the AI component to a specific part of the representation, since the notice here referred to the whole image while the AI image itself contained no clearly recognizable origin cues. Third, the design separates three levels that often coincide in prior studies: the actual, the assumed and the communicated image origin. This separation made it visible that actual image type produced no detectable evaluative difference without labelling, whereas the visible notice changed quality perception even though it did not clearly shift beliefs about origin. A photographic comparison condition and a manipulation check that separately records noticing the notice and believing its content are therefore useful elements for future studies.

Limitations

Several limitations qualify the results. First, the non-significant findings for H1 establish only that no detectable difference was present in this sample, not equivalence of the image types, which would have

required predefined equivalence bounds and an equivalence testing procedure (Lakens, 2017). Second, statistical power was calibrated to medium effects, so that smaller differences, and in particular the interaction between image origin and labelling and the moderation by AI attitude, would have required a larger sample; the moderation effect on perceived product quality lay in the expected direction at $p = 0.066$ without permitting a conclusion either way. Third, the significant labelling effect on perceived product quality is limited in size at $\eta p^2 = 0.030$, and because several hypotheses were tested across three dependent variables without a formal correction for multiple testing being foreseen in the analysis plan, the quality finding should be understood as an initial indication requiring confirmation.

Fourth, the manipulation succeeded with respect to visibility but only partly with respect to origin beliefs, since the notice was noticed by a large majority yet did not significantly change the assessment of whether the image was AI-generated, leaving open whether the quality decline arose from belief in artificial origin, from general uncertainty or from lower perceived credibility of the representation. Fifth, the counterfactual labelling of the photograph, although methodologically necessary to separate the effect of the notice from that of actual image type, does not correspond to a real purchase situation, so that the labelling effect computed across both image types does not fully match the effect of truthful labelling of genuinely AI-generated product images.

Sixth, all results rest on a single image pair, one T-shirt and one male model. The two images were closely aligned in design but did not show the same person, and although the pretest with ten participants indicated high perceived similarity, its small size means that an influence of remaining differences in face, expression or appearance cannot be fully excluded, and transferability to other garments, female models, different image styles or lower-quality AI representations remains open. Seventh, external validity is constrained by sample composition, since participants were recruited via social networks and study platforms, the majority were younger than 35 years and average prior exposure to AI-generated images was comparatively high, so that older or more AI-skeptical target groups may respond differently. Eighth, the study depicts a strongly reduced purchase situation without price, brand, material information, customer reviews or further product shots, and it measured purchase intention rather than actual purchasing behaviour. Ninth, the measures are limited in their informational value, since the scales were derived from existing instruments, translated and adapted to the evaluation of a single product representation; their high alpha coefficients show close association among items within the sample but do not establish that the adapted scales capture all content facets of the constructs, and consumer trust in particular would in a real online shop be shaped more strongly by retailer, brand and further information absent from the experiment.

WOHLLEBE & ROSS

Discover publication opportunities:
www.wr-publishing.org

International Journal of Applied Artificial Intelligence in Business (ISSN: 3056-1892)
Volume: 01 Issue: 01 Year: 2026
https://doi.org/10.51137/wrp.ijaaib.1062

Conclusion

This study examined how AI-generated and photographically produced product images act on consumer trust, perceived product quality and purchase intention in fashion e-commerce, and what influences a visible label as AI-generated exerts. For the unlabeled conditions, no statistically significant difference between the AI-generated product image and the photograph emerged on any of the three outcome variables, so that the AI image used showed no detectable evaluative disadvantage in the purchase situation examined, although no full equivalence can be inferred in the absence of equivalence tests. The visible notice of a purported AI origin significantly lowered perceived product quality, while trust and purchase intention were also rated lower under labelling without reaching the significance level. The labelling effect did not differ detectably between the actual AI image and the photograph, and no sufficient evidence emerged that a more positive attitude toward AI attenuates the negative effect of labelling.

The research question is therefore answered in a differentiated way. For the product image examined, it was not actual image origin that carried a recognizable evaluative disadvantage but the display of a visible origin notice, whose negative effect concentrated on perceived product quality. The results thus indicate that the technical generation of a product image and the communication of its origin must be distinguished. For retailers, this shifts the question away from a fundamental decision against AI-generated product images and toward how a required or voluntary transparency notice is designed and embedded in product communication.

Further research is needed on several fronts. Future studies should use predefined equivalence bounds to test whether practically relevant differences between AI-generated and photographic product images can be excluded and should examine whether the results transfer to other product categories, poses, models and to older or more AI-skeptical target groups. More realistic purchase environments including price, brand, additional product information and actual or near-realistic purchase decisions could show whether the observed quality discount extends beyond measured purchase intention. The label itself deserves particular attention, specifically its wording, placement and depth of explanation, as well as the question of whether a notice referring to individual AI-generated image elements acts differently from a blanket label on the entire image. Finally, it remains open how the effect of the notice develops as AI-generated product images become more commonplace and labelling loses its novelty value.

References

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

WOHLLEBE & ROSS

Discover publication opportunities:
www.wr-publishing.org

Akerlof, G. A. (1970). The market for "Lemons": Quality uncertainty and the market mechanism. The Quarterly Journal of Economics, 84(3), 488–500. https://doi.org/10.2307/1879431

Balaban, D. C., & Phua, J. (2026). Is this real or fake? Examining the effects of disclosure of AI-generated content and source type in Instagram-based travel destination advertising on consumer attitudes and behavioral intentions. Computers in Human Behavior, 183, 109047. https://doi.org/10.1016/j.chb.2026.109047

Blanca, M., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? Psicothema, 29(4), 552–557. https://doi.org/10.7334/psicothema2016.383

Brislin, R. W. (1970). Back-translation for cross-cultural research. Journal of cross-cultural psychology, 1(3), 185–216. https://doi.org/10.1177/135910457000100301

Bui, H. T. (2026). Examining the effect of AI advertising involvement disclosure on advertising value and purchase intentions. Journal of Research in Interactive Marketing, 20(5), 613–632. https://doi.org/10.1108/JRIM-02-2025-0066

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

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. Journal of Economic Behavior & Organization, 81(1), 1–8. https://doi.org/10.1016/j.jebo.2011.08.009

Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. Journal of Marketing Research, 16(1), 64–73. https://doi.org/10.1177/002224377901600110

CNN. (2025). Fashion giant H&M plans to use AI clones of its models. Not everyone is happy. CNN. https://www.cnn.com/2025/03/28/style/h-and-m-ai-models-intl-scli

Cohen, J. (1992). A power primer. Psychological Bulletin, 112(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297–334. https://doi.org/10.1007/BF02310555

Delacre, M., Lakens, D., & Leys, C. (2017). Why psychologists should by default use Welch's t-test instead of Student's t-test. International Review of Social Psychology, 30(1), 92–101. https://doi.org/10.5334/irsp.82

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. Journal of Experimental Psychology: General, 144(1), 114–126. https://doi.org/10.1037/xge0000033

Dodds, W. B., Monroe, K. B., & Grewal, D. (1991). Effects of price, brand, and store information on buyers' product evaluations. Journal of Marketing Research, 28(3), 307–319. https://doi.org/10.1177/002224379102800305

Döring, N., & Bortz, J. (2016). Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften (5th ed.). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41089-5

European Commission. (2026). Guidelines on the implementation of the transparency obligations for certain AI systems under Article 50 of the AI Act. https://digital-strategy.ec.europa.eu/en/library/guidelines-transparency-obligations-providers-and-deployers-ai-systems

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39(2), 175–191. https://doi.org/10.3758/BF03193146

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. Management Information Systems Quarterly, 27(1), 51–90. https://doi.org/10.2307/30036519

Grassini, S. (2023). Development and validation of the AI attitude scale (AIAS-4): A brief measure of general attitude toward artificial intelligence. Frontiers in Psychology, 14. https://doi.org/10.3389/fpsyg.2023.1191628

Grigsby, J. L., Michelsen, M., & Zamudio, C. (2025). Service ads in the era of generative AI: Disclosures, trust, and intangibility. Journal of Retailing and Consumer Services, 84, 104231. https://doi.org/10.1016/j.jretconser.2025.104231

H&M Group. (2025). H&M continues its exploration of creativity with AI. https://hmgroup.com/news/hm-continues-its-exploration-of-creativity-with-ai/

Jiang, Z., & Benbasat, I. (2004). Virtual product experience: Effects of visual and functional control of products on perceived diagnosticity and flow in electronic shopping. Journal of Management Information Systems, 21(3), 111–147. https://doi.org/10.1080/07421222.2004.11045817

Kamali, N., Nakamura, K., Kumar, A., Chatzimparmpas, A., Hullman, J., & Groh, M. (2025). Characterizing photorealism and artifacts in diffusion model-generated images. Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems, CHI '25, 1–26. https://doi.org/10.1145/3706598.3713962

Kim, S. (2026). AI vs. human models in fashion brand advertising: A schema theoretical perspective on consumer responses. Journal of Retailing and Consumer Services, 92, 104820. https://doi.org/10.1016/j.jretconser.2026.104820

Łabuz, M. (2025). A teleological interpretation of the definition of DeepFakes in the EU Artificial Intelligence Act—A purpose-based approach to potential problems with the word "existing". Policy & Internet, 17(1), e435. https://doi.org/10.1002/poi3.435

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. Social Psychological and Personality Science, 8(4), 355–362. https://doi.org/10.1177/1948550617697177

Levene, H. (1960). Robust tests for equality of variances. In Contributions to probability and statistics: Essays in honor of Harold Hotelling (pp. 278–292). Stanford University Press.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

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

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

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. Psychological Methods, 17(3), 437–455. https://doi.org/10.1037/a0028085

Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. Proceedings of the National Academy of Sciences, 119(8), e2120481119. https://doi.org/10.1073/pnas.2120481119

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. Journal of Experimental Social Psychology, 45(4), 867–872. https://doi.org/10.1016/j.jesp.2009.03.009

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. American Psychologist, 17(11), 776–783. https://doi.org/10.1037/h0043424

Otto. (2025, September 17). Künstliche Intelligenz löst Modelfotografie bei OTTO ab. https://www.otto.de/unternehmen/de/presse/künstliche-intelligenz-löst-modelfotografie-bei-otto-ab

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence and Amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA Relevance) (2024). http://data.europa.eu/eli/reg/2024/1689/oj

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. Educational Research Review, 6(2), 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Schilke, O., & Reimann, M. (2025). The transparency dilemma: How AI disclosure erodes trust. Organizational Behavior and Human Decision Processes, 188, 104405. https://doi.org/10.1016/j.obhdp.2025.104405

Shi, Y., & Jiang, Z. (2026). Consumer responses to AI disclosure labels: The role of novelty and authenticity. Sage Open, 16(1), 21582440261417793. https://doi.org/10.1177/21582440261417793

Zeithaml, V. A. (1988). Consumer perceptions of price, quality, and value: A means-end model and synthesis of evidence. Journal of Marketing, 52(3), 2–22. https://doi.org/10.1177/002224298805200302

Zhang, Y., & Liu, C. (2026). AI-driven consumer research in fashion: A systematic and bibliometric review (2022–2025) and future research agenda. Journal of Theoretical and Applied Electronic Commerce Research, 21(3), 74. https://doi.org/10.3390/jtaer21030074

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. Experimental Economics, 13(1), 75–98. https://doi.org/10.1007/s10683-009-9230-z

WOHLLEBE & ROSS

Discover publication opportunities:
www.wr-publishing.org