“Okay, ChatGPT, that was actually funny”. A systematized review of AI-involvement cues in humorous advertising.

Martin Kasperski[a*], Benjamin Österle[b], Jörg Henseler[a,c] and Sue Rossano- Rivero[d]

[a] Faculty of Engineering Technology, University of Twente, Enschede, The Netherlands

[b] Heilbronn University of Applied Sciences, Schwäbisch Hall, Germany

[c] NOVA Information Management School, Universidade Nova de Lisboa, Lisbon, Portugal

*Corresponding author: Martin Kasperski, Ph.D. Candidate, Faculty of Engineering Technology, University of Twente, Drienerlolaan 5, 7522 NB Enschede, The Netherlands.

Email: martin.kasperski@hsnr.de

Funding: This work was supported by national funds through FCT (Fundação para a Ciência e a Tecnologia), under the projects UID/04152/2025 – Centro de Investigação em Gestão de Informação (MagIC)/NOVA IMS (https://doi.org/10.54493/UID/04152/2025; 2025-01-01 to 2028-12-31) and UID/PRR/04152/2025 (https://doi.org/10.54493/UID/PRR/04152/2025; 2025-01-01 to 2026-06-30).

[d] Hochschule Niederrhein University of Applied Sciences, Mönchengladbach, Germany

“Okay, ChatGPT, that was actually funny”. A systematized review of AI-involvement cues in humorous advertising.

Abstract

Despite the rapid adoption of generative AI in advertising, evidence on audience responses to AI-generated humorous ads remains fragmented. This systematised integrative review applies the Theory-Context-Characteristics-Method (TCCM) mapping with stimulus-organism-response (SOR) and elaboration-likelihood-model (ELM) syntheses to explain how AI-involvement cues alter humorous persuasion. Findings converge on four mechanisms: (1) authorship and disclosure cues shape authenticity inferences and source legitimacy. (2) When AI-involvement is apparent, people are more likely to activate persuasion knowledge, increasing scepticism and shifting attention from enjoying the message to scrutinising its underlying intent. (3) Humour outcomes depend on affective appraisal and appropriateness, determining benign mirth versus offence. (4) Platform dynamics translate entertainment value into engagement while accelerating backlash and brand-safety risk. Across mechanisms, effects depend on disclosure salience and framing, humour strength, audience AI-literacy, and cultural offence thresholds. We translate these insights into the Humorous AI advertising (HAIA) model and propositions for experiments on AI-generated humorous ads.

Keywords: AI-generated advertising, humour in AI advertising, AI disclosure, persuasion knowledge, ELM, SOR

Introduction

Humour is a high-impact but highly variable advertising strategy. It can capture attention, improve recall, and foster positive attitudes, yet it can also backfire if audiences perceive a joke as inappropriate, confusing, or distracting (Weinberger & Gulas, 1992; Eisend, 2011; Warren et al., 2021). Online environments amplify both sides of this trade-off because humorous ads are

consumed quickly, publicly commented on, and redistributed through platform algorithms that prioritize engagement over brand safety (Vermeulen et al., 2019)."Because of its complexity, some consider the understanding of humour as the last remaining Turing test for artificial intelligence" (Weinberger et al., 2021, p. 1). Generative AI adds a new dimension: on the production side it reduces creative costs and scales humour, while on the reception side it introduces an authorship cue that can change how audiences interpret the same joke. When AI involvement is apparent, audiences may activate persuasive awareness, reassess effort and accountability, and see the message as more manipulative (Kirk & Givi, 2025; Rezwana & Maher, 2023). Early evidence shows that perceived AI use can shift content evaluations, purchase intentions, and judgments of authenticity, eeriness, and credibility, while introducing distinct risks around offensiveness and norm violations (Gu et al., 2024; Rae, 2024; Shambhavi et al., 2025; Wu et al., 2025). Yet findings remain dispersed across humour-in-advertising research, persuasion knowledge, disclosure research, consumer responses to AI-generated content, and adjacent human-AI interaction work, leaving a gap in understanding when AI-generated humour engages audiences versus triggers scepticism or backlash (Kariyawasam et al., 2024; Karpińska-Krakowiak et al., 2025; Kim & Chilton, 2025; Lu & Yuan, 2024). AI labels and disclaimers can reduce trust, shift authenticity judgments, and activate coping responses (Koning & Voorveld, 2025; Kirkby et al., 2023) — but these effects are not universal: when AI is framed as assistance rather than replacement, disclosure can be neutral or even protective (Haupt et al., 2025). Humour effects depend on consumer judgements and emotional reactions - amusement, surprise, irritation, or moral disgust (Strick et al., 2017; Martin & Ford, 2018). AI cues can influence these appraisals by shaping what audiences perceive as appropriate, authentic, and socially acceptable. This is especially relevant for sensitive humour, where bias or offensiveness can generate backlash that spreads rapidly (Saumure et al., 2025; Voutsa et al., 2025). Although the topic is increasingly relevant for marketing practice, the evidence base remains fragmented across humour, disclosure, deepfake, and platform-engagement research.

Individual papers examine specific stimuli but findings are rarely integrated into a mechanism-based account (Campbell et al., 2022; Rae, 2024). Hence, two questions arise: How do AI involvement cues such as disclosure salience, creator labels, perceived agency, and anthropomorphism shape consumer responses to humorous advertising, and which humour-related mechanisms and boundary conditions determine when AI-generated humour drives engagement versus scepticism or backlash? This paper synthesizes emerging empirical evidence on AI in humorous advertising and offers an integrative explanation of how AI involvement cues influence downstream outcomes. We conduct a systematic integrative review (Snyder, 2019) and combine findings that align the Stimulus-Organism-Response (SOR) framework from environmental psychology (Mehrabian & Russell, 1974) with the route-based processing approach of the elaboration likelihood model (ELM) (Petty & Cacioppo, 1986). Additionally, we draw on persuasion knowledge theory to explain when AI cues lead to scepticism and defensive responses (Friestad & Wright, 1994; Campbell & Kirmani, 2008).

Methodology

We followed a systematised integrative review approach suited to emergent and heterogeneous evidence (Snyder, 2019). The aim was to consolidate early empirical findings on AI involvement in humorous advertising and translate them into a coherent mechanism structure to guide experiments and theory-building. Searches were conducted in Web of Science, Scopus, and Google Scholar using combinations of terms related to generative AI, advertising, and humour. We included empirical studies published 2021–2025 examining audience responses to AI-generated humorous advertising. The searches yielded 319 records in December 2025. After deduplication, two coders independently screened 282 records, 22 full texts were retrieved, and 16 studies met the final inclusion criteria. Inter-rater agreement was Cohen's $\kappa = 0.83$. Disagreements were resolved through discussion, and eligibility criteria were not amended during screening. TCCM and SOR extraction for the 16 included studies was performed by the

first author. The PRISMA flow according to Liberati et al. (2009) is summarised in Figure 1. [Figure 1 near here]

For each included study, we extracted theory, context, characteristics, and method attributes (TCCM) (Paul & Rosado-Serrano, 2019), coded stimulus operationalisations (disclosure salience, cue modality, collaboration framing), organismic appraisals (authenticity, trust, eeriness, moral emotions), and responses (attitudes, intentions, engagement metrics). The extracted evidence was synthesized into four recurring mechanism clusters, then mapped into an ELM-SOR interpretation that explains why the same humour can be processed as playful entertainment in one setting but as suspicious persuasion in another.

Findings: Mechanism clusters

The included studies are recent and methodologically diverse, examining short-form digital ad content (social media posts, parodies, memes) with AI involvement signalled through disclaimers, creator labels, or visible AI artefacts (Koning & Voorveld, 2025; Exner et al., 2025). Across the studies reviewed, four mechanism clusters recur. These explain how AI cues shape how audiences interpret humorous ads, influencing their judgements of legitimacy, their processing of the message, their perception of humour, and platform engagement dynamics. Cluster 1 (C1) captures inferences about the *reliability of the source* (authenticity, credibility, effort, accountability), whereas Cluster 2 (C2) captures inferences about the *persuasive intent behind the message* (manipulation, marketer tactics, ulterior motives). The two clusters are causally ordered rather than overlapping: as specified in P2, source-legitimacy concerns (C1) precede and trigger persuasion-knowledge activation (C2), consistent with Friestad and Wright (1994).

Cluster 1: Authorship and source legitimacy

AI cues often serve as authorship signals and influence how audiences evaluate the source's credibility. When viewers learn that humour was generated by AI, they may infer lower creative effort, weaker intentionality, or reduced accountability. This can reduce perceived authenticity and trust, thereby dampening favourable ad and brand responses (Kirk & Givi, 2025; Koning & Voorveld, 2025). Disclosure effects vary with cue visibility and the way the disclosure is framed. Explicit AI labels harm authenticity when the disclosure suggests replacing a human creator, but not when AI is framed as a tool supporting a consistent brand voice. In prosocial or cause-related contexts, legitimacy can be especially delicate because perceived sincerity and credibility are crucial. AI-generated prosocial ads have been shown to decrease positive attitudes due to lower perceived credibility (Baek et al., 2024). Visual cues also play a role. When the content shows obvious AI artifacts, viewers might associate it with automation even without a label, which can influence perceptions of quality and credibility (Exner et al., 2025). In humour contexts, legitimacy depends on whether the brand is considered entitled to joke: AI authorship can be seen as awkward when the humour relies on social nuance, but as novel when framed as playful experimentation, explaining why some disclosures appear neutral while others increase scepticism.

Cluster 2: Persuasion Knowledge activation and cognitive defence

AI disclosures can activate persuasion knowledge and trigger cognitive defence. When audiences recognise an AI-generated ad, they may interpret it as more strategic or manipulative, shifting from enjoyment to scrutiny. This reduces the likelihood that humour functions as a simple peripheral cue and increases counterarguing or resistance (Friestad & Wright, 1994; Campbell & Kirmani, 2008). Evidence is particularly consistent in contexts where the ad invites moral evaluation, such as cause-related or charitable appeals. In these cases, AI involvement can raise questions about sincerity and underlying motives, which may reduce trust and

behavioural intentions (Arango et al., 2023; Qiu et al., 2025). Humorous disclosures can sometimes buffer this effect by framing the AI cue itself as part of the joke and by signalling transparency in a lighter tone (Karpińska-Krakowiak et al., 2025). Defensive processing intensifies among audiences with higher AI literacy or stronger prior concern about manipulation. In parody or deepfake settings, disclaimers reduce perceived deception when viewers grasp the parody intent, but trigger suspicion when comprehension is low (Lu & Yuan, 2024).

Cluster 3: Affective humour processing and appropriateness

Humour succeeds when audiences understand the unexpected element of the joke and perceive it as harmless and amusing rather than uncomfortable. AI cues can shift this appraisal by changing what is seen as appropriate and by affecting emotional attribution along pleasure, arousal, and dominance dimensions of affective response (Sallaku et al., 2025). If a joke is interpreted as socially tone-deaf or biased, negative emotions can dominate and brand attitudes decline (Warren et al., 2021). Several studies connect AI involvement with heightened sensitivity to offensiveness and bias in humorous ads, and recent evidence indicates that AI-generated jokes can be rated less funny than human-authored ones when authorship is disclosed (Gorenz & Schwarz, 2024). Exploratory work suggests that when humour targets gender or other social categories, AI generations can amplify scepticism and perceived unfairness, especially among younger audiences (Voutsa et al., 2025). Related evidence indicates that humour can expose and reproduce generative AI biases, which can increase moral disgust and backlash risk (Saumure et al., 2025). Appropriateness is context-dependent. Meme-style humour may tolerate roughness or absurdity and can reduce expectations of polished authorship. In contrast, humour tied to sensitive brand values or high-stakes categories offers less room for error. Studies on offensiveness detection in meme ads highlight that subtle wording changes can shift a joke from playful to harmful, suggesting that AI-generated humour

requires careful testing of boundaries before deployment (Shambhavi et al., 2025; Strick et al., 2017). Recent work also indicates that AI may favour self-enhancing humour effective in negative-valence contexts, suggesting that affective appraisal is sensitive to the surrounding emotional frame (Cao et al., 2025)

Cluster 4: Entertainment value and platform engagement

Beyond persuasion, humorous ads are often judged by entertainment value and platform engagement outcomes. This cluster captures how AI involvement can shape attention, sharing, commenting, and algorithmic amplification. In fast-scroll contexts, novelty and humour strength can drive engagement even when deeper trust metrics are mixed (Rae, 2024; Wu et al., 2025), and meme-style social-media-native formats produce engagement effects distinct from those of conventional ad formats (Kim & Chilton, 2025). Platform research on humorous video formats shows that timing and intrusiveness influence acceptance, which translates into design constraints for AI-generated humour (Vermeulen et al., 2019). For example, a prominent AI disclaimer can reduce deception concerns but also interrupt comedic rhythm, lowering enjoyment and sharing. Engagement effects are also culturally bounded. Offence thresholds for humour, especially around gender, race, and political identity, vary substantially across markets, and AI-generated content may amplify cross-cultural mismatches when training data reflects one cultural register but the audience holds another (Saumure et al., 2025; Voutsa et al., 2025). When AI-generated humour produces ambiguous or stereotyped content, backlash can become a platform-native response outcome rather than a rare edge case (Voutsa et al., 2025).

ELM-SOR synthesis

To integrate the clusters, we map AI involvement cues into a Stimulus-Organism-Response framework and specify how these links can vary across ELM routes. In this synthesis, the stimulus is not just humour, but humour combined with an AI involvement cue such as a disclaimer, creator label, collaboration framing, or visual artefact (Koning & Voorveld, 2025;

Exner et al., 2025). Organismic states include the audience's perception of authenticity and legitimacy, activation of persuasion knowledge, and emotional responses to humour such as amusement, eeriness, or moral emotions (Gu et al., 2024; Kirk & Givi, 2025). Responses encompass attitudinal outcomes and platform-specific engagement behaviours (Wu et al., 2025). Table 1 summarizes this synthesis. [Table 1 near here]

ELM helps explain when humour operates as a peripheral cue and when it becomes part of deliberative evaluation. Under low involvement, humour and novelty drive engagement. Under higher involvement, AI cues activate persuasion knowledge or credibility concerns, shifting processing toward the central route and increasing the weight of authenticity and perceived intent (Petty & Cacioppo, 1986; Friestad & Wright, 1994). Across studies, three contingencies recur. First, framing of disclosure matters. Transparency framed as responsible tool use is less likely to trigger resistance than replacement framing (Haupt et al., 2025). Second, humour strength and style matter. Strong, clearly benign humour can buffer scrutiny, whereas edgy humour magnifies the risk of inappropriateness (Warren et al., 2021). Third, audience and cultural conditions matter, including AI literacy and sensitivity to offence (Voutsa et al., 2025). These contingencies are consolidated in Table 2. [Table 2 near here].

Proposition: Humorous AI advertising (HAIA) model

The HAIA model formalises the review synthesis by linking AI involvement cues to outcomes in humorous advertising through four recurring mechanisms. The integration of SOR and ELM in the HAIA framework is more than a notational reconciliation of two parent traditions. It makes two novel contributions explicit: The integration of SOR and ELM in the HAIA framework is more than a notational reconciliation of two parent traditions. It makes two novel contributions explicit: (i) perceived appropriateness is elevated from a stimulus property (in SOR) or a peripheral-cue input (in ELM) to an organism-level switch that determines which processing route dominates and conditions response valence; and (ii) the same configuration of

AI-involvement cues and humour stimulus features can produce either engagement or backlash, depending on that switch, a crossover mechanism that neither parent framework captures on its own. Empirically, appropriateness in studies of AI-generated humour should therefore be measured as a moderator. AI cues shape organism states by influencing inferences of authorship and legitimacy, such as authenticity, credibility, effort, and accountability, and by activating persuasion knowledge, which can increase scepticism and defensive processing. In parallel, the model specifies humour processing as an emotional and cognitive pathway in which comedic content is appraised and experienced, with appropriateness acting as a switch that determines whether humour is enjoyed or rejected. The model highlights three pathways through which effects tend to unfold. The first is a legitimacy pathway where AI cues drive authenticity and credibility judgments that then shape evaluations and intentions. The second is a persuasion knowledge pathway, in which AI cues trigger coping and resistance responses that reduce persuasion and can increase rejection. The third is a humour pathway where amusement and perceived appropriateness drive engagement and sharing, but where low appropriateness can flip the response toward irritation and backlash. Figures 2 and 3 distinguish which of these pathways tends to dominate under lower-effort processing versus higher-effort processing, clarifying why the same AI-generated joke can either amplify entertainment and diffusion or prompt scrutiny and resistance. [Figure 2 and 3 near here].

To move from a descriptive synthesis to a set of empirically tractable claims, we formalise the HAIA framework into five propositions (P1–P5) summarised in Table 3. Each proposition specifies a directional prediction with named mediators or moderators, mapped to the stimulus (S1 = AI-involvement cues; S2 = humour stimulus features; S3 = contextual stimulus features), organism (O1 = source-legitimacy inferences; O2 = persuasion-knowledge activation; O3 = humour appraisal and appropriateness), and response (R1 = persuasion; R2 = resistance; R3 = engagement; R4 = backlash) elements of the model. [Table 3 near here].

Discussion and Conclusion

This review provides a mechanism-based explanation of why AI-generated humour sometimes enhances outcomes and other times backfires. The main contribution is to consider AI-involvement as a meaningful cue that influences the interpretation of humour, rather than merely a production tool. Across the four clusters, effects are driven by perceptions of legitimacy, cognitive coping responses, emotional humour assessment, and platform engagement dynamics. For theory, the synthesis bridges humour research and persuasion theory by showing that AI cues can change the route through which humour is processed. When the cue activates persuasion knowledge or credibility concerns, audiences are more likely to scrutinise intent and authenticity, which reduces the likelihood that humour operates as a straightforward peripheral cue (Petty & Cacioppo, 1986; Campbell & Kirmani, 2008). This perspective clarifies why disclosure effects vary and why some studies report neutrality rather than a consistent AI penalty. In practice, disclosure should be considered part of the message design rather than a compliance afterthought. The HAIA model predicts that disclosure framing ('AI assisted the creative team' vs. 'AI generated this ad') and disclosure placement (pre-roll vs. post-roll) will produce different downstream responses by activating different organism-level pathways. Practitioners testing AI-generated humour should pre-test for perceived appropriateness as an organism-level moderator (P4) rather than treating it as a manipulation check, particularly in offence-sensitive categories such as cause-related advertising (Saumure et al., 2025). For experimental design, holding the humorous stimulus constant while varying only the AI involvement cue isolates which of the three organism pathways drives the effect (Exner et al., 2025; Koning & Voorveld, 2025). Future work should additionally vary humour risk and offence sensitivity as moderators. In conclusion, AI involvement cues shape humorous advertising through four mechanisms that determine whether humour is processed as entertainment or as suspicious persuasion.

References

Arango, L., Singaraju, S. P., & Niininen, O. (2023). Consumer responses to AI-generated charitable giving ads. *Journal of Advertising*, *52*(4), 486-503.

Baek, T. H., Kim, J., & Kim, J. H. (2024). Effect of disclosing AI-generated content on prosocial advertising evaluation. *International Journal of Advertising*, 1-22.

Bower, A. H., & Steyvers, M. (2021). Perceptions of AI engaging in human expression. *Scientific Reports*, *11*(1), 21181.

Campbell, C., Plangger, K., Sands, S., & Kietzmann, J. (2022). Preparing for an era of deepfakes and AI-generated ads: A framework for understanding responses to manipulated advertising. *Journal of Advertising*, *51*(1), 22-38.

Campbell, M. C., & Kirmani, A. (2008). I know what you're doing and why you're doing it. The use of persuasion knowledge model in consumer research. In C. P. Haugtvedt, P. M. Herr, & F. R. Kardes (Eds.), Handbook of consumer psychology (pp. 549-573). New York: Taylor and Francis.

Cao, Y., Cao, J., Hou, Y., & Ji, L. J. (2025). How humorous is AI? Exploring ChatGPT's role in humor generation and human-AI interaction. *Computers in Human Behavior Reports*, *20*, 100807.

Eisend, M. (2011). How humor in advertising works: A meta-analytic test of alternative models. *Marketing letters*, *22*(2), 115-132.

Exner, Y., Hartmann, J., Netzer, O., & Zhang, S. (2025). AI in disguise-How AI-generated ads' visual cues shape consumer perception and performance. *Available at SSRN*, *5096969*.

Friestad, M., & Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. *Journal of consumer research*, *21*(1), 1-31.

Gorenz, D., & Schwarz, N. (2024). How funny is ChatGPT? A comparison of human-and AI-produced jokes. *Plos one*, *19*(7), e0305364.

Gu, C., Jia, S., Lai, J., Chen, R., & Chang, X. (2024). Exploring consumer acceptance of AI-generated advertisements: From the perspectives of perceived eeriness and perceived intelligence. *Journal of Theoretical and Applied Electronic Commerce Research*, *19*(3), 2218-2238.

Haupt, M., Freidank, J., & Haas, A. (2025). Consumer responses to human-AI collaboration at organizational frontlines: strategies to escape algorithm aversion in content creation. *Review of Managerial Science*, *19*(2), 377-413.

Joshi, N. N. (2025, January). Evaluating human perception and bias in AI-generated humor. In *Proceedings of the 1st Workshop on Computational Humor (CHum)* (pp. 79-87).

Kariyawasam, H., Niwarthana, A., Palmer, A., Kay, J., & Withana, A. (2024, March). Appropriate incongruity driven human-AI collaborative tool to assist novices in humorous content generation. In *Proceedings of the 29th International Conference on Intelligent User Interfaces* (pp. 650-659).

Karpińska-Krakowiak, M., Hajduk, G., Mroz-Gorgon, B., & Skowron, L. (2025). Alleviating aversion to artificial intelligence in digital advertising with humorous disclosures. Proceedings of the 58th Hawaii International Conference on System Sciences.

Kim, S., & Chilton, L. B. (2025). AI humor generation: Cognitive, social and creative skills for effective humor. *arXiv preprint arXiv:2502.07981*.

Kirk, C. P., & Givi, J. (2025). The AI-authorship effect: Understanding authenticity, moral disgust, and consumer responses to AI-generated marketing communications. *Journal of Business Research*, *186*, 114984.

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

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

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., ... & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. Annals of internal medicine, 151(4), W-65

Lu, H., & Yuan, S. (2024). "I know it's a deepfake": the role of AI disclaimers and comprehension in the processing of deepfake parodies. *Journal of Communication*, *74*(5), 359-373.

Martin, R. A., & Ford, T. (2018). *The psychology of humor: An integrative approach*. Academic press.

Mehrabian, A., & Russell, J. A. (1974). *An approach to environmental psychology*. The MIT Press.

Paul, J., & Rosado-Serrano, A. (2019). Gradual internationalization vs born-global/international new venture models: A review and research agenda. *International Marketing Review*, *36*(6), 830-858.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In *Advances in experimental social psychology* (Vol. 19, pp. 123-205). Academic Press.

Qiu, X., Wang, Y., Zeng, Y., & Cong, R. (2025). Artificial intelligence disclosure in cause-related marketing: A persuasion knowledge perspective. *Journal of Theoretical and Applied Electronic Commerce Research*, *20*(3), 193.

Rae, I. (2024, May). The effects of perceived AI use on content perceptions. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

Rezwana, J., & Maher, M. L. (2023). User perspectives of the ethical dilemmas of ownership, accountability, leadership in human-AI Co-creation. In *IUI workshops*.

Sallaku, K., Macca, L. S., Santoro, G., & Thu Phan, T. (2025). Advertising in the digital era: Can AI evoke emotions like humans?. *Business Process Management Journal*, *31*(5), 1785-1811.

Saumure, R., De Freitas, J., & Puntoni, S. (2025). Humor as a window into generative AI bias. *Scientific Reports*, *15*(1), 1326.

Shambhavi, Jha, D., & Halder, R. (2025, August). When Humor Crosses the Line: Decoding Offensiveness in Meme-Ads Through Language Model Reasoning. In *International Conference on Computer Information Systems and Industrial Management* (pp. 208-222). Cham: Springer Nature Switzerland.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of business research*, *104*, 333-339.

Strick, M., Holland, R. W., Van Baaren, R. B., Van Knippenberg, A., & Dijksterhuis, A. (2017). Humour in advertising: An associative processing model. *European Review of Social Psychology: Volume 24*, 32-69.

Vermeulen, I., Droog, E., & Burgers, C. (2019). Overlay ads in humorous online videos: it's a matter of timing. *International Journal of Advertising*, *38*(7), 1046-1066.

Voutsa, M. C., Kerkidou, D., & Hatzithomas, L. (2025, May). Laughing at Bias: An Exploratory Study on Generation Z's Persuasion and Skepticism Toward AI-vs. Human-Generated Sexist Humorous Ads. In *International Conference on Persuasive Technology* (pp. 71-85). Cham: Springer Nature Switzerland.

Warren, C., Carter, E. P., & McGraw, A. P. (2021). Being funny is not enough: The influence of perceived humor and negative emotional reactions on brand attitudes. In *Humor in advertising* (pp. 117-137). Routledge.

Weinberger, M. G., & Gulas, C. S. (1992). The impact of humor in advertising: A review. *Journal of advertising*, *21*(4), 35-59.

Weinberger, M. G., Gulas, C. S., & Taylor, C. R. (Eds.). (2021). *Humor in advertising: Classic perspectives and new insights*. Routledge.

Wu, Z., Zeng, L., & Huang, Y. (2025). Influence of the Characteristics of AI-Generated Advertising on Consumers' Purchase Intention. *Journal of Arts and Cultural Studies (ACS)*, *1*(2), 01.

Tables and Figures

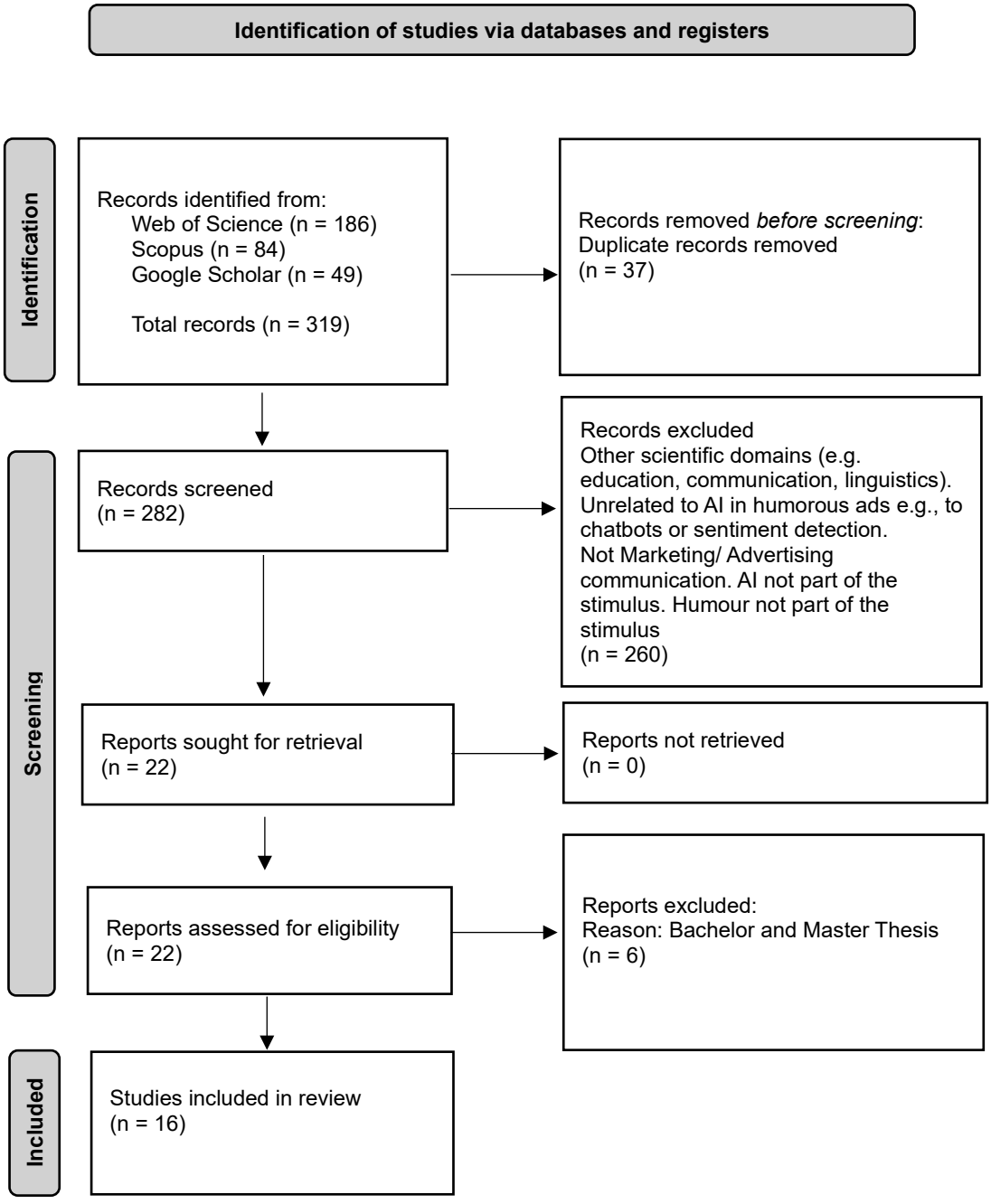

Figure 1: PRISMA flow diagram (Source: Author's own work)

Cluster/ELM Route	Stimulus (S)	Organism (O)	Response (R)
Authorship & Source Legitimacy Central Route	• AI authorship/creator label (human vs AI vs human+AI) (Bower & Steyvers, 2021; Joshi, 2025; Rae, 2024) • Disclosure transparency and framing (explicit vs ambiguous) (Lu & Yuan, 2024; Karpińska-Krakowiak et al., 2025; Voutsa et al., 2025) • Source/agent cues (avatar; anthropomorphism; byline) (Karpińska-Krakowiak et al., 2025; Rae, 2024)	• Credibility and authenticity inferences (Rae, 2024; Voutsa et al., 2025) • Legitimacy and trust judgments (Sallaku et al., 2025; Voutsa et al., 2025) • Attributions of intention and creative effort (Gorenz & Schwarz, 2024; Rae, 2024; Joshi, 2025) • Algorithm aversion / 'machine' heuristic (Bower & Steyvers, 2021; Joshi, 2025; Karpińska-Krakowiak et al., 2025)	• Creator/company favourability and satisfaction (Rae, 2024; Voutsa et al., 2025) • Acceptance and willingness-to-engage intentions (Joshi, 2025) • Reduced baseline receptivity (Voutsa et al., 2025)

	• Source Blindness (Joshi, 2025)	• Beliefs about AI capability (Bower & Steyvers, 2021; Joshi, 2025)	
Persuasion Knowledge & Cognitive Defence Central Route	• AI-involvement cues interpreted as tactics (Voutsa et al., 2025) • Disclosure framing highlighting marketing intent (Voutsa et al., 2025) • High-stakes or sensitive contexts; argumentative features (Lu & Yuan, 2024; Voutsa et al., 2025)	• Persuasion knowledge activation (Voutsa et al., 2025) • Scepticism, counterarguing, and cognitive scrutiny (Lu & Yuan, 2024; Voutsa et al., 2025) • Perceived manipulation/reactance; brand-safety monitoring (Shambhavi et al., 2025; Voutsa et al., 2025) • Moral/ethical appraisal (Saumure et al., 2025; Voutsa et al., 2025)	• Resistance and reduced compliance (Voutsa et al., 2025) • Negative coping responses (Lu & Yuan, 2024; Voutsa et al., 2025) • Amplified downside of norm-violating content (Shambhavi et al., 2025; Voutsa et al., 2025)
Affective Humour	• Humour type and strength (Cao et al., 2025; Karpińska-	• Mirth/enjoyment vs. confusion (Gorenz & Schwarz, 2024; Lu & Yuan, 2024)	• Immediate amusement/laughter vs. irritation (Gorenz & Schwarz, 2024; Lu & Yuan, 2024)

Processing & Appropriateness Peripheral Route	Krakowiak et al., 2025; Voutsa et al., 2025) •Incongruity/ resolution; linguistic fluency (Kariyawasam et al., 2024; Kim & Chilton, 2025) • Topic sensitivity and cultural context (Cao et al., 2025; Saumure et al., 2025)	• Appropriateness / 'Benignness' judgments (Cao et al., 2025) Offensiveness and affective discomfort (Shambhavi et al., 2025; Voutsa et al., 2025; Saumure et al., 2025) • Affective evaluation along pleasure–arousal–dominance dimensions [PAD framework - Pleasure, Arousal, Dominance] (Sallaku et al., 2025)	• Emotional resonance and affect transfer (Sallaku et al., 2025) • Downstream evaluative/behavioural responses (Karpińska-Krakowiak et al., 2025; Voutsa et al., 2025)

Entertainment Value & Social Media Engagement Peripheral Route	• Social-media-native humour formats (meme-ads) (Kim & Chilton, 2025; Shambhavi et al., 2025) • Novelty/virality cues and platform dynamics (Kim & Chilton, 2025) •Interactivity/participation prompts (Kim & Chilton, 2025)	• Perceived entertainment and social value of sharing (Lu & Yuan, 2024) • Social identity/bonding motivation (Kim & Chilton, 2025) • Relatability (Kim & Chilton, 2025)	• Sharing and engagement (likes/comments) (Lu & Yuan, 2024; Kim & Chilton, 2025) • Social diffusion and virality (Lu & Yuan, 2024) • Backlash amplification and reputational harm (Shambhavi et al., 2025)

Table 1: ELM-SOR synthesis of AI involvement in humorous advertising, empirical papers only (Source: Author's own work)

Contingency	**Primary Cluster(s) Affected**	**Mechanism Explanation**
Repeated Exposure / Familiarization	Authorship & Source Legitimacy; Affective Humour Processing	Repeated exposure can reduce algorithm aversion and increase humour appreciation as audiences become familiar with AI-generated humour and update beliefs about AI capability (Bower & Steyvers, 2021; Joshi, 2025).
Humour Strength Threshold	Affective Humour Processing; Source Legitimacy	Disclosure effects depend on joke quality. Weak AI humour amplifies source penalties, whereas strong humour reduces psychological distance and can bring AI outputs to parity with human-authored ads (Bower & Steyvers, 2021; Karpińska-Krakowiak et al., 2025; Kim & Chilton, 2025).
Socio-Political Sensitivity	Persuasion Knowledge & Cognitive Defence	Appropriateness depends on the target's political sensitivity. Models are more likely to reproduce stereotypes for lower-salience traits, whereas high-stakes traits are more constrained, making avoiding punching down especially important (Saumure et al., 2025).
Contextual Valence	Affective Humour Processing; Social Media Engagement	AI humour performance varies with the emotional valence of the context. In negative contexts, AI can outperform humans in humour coping by favouring self-enhancing humour that defuses tension, consistent with a rationality bias toward effectiveness (Cao et al., 2025).

Contingency	Primary Cluster(s) Affected	Mechanism Explanation
Source Awareness	Source Legitimacy; Cognitive Defence	Authorship awareness shifts evaluations. Under blind conditions, AI jokes can be rated as funnier and misattributed to humans. Disclosure can trigger humour discounting and scrutiny unless the humour is strong(Bower & Steyvers, 2021; Joshi, 2025; Karpińska-Krakowiak et al., 2025; Voutsa et al., 2025).
Audience Characteristics	All clusters	Receptivity varies by cohort and expertise. Gen Z is typically more open, while experts discount algorithmic creativity. These traits shape baseline scepticism and bias sensitivity (Bower & Steyvers, 2021; Gorenz & Schwarz, 2024; Joshi, 2025; Voutsa et al., 2025).

Table 2: Contingencies of AI in humorous advertising (Source: Author's own work)

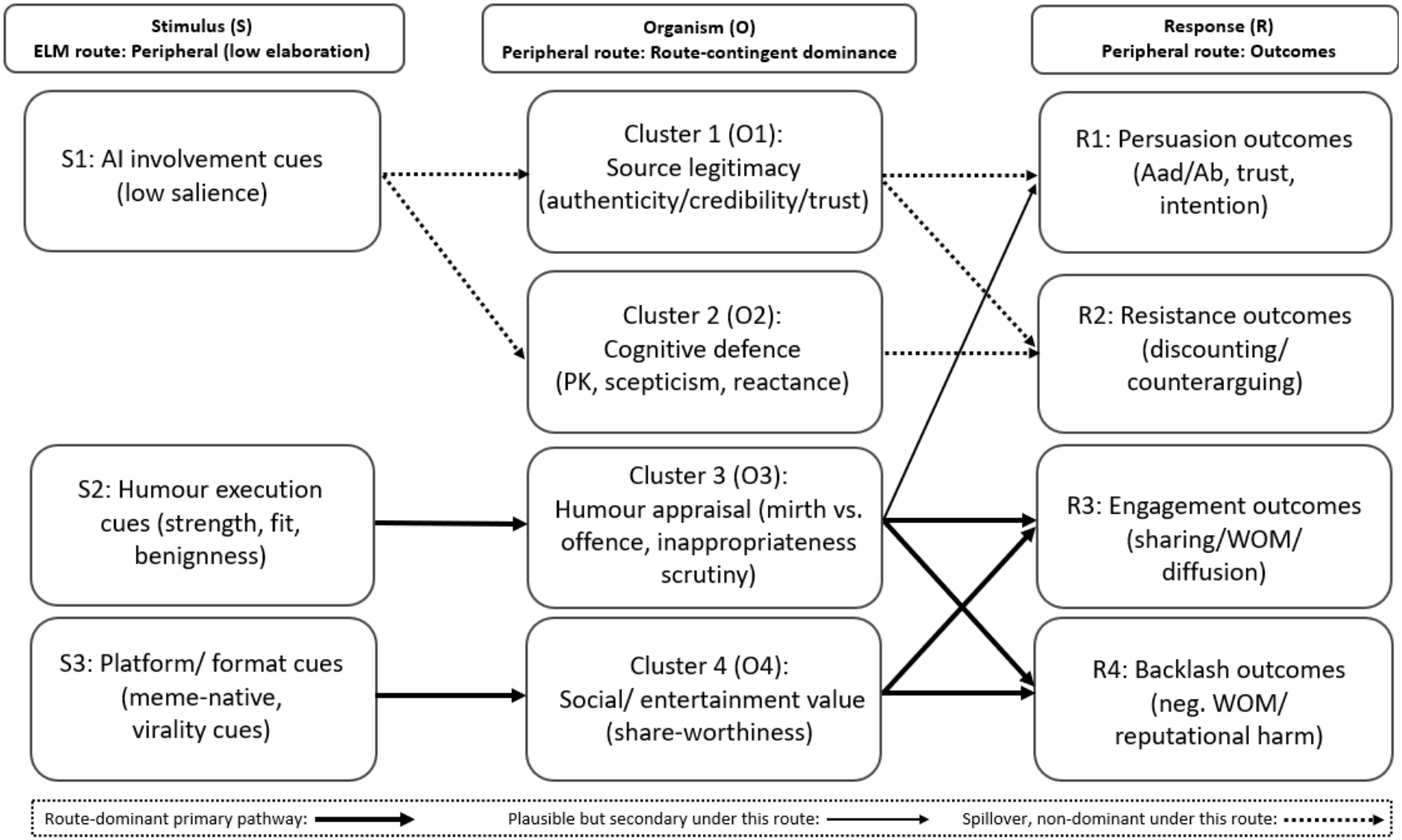

Figure 2: HAIA Model (peripheral route). S2 and S3 primarily activate Cluster 3 and Cluster 4, driving engagement and, secondarily, persuasion. Thick solid arrows indicate route-dominant primary SOR pathways, thin solid arrows indicate plausible but secondary pathways under this route, and dotted arrows indicate spillover, non-dominant paths (Source: Author's own work).

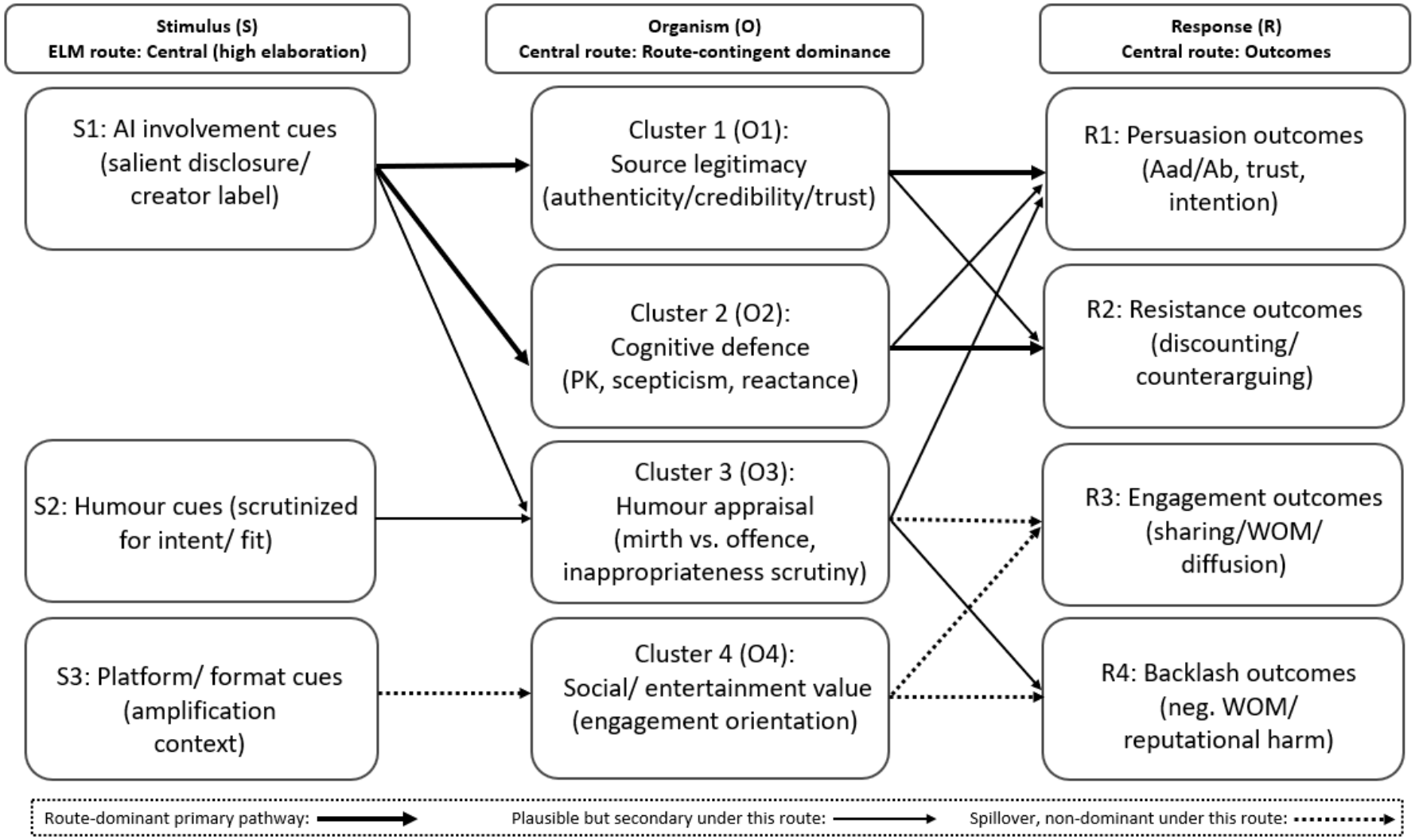

Figure 3: HAIA Model (central route). Salient AI involvement cues S1 primarily activate Cluster 1 and Cluster 2, with Cluster 3 serving as a switch for appropriateness scrutiny that shapes persuasion versus resistance/backlash. Thick solid arrows indicate route-dominant primary SOR pathways, thin solid arrows indicate plausible but secondary pathways under this route, and dotted arrows indicate spillover, non-dominant paths (Source: Author's own work)

ID	Title	Proposition
P1	Route-shift proposition	Salient AI-involvement cues (S1) shift audience processing from the peripheral to the central route. As a result, source-legitimacy inferences (O1) and persuasion-knowledge activation (O2) weigh more heavily on responses, while affective humour appraisal (O3) weighs less.
P2	Serial-mediation proposition	AI-involvement cues (S1) reduce persuasion (R1) and increase resistance (R2) through a serial process: AI cues first lower source-legitimacy inferences (O1), which in turn activate persuasion knowledge (O2), which in turn weakens persuasion and strengthens resistance (S1 → O1 → O2 → R1/R2).
P3	Humour-buffering proposition	Strong, benign humour (S2) buffers the route-shift in P1. When humour is strong, affective humour appraisal (O3) retains substantial influence on persuasion (R1) and engagement (R3) even under salient AI cues (S1).
P4	Appropriateness-switch proposition	Perceived appropriateness (O3) acts as a switch that determines whether AI-generated humour produces engagement or backlash. Under high appropriateness, AI cues (S1) combined with humour (S2) increase engagement (R3). Under low appropriateness, the same combination instead increases backlash (R4).
P5	Context-sensitivity proposition	Context (S3) shapes how strongly AI cues (S1) affect organism-level processing. The effect of AI cues on source-legitimacy inferences (O1) and persuasion-knowledge activation (O2) is weaker in negative-valence contexts (somber, cause-related), and the appropriateness switch in P4 shifts upward in offence-sensitive contexts, narrowing the conditions under which engagement results.

Table 3: Testable propositions about the route shift, serial mediation, humour buffering, appropriateness switch and context-sensitivity (Source: Author's own work)