

The Transparency Paradox: Examining 'Made with AI' Labels' Perspective on Audience Trust and Content Engagement

¹Harshitha S, ²Sanju Xavier

¹Student, ² Assistant Professor

¹Department of Journalism and Mass Communication, ¹Jain (Deemed-to-be University), Bengaluru, India.

Abstract - The use of Generative AI (GAI) within digital media content led to the development of "Made with AI" labels in the form of Instagram and YouTube. Although the use of this label aimed to ensure transparency, its implementation has created a "Transparency Paradox." Considering the Generation Z target market, this research investigates how the use of such disclosure affects trust and engagement. The results highlight the existence of a "Competence Penalty," where the content that bears a label is viewed as having no sweat equity or passion. Thus, an inconsistency arises as the label lowers trust while maintaining engagement rates due to entertainment.

Index Terms: Transparency Paradox, Generation Z, AI Disclosure, Digital Trust, Content Engagement, Generative AI.

I. INTRODUCTION

The "Made with AI" trademark is no longer just a branding statement of platforms but a tracker aspect of content management in the contemporary digital world. With generative artificial intelligence (GAI) making it difficult to differentiate human creative input from computer programming, platforms like TikTok, Instagram, and YouTube have started relying on such labels as digital receipts for their audience. The Made with AI designation, in its simplest form, indicates that the media – whether it be photorealistic images, a voice generated using cloning techniques, or videos developed via the assistance of artificial intelligence has been created using the help of AI applications. Just like how the list of ingredients gives information about the composition of food, the labels will inform people of the nature of the text, whether human-made or machine-made.

The most significant reason why these labels are necessary is to defend digital truth. The use of

deepfakes may affect how people perceive global issues, particularly when combined with artificial intelligence and misinformation that can spread worldwide within seconds. The public continues to battle an acute lack of authenticity, leaving both watchmen and platforms believing that requiring labels on such content will improve both the honesty and safety of the user's experience; however, the implementation of this labelling has created a complex psychological effect, which will be the focus of this study, known as the Transparency Paradox.

One of the largest elements of the Transparency Paradox is the bizarre lack of interrelation between trust and content engagement. Logic dictates that when a user does not trust a post, they will not communicate with it. However, social media indicators typically suggest the opposite. It is the age of Sceptical Consumption, in which users acknowledge that they do not believe an AI-tagged content but still do it, liking, sharing, and, most importantly, watching to the finish.

Artificial intelligence can create a spectacle, an image, or a sound so amazing, strange, or entertaining that it disregards any form of doubt. The person who is watching may know that it was created using artificial intelligence, meaning that it is not real, but he or she will continue watching because it looks amazing. This creates an empty interaction where the person gets numbers such as likes and views, but does not get the authenticity that usually comes with it. This paper seeks to investigate how engagement works in order to find out whether calling something "Made with AI" kills virality or not.

The Digital Native generation of Gen Z and Millennials.

The most affected groups of this paradox are the “digital native generation- Gen Z and Millennials. The Generation Z (14-29, 1997-2012) takes a distinct status in this study. Growing up in the age of photo filters, catfishing and deepfakes, Gen Z is now more hypersensitive to things that appear fake. For this segment of the population, authenticity is the greatest form of social capital. They are the most vocal proponents of transparency and require these labels to ensure they are not misled.

1.2 Background of study

The rapid development of Generative A.I. (GenAI) has greatly changed the way content is created digitally. Many new applications now generate hyper-realistic images, video and text. The last few years have seen the application of Generative A.I. make its way out of experimentation and into daily life, and serve as the foundation for many of today's digital creation tools. As A.I.-generated content is becoming increasingly difficult to identify compared to that created by a human, social media companies (the primary medium used for consuming digital content) are being held to a high standard by regulatory agencies, ethical organisations, and the public in terms of being clear about their use of Generative A.I. tools and their efforts to combat the impact of misinformation.

To respond to this change, major technology companies (Meta [Instagram, Facebook], ByteDance [TikTok], Google [YouTube]) have required that all content created with an application that uses Generative A.I. be labelled "Made with A.I.". These labels intend to provide users with a "digital nutrition label" that will allow them the informed consent required to assess the authenticity of the media they are presented with. Platforms such as Meta, ByteDance and Google assume that if they are transparent about their use of Generative A.I., they will build trust with their users. By requiring the use of these labels, these platforms are ensuring the integrity of the digitally created content by providing a clear distinction between the artificial and the human originals.

1.2 Statement of the problem

In a world filled with misinformation and deepfakes, social media applications such as Instagram use the "Made with AI" labelling as a means of protecting the authenticity of the content posted on these platforms. In terms of their function, the main role that these

labels play is that of ensuring transparency, as this practice can be considered a kind of "digital receipt" through which the user can identify whether media was created by a human or a computer. The concept of being transparent in this context can be seen as an extension of one of the basic premises – transparency increases credibility.

Nevertheless, the increasing incorporation of AI into content generation leads to certain negative effects caused by this labelling practice on Gen Z users, who are often hypersensitive to any changes made. At the root of this phenomenon lies the issue of the "Transparency Paradox," through which the mere fact of mentioning the role played by an AI tool in the creation of content undermines the relationship between the producer and the consumer of this media. Although honesty is demanded by Gen Z members, AI labelling leads to a "competence penalty." Accordingly, the required label functions as a "transparency tax" that punishes people for their transparency by creating a block to their sincere ability to connect emotionally.

1.4 Scope of Study:

It aims to examine psychological and behavioural implications of disclosure policy for AI through the case study of the social network Instagram from the perspective of the "Transparency Paradox". The scope of the study is limited to people born in 1997-2012 years – Generation Z, who can be defined as digital natives and sensitive to the phenomenon of digital authenticity. In particular, the selected demographic is characterised by their birth in an environment which is permanently changing due to the constant development of social media and operations using AI. These people have grown up with the phenomenon of social media that is becoming increasingly enhanced by artificial intelligence, and their perception of labels "Made with AI" reflects the shifting realities of digital society. In this way, it is important to conduct research targeted at the generation concerned in order to understand how those who value realness perceive synthetic content in social media.

1.4 Objective:

1. To analyse whether the audience finds “made with AI”- labelled content credible.
2. To examine the engagement dynamics (i.e., likes, shares, and watch time) for content which is identified as AI

3. To evaluate whether it increases the transparency among Gen Z and how it shapes their perception regarding AI

II. REVIEW OF LITERATURE

Jang, M. (2026). Building trust in AI: The role of technical capacity, social risk, and corporate institutional accountability. *Information*, 17(2), 212. The researcher analysed public perceptions of AI through a nationally representative survey of adults in the United States, distinguishing between trust in AI as a broad technological system and trust in its individual components. The findings revealed that trust was most strongly influenced by perceptions of AI's cognitive capabilities, such as intelligence and problem-solving ability. However, concerns related to social risks including job displacement and privacy violations significantly reduced levels of trust. Importantly, the study demonstrated that trust in AI depended more heavily on corporate and institutional accountability than on abstract ethical claims alone. The researcher concluded that sustainable trust in AI required a shift away from opaque "black box" systems toward governance structures in which developers and organisations were legally and socially accountable for the societal consequences of AI technologies. (Jang, 2026)

Kouroupi, N. (2026). Generative AI for content creation: Opportunities and ethical concerns. In *The ethics of AI in hospitality and tourism* (1st ed., pp. 115–138). Routledge.

The researcher examined how generative AI technologies have reshaped tourism marketing through customised travel itineraries, automated social media management, and immersive visual experiences. While these developments improved operational efficiency and expanded creative opportunities for small and medium-sized tourism businesses, the study also identified several ethical concerns, including the loss of genuine human interaction, the spread of AI-generated false reviews, and uncertainty surrounding copyright and ownership of AI-generated content. The researcher argued that long-term success in AI-driven tourism marketing depended on adopting a "human-in-the-loop" approach, in which human oversight ensures that automated content remains authentic, culturally sensitive, and engaging for diverse travellers. (Kouroupi, 2026)

Escolano, V. J. C., Yee, Y.-M., Shiang, W.-J., Hernandez, A. A., & Van Nang, D. (2026). Generative AI recommendations for environmental sustainability: A hybrid SEM–ANN analysis of Gen Z users in the Philippines. *Information*, 17(2), 203. The researchers examined how perceptions of usefulness and personalised recommendations influenced the adoption of generative AI for eco-friendly practices among Filipino Gen Z users. Using a hybrid analytical approach combining structural equation modelling (SEM) and artificial neural networks (ANN), the study integrated the theory of planned behaviour with sustainability-focused theoretical frameworks. The findings demonstrated that Filipino Gen Z users were motivated to adopt generative AI because of its perceived usefulness and ability to provide personalised sustainability recommendations. However, levels of trust in AI systems and perceptions of associated risks significantly influenced the extent of adoption, with "environmental consciousness" identified as a key factor connecting technological engagement with sustainable behavioural action. (Escolano et al., 2026)

Brandão, P. R. (2025). The impact of artificial intelligence on modern society. *AI*, 6(8), 190.

The researcher discussed the transformative influence of AI across multiple sectors while acknowledging the broader societal challenges associated with its expansion. The paper examined positive contributions of AI in areas such as healthcare, employment, and data processing, including the role of Aidoc in supporting accurate medical diagnosis and argued that AI technologies enhanced efficiency, processing speed, and analytical precision, particularly in data-intensive industries. At the same time, the paper identified several risks, including algorithmic bias, the opacity of "black-box" decision-making systems, and cybersecurity vulnerabilities. The researcher maintained that the sustainable integration of AI into society required proactive governance approaches grounded in ethics and accountability, encouraging governments and institutions to develop regulatory frameworks that support technological innovation while protecting human agency, social equality, and public trust. (Brandão, 2025)

Pizzi, G., Matta, S., Caboni, F., & Stewart, D. W. (2025). Immersive and generative technology: New

tools for marketing, new tools for consumer response. *Psychology & Marketing*, 42(12), 3236–3250.

The researchers examined the impact of immersive technologies including augmented reality (AR), virtual reality (VR), extended reality (XR), and generative artificial intelligence (GenAI) — on marketing and consumer experience. The authors argued that these technologies increased consumer engagement through deep personalisation, interactive environments, and multi-sensory feedback. Identified challenges included issues of accessibility, digital fatigue, and the preservation of authenticity in consumer interactions. The study further highlighted a significant shift in consumer behaviour as the boundaries between physical and digital spaces increasingly merged, raising critical research concerns related to attribution, co-creation, and transparency within digital marketing environments. (Pizzi et al., 2025)

2.2 Theoretical Framework:

According to Agenda-Setting Theory, it is not the content but the topics media decides for audiences to be concerned about. In this research, the "Made with AI" logo serves as an agenda setter, as its use prompts audience concern about the themes of "authenticity" and "origin" of posts. Such salience causes skepticism since 131 participants admitted that the label created a skeptical reaction towards content. As a result, users see the "AI transparency" issue as important, which affects their perception of the post.

The Uses and Gratifications (U&G) Theory assumes that audiences actively choose media according to particular needs, and this idea relates to the "Utility Buffer" introduced above. While Generation Z individuals usually refrain from sharing AI-generated content since it does not comply with such user requirements as "social status" and "authenticity," they often choose to "save" such content for their "utility" or "informational" value. Thus, for example, posting a study guide generated by AI can satisfy educational needs despite lacking social value.

2.3 Research Gap:

The Intentionality vs. Automation Gap: A lack of empirical data on when Generation Z views AI as a legitimate collaborative tool versus a "lazy," passionless alternative.

Long-Term Creator Sustainability: Failure to explore how the disjuncture between short-term

entertainment value and long-term scepticism impacts a creator's reputation and audience trust.

Visual Platform Context (Instagram): A research void regarding how the "Transparency Paradox" operates on highly visual platforms, as most existing literature focuses strictly on text-based "hard news."

The Creator "Transparency Tax": Insufficient evidence on whether mandatory platform labelling acts as a reputational penalty or if it can be leveraged to build authentic audience partnerships.

Gen Z's Psychological Dilemma: A significant lack of understanding concerning the internal cognitive dissonance and psychological dilemmas that AI labels impose specifically on Generation Z users.

3.1 Research Design

The research design selected for this study is quantitative, being cross-sectional. Cross-sectional designs are chosen because of their effectiveness in offering an accurate "snapshot" of perceptions and behavioural tendencies of the Generation Z generation at a certain point in time. What makes the design unique is the ability to conduct an investigation into the transparency paradox behavioural response to the revelation of the technologies involved in producing AI vs. actual content. By employing a comparative design and presenting the participants with two versions, one with "Made with AI" tags and one without, they will allow performing statistics on the changes in the level of trust and engagement.

3.2 Population and Sampling

Strict criteria have been established in selecting the target population, which has been limited to only those individuals belonging to Generation Z (those born between 1997 and 2012). The reason for selecting this particular age bracket is that people aged between 11 and 24 years form the native users of digital technology, who are the primary consumers of Instagram visual images. Data will be collected using purposive and snowball sampling techniques among a sample size of N=250. Recruitment of participants will be done through Instagram (Stories & DM).

3.3 Data Collection Procedures

The procedure involves the use of a standardised questionnaire, which is administered via Google Forms. The procedure is divided into two phases, which are the following:

Screening: In terms of screening, the first question posed in the procedure is the “age gate,” which only allows individuals born between 1997 and 2012 to complete the survey.

Exposure to Stimuli: Visual stimuli in the form of Instagram posts will be shown to participants.

3.4 Data Analysis Technique

In order to assess the effect that the AI disclosure has on user behaviour, quantitative methods will be employed to analyse the gathered data. Upon collecting the data on the basis of the N=250 surveys, descriptive statistics will be calculated in order to establish whether the users have any patterns in trusting the information provided by the AI algorithms, as well as whether they see the content producer as a credible source of information. To

perform a comparative analysis of how users behave based on their intention to interact, depending on whether the label is added to the content, will be another step in the analysis.

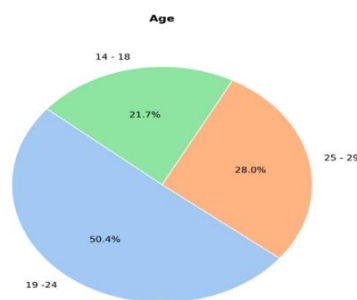
3.5 Ethical Considerations:

The ethical issues related to the data gathering process are ensured by making sure that the participants are aware that the survey is meant to be carried out only for an academic purpose on "informed consumption." In order to make sure that the privacy of the participants remains intact, especially considering the privacy-personalisation dilemma involved in social media research studies, no personal identifying details will be sought from the participants.

IV. ANALYSIS AND CONCLUSION

Data Analysis

Option	Count
19 -24	128
25 - 29	71
14 - 18	55

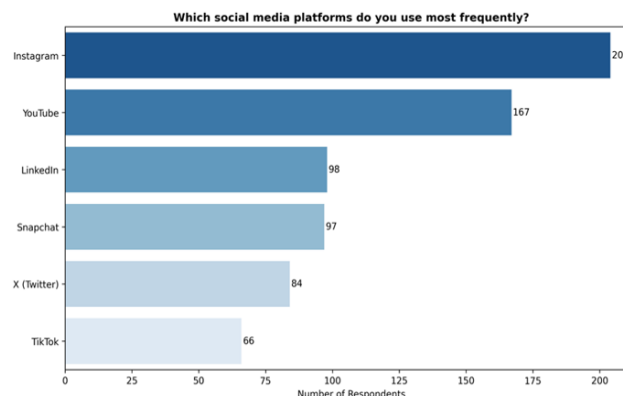


Respondents Distribution by Age Groups

It is important to note that the emphasis has been laid on the "digital natives" demographic since the biggest portion of the total N = 250 respondent sample falls into the category of Generation Z. In particular, the respondents aged 19-24 years account for 50.4%, respondents aged 25-29, 28% of the

sample, and those aged 14-18, 21.7%. It is important to note that the above-mentioned distribution of respondents is crucial to ensure the validity of research findings, since it implies that they would represent the view of a mature "digital native" audience.

Option	Count
Instagram	204
YouTube	167
LinkedIn	98
Snapchat	97
X (Twitter)	84
TikTok	66



Regarding platform frequency, Instagram stands out as the most prevalent platform chosen for the analysis because 204 of the 250 participants have higher usage

frequency on this platform compared to others. This is followed by YouTube, where 167 participants use the platform, and then LinkedIn, which has 98

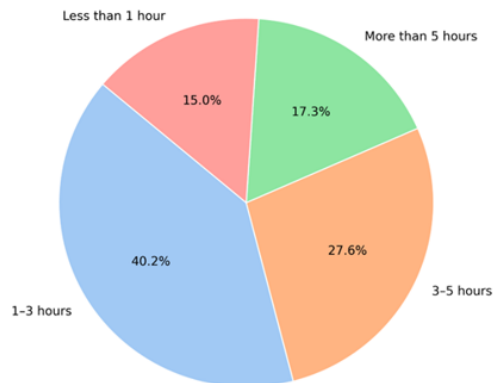
participants, while TikTok and X show much lower platform frequencies. With this platform being highly

utilised by the participants in the study, there is a clear justification for focusing on Instagram.

Option	Count
1–3 hours	102
3–5 hours	70
More than 5 hours	44
Less than 1 hour	38

In terms of behaviour associated with content consumption, there is an indication of a fairly high level of digital immersion on the part of participants in the research. From the analysis above, it is clear that close to 40% of the sample spend 1-3 hours per

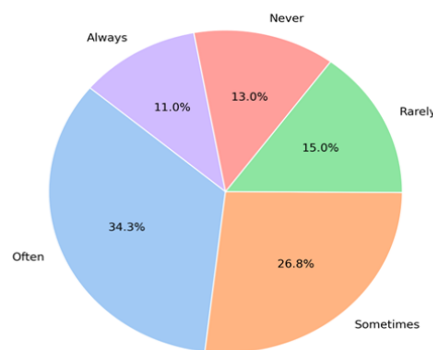
How many hours per day do you spend consuming digital content on average?



day on consuming digital content, while 27.6% consume it for 3-5 hours daily, and another 17.3% use more than 5 hours on content consumption each day. Thus, one can say that about half of the sample consists of heavy users.

Option	Count
Often	87
Sometimes	68
Rarely	38
Never	33
Always	28

How often do you notice "Made with AI" or "AI-generated" labels on content you consume?

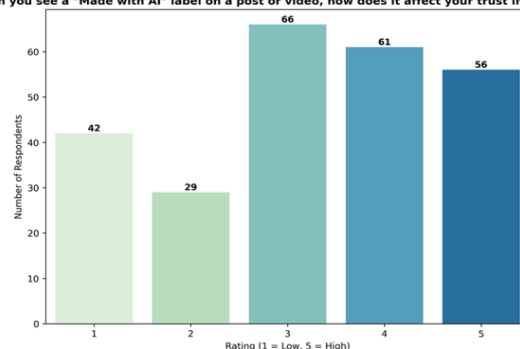


On the matter of how visible the transparency mandates were, when considering those who noticed either "Made with AI" or "AI-generated" labels, a total of 45.3% reported seeing the labels "Always" (11%) or "Often" (34.3%). In this regard, one can say that the awareness campaign by Instagram is at an

advanced stage of saturation among the Gen Z crowd. Nonetheless, 26.8% reported seeing the labels only "Sometimes," while 28% said they did not see them "Rarely" or "Never," suggesting a possible problem with the visibility of the label itself or the users' attention span.

Option	Count
1(Increases trust)	42
2	29
3	66
4	61
5(decreases trust)	56

When you see a "Made with AI" label on a post or video, how does it affect your trust in that content



Data on the influence of "Made with AI" on trust demonstrate that this example strongly supports the concept of the "Transparency Paradox." It is clearly visible from the graph below, which presents a "skew" in terms of trust depletion since the total number of respondents who found the label harmful to trust (a rating of 4, or 61 respondents) and those who said it decreased their trust (a rating of 5, or 56

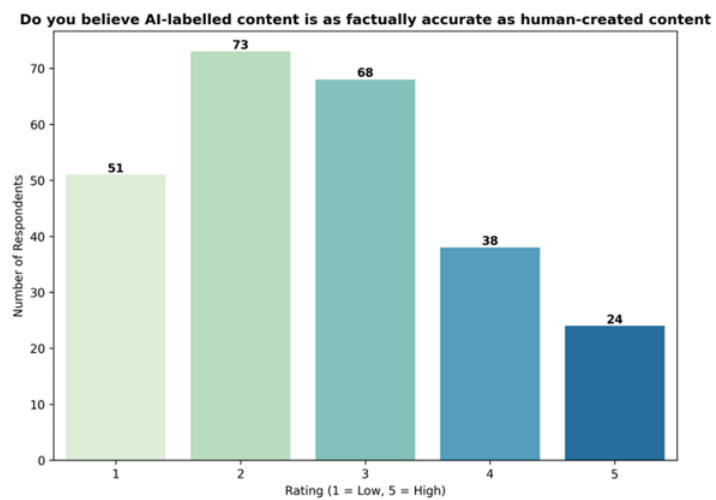
Option	Count
1(Strongly agree)	51
2	73
3	68
4	38
5 (Strongly disagree)	24

For the question on whether the accuracy of AI-produced material is comparable to that of human-produced material, there is clearly an overall predisposition towards disbelief, but it is worth noting that a significant group of participants indicated a "soft agreement." The total number of participants who chose ratings 1 and 2 (124, about

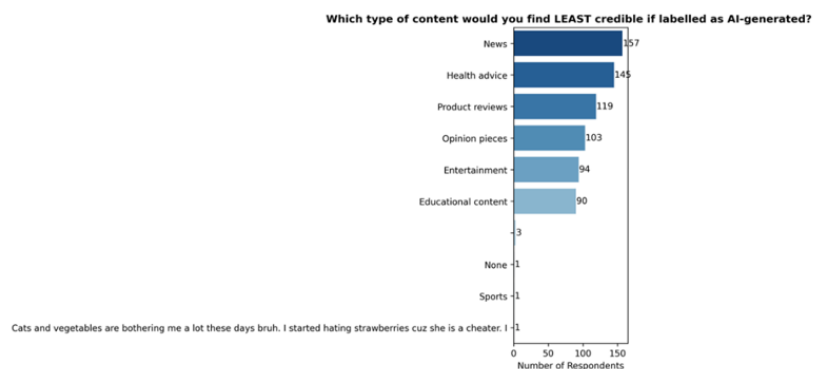
News	157
Health advice	145
Product reviews	119
Opinion pieces	103
Entertainment	94
Educational content	90
Others	2
None	1
Sports	1

The statistical data about how "Made with AI" impacts the user's trust confirms the existence of the "Transparency Paradox." As follows from the given chart, there is a "skew" of trust depletion in favor of

respondents) totals 117 respondents (or 46.8%). The largest number of respondents (66, or 26.4%) were neutral about the use of the label (their rating was 3). Only 42 respondents (16.8%) indicated that the "Made with AI" label improved their trust. This finding can be considered a key point of the study, since it proves that the required transparency from Instagram can work against itself.



50%) reflects the fact that they have confidence in the accuracy of AI-produced material as compared to that produced by humans. That said, the highest number of participants' ratings (73) belongs to the second category, indicating that although they do not agree with the notion of the accuracy of AI, they believe in it much more than not agreeing with it.



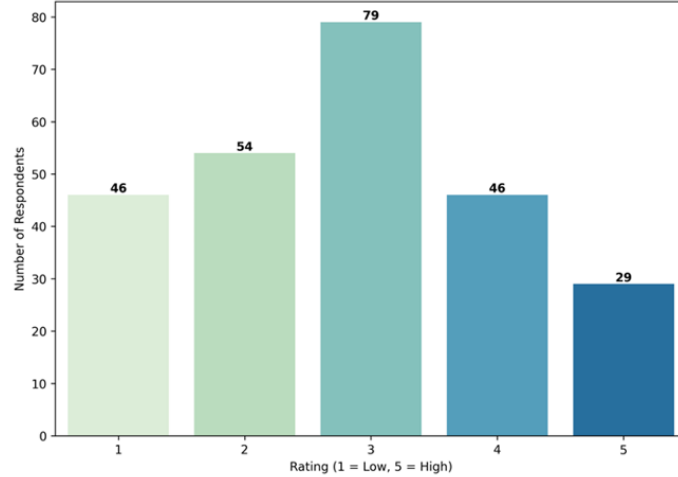
the latter, which means that 117 users (46.8%) think that "Made with AI" decreases their trust level (5 points, 56 respondents) or influences it negatively (4 points, 61 respondents). It should be stressed that the

largest percentage was the neutral one (3 points), including 66 participants of the experiment. Consequently, one should conclude that it is impossible to say that "Made with AI" is a clear indication for Gen Z individuals. In addition, 16.8%

(42 respondents) said that this term increases their trust. It is a significant result proving the research hypothesis since Instagram's transparency measures might fail because of "red flagging."

Option	Count
1(Yes, significantly)	46
2	54
3	79
4	46
5(No, not at all)	29

Does knowing a content creator used AI tools make you question their authenticity or originality

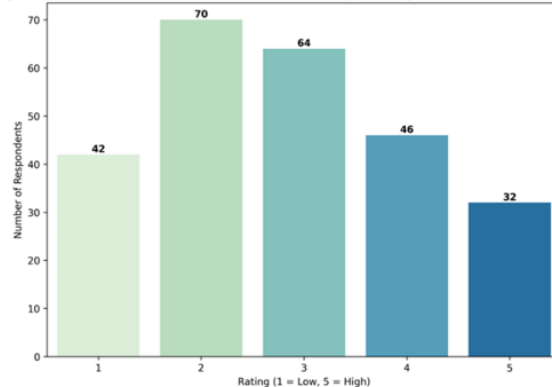


Whereas when asked about the implications of the awareness of usage of AI tools by a content creator on the legitimacy and creativity of such, it can be seen there was a clear spike in those who either felt skeptical or neutral towards the issue. In other words, the maximum number of responses (79) was received under Rating 3, indicating the hesitation or the

"evaluative uncertainty" of respondents regarding the humanity of the creative process. Nonetheless, it should be mentioned that 40% of the respondents altogether (100) belonged to Ratings 1 and 2, showing that many Generation Z members find it hard to believe in the actual creativity of a person due to involvement of AI.

Option	Count
1(Strongly agree)	42
2	70
3	64
4	46
5(Strongly disagree)	32

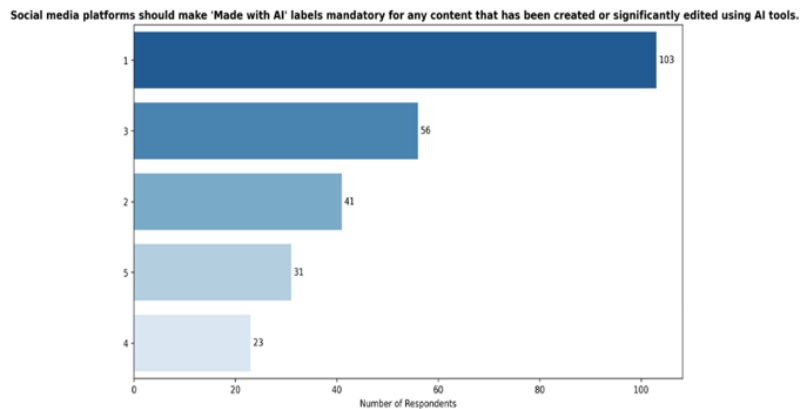
I am more likely to trust content labelled as 'Made with AI' if it is posted by a verified or well-known content creator.



The survey evaluated the degree of acceptance of the "Made with AI" badge for content produced by a reputable source. In this example, the bias towards the approval of the statement is evident since 44.8% of respondents chose Ratings 1 and 2. It is possible to say that the "Source Credibility" variable to some extent counterbalances the issue of distrust toward the label associated with artificial intelligence. Namely, in the particular case under consideration,

the credibility of the producer becomes the second choice for the consumer: individuals tend to believe in the notification because they had a chance to develop a relationship with the producer in advance. However, the relatively high numbers of neutral (64) and negative (78) answers suggest that verification status does not serve as a "silver bullet" in addressing the transparency paradox.

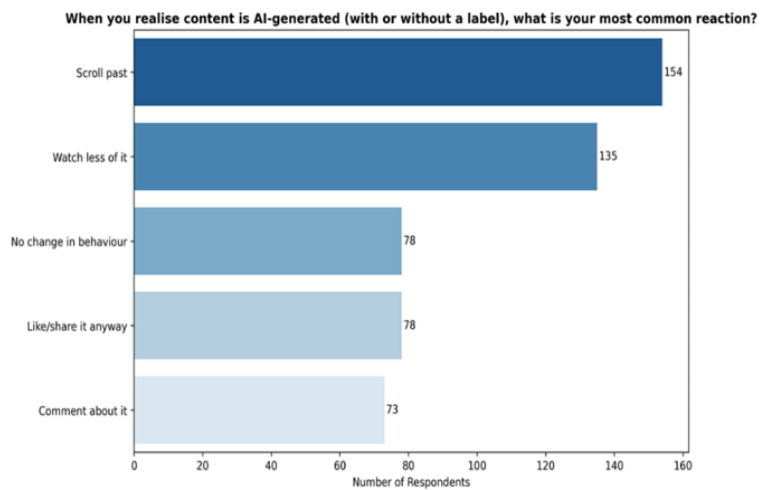
Option	Count
1 (Strongly agree)	103
2	41
3	56
4	23
5 (Strongly disagree)	31



From the survey results, it can be seen that the approval rating regarding the transparency of institutions concerning the implementation of artificial intelligence by Generation Z is almost universal. In the context of whether social media platforms must mandatorily label their posts "Made with AI" if either of them is created or edited using artificial intelligence, the answer is positive. For instance, out of 250 participants, 103 of them (41.2%)

were strongly supportive of the idea, while 144 people (57.6%) were considered part of Rating 1 and Rating 2. This shows that there is a moral responsibility in terms of sharing information in the digital world. However, it also poses a contradiction in the study since users demand the labelling of posts based on ethical considerations and fairness (FATE), but they do not trust the posts' value because of this practice.

Option	Count
Scroll past	154
Watch less of it	135
No change in behaviour	78
Like/share it anyway	78
Comment on it	73

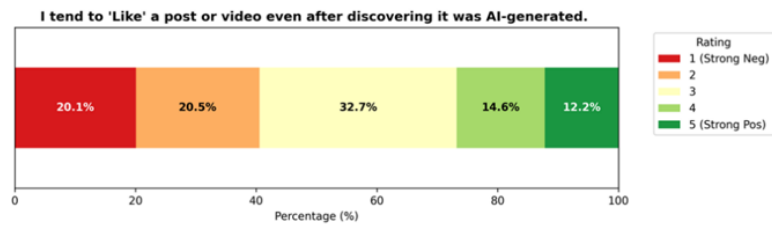


The behavioural statistics show a strong pattern of "Avoidance" among Generation Z participants upon learning that the content was created using AI. By far, the most popular option was "Scroll past" (154), with close behind it being "Watch less of it" (135) close behind. Such an outcome shows that the presence of AI in the content creates a passive avoidance strategy regardless of the quality of the latter. As for the other behaviours, one should note that 78 participants had

no change in behaviour, yet another 78 decided that "they would like/share it anyway," while 73 would choose to comment on it. Obviously, the above facts prove that there is the "Sceptical Consumption Gap," as while some of the audience members ignore the content, others keep engaging with it. As can be seen from the above, AI does harm to the "evaluative trust" but does not influence the "functional engagement."

Option	Count
1 (never)	51
2	52
3	83
4	37
5 (Very Frequently)	31

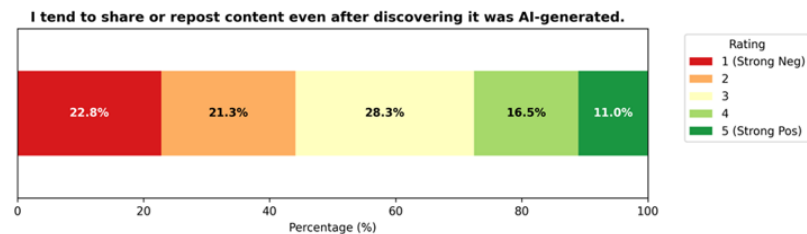
The statistics on the user "Likes" after the introduction of artificial intelligence show that there is a certain consistency in passive interaction, despite the established mistrust. While a total of 40.6% of users (Ratings 1 and 2) state that they rarely or never "Like" the content after learning about AI implementation, the highest number belongs to the



neutral category (Rating 3), 32.7%. Moreover, 26.8% of the participants (Rating 4 and Rating 5) still find the visual material interesting enough to give it likes. In essence, this evidence suggests that there is a problem with the "Sceptical Consumption" phenomenon because Gen Z can enjoy the visuals despite the mistrust.

Option	Count
1 (Never)	58
2	54
3	72
4	42
5 (Very Frequently)	28

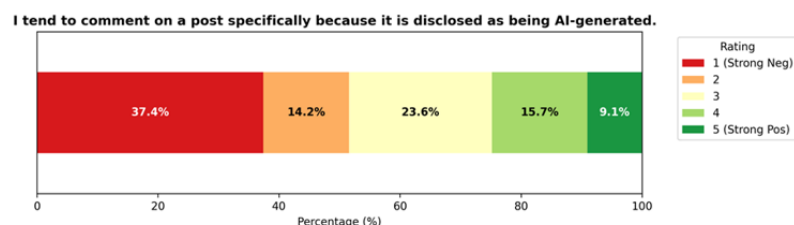
Whereas active forms of engagement, such as sharing and retweeting, involve more friction than passive forms like likes, 44.1% of participants (112 people) prefer the options "Never" and "Rarely" when asked about whether they share content that they recognise as being created by artificial intelligence. While it seems plausible that Generation Z would privately consume or even privately like AI content, it appears



less likely that they would actively use their own personal online identity in conjunction with AI content by posting it themselves. The "Transparency Paradox" becomes especially apparent in this scenario: transparency efforts designed to promote truthfulness instead serve as a social obstacle because the social value of AI content is viewed as relatively low.

Option	Count
1 (Never)	95
2	36
3	60
4	40
5 (Very Frequently)	23

From the results, it can be concluded that AI labelling is seen more as a behavioural inhibitor than a behaviour enhancer among the majority of respondents in this research. About half of the sample (51.6%) responded to the question concerning their



behaviour upon spotting such content labelled with the label 'AI', using the responses "Never" and "Rarely," with the first one being the answer chosen by the majority (37.4%). In other words, this label does not provide enough social meaning or

importance to motivate young people to comment on such AI-created content. At the same time, another segment of the respondents (24.8%) commented on such posts from time to time. The implication behind

Option	Count
1 (Strongly Agree)	57
2	73
3	70
4	29
5 (Strongly disagree)	25

Concerning the valuation of content, a clear pattern emerges of its "devaluation" in cases where AI was involved in the process. If asked whether an AI label causes content to be viewed as "less valuable" or "less impressive," a substantial number of respondents answered positively, and 54 of them (21.6%) chose Rating 1 (Strongly Agree), along with 60 others

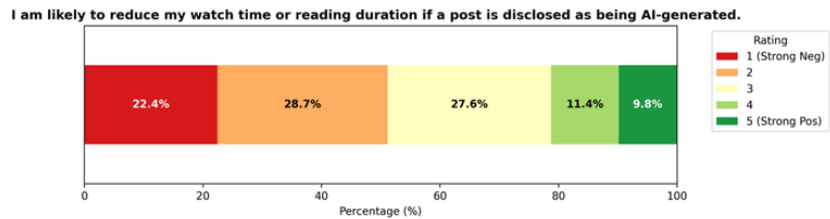
Option	Count
1 (never)	56
2	59
3	74
4	33
5 (Very Frequently)	32

The initial findings on the trust comparison data show a strong tendency towards the use of unlabelled information among the targeted population. The responses on how much the respondents trusted the unlabelled post compared to the AI-labelled post showed that more than half of the respondents (46.8%) agreed with their level of trust being high. Out of this figure, 46 respondents chose "Strongly

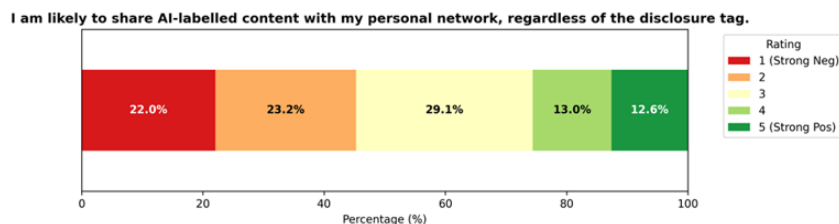
Option	Count
1 (Most likely)	48
2	59
3	64
4	36
5 (Less likely)	46

As for the findings concerning users' intentions to save/bookmark AI-labelled posts, there seems to be

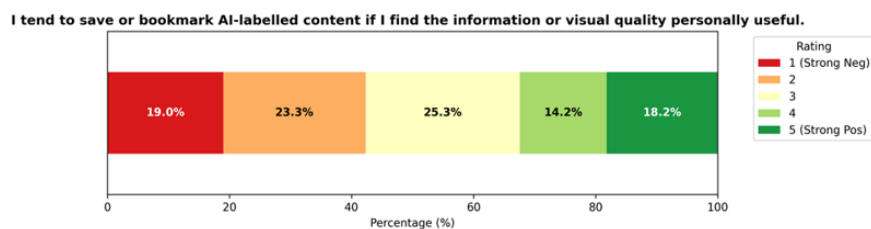
such findings is that despite AI labelling leading to behavioural inhibition, it can still provoke some people into engaging in discussion.



(24%), who selected Rating 2. In total, 45.6% of the respondents agree with the claim that AI labelling decreases the inherent value of content. While the highest number of participants (72 out of 250) selected the middle option (neutral), meaning that some people remain unsure, it should be noted that there are only 25.6% who disagree with such claims.



Agree," while 71 respondents chose Rating 2. On the other hand, only 19.6%, or 49 respondents, were able to agree that labelling made them more trusting. These figures serve as a foundation for the Transparency Paradox. It implies that the current generation tends to trust an unlabelled post more than a labelled one.



an interesting contradiction between usefulness and lack of trust. Specifically, when asked whether one

would save posts because of their information or visual appeal, regardless of being made by an AI, 107 people (totalling 42.3% of the sample) showed their agreement on the scale, including 48 individuals who chose the option "Most Likely" and 59 who chose Rating 2. While it shows that the usefulness of the AI

Option	Count
1 (Never)	56
2	66
3	80
4	29
5 (Very frequently)	23

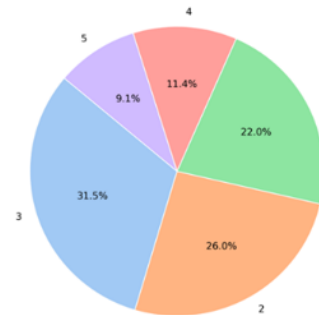
The conclusions regarding the effectiveness of the "Made with AI" label in causing curiosity to boost interactions can be drawn from the analysis of the survey data. The percentage of people who answered the question concerning the level of curiosity caused by the "Made with AI" label reveals that the neutral category was represented by the highest number of respondents, 80 people, or 31.5%, indicating Rating 3. However, from the analysis of the results, it is

Option	Count
Depends on the content type	85
The unlabelled one	66
The AI-labelled one	57
No preference	46

In the task where the participants were asked to directly choose one out of two posts which were identical apart from the label, a quite complicated reaction was observed. The majority, or 33.5% (n=85) of all respondents admitted that the engagement will "depend on the content type," meaning that simply by adding the label, the message is not automatically dismissed – the decision is always made individually and on a situational basis. The second group of participants who expressed a

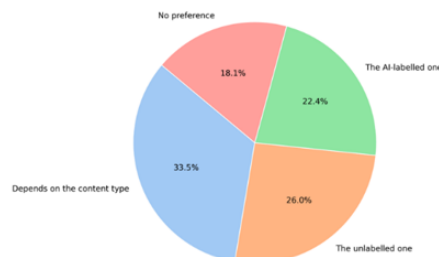
product is to some extent capable of overcoming distrust, another third of those surveyed, namely 32.5% (82 people), admitted that they were likely to dismiss such posts, while 25.3% (64 people) stayed undecided.

I have engaged more with certain content specifically because the 'Made with AI' label sparked my curiosity.



evident that the "Made with AI" label is not entirely a positive aspect that prompts increased engagement, considering that the majority of the respondents belong to the "Never" category, 56 people, or 48%, and Rating 2, 66 people. Only 20.5% claimed that the label causes frequent interest, with 29 people selecting Rating 4 and only 23 responding with "Very Frequently."

If two identical posts appeared; one labelled "Made with AI" and one not which would you be more likely to engage with?



preference for any of the messages chose the post which did not contain the label more often than the one containing it: 26% (n=66) versus 22.4% (n=57). 18.1% (n=46) reported having no preference at all when choosing between the two posts. From this, it could be concluded that even with some preference towards a post created by a person rather than an algorithm, the major factor which impacts Generation Z engagement would be information/visual content type.

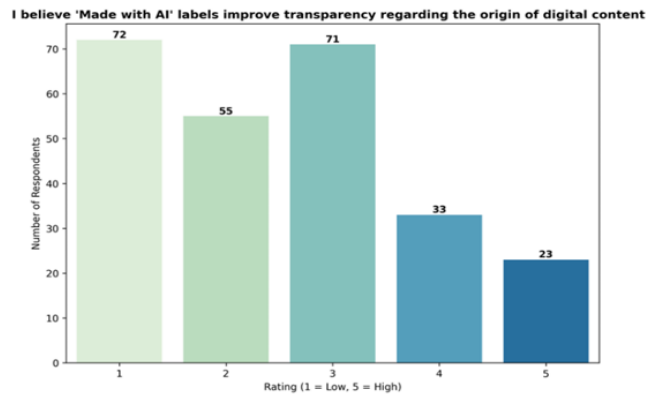
Option	Count
1 (Strongly agree)	72
2	55
3	71
4	33
5 (Strongly disagree)	23

However, opinions on the efficiency of 'Made with AI' labelling in promoting transparency have a clear trend in favour of agreeing with the statements. Indeed, the largest number of respondents (72 respondents) selected "Strongly Agree," while 55 of them agreed with the statement; therefore, they see the 'Made with AI' approach as efficient for achieving digital transparency. Meanwhile, there is a noticeable number of respondents whose stance is neutral. This

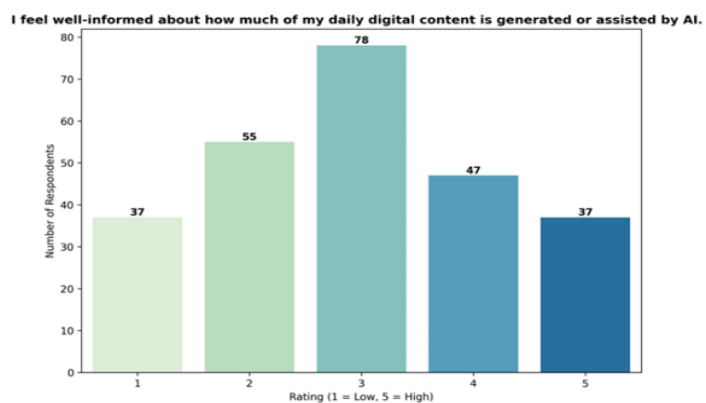
Option	Count
1 (Very Uninformed)	37
2	55
3	78
4	47
5 (Very informed)	37

As for the way people see themselves being informed about the existence of AI in their feed, we can say that the distribution of this data is fairly balanced but ambiguous at the same time. The biggest group of 78 participants belongs to the neutral category, which is "Neutral". This means that quite a lot of people are neither sure about the existence of AI in their feed nor unaware of it. As for people who took a definite

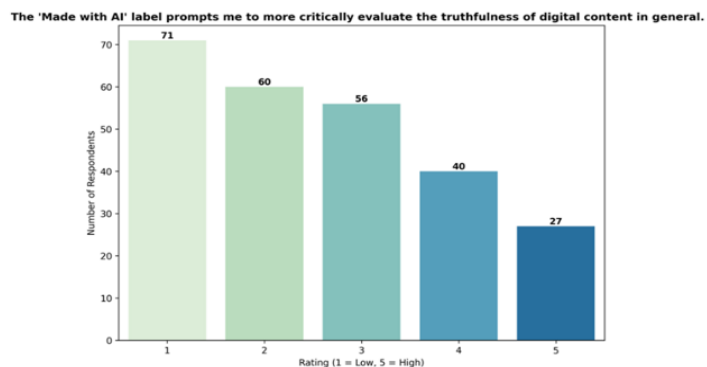
Option	Count
1 (Strongly agree)	71
2	60
3	56
4	40
5 (Strongly disagree)	27



number is 71 respondents. On the other hand, respondents who disagreed with the statement became fewer (33 respondents), and those who strongly disagreed became even fewer (23 respondents). Hence, one may conclude that Gen Z members find transparent approaches efficient but demonstrate behaviour in other cases showing their disagreement.



stand, the data is also rather even: 36.2% of participants feel that they are less informed (Rating 1: 37; Rating 2: 55), while 33.1% of respondents believe that they are well-informed enough (Rating 4: 47; Rating 5: 37). Thus, no clear majority on the side of informed people is evident, which proves the initial hypothesis concerning the necessity of institutions' labeling of AI materials.



The information about the impact of the 'Made with AI' disclosure on consumer scepticism shows that disclosure is the main trigger of critical thinking. The majority of participants have indicated that disclosure of the logo makes them think critically about the validity of information in the online environment. Indeed, 71 respondents chose the "Strongly Agree" option, whereas 60 respondents marked Agree, which

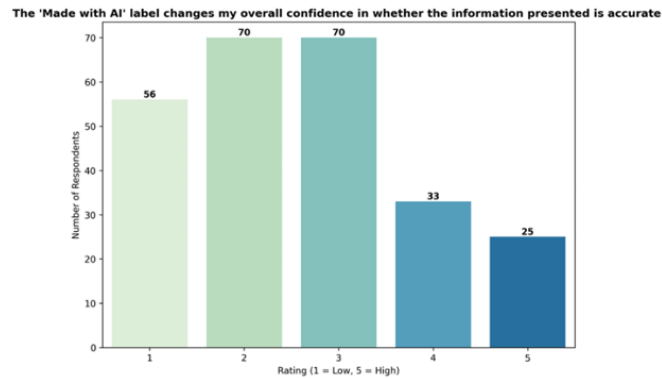
Option	Count
1 (Strongly agree)	56
2	70
3	70
4	33
5 (Strongly disagree)	25

Based on the survey responses concerning the influence of the 'Made with AI' label on informational confidence, it can be said that disclosure serves as an effective trigger in shifting the trust. In total, there were 126 people who stated that the use of the label affected their confidence in the accuracy of information. Among those, 56 people chose "Strongly agree," and 70 people selected "Agree". However, although a substantial neutral group of 70

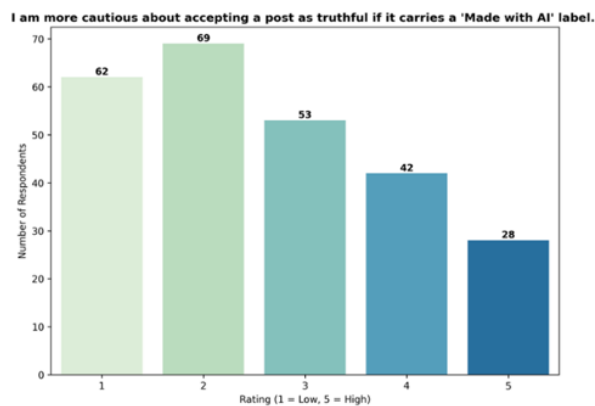
Option	Count
1 (Strongly agree)	62
2	69
3	53
4	42
5 (Strongly disagree)	28

The statistics from the survey on user caution provide an insight into a noticeable change in how the content would be scrutinised once the AI disclosure was introduced. Out of 178 individuals asked whether they were more careful in accepting the post as being true when there was an AI label attached to it, the majority chose to be more careful. In particular, 62 respondents strongly agreed with the aforementioned question, while 69 individuals out of the whole

accounts for the largest share of all responses. At the same time, only a very small number of participants disagreed, since 40 individuals gave Rating 4, whereas 27 others strongly disagreed. Also, there were 56 neutral responses. The disclosure requirement is proven to be effective in achieving its purpose, increasing the level of media scepticism.

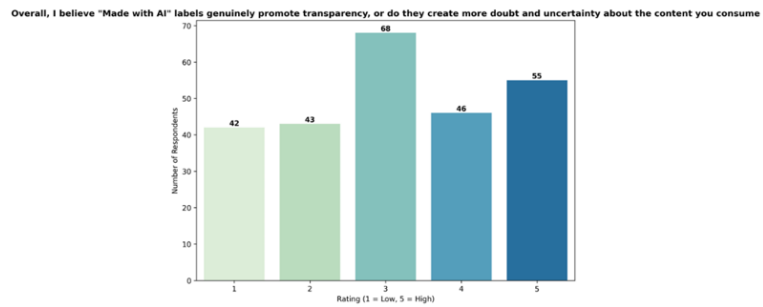


people was found (Neutral), a relatively insignificant number of people in the sample indicated no impact of the label on informational confidence. There were 33 people (Disagree) and 25 people ("Strongly disagree") who expressed that opinion. Therefore, AI disclosure does not have a passive effect but rather plays an essential role in the generation's judgment of information accuracy.



sample gave their agreement to this question, which sums up to a total of 51.5%. At the same time, only 28% of the sample was on lower levels of caution, with those being disagree (42) and strongly disagree (28). 53 respondents remained neutral to this question. Thus, the statistics prove that the first reaction to seeing the 'Made with AI' label on any post is verification.

Option	Count
1 (Creates More Doubt)	42
2	43
3	68
4	46
5 (Genuinely Promotes Transparency)	55



The final measurement considered the overall sentiment about whether 'Made with AI' labels actually support transparency or create doubt and uncertainty. According to the data collected, there was a wide diversity in the sentiment among participants, and the biggest cohort of 68 people remained neutral on the matter. As for those who had a definite opinion on the issue, slightly more people believed that the labels supported transparency; 55 of the participants were sure that the labels promoted transparency indeed, and 46 chose Rating 4. Nevertheless, these results are matched almost equally by those respondents who consider the labels a source of scepticism, since 42 of them think that the labels generate doubt, and 43 participants chose Rating 2. Such results show that the key paradox of AI labelling is that, despite recognising the importance of truthfulness, the labels' actual effect is as likely to be distrust as transparency.

Future outlook analysis reveals a core contradiction between technological progress and human value retention. Respondents predict "Made with AI" labels will become normalized "nutrition facts" within a "two-tier content economy," where human-made content commands a premium. While AI is accepted for entertainment, educational use demands absolute verification. Socio-economic anxieties include creator displacement, algorithmic "brainrot," and unoriginal "AI slop" eroding platform trust. Conversely, some argue Gen Z will become immune to the labels as AI normalizes. Ultimately, transparency is seen as a crucial "baseline" to combat misinformation and safeguard human values in an increasingly realistic digital landscape.

V. RESULTS AND DISCUSSION

This study of 253 Generation Z participants reveals that while platform-mandated "Made with AI" labels

achieve institutional transparency, they fundamentally alter media consumption, trust, and social behaviour.

- **Perceived Credibility & Critical Thinking:** The label acts as an obstacle to blind trust. A significant portion of respondents (126) stated it directly diminishes content reliability, while 131 participants noted the label triggers immediate critical thinking. For Gen Z, this "healthy scepticism" shifts consumption from passive to provisional; AI content is accepted for entertainment but strictly verified via human sources for educational purposes.
- **Engagement Dynamics & The "Utility Buffer":** The label introduces social friction, drastically reducing content's "social currency." Only 20.5% felt the label sparked curiosity, while nearly half expressed negative intentions regarding public engagement; 45.2% refused to share AI-labelled content to protect their online reputation. However, a private "utility buffer" exists: 42.3% routinely save AI content for personal use if it provides functional or informational value.
- **Long-Term Transparency & Future Outlook:** A strong majority (127) agreed that labels successfully fulfil ethical disclosure needs. However, this transparency inadvertently penalizes creators by coding AI as "less creative."

Looking ahead, Gen Z views labelling as an inevitable industry standard—akin to digital "nutrition labels" giving rise to a two-tier content economy where human-made content commands a premium. While fears persist over platform-wide "AI slop" and creator displacement, some anticipate a future "immunity" where users eventually desensitize to the tag. Ultimately, the label serves as a critical baseline regulator, ensuring Gen Z avoids cognitive

manipulation while navigating an unverified digital landscape.

VI. CONCLUSION

The "Made with AI" label has fundamentally rewritten the digital social contract for Generation Z, transforming from a simple transparency marker into a critical psychological barrier. It forces a transition from passive consumption to a verification-based approach, cultivating a "healthy suspicion" where content is doubted until human authenticity is proven. This creates a distinct behavioral dichotomy. To preserve their digital reputation, Gen Z denies AI content public "social currency" by refusing to share what is perceived as "low-effort" material. However, a private utility loop remains active; users routinely bookmark and save AI content for its functional, time-saving benefits.

Ultimately, this shifts social media into a structured two-tier content economy. "Human-Made" content has been elevated to a luxury status signifying true creativity and passion, while AI content serves as the baseline for mass production and functional utility. Though users may eventually develop an "immunity" to the tags treating them like standard Terms & Conditions they still view them as an essential right. In an era threatened by "AI slop," these labels empower Gen Z to act as "digital auditors," leveraging transparency to safely navigate an automated landscape.

Limitation:

1. Self-selection and sampling bias

Participants in this research were recruited via Instagram Stories and private messages, which implies that only active Instagram users could become a part of this research project. Such a selection method tends to over-represent people with high levels of opinion-expressiveness towards AI technology, thus possibly leading to an overestimation of the level of Generation Z scepticism.

2. Cross-sectional approach and temporal limitations

The single point of measurement used by researchers is not sufficient for examining changes in attitude due to the rapid normalisation of AI systems and their increased usage in social media applications. The rapid development of platforms' policies and generative AI technology may affect the validity of conclusions made by researchers.

3. Self-reporting biases

In addition to using social desirability biases as a potential confounding factor when measuring attitudes, researchers rely on self-reported intentions of participants to assess the extent of AI content avoidance in Gen Z individuals. It is possible that respondents were unwilling to report scrolling past posts generated via AI.

VII. RECOMMENDATION

Considering the finding that Generation Z perceives honesty as a core value yet faces "social friction" in interactions with the content generated through an automation process, the first recommendation is that social media platforms shift towards implementing a "Context-Specific Labelling System." Nowadays, the same "Made with AI" label applies to any piece of content from a filtered selfie to a deepfake news post. According to this research, such a "one-size-fits-all" policy makes individuals unnecessarily distrustful of creative work. Instead of it, the proposed solution is to have a hierarchical system in place, which will identify whether the post is "AI-Assisted," i.e., human-created with a little bit of help from AI, or "AI-Generated," fully machine-produced. In this way, social media platforms will be able to protect the concept of "Human Premium," giving credit to the creators who use AI merely as a tool, while giving the necessary warnings to the other posts.

Secondly, the platform must connect suspicion and truth by integrating "Instant Verification Tools" within the label itself. As has been proven, when young adults see the label, they enter the stage of "Healthy Suspicion" where they wish to verify the information at hand. However, due to difficulties of leaving the app, users often fail to do so. It is advised to have the "Made with AI" label to be interactive in that clicking on it would open up either the "Source Summary" or the link to other articles that were written by humans but still cover the same topic. This way, the mere label would serve as a tool for increasing digital literacy. This will not only allow the user to use their instinct of doubt, but also will ensure that the utility buffer, discovered in the research, is the one in which people keep the AI content, because it gives them valuable information that is verified.

REFERENCES

- [1] Labelling AI-generated media online - PMC. (n.d.). PMC. Retrieved February 16, 2026, from <https://pmc.ncbi.nlm.nih.gov/articles/PMC12166545/>
- [2] Kouroupi, N. (2026). Generative AI for Content Creation. In *The Ethics of AI in Hospitality and Tourism* (pp. 115–138). <https://doi.org/10.4324/9781003643951-7>
- [3] Pizzi, G., Matta, S., Caboni, F., & Stewart, D. W. (2025). Immersive and generative technology: new tools for marketing, new tools for consumer response. *Psychology and Marketing*, 42(12), 3236–3250. <https://doi.org/10.1002/mar.70036>
- [4] Raghupathi, W., Ren, J., & Kulkarni, T. (2026). Exploring the topics and sentiments of AI-Related Public Opinions: an advanced Machine learning text analysis. *Information*, 17(2), 134. <https://doi.org/10.3390/info17020134>
- [5] Escolano, V. J. C., Yee, Y., Shiang, W., Hernandez, A. A., & Van Nang, D. (2026). Generative AI Recommendations for Environmental Sustainability: A hybrid SEM–ANN analysis of Gen Z users in the Philippines. *Information*, 17(2), 203. <https://doi.org/10.3390/info17020203>