

**Project Report on “AI vs. Human Travel Reviews:
An Exploratory Study of Perceived Authenticity
and Emotional Tone”**

AI vs. Human Travel Reviews: An Exploratory Study of Perceived Authenticity and Emotional Tone

Submitted to:

Name: Ahmed Imran Kabir

Designation: Lecturer, School of Business and Economics

Major: Management Information System

Submitted by:

Name: Humayra Afra Ibnat Sneha

ID: 111212069

Major: Management Information System

Registration Trimester: Spring 2024



School of Business and Economics
United International University

Date of submission: July 28, 2025

Letter of Transmittal

July 28, 2025

Ahmed Imran Kabir

Lecturer

School of Business and Economics

United International University (UIU)

Email: ahmedimran@bus.uiu.ac.bd

Subject: Submission of project report.

Dear Sir,

With due respect, I am submitting my project report titled “**AI vs. Human Travel Reviews: An Exploratory Study of Perceived Authenticity and Emotional Tone**” with the aim of completing the degree of Bachelor of Business Administration, Major in Management Information Systems.

This paper explores the possibility of AI to create travel reviews that sound genuine and contrast them with actual user reviews in terms of emotional analysis and perception. I hope this paper meets your expectations and is an addition which will contribute to the existing literature in the area of MIS and AI.

Thank you for your continuous guidance and support throughout this project.

Sincerely,

Humayra Afra Ibnat Sneha

Bachelor of Business Administration

Management Information System

ID: 111212069

United International University

Certification Of Similarly Index

AI vs. Human Travel Reviews: An Exploratory Study of Perceived Authenticity and Emotional Tone

ORIGINALITY REPORT

18%

SIMILARITY INDEX

16%

INTERNET SOURCES

12%

PUBLICATIONS

11%

STUDENT PAPERS

PRIMARY SOURCES

1	www.coursehero.com Internet Source	2%
2	h20000.www2.hp.com Internet Source	1%
3	Submitted to University of Newcastle Student Paper	1%
4	kmr.kasba.or.kr Internet Source	1%
5	www.ssoar.info Internet Source	1%
6	Aarni Tuomi, Husna Zainal Abidin, Pasi Tuominen, Mário Passos Ascensão. "Chapter 36 Convincingness of AI-Generated Restaurant Reviews", Springer Science and Business Media LLC, 2025 Publication	<1%
7	Submitted to University of Edinburgh Student Paper	<1%
8	Daniel Belanche, Sergio Ibáñez-Sánchez, Pau Jordán, Sergio Matas. "Customer reactions to generative AI vs. real images in high-involvement and hedonic services", International Journal of Information Management, 2025 Publication	<1%
9	Submitted to PSB Academy (ACP eSolutions) Student Paper	<1%

10	www.jotmi.org Internet Source	<1 %
11	Submitted to Australian National University Student Paper	<1 %
12	Submitted to University of Portsmouth Student Paper	<1 %
13	tmstudies.net Internet Source	<1 %
14	Submitted to Kaplan College Student Paper	<1 %
15	ruor.uottawa.ca Internet Source	<1 %
16	c.coek.info Internet Source	<1 %
17	Submitted to universititeknologimara Student Paper	<1 %
18	Submitted to ICTS Student Paper	<1 %
19	burjcdigital.urjc.es Internet Source	<1 %
20	www.blacktietradelines.com Internet Source	<1 %
21	dspace.uiu.ac.bd Internet Source	<1 %
22	www.aau.edu Internet Source	<1 %
23	www.theseus.fi Internet Source	<1 %
24	Submitted to South University Student Paper	<1 %

ia802904.us.archive.org

25	Internet Source	<1 %
26	Submitted to Queensland University of Technology Student Paper	<1 %
27	Submitted to University of Malta Student Paper	<1 %
28	www.apacchrie.org Internet Source	<1 %
29	www.malque.pub Internet Source	<1 %
30	"Artificial Neural Networks and Machine Learning – ICANN 2024", Springer Science and Business Media LLC, 2024 Publication	<1 %
31	Danny McCarroll. "Simple Statistical Tests for Geography", Chapman and Hall/CRC, 2019 Publication	<1 %
32	ir.kiu.ac.ug Internet Source	<1 %
33	elibrary.tucl.edu.np Internet Source	<1 %
34	Submitted to University of South Florida Student Paper	<1 %
35	Submitted to BB9.1 PROD Student Paper	<1 %
36	e-repository.unsyiah.ac.id Internet Source	<1 %
37	Zining Luo, Yang Qiao, Xinyu Xu, Xiangyu Li et al. "Cross sectional pilot study on clinical review generation using large language models", npj Digital Medicine, 2025 Publication	<1 %

38	www.ncbi.nlm.nih.gov Internet Source	<1 %
39	Submitted to United International University Student Paper	<1 %
40	Submitted to Weatherford College Student Paper	<1 %
41	infostatconsulting.wordpress.com Internet Source	<1 %
42	research-system.siam.edu Internet Source	<1 %
43	Nuckchady, Girish. "Courage in Leadership – A Narrative Study", University of the Witwatersrand, Johannesburg (South Africa), 2025 Publication	<1 %
44	dspace.cuni.cz Internet Source	<1 %
45	pmc.ncbi.nlm.nih.gov Internet Source	<1 %
46	repositorio.ulatina.ac.cr Internet Source	<1 %
47	meral.edu.mm Internet Source	<1 %
48	repository.ub.ac.id Internet Source	<1 %
49	repository.up.ac.za:8080 Internet Source	<1 %
50	academic.oup.com Internet Source	<1 %
51	arxiv.org Internet Source	<1 %

52	d-nb.info Internet Source	<1 %
53	etd.repository.ugm.ac.id Internet Source	<1 %
54	etd.uwc.ac.za Internet Source	<1 %
55	hdl.handle.net Internet Source	<1 %
56	ir.lib.uwo.ca Internet Source	<1 %
57	www.researchsquare.com Internet Source	<1 %
58	www.scribd.com Internet Source	<1 %
59	www.um.edu.mt Internet Source	<1 %
60	Feng, Vivian. "The Relationship Between Biomarkers of Gut Barrier Integrity and Cognition in Cognitively Vulnerable Adults.", University of Toronto (Canada), 2024 Publication	<1 %
61	Schmidt, Robin Karl. "Automatic Document Classification in Technical Logbooks: A Comparative Study of Supervised, Weakly Supervised and Unsupervised Machine Learning Approaches", Universidade NOVA de Lisboa (Portugal), 2024 Publication	<1 %
62	californiaagriculture.ucanr.edu Internet Source	<1 %
63	digitum.um.es Internet Source	<1 %

64	dokumen.pub Internet Source	<1 %
65	eprints.whiterose.ac.uk Internet Source	<1 %
66	thesis.unipd.it Internet Source	<1 %
67	www.davidpublisher.com Internet Source	<1 %
68	Balázs Kovács. "The Turing test of online reviews: Can we tell the difference between human-written and GPT-4-written online reviews?", Marketing Letters, 2024 Publication	<1 %
69	Tshuma, Mbongeni. "Facebook Reviews and Intentions to Purchase Hotel Accommodation in South Africa", University of the Witwatersrand, Johannesburg (South Africa) Publication	<1 %
70	Li, Zhi. "Informing Evaluation Practice Through Research on Evaluation", The University of North Carolina at Charlotte, 2024 Publication	<1 %
71	de Almeida Vasconcelos, Maria. "The Influence of User-Generated Content on Tourism Organizations - The Case of Portuguese Center Region", Instituto Politecnico de Viseu (Portugal), 2024 Publication	<1 %

Declaration of the Student

I, Humayra Afra Ibnat Sneha, hereby declare that the research paper titled “**AI vs. Human Travel Reviews: An Exploratory Study of Perceived Authenticity and Emotional Tone**” is the result of my own work, which was completed sincerely and independently to fulfill the degree requirement of Bachelor of Business Administration, major in Management Information Systems, of United International University (UIU).

I verify that all the information is properly cited and referenced according to the APA 6th edition citation style. My supervisor was Ahmed Imran Kabir who gave me directions and assisted me throughout the research.

This paper is not presented in any other university or institution in order to obtain any degree or diploma. I hold exclusive responsibility on the contents and that no plagiarism, and academic dishonesty was involved.

Sincerely,

Humayra Afra Ibnat Sneha

Bachelor of Business Administration

Management Information System

ID: 111212069

United International University

Acknowledgement

I would like to deeply appreciate my supervisor, Ahmed Imran Kabir, who has given priceless guidelines, positive feedback as well as unending support in the research process. This project would have not been achieved without their support and guidance.

I would like to express my gratitude to every participant who attended the survey. Their feedbacks played a vital role in concluding this research.

I also would like to acknowledge the faculty members of the Department of School of Business and Economics at United International University, who supported me in my academic process.

Finally, I would like to express my utmost thanks to my family and close ones who supported me, inspired, and motivated me all the time.

Abstract

Over the last few years, the use of artificial intelligence (AI) to create new content has exponentially increased, as big language models such as GPT-4 allow producing realistic and human-feeling text at a scale. The increase of AI-generated reviews on the travel platforms provokes important questions regarding trust and authenticity and the capacity of users to be able to differentiate between the authentic and machine-written review. This exploratory study aims to examine perceived authenticity and emotional tone of AI-generated versus human Airbnb review to determine its implications on consumer trust to digital travel platforms.

The research study constituted a mixed method design with a perception-based survey integrated with computation linguistics. As a result, 150 real Airbnb guest reviews and 40 reviews created by AI were analyzed through the natural language processing (NLP) tools including VADER sentiment analysis and Plutchik emotion wheel via NRC lexicon. Textual difference was determined using sentiment scores, emotional distributions, statistical tests, Mann-Whitney U tests, and Chi-square tests. Also, 76 participants were presented with five pairs of reviews (one actual, and one AI-generated) and they were asked which one was “human-like”. The results showed that actual reviews were overall more emotional, with the emotional range that includes joy and fear, whereas AI reviews were more biased toward trust, surprise, and sadness. Participants in 4 out of 5 pairs found the human-rated reviews to be more human-like, which confirms the hypothesis that the reviews written by humans are considered more authentic. However, out of AI-generated reviews, one was significantly chosen, which proves that even the text written by a machine can mislead the reader.

This study adds to the rising discussion of AI credibility within digital context and requires an increase in content labeling, user literacy and emotional modeling throughout AI review generation.

Keywords: AI-Generated Reviews, Human-Written Reviews, Airbnb, Travel Reviews, Sentiment Analysis, Human Perception, User Trust.

Table of Content:

Letter of Transmittal	iii
Certification Of Similarly Index	iv
Declaration of the Student	x
Acknowledgement	xi
Abstract	xii
List of Figures:	xv
List of Tables:	xvi
List of Equations	xvii
List of Acronyms & Abbreviations	xviii
CHAPTER I:	1
INTRODUCTION	1
1.1 Background of the Study	2
1.2 Statement of the Problem.....	2
1.3 Objectives of the Study	2
1.4 Scope and Limitations of the Study	3
1.5 Definition of Key Terms	4
1.6 Organization of the Remaining Chapters.....	5
CHAPTER II:	6
REVIEW OF THE LITERATURE	6
2.1 Relevant Theory.....	7
2.2 Literature Survey	7
2.3 Hypothesis Development	9
CHAPTER III:	10
RESEARCH METHODS	10
3.1 Research Design	11
3.2 Sample	11
3.3 Questionnaire Development	12
3.4 Data Collection.....	13
3.4.1 Tools and Techniques:	14
3.5 Data Analysis Plan	15
CHAPTER IV:	19

RESEARCH FINDINGS & ANALYSIS	19
4.1 Text-Based Analysis of AI and Real Reviews.....	20
4.1.1 Word Cloud Visualization.....	20
4.1.2 Sentiment Count and Comparison (RQ 2a).....	21
4.1.3 Emotion Count and Comparison (RQ 2b).....	23
4.1.4 Radial Plutchik Emotion Wheel Visualization.....	25
4.2 Survey-Based Analysis: Human Perception of Review Authenticity.....	26
4.2.1 Human-Like Preference Summary (RQ 1).....	26
4.2.2 Percentage Preference.....	27
4.2.3 Demographic Influence Analysis.....	28
4.2.4 Visual Insights.....	28
4.3 Summary of Key Insights.....	29
CHAPTER V:	30
DISCUSSION	30
5.1 Conclusions.....	31
5.1.1 Hypothesis Evaluation.....	32
5.2 Scope for Future Study.....	33
5.3 Recommendations.....	34
Reference	35
Appendix-A:	38

List of Figures:

Figure 1: Review Preference Real vs. AI.....17

Figure 2: Emotion Distribution Real vs. AI.....18

Figure 3: AI-Generated Reviews Word Cloud20

Figure 4: Real Reviews Word Cloud21

Figure 5: Sentiment Comparison (%)22

Figure 6: Emotion Comparison (%)23

Figure 7: Plutchik Emotion Wheel: AI vs. Human Reviews24

Figure 8: Radial Plutchik Wheel25

Figure 9: Preference % Real vs. AI27

Figure 10: Choices by Review Reading Habit28

Figure 11: Review Preferences29

List of Tables:

Table 1: Key terms	4
Table 2: Human-like Preference (%) by Age Group	12
Table 3: Human-like Preference (%) by Gender.....	12
Table 4: Human-like Preference (%) by Review Reading Habit.....	12
Table 5: Human-like Preference Percentage by Review Pair	13
Table 6: Shapiro-Wilk Normality Test Results	16
Table 7: Mann-Whitney U Test Results	16
Table 8: Chi-Square Test	17
Table 9: Emotion Distribution Comparison	17
Table 10: Sentiment Distribution Comparison (%): AI vs. Real Reviews	22
Table 11: Emotion Count and Comparison.....	23
Table 12: Dominant Emotion Human vs. AI.....	24
Table 13: Effect sizes from small to medium	26
Table 14: Preference % Real vs. AI	27

List of Equations

No.	Equations	Description
1	$\phi = \sqrt{\chi^2 / N}$	Phi coefficient – used to measure effect size for Chi-square goodness-of-fit test.
2	Cramér’s V = $\sqrt{\chi^2 / (N \times (k - 1))}$	Cramér’s V – effect size measure for Chi-square test of independence.
3	$r = Z / \sqrt{N}$	Rank biserial correlation – used as the effect size for Mann-Whitney U test.
4	Sentiment Range: –1 to +1	VADER compound sentiment score scale: from negative (–1) to positive (+1).
5	8 × 2 contingency table	Format used in Chi-square test of independence for 8 emotions × 2 review types.
6	$p < .05$	Standard significance level used in hypothesis testing (implied across all tests).
7	$\chi^2 (7, N = 1095) = 18.23, p = .011$	Result of the Chi-square test on Plutchik emotion distribution differences.
8	Mann Whitney U = 4307, p < .001, r = 0.31	Statistical result comparing sentiment scores between human and AI reviews.
9	Mean compound score (Human) = 0.776 ± 0.17	VADER sentiment average for human-written reviews.
10	Mean compound score (AI) = 0.409 ± 0.31	VADER sentiment average for AI-generated reviews.

List of Acronyms & Abbreviations

Acronym / Abbreviation	Full Form
AI	Artificial Intelligence
LLM	Large Language Model
GPT-4	Generative Pre-trained Transformer 4 (by OpenAI)
NLP	Natural Language Processing
VADER	Valence Aware Dictionary and Sentiment Reasoner
NRC	National Research Council Canada – Emotion Lexicon
χ^2	Chi-Square (Statistical Test Symbol)
ϕ (phi)	Phi Coefficient (Effect Size for Chi-Square Test)
Cramér's V	Cramér's V (Effect Size Measure for Chi-Square Test of Independence)
r	Rank Biserial Correlation (Effect Size for Mann-Whitney U Test)
RQ	Research Question
eWOM	Electronic Word of Mouth
U	Mann-Whitney U Statistic

**CHAPTER I:
INTRODUCTION**

1.1 Background of the Study

Reviews generated by users on online platforms have become a major information source in travel planning ((Chong, Khong, Ma, McCabe, & Wang, 2018). Websites such as Airbnb, TripAdvisor, and Google Reviews now play an important role in choosing a destination, where people would like to visit and stay, the host's reputation for the stay, and local businesses ((Zervas, Proserpio, & Byers, 2020). At the same time with the development of large language models (LLMs) like GPT-4 by OpenAI makes it possible to generate persuasive human-like text at scale automatically (Acerbi & Stubbersfield, 2023). According to industry reports, AI-generated Google reviews have increased by 279% since 2019, and different e-commerce sites now label content as "AI assisted" (Lambert,2025). This convergence creates pressing concern about authenticity, consumer trust, and the effectiveness of existing review moderation systems.

1.2 Statement of the Problem

Despite the widespread use of machine-written reviews, there is a lack of data on experimental evidence on the reliability of consumers, whether they can distinguish between AI-generated text and genuine travel experiences (Christensen, Hansen, & Wilson, 2024). If AI-generated reviews are indistinguishable, customers may rely on AI-generated artificial stories while making booking decisions, without considering the authentic informational reviews of peer feedback, which may hurt the platforms' credibility. On the other hand, if AI-generated text is easily identified, there will be reputational penalties for businesses that use it. Thus, it is necessary for a systematic investigation to determine:

- (i) The perceptual gap between AI-generated reviews and human reviews, and
- (ii) The linguistic features on which credibility judgments are founded.

1.3 Objectives of the Study

The research aims to achieve the following three objectives:

- i. **Assess perceived authenticity:** Evaluates the perception of whether human-written Airbnb reviews are judged as more “human like” as compared to those created by AI.
- ii. **Compare linguistic sentiment and emotion:** Quantitatively analyze the differences or similarities in a measure of VADER sentiment scores and in a measure of Plutchik emotion distributions between human-written and AI-generated reviews.
- iii. **Explore the implications in terms of trust and decision-making:** Discuss how the findings impact the platform policy, applicable user literacy, and future detection tools.

1.4 Scope and Limitations of the Study

This study analyzed a sample set including 150 real Airbnb guest reviews from New York City (obtained from Kaggle) and 40 AI-generated reviews written using GPT-4 to resemble the context of the authentic reviews sourced from Kaggle. The responses of the participants were collected from 76 individuals through university networks and travel-oriented Facebook groups; thus, the results are exploratory and do not represent on a national level. This study used a combination of NLP (natural language processing)-driven text analysis tools, which includes the word clouds, the VADER sentiment scores, and Plutchik/NRC-based emotion classifications, and the inferential statistical tests, including the Chi-square goodness-of-fit test, Mann-Whitney U test, and 8x2 Chi-square test of independence.

A number of limitations are to be mentioned. The generalizability of results of other demographics and languages may be limited due to the use of convenience sampling and English-only reviews, and trust was measured indirectly via participants’ judgments of human-likeness rather than actual booking behavior, and the responses were analyzed on short-term perceptual responses with little attention to the long-term consumer trust or decision-making. The finding should be analyzed and interpreted with caution since a small respondent pool was used (N = 76) and recruited through non-random means.

However, as an exploratory study, the primary objective was to identify patterns that could guide larger, confirmatory research. Even with this small sample size, significant trends were still revealed. People prefer to read reviews based on human writing and that AI content is sometimes particularly convincing, highlighting the practical usefulness of disclosures on travel review websites.

1.5 Definition of Key Terms

Term	Operational Definition in This Study
AI generated review	A text passage describing an accommodation experience, created entirely by GPT-4 without first-hand human experience.
Human written review	A user-generated Airbnb comment obtain from genuine guest feedback.
Perceived human likeness	Binary survey choice indicating which of two reviews “feels more human.”
VADER compound score	A numeric sentiment metric ranging from -1 (negative) to $+1$ (positive).
Plutchik emotions	Eight categorical emotions—joy, trust, fear, surprise, sadness, disgust, anger, anticipation—extracted via the NRC Emotion Lexicon.
ϕ (phi) / Cramér’s V	Effect size measures for χ^2 tests; values near 0 = negligible, $\geq 0.3 \approx$ medium.
r (rank biserial)	Effect size for Mann-Whitney U; $r \approx 0.3$ denotes a medium effect.

Table 1: Key terms

1.6 Organization of the Remaining Chapters

- **Chapter II – Literature Review:** Discusses Source Credibility Theory, eWOM Theory, and empirical studies on AI- generated text and traveler behavior, which leads to two testable hypotheses.
- **Chapter III – Research Methods:** Outlines data sources, survey design, methods of computational techniques, and statistical analysis.
- **Chapter IV – Results and Analysis:** Provides descriptive patterns and inferential statistics in response to each research question.
- **Chapter V – Discussion, Conclusions, and Recommendations:** Interprets findings in relation to theory, evaluates the hypotheses, acknowledges limitations, and discusses practical implications for platforms, developers, and future research.

CHAPTER II:
REVIEW OF THE LITERATURE

2.1 Relevant Theory

This study is based on “*Source Credibility Theory*” which highlights that human are more inclined to believe what they have experienced and seen rather than artificial content, and persuasion is caused by the trustworthiness and competence of the source (Hovland & Weiss, 1951). Moreover, “*Electronic Word-of-Mouth Theory*” explains the role of how user-generated reviews reduce uncertainty in high-stakes services like travel (Park & Lee, 2008). These frameworks guide in reviewing whether AI-generated reviews can compete with human-written reviews in building trust and authenticity

2.2 Literature Survey

Online Reviews in Travel Decision Making

Online reviews play a significant role in travel behavior. According to a study “Use and Impact of Online Travel Reviews”, narrative depth and trustworthiness in travel reviews have a strong correlation with booking intention (Gretzel & Yoo, 2008). This is supported by a recent finding, where a correlation ($r = 0.563$) between review valence and purchase decision was found (Qiu & Zhang, 2023). These findings emphasize how review content, tone, and credibility play a significant role in customer choices.

The Rise and Reception of AI-Generated Content

With the development of large language models (LLMs) like GPT-4, the creation of artificial content has increased. According to a study there has been an increase of 279% in AI-generated Google reviews since 2019 (Lambert,2025), and a study shows that the responses generated by AI in service platforms can be similar to human responses, which is why some platforms adopted “AI-assisted” labels (Lim & Schmäzle, 2024). Even with this, there is still doubt among those who use it.

The “AI aversion” effect was recorded in a study “*The great AI witch hunt*”, where it shows an effect of reduced trust in text that users suspect of being machine-generated (Hadan

et al., 2024). However, to identify AI-generated content, stylometric detection methods (Schuster, Schuster, Shah, & Barzilay, 2020) and models such as DetectGPT have been established, but their effectiveness remains limited (Mitchell, Lee, Khazatsky, Manning, & Finn, 2023). This inconsistency in detection creates uncertainty in users' capability to judge the authenticity of online reviews.

Perceptions, Emotions, and the Challenge of Detection

The more advanced AI-generated reviews are, it would become more difficult to tell the difference between AI-generated reviews and the reviews written by people. Studies such as *"Convincingness of AI-Generated Restaurant Reviews"* (Tuomi, Abidin, Tuominen, & Ascensão, 2025), and *"AI Can Write a More Believable Restaurant Review Than a Human Can"* confirms that when it comes to the detection of AI-generated restaurant reviews, users often struggle, mostly when the content is emotionally positive (Kovács, 2024).

One of the main elements of credibility in reviews is emotional expressiveness. A study, *"Emotions and Cognitions When Reading Online Reviews: Effect on Tourism Service Image"* indicates that emotion-rich reviews tend to enhance trustworthiness (Hidalgo-Alcazar, Ruiz, & Sicilia, 2021). Similarly, in a study to examine how emotional expressions in reviews impact consumer perceptions, the Plutchik emotion wheel was used, which shows that emotional tone contributes to perceived authenticity and trust (Li, Scott, & Walters, 2014).

Fake Review Practices and Platform Ethics

Fake reviews as a strategic weapon are not something new. A study shows that, companies tend to use the deceptive content to manipulate the perceptions (Mayzlin, Dover, & Chevalier, 2014). These practices may become harder to trace, as with the help of AI tools it is easier to mass-produce fake reviews.

As a result, concerns around transparency and ethical content labeling have grown. A recent study proposed platforms to identify and clearly reveal the machine-generated content in order to maintain the trust of the users (Floridi & Chiriatti, 2020). Reflecting this trend, platforms began labeling certain reviews as “AI generated” in, showing the industry is shifting towards more transparency.

2.3 Hypothesis Development

Based on these theories and empirical findings, the study’s hypotheses are:

- **H1:** In paired comparisons, human-written reviews will be perceived as more reliable than AI-generated reviews.
- **H2:** Participants who frequently read reviews before booking will be able to detect review authenticity more accurately.

These hypotheses were tested using statistical analyses (chi-square tests, Mann-Whitney U test) and emotion analysis (Plutchik/NRC framework).

**CHAPTER III:
RESEARCH METHODS**

3.1 Research Design

This study employed a quantitative exploratory study design to test the understanding of humans about the AI-generated vs. genuine Airbnb reviews. A between-subjects online survey was designed to present with five pairs of reviews to the participants; each pair contained one AI-generated and one authentic human-written review. Respondents chose which review felt more “human-like.” The goal was to assess perception patterns as well as the human-likeness preferences (Research Question 1 – Perceived Authenticity).

At the same time, the study used computational text analysis to examine linguistic patterns of the reviews. This dual-method approach offers an in-depth understanding of both perception (survey) and objective differences in texts (NLP analyses), addressing *Research Question 2a* (Sentiment Differences) and *Research Question 2b* (Emotion Differences).

3.2 Sample

Review Data

- **Human reviews:** 150 actual Airbnb reviews from New York City (obtained from a Kaggle dataset).
- **AI-generated reviews:** 40 reviews created using GPT-4, intentionally made to resemble the human review dataset in both tone and structure.

Survey Participants

- **Total responses:** 76 valid survey responses, which were obtained via convenience sampling (Facebook travel groups and university networks).
- **Demographics captured:** Age, gender, and frequency of reading reviews before booking accommodation. This allowed cross-tabulation to explore whether demographic factors influenced perception.

Age Group	Pair 1 (A/B)	Pair 2 (A/B)	Pair 3 (A/B)	Pair 4 (A/B)	Pair 5 (A/B)
Below 20	75.00 / 25.00	50.00 / 50.00	62.50 / 37.50	75.00 / 25.00	87.50 / 12.50
20–25	38.46 / 61.54	38.46 / 61.54	35.90 / 64.10	66.67 / 33.33	56.41 / 43.59
26–30	50.00 / 50.00	25.00 / 75.00	25.00 / 75.00	50.00 / 50.00	50.00 / 50.00
31–35	25.00 / 75.00	25.00 / 75.00	0.00 / 100.00	50.00 / 50.00	50.00 / 50.00
36–40	25.00 / 75.00	25.00 / 75.00	37.50 / 62.50	62.50 / 37.50	75.00 / 25.00
41–45	33.33 / 66.67	100.00 / 0.00	66.67 / 33.33	33.33 / 66.67	100.00 / 0.00
46–50	33.33 / 66.67	33.33 / 66.67	33.33 / 66.67	66.67 / 33.33	66.67 / 33.33
50+	71.43 / 28.57	28.57 / 71.43	42.86 / 57.14	71.43 / 28.57	100.00 / 0.00

Table 2: Human-like Preference (%) by Age Group

Gender	Pair 1 (A/B)	Pair 2 (A/B)	Pair 3 (A/B)	Pair 4 (A/B)	Pair 5 (A/B)
Female	42.00 / 58.00	40.00 / 60.00	32.00 / 68.00	66.00 / 34.00	70.00 / 30.00
Male	44.00 / 56.00	32.00 / 68.00	52.00 / 48.00	60.00 / 40.00	64.00 / 36.00
Other	100.00 / 0.00	100.00 / 0.00	0.00 / 100.00	100.00 / 0.00	0.00 / 100.00

Table 3: Human-like Preference (%) by Gender

Review Reading Habit	Pair 1 (A/B)	Pair 2 (A/B)	Pair 3 (A/B)	Pair 4 (A/B)	Pair 5 (A/B)
Always	31.82 / 68.18	36.36 / 63.64	36.36 / 63.64	61.36 / 38.64	63.64 / 36.36
Never	25.00 / 75.00	25.00 / 75.00	25.00 / 75.00	50.00 / 50.00	75.00 / 25.00
Sometimes	64.29 / 35.71	42.86 / 57.14	42.86 / 57.14	71.43 / 28.57	71.43 / 28.57

Table 4: Human-like Preference (%) by Review Reading Habit

3.3 Questionnaire Development

The online questionnaire was created using Google Forms, which contained:

- **Review Evaluation Section:** Five sets of review pairs labeled A/B; each pair contained one AI-generated and one human-written review. The key question was:

“Which review feels more human-like?” Pair order and labeling were randomized to reduce bias.

Review Pair	Option Preferred More	A (%)	B (%)
Pair 1	B	43.42	56.58
Pair 2	B	38.16	61.84
Pair 3	B	38.16	61.84
Pair 4	A	64.47	35.53
Pair 5	A	67.11	32.89

Table 5: Human-like Preference Percentage by Review Pair

This section addressed RQ1 (Can readers tell apart AI vs. real reviews?).

- **Demographic Section:** Participants answered multiple-choice questions covering:
 - i. Age range (e.g., below 20, 20–25, etc.)
 - ii. Gender
 - iii. Frequency of reading reviews before booking (“Always”, “Sometimes”, “Never”)

All items were close-ended for ease of quantitative analysis.

3.4 Data Collection

- **Real Reviews:** Scraped anonymously from Airbnb listing reviews in the Kaggle dataset.
- **AI Reviews:** Generated using GPT-4 with prompts crafted to resemble the actual review in content, length, and style.
- **Survey Deployment:** Google Form was shared online and remained open for several days. All 76 completed responses were included; incomplete responses were excluded. Reviews were anonymized to eliminate personal identifiers like names or locations.

3.4.1 Tools and Techniques:

The study employed a combination of computational and statistical tools to analyze linguistic patterns and survey responses. The following is a brief description of each method:

- **Word Cloud:** Visual representation of word frequency with words used most. Highlights dominant terms in textual data.
- **VADER Sentiment Analysis:** A rule-based sentiment analyzer specially tuned for short, informal text (like reviews), returning polarity scores from -1 (negative) to +1 (positive) (Barik & Misra, 2024).
- **Plutchik Emotion Model (via NRC Lexicon):** This model classifies text into eight basic emotions (joy, trust, fear, surprise, sadness, disgust, anger, anticipation) using pre-labeled emotion-word associations.
- **Shapiro-Wilk Test:** A statistical test to determine whether a dataset is normally distributed (Dudley, 2012).
- **Mann-Whitney U Test:** A non-parametric test comparing differences between two independent groups when the data does not follow a normal distribution (“Mann–Whitney U Test,” 2010b).
- **Chi-square Goodness-of-Fit Test:** Used to assess whether observed categorical frequencies deviate from an expected distribution (e.g., 50:50 human vs. AI choices).
- **Chi-square Test of Independence:** Determines whether two categorical variables are associated (e.g., review type x emotion).

Effect Size Metrics:

- **Rank Biserial Correlation (r):** Measures effect size for Mann-Whitney U test.
- **Cramér’s V and Phi (ϕ):** Used with Chi-square tests to indicate the strength of association.

Additional Analytical Techniques:

- ***Survey Preference Counts:*** A simple tally of how many participants preferred the AI-generated or human-written review in each pair.
- ***Demographic Cross-tabulation:*** A method for comparing preference choices across different demographic groups (age, education, review-reading frequency).
- ***Bar Charts and Visualizations:*** Used to clearly present survey choices and emotional/sentiment distributions.

3.5 Data Analysis Plan

Python was used as a data analysis tool with libraries: pandas, seaborn, matplotlib, nltk, and scipy.

A. Textual Analysis

- ***Word Cloud:*** Visual frequency analysis of terms in AI vs. human reviews.
- ***Sentiment Analysis:*** Used VADER to extract compound sentiment scores (for RQ2a).
- ***Emotion Analysis:*** Plutchik's eight emotions were mapped using the NRC lexicon (for RQ2b).
- ***Visualizations:*** Word clouds, sentiment histograms, and radial emotion plots provided visual comparison.

B. Survey Data Analysis

- ***Preference Counts:*** Counted how many participants chose AI or human review for each pair (*for RQ1*).
- ***Demographic Cross-tabulation:*** Analyzed whether factors like review-reading frequency influenced choices.

- **Visual Representations:** Bar charts were prepared to illustrate choice distributions.

C. Statistical Testing

In order to determine the difference and establish the validity of the results, the following inferential tests were conducted ($\alpha = 0.05$):

1. A **Shapiro-Wilk normality test** was used to assess the distribution of sentiment scores, confirming non-normality.

Group	Shapiro-Wilk p-value	Normal
Human	0.0000	No
AI	0.0000	No

Table 6: Shapiro-Wilk Normality Test Results

2. To answer RQ2a (linguistic sentiment differences), a **Mann-Whitney U** test was conducted to compare VADER compound sentiment scores between human ($n = 150$) and AI-generated ($n = 40$) reviews. Effect size was reported using rank biserial correlation ($r = Z/\sqrt{N}$).

Test Used	U Statistic	p-value	Effect Size (r)	Median (Human)	Median (AI)
Mann-Whitney U	4307.00	0.0000	0.307	0.899	0.717

Table 7: Mann-Whitney U Test Results

3. For RQ1 (perceived authenticity), a **Chi-square goodness-of-fit test** was applied to each review pair to determine whether participant choices between "human" and "AI" reviews significantly deviated from a 50:50 chance. Effect size was reported as phi ($\phi = \sqrt{\chi^2 / N}$).

Pair	Chosen Real	Chosen AI	% Real	% AI	χ^2	p-value	ϕ
Pair 1	33	43	43.4	56.6	1.32	0.2513	0.13
Pair 2	47	29	61.8	38.2	4.26	0.0389	0.24
Pair 3	47	29	61.8	38.2	4.26	0.0389	0.24
Pair 4	27	49	35.5	64.5	6.37	0.0116	0.29
Pair 5	51	25	67.1	32.9	8.89	0.0029	0.34

Table 8: Chi-Square Test

Figure 1: Review Preference Real vs. AI

4. To explore RQ2b (emotion distribution differences), a **Chi-square test of independence** was conducted using an 8x2 contingency table (eight Plutchik emotions x two review types). Effect size was reported using Cramér's V.

Emotion	Human Count	AI Count
Joy	252	40
Trust	297	50
Fear	40	5
Surprise	89	12
Sadness	65	10
Disgust	20	11
Anger	27	9
Anticipation	199	24
Chi² (df=7)		18.23
p-value		0.0110
Cramér's V		0.048

Table 9: Emotion Distribution Comparison

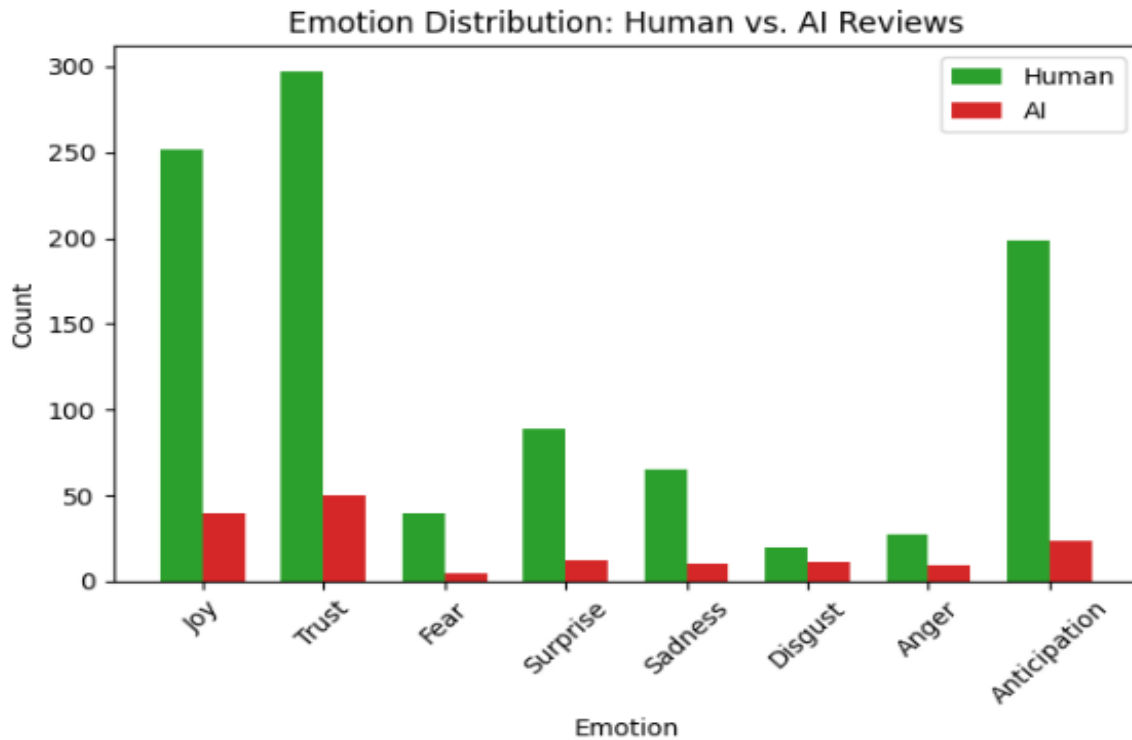


Figure 2: Emotion Distribution Real vs. AI

These sample testing provides statistical support for RQ1 (participants' ability to identify human vs. AI reviews), RQ2a (differences in sentiment polarity and intensity), and RQ2b (differences in emotional expressions).

Together, these findings offer insight into how perception of writing language and perceived authenticity may shape trust and perception toward AI-generated travel reviews.

CHAPTER IV:
RESEARCH FINDINGS & ANALYSIS

Sentiment	AI Reviews (%)	Real Reviews (%)
Positive	70.00	94.00
Negative	20.00	2.67
Neutral	10.00	3.33

Table 10: Sentiment Distribution Comparison (%): AI vs. Real Reviews

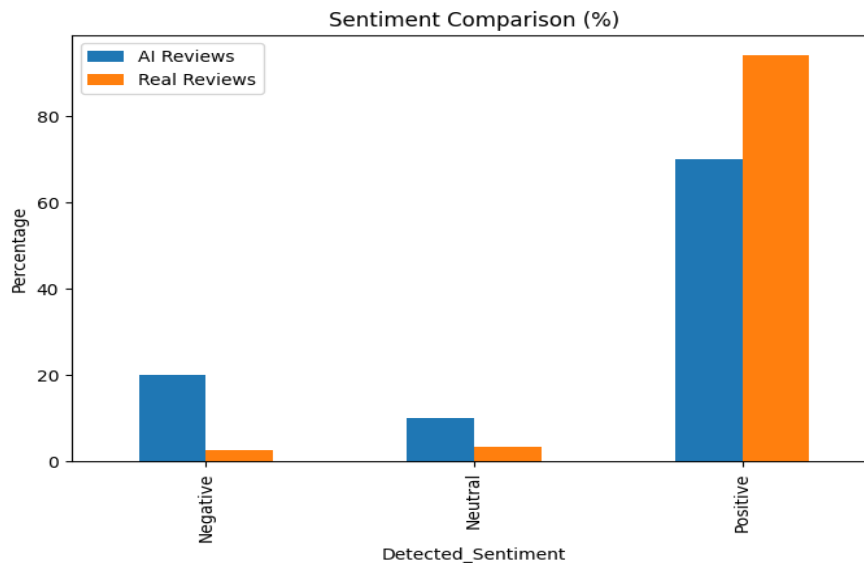


Figure 5: Sentiment Comparison (%)

This confirms that AI tends to generate overly optimistic language, potentially reducing realism, while real human feedback allows for nuance and imperfection.

Statistical Summary:

Human reviews were significantly more positive than AI reviews (*median* = 0.83 vs 0.38), *Mann-Whitney* $U = 4\,307$, $p < .001$, $r = 0.31$ (*medium effect*).

Mean compound scores:

- Human: 0.776 ± 0.17
- AI: 0.409 ± 0.31

4.1.3 Emotion Count and Comparison (RQ 2b)

Applying the wheel of emotions by Plutchik using the NRC lexicon in Python, every review was assigned to zero or more emotional classes. The ratio of each of the main emotions was compared in the AI-generated and real reviews.

Emotion	AI Reviews (%)	Real Reviews (%)
Positive	50.00	54.67
Trust	32.50	24.00
Negative	7.50	2.67
Fear	5.00	15.33
Surprise	2.50	0.67
Anger	2.50	2.00
Anticipation	—	0.67

Table 11: Emotion Count and Comparison

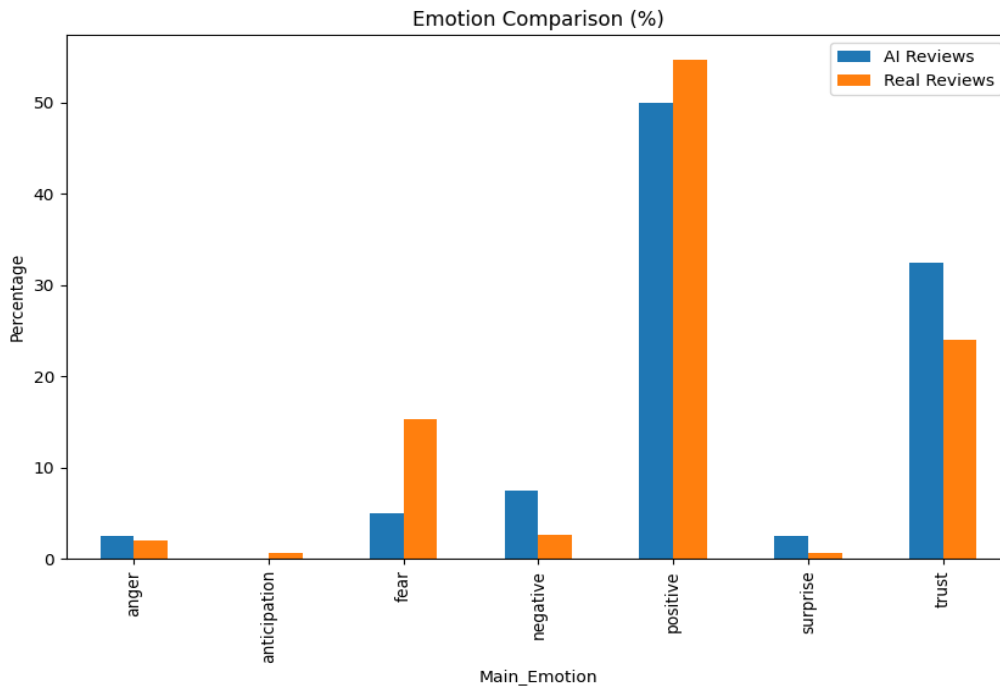


Figure 6: Emotion Comparison (%)

Emotion detection mapped eight core emotions: *joy*, *trust*, *anticipation*, *surprise*, *fear*, *sadness*, *disgust*, and *anger*.

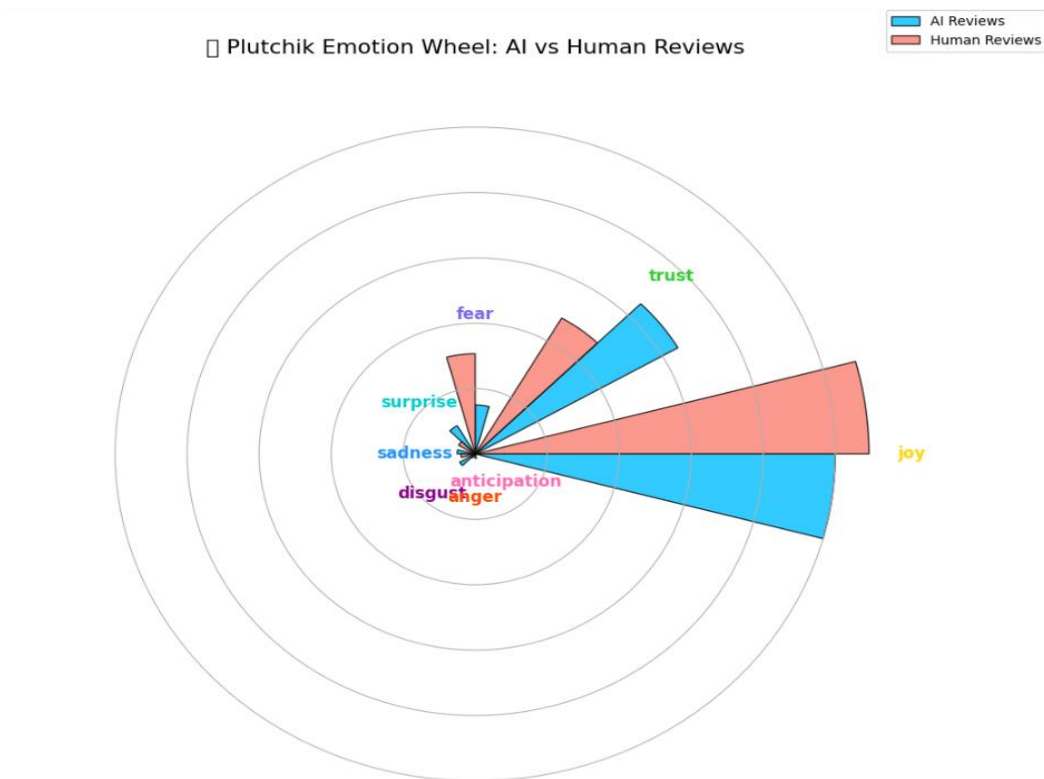


Figure 7: Plutchik Emotion Wheel: AI vs. Human Reviews

Emotion	Dominant in
Joy	Human
Trust	AI
Fear	Human
Surprise	AI
Sadness	AI

Table 12: Dominant Emotion Human vs. AI

Real reviews conveyed more varied emotional depth (e.g., *joy*, *fear*), while AI reviews focused heavily on positive tones like *trust* and *surprise*.

Statistical Summary:

An 8×2 χ^2 test showed a statistically significant but small difference in emotion profiles, $\chi^2(7, N = 1,095) = 18.23, p = .011, \text{Cramér's } V = 0.05$.

- *Human*: more *joy* (252 vs 40), *anticipation* (199 vs 24)
- *AI*: more *trust* (297 vs 50), *surprise* (89 vs 12)

4.1.4 Radial Plutchik Emotion Wheel Visualization

A radial chart highlighted the emotional distribution of each review type:

- **AI Reviews:** High level of concentration around *trust* with little spread to negative emotions.
- **Real Reviews:** More evenly distributed across both positive and negative emotional fields (e.g., *fear*, *joy*).

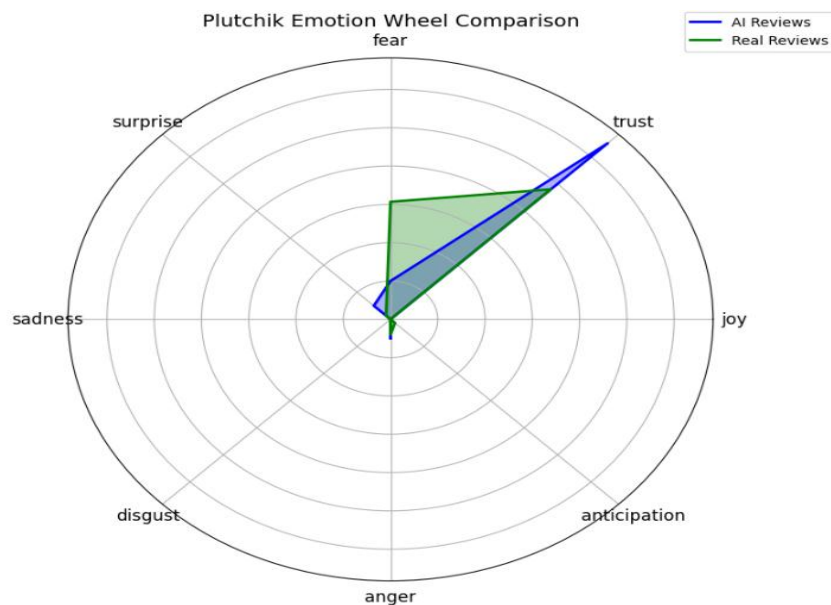


Figure 8: Radial Plutchik Wheel

This visualization reinforces those real reviews provide a richer emotional pattern.

4.2 Survey-Based Analysis: Human Perception of Review Authenticity

Participants were presented with five review pairs (each containing one AI-generated and one real review) and asked which one felt more “human-like.”

4.2.1 Human-Like Preference Summary (RQ 1)

Across 76 total responses:

- In four out of five pairs, the real review was more often seen as human-like.
- In only pair, the AI review was chosen more frequently.

This indicates that although real reviews are generally preferred, AI reviews can sometimes be convincing enough to appear authentic and mislead readers.

A χ^2 goodness-of-fit test was applied to each pair ($n = 76$).

Four of five pairs deviated significantly from chance ($p < .05$).

Effect sizes (ϕ) ranged from small to medium.

Pair	Human chosen	AI chosen	χ^2	p	ϕ	Interpretation
1	43	33	1.32	.251	.13	Human >>
2	47	29	4.26	.039	.24	Human >>
3	47	29	4.26	.039	.24	Human >>
4	27	49	6.37	.012	.29	AI >>
5	51	25	8.89	.003	.34	Human >>

Table 13: Effect sizes from small to medium

4.2.2 Percentage Preference

The following table summarizes participant preferences:

Review Pair	% Chose Real	% Chose AI
Pair 1	56.6%	43.4%
Pair 2	61.8%	38.2%
Pair 3	61.8%	38.2%
Pair 4	35.5%	64.5%
Pair 5	67.1%	32.9%

Table 14: Preference % Real vs. AI

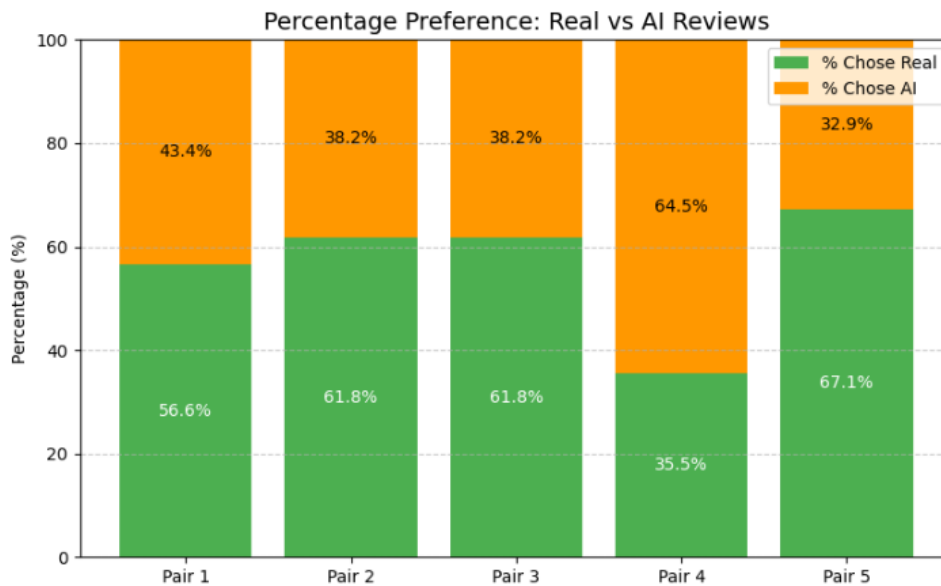


Figure 9: Preference % Real vs. AI

The exception is Pair 4, where most of the respondents considered AI-written review more human than a real one which is an important observation that should be used to justify better review labeling policies.

4.2.3 Demographic Influence Analysis

Demographics included age, gender, and habitual review-reading behavior.

- Frequent review readers were slightly better at choosing real reviews.
- No consistent pattern was found across age or gender.

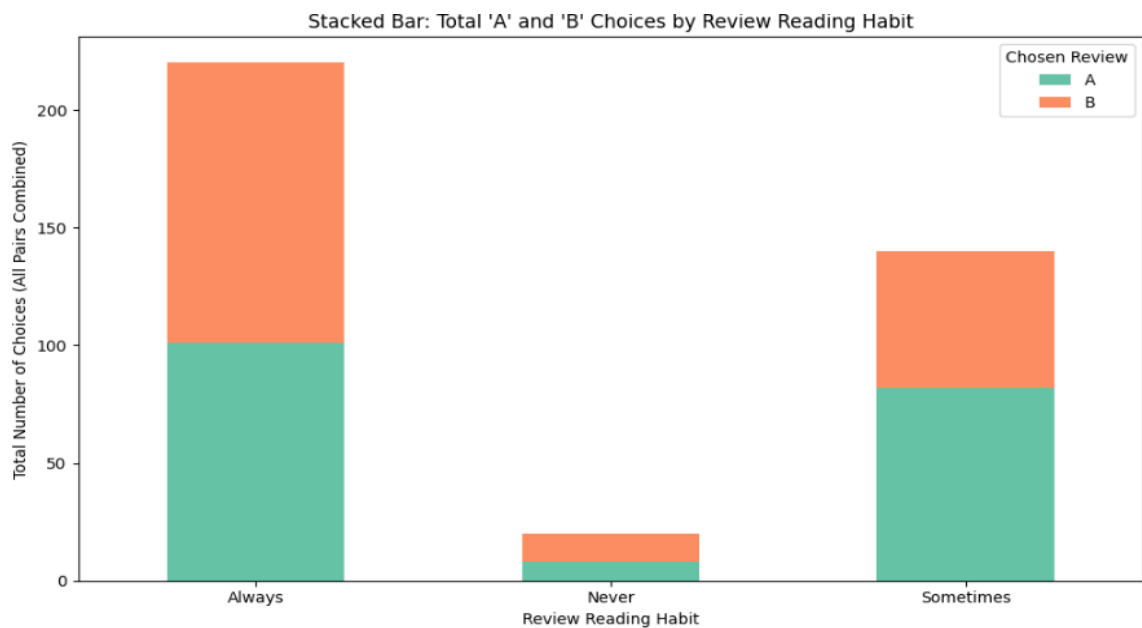


Figure 10: Choices by Review Reading Habit

4.2.4 Visual Insights

Python visualizations (bar plots and count plots) showed a tendency toward variability in choice across pairs. There was never one style of review that prevailed, highlighting the fluent nature of writing created by AI.

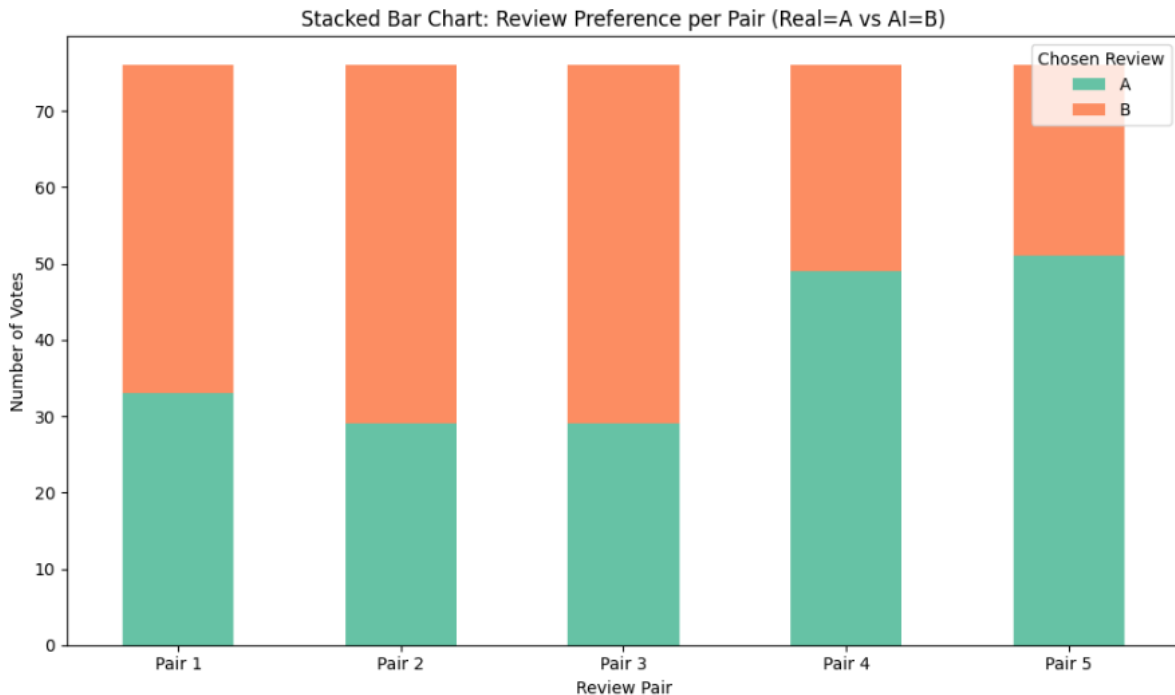


Figure 11: Review Preferences

4.3 Summary of Key Insights

- AI-generated reviews often show positive tone (*trust, surprise*), but lack emotional diversity.
- Real reviews are more emotionally detailed and authenticated by being more specific.
- Survey participants preferred real reviews in most cases, but AI reviews occasionally appeared more “human.”
- Frequent review readers showed a small edge in detection accuracy.

These findings showcase how AI text can more closely mimic human interaction in a constantly developing way and how users might struggle to distinguish natural responses against those produced synthetically. The results are consistent with the ongoing demands of content platform transparency, labels, and digital literacy.

**CHAPTER V:
DISCUSSION**

5.1 Conclusions

This study examined how authentic and emotional AI-generated reviews and real user reviews seemed in the travel and accommodation industry. By combining computational analysis (e.g., sentiment and emotion detection using NLP) with a perception-based survey, the research aimed to understand whether AI-generated content was convincing enough in providing the same human-written travel experience, and how such AI-generated reviews would be received by readers.

The computational text analysis showed that although AI-generated reviews were fluent and grammatically perfect, they had a more limited emotional range. Real Airbnb reviews showed more diverse and experience-oriented language, such as “place,” “great apartment,” “location,” and names like “Garon” and “Jennifer,” which indicate personal descriptions. Sentiment analysis showed both review types skewed positively, although AI-generated reviews were more balanced with positivity, with an absence of negative or neutral tones. Emotion analysis using Plutchik’s Wheel of Emotions framework revealed different emotional signatures between AI-generated reviews and human-written reviews. AI-generated reviews more frequently expressed trust, surprise, and sadness, whereas human-written reviews mostly expressed joy and fear. This difference highlights that although AI is capable of mimicking emotional language, it is limited in terms of the natural diversity of emotions found in real-life experiences.

Survey results confirmed that while real reviews were more frequently perceived as human-like (preferred in 4 out of 5 pairs), AI reviews were still selected as more authentic in one case. Though human-written reviews are still the more reliable in every aspect, AI-generated content can still sometimes appear convincingly natural to mislead or convince readers. Additionally, respondents who always read reviews prior to booking the accommodations indicated a little bit higher ability to identify the real review, which means that habitual exposure to review material might improve the ability to read the reviews.

In conclusion, the study presents not only theoretical implications but practical ones: it supports previous literature on the importance of emotional richness and authenticity in

travel reviews while revealing the new challenges that may arise as AI-generated content grows more complex. With the increasing advancement of the AI tools, it may become harder to tell the difference between human and artificial feedback, which gives rise to the questions about governance, trust design, and the ethics of transparency of the platforms.

Statistical testing strengthened these insights. The Chi-square test confirmed H1, indicating that human-written reviews are more often perceived as authentic, although Pair 4 proved that AI can occasionally perform better. The Mann-Whitney U test ($r = .31$) partially supported H2, confirming a measurable sentiment gap. Contrary to prior assumptions about LLMs being “overly positive,” the AI reviews in this study leaned toward more neutral tones. The Chi-square test of the differences in the numbers of emotions was found to be statistically significant, but the effect size (Cramer's $V = .05$) indicated that the emotional palette was reasonably similar across review types.

5.1.1 Hypothesis Evaluation

The findings allow for the evaluation of the hypotheses proposed in Chapter II:

- **H1:** In paired comparisons, human-written reviews will be perceived as more reliable than AI-generated reviews.

Supported. In 4 out of 5 review pairs, participants selected the real (human-written) review as more human-like and trustworthy, indicating a clear preference for real reviews over AI-generated ones.

- **H2:** Participants who frequently read reviews before booking will be able to detect review authenticity more accurately.

Partially supported. While frequent review readers performed slightly better in identifying real reviews, the difference was relatively small. However, there was an emerging pattern indicating that the accuracy of the perception could be affected by review-reading habits.

5.2 Scope for Future Study

While this study provided insightful findings, several limitations provide opportunities for future research.

Since the sample was comparatively small ($n = 76$) and was collected using non-probability convenience sampling techniques, the findings should be interpreted with caution, and it is not reasonable to apply these findings to the general population. The demographic scope was also limited, primarily reflecting responses from university students and online travel communities.

However, it is important to note that this study was exploratory in nature, and its aim was to reveal the patterns of perceptions and emotional indicators instead of statistically meaningful generalizations. The goal was to identify trends and raise questions about the evolving role of AI in digital review spaces, not to offer definitive conclusions.

Even with a limited sample, meaningful trends were observed, such as a general preference towards human-written reviews, and the AI-generated reviews became occasionally believable. These results indicate that AI content is becoming more sophisticated and can affect the trust of the user.

Future studies should consider:

- Collecting larger and more diverse samples across different cultural and linguistic groups.
- Carrying out longitudinal research on how AI perception evolves with time.
- Including behavioral tracking techniques (e.g., eye-tracking or response time) to discover more cognitive activity used in evaluating the authenticity of reviews.
- Analyzing platform-specific influences, such as visual design or reviewer metadata, on trust perception.

Analysis of trust was assessed indirectly through “human-likeness” choices, not through a validated trust scale. Future studies should include direct trust metrics. Similarly, the

actual impact on user decision-making (e.g., willingness to book) was not assessed, and it is a path that remains uncovered.

5.3 Recommendations

For Online Travel Platforms:

- AI labeling policies: Platforms like Airbnb could implement soft AI labeling (e.g., “AI-assisted”) to promote transparency without compromising the user trust.
- Encouraging detailed reviews: Platforms should encourage users to add as many details and emotions as possible to help maintain authenticity in human-written reviews.

For Researchers:

- Incorporate behavioral markers: Future experiments should include eye-tracking, reading time, or cognitive-load measurements to further investigate how users assess authenticity.
- Broader emotion modeling: Use deeper sentiment/emotion models (e.g., transformer-based emotion classifiers) to better capture tone in AI vs. human narratives.

For Developers of LLMs:

- Add emotional nuance: Training models to mimic emotional inconsistency and ambiguity may improve the sense of reality of AI-generated reviews.
- Anti-positivity measures: To gain credibility, AI reviews should not repeat all the positive features and leave some natural flaws.

For Users and Readers:

- Be cautious of overly generic praise: AI-generated reviews are usually positive and lack experiential specificity. Users should be encouraged to look for concrete details and story depth as a sign of authenticity.

Reference

- Chong, A., Khong, K. T., McCabe, S., & Wang, Y. (2018). Analyzing key influences of tourists' acceptance of online reviews in travel decisions. *Internet Research*, 28, 564–586. <https://doi.org/10.1108/INTR-05-2017-0212>
- Zervas, G., Proserpio, D., & Byers, J. (2015). A first look at online reputation on Airbnb, where every stay is above average. *Marketing Letters*, 32, 1–16. <https://doi.org/10.1007/s11002-020-09546-4>
- Acerbi, A., & Stubbersfield, J. (2023). Large language models show human-like content biases in transmission chain experiments. *Proceedings of the National Academy of Sciences of the United States of America*, 120. <https://doi.org/10.1073/pnas.2313790120>
- Lambert, M. (2025, January 20). From 2019 to 2024: AI-generated Google reviews increased by 279.2%. *Originality.AI*. <https://originality.ai/blog/ai-google-reviews-study>
- Christensen, J., Hansen, J., & Wilson, P. (2024). Understanding the role and impact of generative artificial intelligence (AI) hallucination within consumers' tourism decision-making processes. *Current Issues in Tourism*, 28, 545–560. <https://doi.org/10.1080/13683500.2023.2300032>
- Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, 15(4), 635–650. <https://doi.org/10.1086/266350>
- Park, C., & Lee, T. M. (2008). Information direction, website reputation and eWOM effect: A moderating role of product type. *Journal of Business Research*, 62(1), 61–67. <https://doi.org/10.1016/j.jbusres.2007.11.017>
- Gretzel, U., & Yoo, K. H. (2008). Use and impact of online travel reviews. In *Information and Communication Technologies in Tourism 2008* (pp. 35–46). Springer. https://doi.org/10.1007/978-3-211-77280-5_4
- Qiu, K., & Zhang, L. (2023). How online reviews affect purchase intention: A meta-analysis across contextual and cultural factors. *Data and Information Management*, 8(2), 100058. <https://doi.org/10.1016/j.dim.2023.100058>

- Lim, S., & Schmäzle, R. (2024). The effect of source disclosure on evaluation of AI-generated messages: A two-part study. *Computers in Human Behavior: Artificial Humans*, 2(1), 100058. <https://doi.org/10.1016/j.chbah.2024.100058>
- Hadan, H., Wang, D. M., Mogavi, R. H., Tu, J., Zhang-Kennedy, L., & Nacke, L. E. (2024). The great AI witch hunt: Reviewers' perception and (mis)conception of generative AI in research writing. *Computers in Human Behavior: Artificial Humans*, 100095. <https://doi.org/10.1016/j.chbah.2024.100095>
- Schuster, T., Schuster, R., Shah, D. J., & Barzilay, R. (2020). The limitations of stylometry for detecting machine-generated fake news. *Computational Linguistics*, 46(2), 1–18. https://doi.org/10.1162/coli_a_00380
- Mitchell, E., Lee, Y., Khazatsky, A., Manning, C. D., & Finn, C. (2023, July 3). DetectGPT: Zero-shot machine-generated text detection using probability curvature. *Proceedings of Machine Learning Research*. <https://proceedings.mlr.press/v202/mitchell23a.html>
- Tuomi, A., Zainal Abidin, H., Tuominen, P., & Passos Ascenção, M. (2025). Convincingness of AI-generated restaurant reviews. In *Springer Proceedings in Business and Economics* (pp. 437–448). https://doi.org/10.1007/978-3-031-83705-0_36
- Kovács, B. (2024, June 14). AI can write a more believable restaurant review than a human can. *Association of American Universities (AAU)*. <https://www.aau.edu/research-scholarship/featured-research-topics/ai-can-write-more-believable-restaurant-review-human>
- Hidalgo-Alcazar, C., Ruiz, S., & Sicilia, M. (2021). Emotions and cognitions when reading online reviews: Effect on tourism service image. *Journal of Technology Management & Innovation*, 16(4), 11–21. <https://doi.org/10.4067/S0718-27242021000400011>
- Li, S., Scott, N., & Walters, G. (2014). Current and potential methods for measuring emotion in tourism experiences: A review. *Current Issues in Tourism*, 18(9), 805–827. <https://doi.org/10.1080/13683500.2014.975679>
- Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, 104(8), 2421–2455. <https://doi.org/10.1257/aer.104.8.2421>

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681–694. <https://doi.org/10.1007/s11023-020-09548-1>

Barik, K., & Misra, S. (2024). Analysis of customer reviews with an improved VADER lexicon classifier. *Journal of Big Data*, 11, 1–29. <https://doi.org/10.1186/s40537-023-00861-x>

Dudley, R. (2012). The Shapiro-Wilk test for normality. Massachusetts Institute of Technology. <https://math.mit.edu/~rmd/46512/shapiro.pdf>

McKnight, P., & Najab, J. (2022). Mann–Whitney U test. In *The SAGE Encyclopedia of Research Design*. <https://doi.org/10.4135/9781412961288.n228>

Appendix-A:

- 1) How often do you read reviews before booking accommodation?
- 2) You will be shown several pairs of reviews about the same property. For each pair, please select the one you believe is more likely to be written by a real person.

Pair 1:

A: Jennifer's studio is exactly what I was hoping for—clean, well-lit, and cozy. The location was a dream, just minutes from Times Square and Penn Station. I especially loved the skylight that flooded the room with natural light during the day. Would definitely stay again!

B: Jennifer's space was such a lovely experience. PreCheckin and checkin was detailed and seamless, she made it easier on us and answered all questions so quickly. The space itself is such a unique, clear & zen energy- we loved it. Not to mention it was very clean & close to everything! Definitely be prepared for the 4flight workout (we appreciated that)! We enjoyed the space and location & would recommend this to anyone coming to explore the city!!

Pair 2

A: Garon's home was a peaceful retreat after a long day in the city. His cats were adorable, and the location was super convenient—close to the G train and some great cafés. The vibe of the apartment was warm and artsy. Garon himself was friendly and easy to talk to. Would recommend!

B: Garon is one of the coolest individuals I've met in a long time. The room is great close to both the G and A C trains, and the neighborhood has an organic market right around the corner. Laundry is close by also. There are two beautiful cats that lounge around all day adding to the chillness of the place.

Pair 3

A: This Midtown apartment was in a great location for exploring NYC. It's simple, clean, and provided everything I needed. The bed was comfy, and the subway was just around the corner. If you're looking for a good value stay, this place is it!

B: Great location at an affordable price. The bathroom could have been cleaner, but for the price, you're mainly paying for the location, which is super convenient. Otherwise, very easy check in, safe area, spacious room with privacy.

Pair 4

A: The apartment had a charming, vintage feel, but some parts showed wear and tear. The host responded quickly to our concerns and was very accommodating. The location was unbeatable, and having a deli and subway nearby was super helpful. Might not be ideal for allergy-sensitive guests.

B: This apartment is like a real castle old and unique. The age-related stains on the bathroom floor and dark discolorations in the carpet too indeed indicate that the building was built a long, long time ago. We weren't happy with the way the apartment was cleaned. Jennifer, the host sent in a cleaning lady right the next morning...

Pair 5

A: Jennifer's place is beautiful, cozy and very well located. We were able to walk to every key point of the city from here. The studio is very clean and spacious. Thank you for your hospitality. We had an amazing time in NYC!

B: This apartment was a hidden gem! It had a gorgeous skylight, and the place felt airy and bright. The host was attentive, and even left us thoughtful touches like a welcome note. Being walking distance from the Empire State Building was a huge plus!