

Decision-Making Preferences in AI Era: Human Perception of AI in Text-Based Tasks Including Emotional Context

Yuliya PETRENKO*

Bratislava University of Economics and Business, Bratislava, Slovakia

*Corresponding author, yuliya.petrenko@euba.sk

Renata KOSÍKOVÁ

Masaryk University, Brno, Czech Republic

renata.kosikova@mail.muni.cz

András SÁRI

Central European University, Wien, Austria

Sari_Andras@phd.ceu.edu

Andrea VERGEINER

Vienna University of Economics and Business, Wien, Austria

andrea.vergeiner@s.wu.ac.at

Abstract. Significant advancements have been made in the field of artificial intelligence (AI) in various domains, ranging from decision making to creative tasks. This study embarks on a comprehensive exploration of human–AI interaction, trust in AI-generated content and the capabilities of AI in emotional tasks. Applying an economic experiment, the paper assesses the preferences between human- and AI-generated responses in email communication with emotional context. Subjects made decisions under conditions of rating and comparing the emails. The scores given to the emails provided a metric for capturing the subjects' preferences. Three hypotheses are explored: first, that humans surpass AI in understanding complex emotions; second, that knowledge of AI origins leads to evaluator bias against it, even when AI exceeds human performance; and third, that AI's content, especially in emotional contexts, is rated lower due to doubts about its authenticity. The results confirmed the first two hypotheses, highlighting humans' superior generation of emotional content and an evident bias against AI when its origins are known. Our findings open new avenues for future research, emphasising the multifaceted relationship between AI, human cognition, and emotion in text-based tasks, and shows that while AI's potential in text generation is vast, its limitations become evident in the intricate terrain of human emotion.

Keywords: artificial intelligence, decision making, emotions, human perception.

Introduction

The ever-advancing landscape of AI has ignited transformative changes across industries and societies, ushering in an era where machines simulate human-like intelligence and capabilities. This technological revolution, underscored by innovations like ChatGPT, prompts profound inquiries into the intricacies of human-AI interaction. This paper explores two pivotal dimensions within this interaction: *the nuanced interplay between users' preferences for AI-generated content and AI's competence in tasks involving the intricate tapestry of human emotions*. By delving into these dimensions, we seek to unravel the complex fabric of trust in AI, AI's efficacy in emotionally charged tasks, and the implications for AI's integration across many domains.

DOI: 10.2478/picbe-2026-0049

© 2026 Y. Petrenko; R. Kosíková; A. Sári; A. Vergeiner, published by De Gruyter.

This work is licensed under the Creative Commons Attribution 4.0 License.

In today's landscape, AI-powered technologies have permeated various facets of human existence, automating tasks and decision-making processes once reserved for human experts. As AI algorithms evolve and AI-driven applications become ubiquitous, a central question emerges: *How do users perceive and trust AI-generated content compared to human-generated content?* The answer holds critical significance for a myriad of sectors and applications. From automated customer service chatbots to content creation platforms, understanding user preferences and the factors that shape trust in AI-generated content is paramount (Wilson & Daugherty, 2018). This inquiry delves into the psychological underpinnings of user interactions with AI, shedding light on the factors determining user trust and choice between AI and human-generated content. Deconstructing these determinants has the potential to guide organizations in harnessing AI's capabilities effectively while ensuring a seamless and engaging user experience (Allen & Choudhury, 2022; Glavas et al., 2025; UNDP, 2025).

Concurrently, our exploration delves into AI's capacity in tasks that entail the intricate understanding and expression of emotions – a hallmark of human intelligence. The domain of emotion holds a unique place in human interactions, shaping how individuals communicate, understand, and respond to each other. *Can AI replicate this nuanced facet of human cognition and communication? Can AI understand and convey emotions as effectively as humans?* These questions become particularly relevant as AI systems encroach upon domains where emotional intelligence is paramount, such as mental health support, customer service, and creative content generation (Deshpande & Rao, 2017; Boz & Kose, 2018; UNDP, 2025). By scrutinizing AI's ability to navigate emotions, our investigation seeks to unravel how AI can replicate and respond to the intricate spectrum of human emotions, offering insights into potential applications and limitations.

We have devised a meticulous two-phase experimental framework to address these multidimensional facets comprehensively. In our experiment, sixty participants were asked to rate the quality of email responses generated by humans and a generative AI ChatGPT. In the initial phase, participants evaluate a series of email responses without prior knowledge of their origin – whether generated by AI or human experts. The subsequent stage introduces participants to AI-generated responses, some explicitly labelled as such, prompting participants to express their preferences among pairs of responses. Augmenting these insights, a post-experiment questionnaire captures participants' demographic information, AI experience, and ability to identify and interpret emotions.

We dissect the intricate dynamics of AI's evolving role through an empirical lens, contextualizing our findings within the broader landscape of AI research. Our work endeavours to chart the path forward for harnessing AI's potential while navigating the complex interplay of technology and human cognition.

Literature review

The ongoing conversation about the comparison of AI and human abilities across diverse tasks uncovers insights that are tied to the characteristics of the individual tasks. There are certain domains, where humans have consistently outperformed AI, including creativity, learning and adaptation, emotional intelligence, physical abilities, ethical judgment, social intelligence (Glavas et al., 2025; UNDP, 2025). In poetry creation as an example for creativity domain, humans continue to outperform AI, pointing to the unique aspects of human cognition and creativity that AI is yet to capture fully (Köbis & Mossink, 2021). Similarly, human-generated logos were found to be

more appealing than those created by AI, once again emphasizing the human touch in creative tasks (Weingarten et al., 2020).

However, a body of work has emerged demonstrating parity between AI and human performance in certain areas. AI was found to match human accuracy in the classification of legal texts, underscoring the strides AI has made in understanding complex, domain-specific language (Orosz et al., 2021). Similarly, AI demonstrated equivalent accuracy with dermatologists in detecting skin cancer, suggesting its potential as an aid in clinical diagnoses (Beltrami et al., 2022). Furthermore, AI has matched human abilities in language translation tasks, indicating the robustness of AI in natural language processing (Hassan et al., 2018).

On the other hand, there are areas where AI has demonstrated superiority over human performance, especially in tasks that demand rapid and accurate processing of vast data sets. AI's ability to outperform humans in extreme weather events from large-scale climate data is a notable example (McGovern et al., 2017). In the complex domain of protein folding prediction, AI has not only surpassed the accuracy of other systems but has also significantly outperformed human expertise, as evidenced by its groundbreaking performance in the Critical Assessment of Protein Structure Prediction (CASP) competition (Senior et al., 2020).

The dynamic and multifaceted nature of this comparative field necessitates ongoing research to fully comprehend the interplay between AI and human abilities across different task domains. This growing body of literature not only elucidates the current capabilities of AI but also sheds light on the potential trajectory of AI evolution and its implications for human roles across various fields.

Based on the above, we propose the following hypothesis. *Hypotheses 1. Humans will outperform artificial intelligence (Chat GPT) in text generation connected to strong emotions.*

This hypothesis is based on research from the fields of computer science, psychology, and linguistics. The primary rationale is rooted in the understanding of AI's current limitations in emotion understanding and generation. Despite substantial progress in emotion recognition and classification from various sources such as facial expressions, speech patterns, and text data (Deshpande & Rao, 2017; Boz & Kose, 2018; Khurana et al., 2023), AI's capabilities in generating text that accurately and authentically represents strong human emotions remains a challenging endeavor (Hovy, 2015; Glavas et al., 2025).

Underlying this challenge are the intricacies of human emotions, which are subjective, multi-faceted, and highly context-dependent (Ekman et al., 1999). Human emotional experiences often contain subtle nuances and rich complexities that are deeply tied to our unique personal histories and socio-cultural contexts (Burkitt, 2021). These complexities can pose substantial challenges to AI systems, which often rely on predefined emotion categories and may struggle to capture the full spectrum of human emotional experiences (Plutchik, 2001).

While AI has demonstrated promising capabilities in specific tasks such as pattern recognition and prediction (e.g., Dietvorst et al., 2018; Ashoori & Weisz, 2019), human intuition and judgment continue to hold sway in tasks requiring high-level cognitive processes (Castelo et al., 2019; Vaccaro, 2024; Glavas et al., 2025). Text generation connected to strong emotions is one such task, requiring not just understanding of emotional states but also the ability to construct coherent and emotionally resonant narratives, a task at which humans excel (Hertz & Wiese, 2019).

Our first hypothesis is based on the current state of AI research and our understanding of human emotional complexities, and invites rigorous experimental verification, a pursuit that would provide critical insights into the intersections of AI, human cognition, and emotions.

Artificial-intelligence-based technologies are more and more being used to automate low-stakes routine tasks, but decision makers in corporations are typically not replaced by AI, though they are increasingly provided with the possibility to be supported by technology to make better informed and more efficient decisions (Allen & Choudhury, 2022; Vaccaro et al., 2024). The accuracy of AI algorithms is rapidly improving, but empirical studies show that people are still more likely to trust their own judgement or a human expert opinion than the advice generated from AI (Commerford et al., 2022; Glavas et al., 2025).

Apart from other factors, which can result in an aversion against advice provided by AI, trustworthiness of an AI-infused decision-making process has an impact on whether people rely on the provided advice or still focus on their own judgement or human expert opinion. The list of factors, which influence the (dis)trust of decision makers in an AI, is long and has been investigated by a stream of research on “algorithm aversion”. Algorithm aversion refers to the phenomenon where humans tend to underestimate the abilities of AI and overvalue human decision making, often leading to a negative bias against AI-generated solutions (Dietvorst et al., 2015; Mariadassou et al., 2024). This bias can occur even when the AI demonstrably outperforms humans, particularly if the AI makes a noticeable mistake, leading to lower ratings (Efendić et al., 2020). Importantly, the capacity to modify the AI system can mitigate AI aversion (Dietvorst et al., 2018).

Trust in AI is depending on the respective task. Decision makers have less trust in AI on subjective compared to objective topics and they would rather trust an AI on analytical than on social exercises (Hertz & Wiese, 2019). Working under time pressure leads to less AI aversion of decision makers (Jung & Seiter, 2021). Finally, decision makers are considering the consequences of the decision and tend to trust AI advice more in lower-stake scenarios (Ashoori & Weisz, 2019).

In conclusion, while AI has made significant strides in decision-making applications, trust remains a critical issue, underscored by the persistence of “AI aversion.” Through further understanding and addressing these influencing factors, researchers and developers can enhance trust in AI, facilitating its broader acceptance and utility in decision-making processes.

The above review enables us to formulate the following hypothesis. *Hypothesis 2. Artificial intelligence is rated lower than humans, if evaluators know that the answer is coming from an artificial intelligence.*

The hypothesis postulates that the evaluation of the performance of AI is negatively biased when evaluators are aware that the output is derived from an AI. This hypothesis is grounded in the concept of “AI aversion” (Dietvorst et al., 2015; Castelo et al., 2019; Efendić et al., 2020). Further support for this hypothesis relies on considerations that evaluators’ assessments of AI might not solely be based on the objective quality of the AI’s output, but are also influenced by a variety of psychological and contextual factors as time pressure, the perceived stakes of the decision (Ashoori & Weisz, 2019; Jung & Seiter, 2021) or the type of task (Hertz & Wiese, 2019).

Considering the literature on AI, humans’ perception of AI, and AI aversion, our second hypothesis offers an important extension to these research streams by specifically examining how the evaluators’ awareness of the source of the output (AI vs. human) might influence their evaluation. Such research would provide valuable insights into how to mitigate bias against AI and facilitate more objective evaluations of AI performance.

In the rapidly evolving field of AI, one of the most intriguing areas of research pertains to the identification, analysis, and processing of human emotions. This multifaceted process involves the interpretation of different human outputs, like textual, facial, and vocal cues.

One of the primary areas of this research is textual data analysis, where AI systems are trained to recognize and measure emotions through the processing of social media content. These

AI models are capable of accurately detecting sentiments such as joy, anger, and sadness from textual interactions, highlighting the critical role of natural language processing in emotion recognition (Khurana et al., 2023). The scope of text analysis has further been extended into the field of mental health. Evidence of this comes from a study that used AI to identify symptoms of depression from Twitter feeds, underscoring the potential applications of AI in mental health monitoring (Deshpande & Rao, 2017).

AI's ability to understand emotions extends beyond textual data, reaching into the realm of facial expressions. Research demonstrates that AI, utilizing convolutional neural networks, can accurately discern basic human emotions such as happiness, sadness, and surprise from facial expressions in photos. This has important implications for sectors requiring emotion-sensitive systems, including customer satisfaction analysis and security applications (Boz & Kose, 2018).

The detection and analysis of emotions through voice and speech pattern processing is another promising area of AI research. A system was shown to be capable of distinguishing between different emotional states using machine learning techniques. This points towards the potential role of AI in providing more nuanced and human-like interactions in settings such as voice assistants or call center operations (Chamishka et al., 2022).

Collectively, these studies highlight the advancing capabilities of AI in the field of emotion detection. Yet, the ethical considerations and privacy implications of such emotionally aware systems demand equally rigorous investigation, emphasizing the need for a balanced, considered approach as this exciting field continues to evolve.

Taking the above into consideration, we propose the following hypothesis. *Hypothesis 3. In emotional tasks artificial intelligence is rated lower, if evaluators know that the answer is coming from an artificial intelligence.*

This hypothesis stems from prior research on AI aversion and scepticism about AI's capacity to handle emotional tasks convincingly (Dietvorst et al., 2015; Hertz & Wiese, 2019). Such a negative bias against AI could become more pronounced in the context of emotional tasks, as humans may hold a perception that AI lacks the capacity to understand and respond to emotions authentically, despite the advances in emotion recognition (Deshpande & Rao, 2017; Boz & Kose, 2018; Khurana et al., 2023; Glavas et al., 2025). Moreover, people might inherently question AI's ability to fully comprehend and replicate human emotional experiences as trust in AI is significantly lower for social tasks compared to analytical tasks (Hertz & Wiese, 2019).

The third hypothesis seeks to extend the current literature on AI aversion by investigating the specific context of emotional tasks. By understanding how knowledge of an output's source influences evaluators' ratings, particularly in emotional tasks, this research could offer significant insights into human-AI interaction and contribute to strategies aimed at reducing bias against AI.

Methodology

We used an experimental method to reveal people's preferences for human- and AI-generated content in an emotional context. In the experiment, subjects made decisions about email responses, some of which had an emotional context and were explicitly generated by AI. The scores given to the emails provided a metric for capturing the subjects' preferences. Subjects made their decisions by rating and comparing the emails.

Experimental design

The experimental setup consists of two experimental parts, the post-experiment questionnaire, the presentation of results, and the payoff. The following email provided the context for the

experimental exercise: *“Kim Smith, a strategy consultant, has just returned to their office after presenting the interim results and recommendations for their project at the XYZ company. The presentation was attended by three division heads who will be affected by the project’s results, and their input is crucial for its success. Kim has received an email from each division head and is now tasked with responding to them directly.”* Each response suggested by each of the three division heads corresponded to a particular emotional writing style: neutral, angry, or cheerful.

The email responses used in the experiment were of both ChatGPT and human origin. The ChatGPT from OpenAI was used as a generative AI tool to generate the ChatGPT responses. The ChatGPT was given the context, the emails from the three division heads, and a response limit of 150 words maximum. Six response emails were generated for each emotion type email, resulting in a total of 18 email responses. To generate human responses, ten professional consultants from an international consulting firm were contacted via email. They were given the context, the three emails from the division heads, and the task of responding to each of the three emails. The feedback from six consultants included their completed exercises, which were further checked for spelling and errors (e.g., replacing their name with the consultant’s name Kim Smith, etc.).

In the first part, subjects were asked to rate six email responses (two responses to each type of emails: neutral, angry, and cheerful) on a scale of 0 to 10 (where 0 means “really bad” and 10 means “really good”) using only even numbers. One of the emails in each pair was written by a generative AI (ChatGPT) and another one – by a human. The subjects were unaware of the source of the emails.

In the second part, the subjects rated three pairs of email responses, one pair for each emotion type (neutral, angry, and cheerful). The subjects were shown two emails of one emotional type at a time, side by side. The first was labelled as generated by ChatGPT, while both emails were generated by ChatGPT¹. Subjects were asked to distribute 10 points between the two emails, indicating their preferred email with higher points for three pairs of response emails using a slider², where all points had to be allocated. The participants choose between response A and response B on a scale from 0 to 10, where 0 means they like A and 10 means they like B. Participants can only choose even numbers, so they cannot choose 5-5.

As an incentive, each subject received a show-up fee of 5 euro and a payoff based on the experimental results. The payoff depended on how close the subject’s rating was to the benchmark represented by the other students’ responses. Prior to the experiment, the students at the Vienna University of Economics and Business were asked to review the study and submit their ratings. Subjects received an additional payment of 3 euro for each rating that matched the rating chosen by most of the aforementioned students. Thus, subjects could earn up to 9 ratings $\times$ 3 euro = 27 euro. Each subject would have received an additional payment of 10 euro, if the average payoff of the session was below 20 euro.

Implementation

The experiment was programmed with o-Tree. All sessions were conducted at the WULABS-A laboratory at the Vienna University of Economics and Business, Austria. The subjects were mainly undergraduate students from all departments of the Vienna University of Economics and Business. Upon arrival, subjects were randomly seated at computer terminals separated by blinds. Sixty

¹ The grouping was made in a way that matching group’s participants would see that the other email was generated by ChatGPT.

² We used a modified version of Max R.P. Grossmann’s o-Tree slider with no default value.

subjects participated in two sessions of thirty subjects each. During each session, subjects were randomly and equally assigned to one of six matching groups. Each matching group received different response emails to rate from the pool of ChatGPT and human response emails. Subjects did not receive the same email response more than once.

In the first part of the experiment, subjects received the instructions for Part 1 on paper, including general instructions and the context of the experimental task. After reading the instructions, subjects continued with the Part 1 rating exercises on their computer screens. After completing Part 1, subjects received on-screen instructions for Part 2, including a brief introduction to ChatGPT (the instructions are available upon request).

After the experiment, subjects were asked to answer several questions on demographics, English knowledge³, experience and willingness to use ChatGPT⁴, their education⁵ and work background⁶, and identification of emotions (summary statistics based on the post-experiment questionnaire is given in Table 1). Then the subjects were shown their payoffs.

Table 1. Summary statistics

Variables	count	mean	sd	min	max
Age	60	22.73	2.81	18	31
Gender (1=female) ⁷	60	1.42	0.50	1	2
English knowledge	60	3.58	0.70	2	5
Used ChatGPT before	60	3.53	0.79	2	5
Education	60	1.53	0.65	1	4
Employment	60	1.53	0.54	1	3

Source: Authors' own research.

Each of our session was run in a batch of four experiments taking 2 hours (30 minutes for each), of which only one random experiment's outcome is paid for all of the participants. One could argue, that this made our participants more tired, and less focused during the experiment, compared to the alternative option in which they only participate in our experiment. As our experiment mainly consists of reading and rating emails, we believe this represents more the state of mind in which people engage in those activities, thus it poses no threat to the conclusions drawn.

Hypothesis testing

A linear regression model is used to test the first hypothesis⁸. Our regression model is the following:

$$Y_i = \alpha + \beta_1 \times D_{M_i} + \beta_2 \times D_{A_i} + \beta_3 \times D_{C_i} + \beta_4 \times D_{M_i} \times D_{A_i} + \beta_5 \times D_{M_i} \times D_{C_i} + \epsilon_i, \quad (1)$$

where observations are indexed by i , Y is the score participants gave in part 1 to the emails on a scale of 0 to 10. D_M dummy variable's value is 1 if the email response was machine generated. D_A and D_C are dummy variables with the value 1 if the email was a response to an angry and

³ 2: Limited English knowledge, 3: Professional working, 4: Full professional 5: Native or bilingual

⁴ 1: I haven't heard about chatGPT, 2: I have heard about ChatGPT before, but haven't ever used it, 3: I have tried ChatGPT before, but haven't used it often, 4: I use ChatGPT often, 5: I use ChatGPT on everyday basis.

⁵ 1: High school, 2: Bachelor's degree, 3: Master's degree, 4: Other.

⁶ 1: Not employed, 2: Part-time employed, 3: Full-time employed.

⁷ A mean of 1.42 indicates that 58% of our sample is female.

⁸ Although our outcome variable is discrete, this is our preferred model due to its ease of interpretation. The results of the ordered logit model are similar to those of the linear regression model. The calculations are available upon request.

cheerful emails, respectively⁹. A single observation corresponds to one rating by one participant¹⁰. As the outcome variable is a discrete variable with just a few possible values, we assume the error term to be heteroscedastic. Standard errors are clustered on the participant level.

To test our first null-hypothesis that is chatGPT performs just as good as humans in emotional tasks, we conducted a Wald-test with the following two restrictions separately:

$$\beta_1 + \beta_4 = 0, \quad (2)$$

$$\beta_1 + \beta_5 = 0. \quad (3)$$

To test the second null-hypothesis, that is AI generated answers are rated just as good as human generated answers if participants know the answer is AI generated, we consider part 2 of our experiment. As participants were randomized into treatment groups where the AI prompts were reversed equally, if the average score is significantly different from 5, then we can reject the second null-hypothesis.

To test the third hypothesis, we used the following model:

$$Y_i = \alpha + \beta_1 \times D_{A_i} + \beta_2 \times D_{C_i} + \epsilon_i, \quad (4)$$

where observations are indexed by i , Y is the rating for the response, D_A and D_C are dummy variables with the value of 1 if the response was to an angry or cheerful email, respectively. If the coefficients are according to our hypothesis, their signs should be positive, meaning a higher score for a response not marked as AI-generated if it is a response to an email with emotional content.

Results and discussions

Regarding our *first hypothesis*, we found that humans outperform artificial intelligence (chatGPT) in text generation connected to strong emotions.

Table 2. Results of part 1

Variables	(1)
machine	1.100** (0.437)
angry	1.567*** (0.329)
cheerful	0.667* (0.344)
machine × angry	-1.967*** (0.569)
machine × cheerful	-2.067*** (0.633)
Constant	7.300*** (0.266)
Observations	360
R-squared	0.086

Robust standard errors in parentheses, clustered on the participant level.

*** p<0.01, ** p<0.05, * p<0.1

Source: Authors' own research.

⁹ We asked participants to categorize the emails in the post-experiment questionnaire, and all participants managed to categorize the angry email as angry, and only one participant failed to categorize the cheerful email as cheerful.

¹⁰ One participant rated multiple emails.

Using Wald-test with the regression model (1) coefficients, we can reject the first null-hypothesis at the 5% significance level and conclude, that on average human email responses score roughly one point better compared to chatGPT, when responding to an emotional email, applying our data from the first part of the experiment¹¹ (Table 2).

Regarding our *second hypothesis*, we found that artificial intelligence is rated lower than humans, if evaluators know that the answer is coming from an artificial intelligence.

On the data from the second part of the experiment, we rejected the second null-hypothesis at the 10% significance level as average score is significantly different from 5 (Table 3).

Table 3. Testing hypothesis 2

Variables	(1)
Average score	5.444* (0.24)
Observations	180

Robust standard error in parentheses, clustered on the participant level.

*** p<0.01, ** p<0.05, * p<0.1, of the t-test if the average is 5.

Source: Authors' own research.

Regarding our *third hypothesis*, we did not find enough evidence to justify that in emotional tasks artificial intelligence is rated lower, if evaluators know that the answer is coming from an artificial intelligence. As Table 4 shows, we could not reject the third null hypothesis as the coefficients were negative, meaning a lower score for a response to an email with emotional content not marked as AI-generated.

Table 4. Testing hypothesis 3

Variables	(1) Clustered SE
angry	-0.367 (0.429)
cheerful	-0.500 (0.472)
Constant	5.733*** (0.251)
Observations	180
R-squared	0.006

Robust standard error in parentheses, clustered on the participant level.

*** p<0.01, ** p<0.05, * p<0.1, of the t-test if the average is 5.

Source: Authors' own research.

One could speculate that, maybe with more observations one could measure a significant effect and reject the third null-hypothesis, but it is also possible that our email response setting, while reaches the threshold where the difference between AI and human text generation quality becomes salient, does not reach the emotional threshold, where humans care in differing degrees of algorithm-aversion. For further research we would suggest expanding the emotional spectrum.

¹¹ We conducted a Wald-test with the restrictions (2) and (3) separately. With a p-value of 0.015 and 0.016 respectively, we rejected the null-hypothesis at the 5% significance level.

Conclusion

This paper explored the intricate interplay between AI and human cognition within the context of emotion-infused tasks. The study's foundation lay in two critical dimensions: the aversion exhibited by humans toward AI-generated content and AI's efficacy in tasks that involve understanding and expressing complex human emotions. Through our experimental design, this investigation sought to unravel users' preferences, trust in AI-generated content, and AI's capabilities in emotional tasks.

The outcomes provided within the experimental framework are insightful. The first outcome revealed that humans outperformed AI in generating emotionally charged text, highlighting the intricate nature of human emotional understanding and expression that AI is yet to fully replicate. The second outcome indicated that AI's performance ratings were lower when participants were aware that the responses were AI-generated, further corroborating the notion of AI aversion. However, the third hypothesis, focusing on the rating differences between AI and human-generated responses in emotional tasks, failed to garner significant evidence, suggesting a complex interplay of factors influencing such evaluations.

The study's comprehensive exploration sheds light on the evolving landscape of AI-human interaction and underscores the need for nuanced investigations into AI's role in emotional tasks and user preferences. The findings have implications for organizations aiming to integrate AI into their workflows, emphasizing the importance of trust-building strategies, task-specific contexts, and ethical considerations. As AI continues to redefine human-technology interfaces, this study contributes valuable insights to shape a future where humans and AI collaborate seamlessly and effectively.

Acknowledgements

We would like to express our sincere thanks to Prof. Dr. Ben Greiner, for his guidance throughout the entire research, financial support of the experiment, and his advice during the preparation of the paper. This work was supported by the Early Stage Grants EUBA, project number 09I03-03-V05-00003 (A-25-101/3020-16).

During the preparation of this manuscript, the authors utilised DeepL Write tool to enhance readability and language clarity. Following the use of this tool, the authors meticulously reviewed and revised the content as necessary and assumes full responsibility for the final content of the publication.

References

- Allen, R., Choudhury, P. (2022). Algorithm-augmented work and domain experience: The countervailing forces of ability and aversion. *Organization Science*, 33 (1), 149–169.
- Ashoori, M., Weisz, J. D. (2019). In AI We Trust? Factors that Influence Trustworthiness of AI-infused Decision-Making Processes. *arXiv:1912.02675*.
- Beltrami, E., Brown, A., Salmon, P., Leffel, D., Ko, J., & Grant-Kels, J. (2022). Artificial intelligence in the detection of skin cancer. *Journal of the American Academy of Dermatology*, 87 (6), 1336–1342.
- Boz, H., & Kose, U. (2018). Emotion extraction from facial expressions by using artificial intelligence techniques. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, 9 (1), 5–16.

- Burkitt, I. (2021). The emotions in cultural-historical activity theory: Personality, emotion and motivation in social relations and activity. *Integrative Psychological and Behavioral Science*, 55, 797–820.
- Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research*, 56 (5), 809–825.
- Chamishka, S., Madhavi, I., Nawaratne, R., Alahakoon, D., De Silva, D., Chilamkurti, N., & Nanayakkara, V. (2022). A voice-based real-time emotion detection technique using recurrent neural network empowered feature modelling: Futuristic trends and innovations in multimedia systems using big data, IoT and cloud technologies (FTIMS). *Multimedia Tools and Applications*, 81, 35173–35194.
- Commerford, B. P., Dennis, S. A., Joe, J. R., & Ulla, J. W. (2022). Man versus machine: Complex estimates and auditor reliance on artificial intelligence. *Journal of Accounting Research*, 60 (1), 171–201.
- Deshpande, M., & Rao, V. (2017). Depression detection using emotion artificial intelligence. *2017 international conference on intelligent sustainable systems (ICISS)*, 858–862.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144 (1), 114.
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management science*, 64 (3), 1155–1170.
- Efendić, E., Van de Calseyde, P. P., & Evans, A. M. (2020). Slow response times undermine trust in algorithmic (but not human) predictions. *Organizational Behavior and Human Decision Processes*, 157, 103–114.
- Ekman, P., et al. (1999). Basic emotions. *Handbook of cognition and emotion*, 98 (45-60), 16.
- Glavas, D., Grolleau, G., Mzoughi, N. (2025) IT professionals trust in artificial intelligence vs. human experts for achieving sustainable development goals. *Sustainable Futures*, Vol. 10, 101153. <https://doi.org/10.1016/j.sft.2025.101153>
- Hassan, H., Aue, A., Chen, C., Chowdhary, V., Clark, J., Federmann, C., Huang, X., Junczys-Dowmunt, M., Lewis, W., Li, M., et al. (2018). Achieving human parity on automatic Chinese to English news translation. URL <https://arxiv.org/abs>.
- Hertz, N., & Wiese, E. (2019). Good advice is beyond all price, but what if it comes from a machine? *Journal of Experimental Psychology: Applied*, 25 (3), 386.
- Hovy, D. (2015). Demographic factors improve classification performance. *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (Volume 1: Long papers)*, 752–762.
- Jung, M., & Seiter, M. (2021). Towards a better understanding on mitigating algorithm aversion in forecasting: An experimental study. *Journal of Management Control*, 32 (4), 495–516.
- Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: State of the art, current trends and challenges. *Multimedia Tools and Applications*, 82, 3713–3744.
- Köbis, N., & Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in human behavior*, 114, 106553.
- Mariadassou, S., Klesse, A.-K., Boegershausen, J. (2024). Averse to what: Consumer aversion to algorithmic labels, but not their outputs? *Current Opinion in Psychology*, 58, 101839.

- McGovern, A., Elmore, K. L., Gagne, D. J., Haupt, S. E., Karstens, C. D., Lagerquist, R., Smith, T., & Williams, J. K. (2017). Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bulletin of the American Meteorological Society*, 98 (10), 2073–2090.
- Orosz, T., Vági, R., Csányi, G. M., Nagy, D., Üveges, I., Vadász, J. P., & Megyeri, A. (2021). Evaluating human versus machine learning performance in a LegalTech problem. *Applied Sciences*, 12 (1), 297.
- Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, 89 (4), 344–350.
- Senior, A.W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Židek, A., Nelson, A. W., Bridgland, A., et al. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, 577 (7792), 706–710.
- United Nations Development Programme (UNDP, 2025). Human Development Report 2025: A matter of choice: People and possibilities in the age of AI. <https://report.hdr.undp.org>
- Vaccaro, M., Almaatouq, A. & Malone, T. (2024) When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, 8, 2293–2303.
- Weingarten, E., Meyer, M. W., Ashkenazi, A., & Amir, O. (2020). Human experts outperform technology in creative markets. *She Ji: The Journal of Design, Economics, and Innovation*, 6 (3), 301–330.
- Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review*, 96 (4), 114–123.