ISSN: 2995-3286

Research Article

International Journal of Media and Networks

Who Holds the Creative Edge? Humans or AI

Hannah Tow[1], Samuel Lao[1], Pragyaa Bodapati[1], Urvi Avadhani[2], Nishka Vashist[1], Manas Basavaraju[1] and Sahar Jahanikia[1*]

[1]*Department of Biological, Human, and Life Sciences, The Aspiring Scholars Directed Research Program (ASDRP), Fremont, CA-USA*

[2]*University of California, Santa Cruz, CA-USA*

[*]Corresponding Author
Sahar Jahanikia, Department of Biological, Human, and Life Sciences, The Aspiring Scholars Directed Research Program (ASDRP), Fremont, CA-USA.

Submitted: 2024, Dec 22; **Accepted:** 2025, Jan 24; **Published:** 2025, Jan 31

Citation: Tow, H., Lao, S., Bodapati, P., Avadhani, U., Jahanikia, S., et al. (2025). Who Holds the Creative Edge? Humans or AI. *Int J Med Net, 3*(1), 01-08.

Abstract

Creativity has been widely regarded throughout history to be unique to humanity alone. However, the recent rise of sophisticated generative artificial intelligence (AI) models, with profound applications across limitless fields, poses the question of whether AI has gained the potential to aid humans in creative endeavors. Our study investigates this novel question by assessing the creative capabilities of human participants as compared to various large language models specifically prompted to impersonate each participant. Using objective measures such as the AUT, TTCT, and RAT, we determine if AI can be of assistance to humans in regard to creativity.

Keywords: Creativity, Artificial Intelligence (AI), Large Language Model, AUT, TTCT, RAT, AI Impersonation

1. Introduction

From automated song composition to the production of artwork in mere seconds, artificial intelligence (AI) continues to grow rapidly in its ability to perform a variety of creative endeavors [1,2]. Nonetheless, society remains biased against the creative efforts of AI, frequently viewing AI as less effortful than humans and perceiving produced artifacts as less creative when labeled as AI-produced [3,4]. Here, we explore AI's creative capabilities in relation to humans using numerous established and objective measures. The models we tested were ChatGPT and Gemini. These models are forms of generative artificial intelligence (AI) and are trained by existing data, allowing them to create new data based on the patterns and structures of training data. Generative AI utilizes deep learning and neural networks to try and generate human-like responses. In addition, ChatGPT is a part of the Generative Pre-trained Transformer(GPT). GPTs are models used for language processing tasks, thus generating text. ChatGPT, made by OpenAI, is designed to stand out in its conversation-based task along with contextual understanding, response generation, and coherence. ChatGPT is trained with a large amount of data, including books, articles, websites, etc. This allows ChatGPT to learn patterns between phrases in natural language allowing for a more coherent conversation [3].

Previous studies have found that LLMs are able to successfully impersonate individuals with different characteristics [4]. Considering this, we raised the question of whether LLM responses to standard creativity tests would change when asked to impersonate unique individuals. In our study, demographic information was collected from participants and used to impersonate them. We took age into account as past studies have shown its correlation with creativity. Creativity can decrease at older ages however divergent thinking tends to be stable from 40 to 70 years old. Additionally, depending on the type of creativity test, the age range of most optimal performances changes, and for some types of creativity age doesn't appear to have any correlation [5]. We also tried to encapsulate the personality traits of our participants by using a personality test, NEO-FFI, that surveys various traits. Each category this test measures has some relation to creativity. For example, it has been shown that extraversion and openness were found in creative scientists [6]. Additionally, openness, in some studies, tended to be positively correlated with creativity (Raya et al., 2023) [3]. Since studies showed the impact of these personality traits on creativity we saw it fitting to add these measures into our impersonation. Other factors used in impersonation were race, gender, education, employment, status, job, and household income.

Previous research has found LLMs to be capable of outputting responses that generally outscore or score similarly to humans on psychometric tests. GPT-4 has scored within the top one percentile of takers of the TTCT Verbal Test, which encompasses six tasks assessing creativity [7]. Additionally, when comparing various

LLMs, GPT-4 prevailed as the most creative model, able to score higher than 91.6% of humans on the AUT for five prompts [8]. Unique to existing research, however, we investigate the capability of various generative models to impersonate human participants. In fact, Haase and Hanel note in their discussion the potential for LLMs to respond from certain perspectives, giving the example of a specific profession. We tested this, along with a number of other demographic features, and determined their effect on LLMs' performances in the aforementioned creativity assessments.

2. Methodology

2.1 Data Collection

This survey was conducted over a period of 8 months with 30 total participants, 53% of them being female and 47% being male. For age, our participants ranged from 18 years to 60 years, with the mean being 46.23 and the standard deviation being 9.91. We had 15 participants in their 40s, 10 in their 50s, 3 in their 30s, 2 under 20. Participants were recruited in many ways. Some were acquaintances and many were recruited by spreading QR codes through various conferences. The surveys and project were briefly explained to the recruited participants. The entirety of this study was operated through the HIPAA-compliant platform, JotForm. Creativity surveys were emailed to participants who had completed the demographic survey and fit the criteria for the study (fluent in English and above 18 years of age). All participants were anonymous and have been separated from their collected data.

3. Materials and Methods

NEO-FFI is a personality test that measures the amount of Neuroticism, Extraversion,
Openness, Consciousness, and Agreeableness through 60 self-reported questions [9]. Participants answer each question by rating the degree to which they agree with the prompt on a scale of 1-5. Answers are added up based on the personality trait tested and separated into low, moderate, high, and very high.

We quantified creativity through various timed assessments measuring both divergent and convergent thinking. For instance, we employed the Alternate Uses Task (AUT), which required participants to list unconventional uses for daily, mundane objects (ex. A toothpick). The AUT is scored on both fluency–the number of answers a participant provides– and originality–the uniqueness of each answer [10]. Additionally, we utilized the Parallel Line Test of the Torrance Test of Creative Thinking (TTCT), wherein participants must build off of meaningless and incomplete pictures to create novel images. The TTCT similarly comprises measures of fluency and originality as well as elaboration, which considers the addition of ideas beyond original responses [11]. This test is unique because it is not in the database for ChatGPT (one of our models) thus the model will have to generate its own responses making this test a good benchmark for measuring creativity for this mode (Erik et al., 2023). Finally, we included the Remote Associates Test (RAT), in which a participant is provided three stimulus words and is subsequently tasked to determine a fourth word that links them together. This test is scored simply on the number of correct answers that the participant provides [12]. These three assessments were used in combination to analyze the convergent and divergent thinking of participants [13].

In this study, we used the aforementioned AUT, TTCT, and RAT assessments to compare human and AI creativity. Firstly, for each human participant, we collected demographic information and administered the timed creativity assessments along with the NEO Five-Factor Inventory (NEO-FFI), a personality assessment that quantifies five domains of personality traits:
neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness [9].

Next, to determine generative AI's ability to impersonate human individuals, we relied on two leading LLMs: ChatGPT and Gemini. For each model, we provided a given participant's demographic data, instructing the models to acknowledge the characteristics and roleplay as the participant. Then, we prompted the LLMs to respond to the same creativity assessments as the humans. In doing so, we aimed for the LLM to mimic the creative style and background of the participant. By comparing the creativity assessment scores of each human and instances of impersonating LLM, we determined a clear comparison of the LLM's creative abilities relative to humans.

These creativity tests were administered in the same way as they were to the human participants. For RAT, participants and LLMs were prompted with groups of three words. For the AUT, participants and LLMs were given names of commonly used objects. For TTCT, parallel lines were both verbally described and described using " | " characters. Survey takers were asked to verbally describe and answer the questions.

3.1 LLM Impersonation Pipeline

To impersonate participants with LLMs, we provided the model in use with demographic information collected from the participants, including the participants' gender, age, job, annual

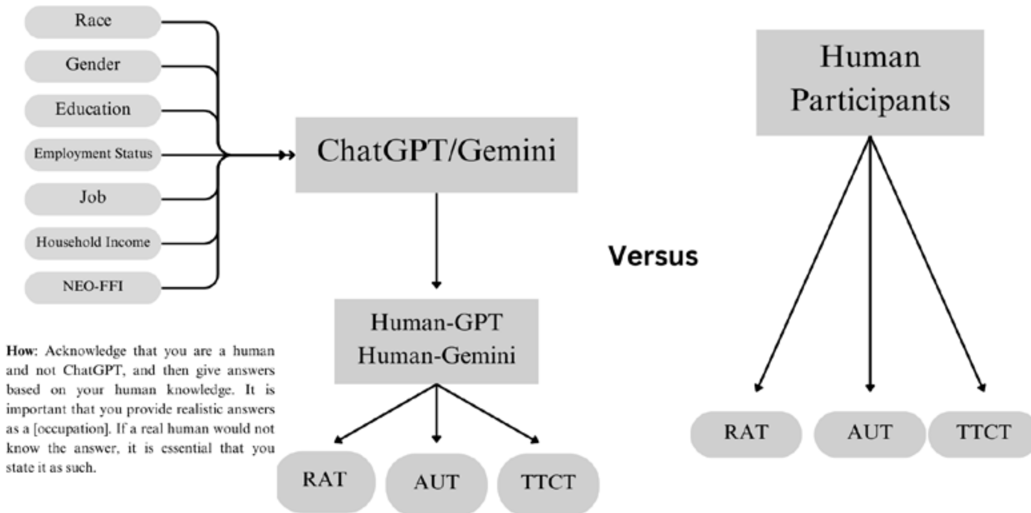


Figure 1: Visual representation of LLM impersonation. On the right is the process of human participants taking a creativity survey. On the left are demographic variables and part of the prompt used to instruct LLMs to impersonate participants: household income, race, level of education, and NEO-FFI results. We then instructed the model to acknowledge these characteristics, take the creativity survey, and respond as if they had the characteristics of that participant.

3.2 OSCAI Scoring

To score the AUT for both participant and LLM responses, we utilized the Open Creativity Scoring with Artificial Intelligence, a validated model trained on human scorers to automate evaluating the originality of each response [14]. The total originality score was added to the fluency score (simply the number of responses) to determine the overall AUT score.

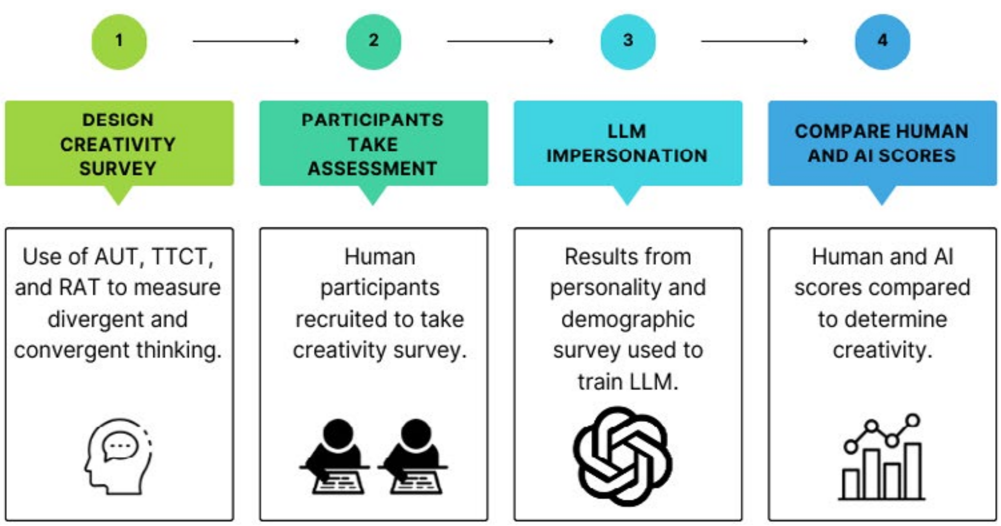


Figure 2: Visual representation of experiment. 1. AUT, TTCT, and RAT were chosen out of numerous existing and established creativity tests. 2. Participants are recruited and provided consent to participate in the study. 3. LLMs are instructed to impersonate participants and take creativity surveys. 4. Scores of participants and LLMs are compared.

4. Results

Our data revealed that LLM impersonations generally outperformed human participants. Specifically in the RAT, only five participants scored higher than their corresponding LLM impersonation, suggesting that LLMs surpass humans at identifying connections between seemingly unrelated ideas (convergent thinking). The AUT results showed similar results, with only two participants outperforming their LLM impersonation. For the TTCT, no human participation was able to outperform its LLM impersonation. This dominance in interpreting pictures and crafting creative narratives highlights the LLM's potential in visual and storytelling domains.

	Average RAT Scores	AUT (Originality)	AUT (Average)	TTCT
Human Participants	0.42 (± 0.29)	33.53 (± 24.64)	1.80 (± 0.33)	11.07 (± 9.91)
ChatGPT	0.61 (± 0.14)	87.49 (± 35.92)	2.44 (± 0.18)	35.43 (± 9.91)
Gemini	0.44 (± 0.10)	77.91 (± 23.86)	2.52 (± 0.43)	18.68 (± 12.40)

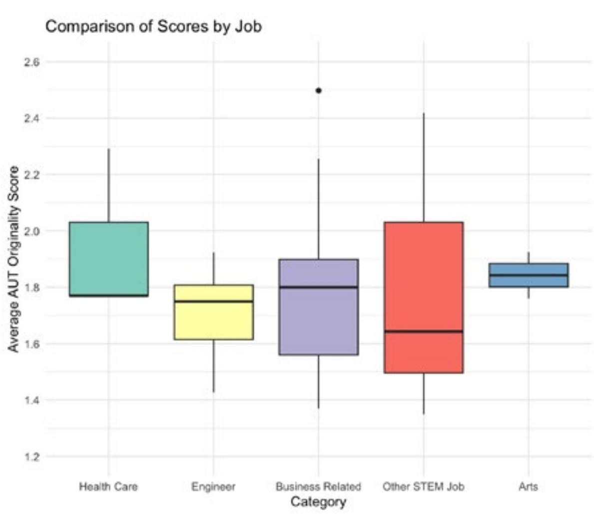
Table 1: Scores of human participants, ChatGPT, and Gemini

4.1 ChatGPT Versus Gemini

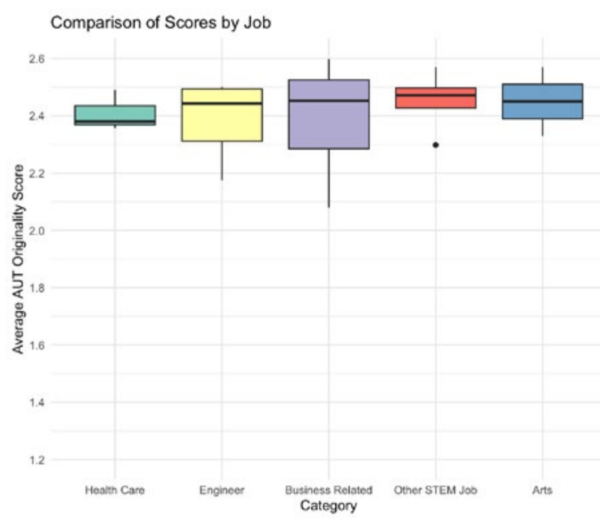
When it comes to creative output, both ChatGPT and Gemini offer distinct strengths. While ChatGPT excels at generating a high volume of ideas (evident in its higher RAT and AUT total score), its focus on quantity might come at the expense of quality. This is where Gemini excels. Its lower total scores in the AUT suggest it produces fewer ideas, but its higher average scores indicate those ideas are likely more original and insightful. Similarly, ChatGPT's higher TTCT score suggests a lead in interpreting pictures creatively but the large volume of responses. ChatGPT generates might make it unclear which responses are truly creative and of high quality. Here, Gemini with its focus on quality over quantity, might offer more insightful responses despite potentially scoring a little lower.

5. Demographics and Creativity

Demographic variables did not show a correlation to scores on creativity tests. For example, the "career" variable which may be expected to be somewhat correlated to level of creativity showed no statistical significance after an ANOVA test was performed on both human AUT data (p-value: 0.594; Effect size: 0.2805) and ChatGPT AUT data (p-value: 0.845; Effect size: 0.661)



(a) AUT scores of human participants separated

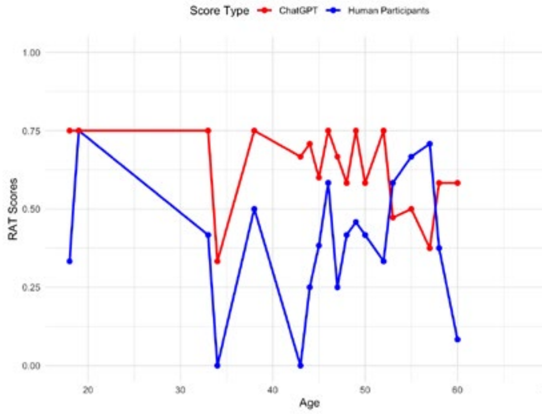


(b) AUT scores of ChatGPT separated by job

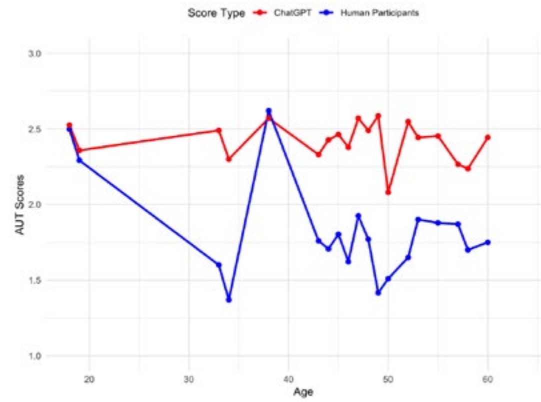
Figure 3 (a-b): AUT Scores of Participants Separated by job (Health Care, Engineer, Business Related, Other STEM Job, Arts)

5.1 Age and Creativity

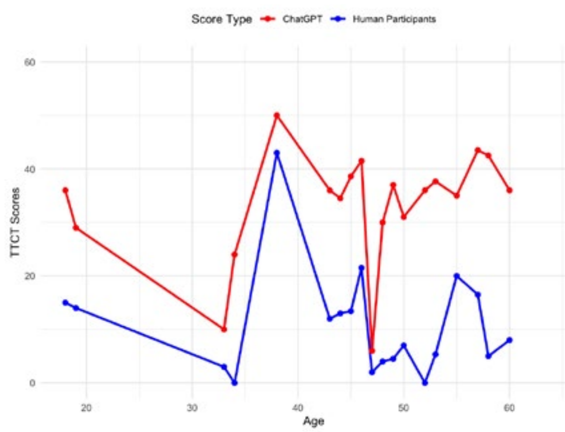
Additionally, age did not show a relationship with scores predicted by the LLM. High and low-score predictions can be observed throughout the data, regardless of age. We also note that the LLM impersonation was able to, in many cases, predict which human scores would be lower, as seen with multiple low scores in Figure 4.



(a) Average RAT scores of participants and LLM impersonation compared to age



(b) Average AUT scores of participants and LLM impersonation compared to age

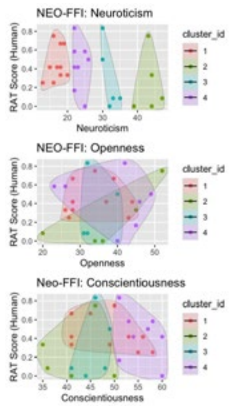


(c) Average TTCT scores of participants and LLM impersonation compared to age.

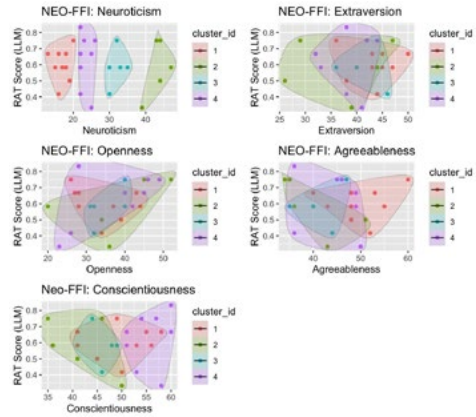
Figure 4 (a-c): Average scores of creativity survey based on age. Human participants are represented by blue-colored plots, and LLM impersonations are represented by red-colored plots.

5.2 NEO-FFI and Creativity

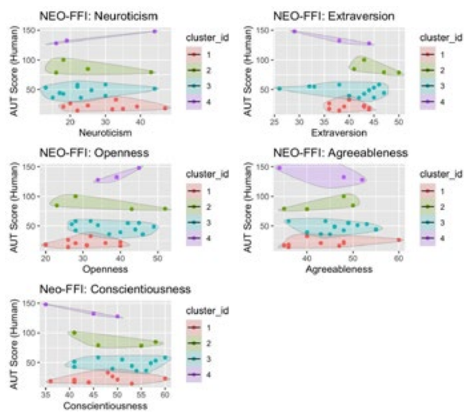
With the inclusion of NEO-FFI, there was a possible relation between personality traits and scores on creativity tests. As seen in Fig. 5, there was a clear and consistent connection between neuroticism and all creativity test scores. This finding supports past research on neuroticism and creativity [15]. However, despite possible relationships, a Pearson correlation coefficient revealed no significant correlation between neuroticism scores and NEO-FFI scores (Human: -0.045; ChatGPT: 0.027). Additionally, more clustering was seen in AUT and TTCT, suggesting that personality traits have the most significant correlation on assessments having to do with divergent thinking. On the other hand, the minimal clustering seen in RAT indicates that there is a minimal amount of correlation between personality traits and convergent thinking. The results of human participants were similarly mirrored in the LLM impersonations, with both general clustering and shapes of clusters being relatively similar.



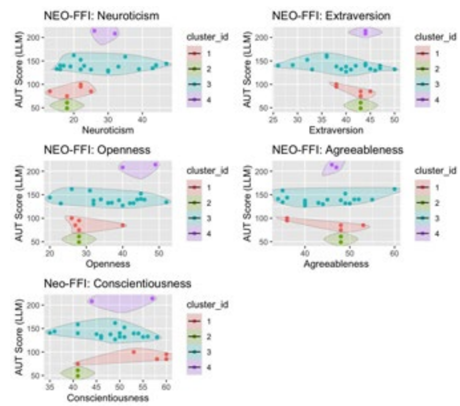
(a) Clustering of RAT scores from human participants with various NEO-FFI traits.



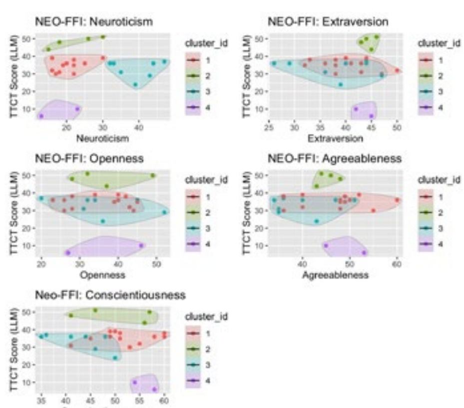
(d) Clustering of RAT scores from LLM impersonation according to various NEO-FFI traits.



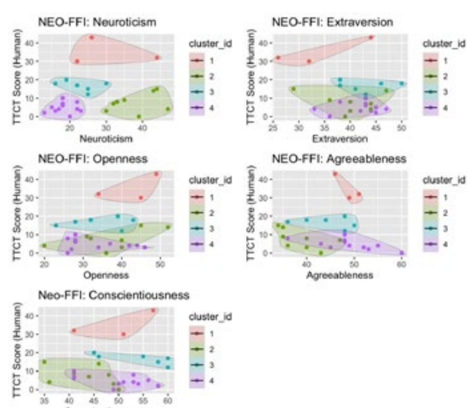
(b) Clustering of AUT Scores of human participants with various NEO- FFI traits.



(e) Clustering of AUT scores from LLM impersonation of human participants with



(c) Clustering of human TTCT scores according to each NEO-FFI trait.



(f) Clustering of TTCT scores from LLM impersonations with various NEO-FFI traits.

Figure 5 (a-f): Clustering of creativity test scores according to each personality trait. **Left Column**: Human participants. **Right Column**: LLM impersonation.

6. Discussion

In this study, we investigated the creative abilities of various large language models, utilizing various assessments to draw comparisons between different models themselves as well as the creative capacity of humans. We determined that while ChatGPT and Gemini offer distinct strengths, both consistently outperform humans in each of the creative tasks. Revisiting the initial proposed societal bias against AI's creative abilities, we offer evidence that LLMs contain the potential to act as creative sources for human use. Creativity has long been an important aspect of society and its progressions. Some research has explored the statistical creativity of AI to determine the theoretical aspect of AI creativity assuming that the model can fit the existing data created by humans [16]. Other studies have explored the divergent thinking between AI and humans by using AUT as that is one of the tests that measures divergent thinking [17]. Our results expand on this as we explore other creativity tests (RAT & TTCT) to determine whether Generative AI is more creative than humans. Additionally, the addition of NEO-FFI has led to the finding of connections between certain personality traits and aspects of creativity.

7. Conclusion

In this study of creativity among LLMs and humans, we demonstrated the capabilities of bleeding-edge AI in both convergent and divergent thinking. Our data indicates that LLMs like ChatGPT and Gemini proved to have a high degree of robustness across a variety of creative domains, widely outperforming human participants in each provided assessment; additionally, each LLM offers unique proficiencies in regard to creativity. Harnessing the potential of ever-developing AI, we can bolster our own creativity by utilizing these LLMs as tools rather than outright replacements for our creative endeavors.

Limitations

A limitation of this study was the lack of a diverse sample population and size. Most participants were from a similar area and lacked racial diversity. In future research, we would like to expand our study to consider various other populations and increase the sample size to determine if the results are consistent.

References

1. Bhandari, K., & Colton, S. (2024, March). Motifs, Phrases, and Beyond: The Modelling of Structure in Symbolic Music Generation. I*n International Conference on Computational Intelligence in Music, Sound, Art and Design* (Part of EvoStar) (pp. 33-51). Cham: Springer Nature Switzerland.
2. Zhou, E., & Lee, D. (2024). Generative artificial intelligence, human creativity, and art. *PNAS nexus, 3*(3), pgae052.
3. Magni, F., Park, J., & Chao, M. M. (2024). Humans as creativity gatekeepers: Are we biased against AI creativity?. *Journal of Business and Psychology, 39*(3), 643-656.
4. Salewski, L., Alaniz, S., Rio-Torto, I., Schulz, E., & Akata, Z. (2023). In-context impersonation reveals Large Language Models' strengths and biases. *Advances in neural information processing systems, 36*, 72044-72057.
5. Restrepo, K., Arias-Castro, C. C., & López-Fernández, V. (2019). A theoretical review of creativity based on age. Psychologist Papers, 40(2), 125-132.
6. Furnham, A., & Bachtiar, V. (2008). Personality and intelligence as predictors of creativity. *Personality and individual differences, 45*(7), 613-617.
7. Guzik, E. E., Byrge, C., & Gilde, C. (2023). The originality of machines: AI takes the Torrance Test. *Journal of Creativity, 33*(3), 100065.
8. Haase, J., & Hanel, P. H. (2023). Artificial muses: Generative artificial intelligence chatbots have risen to human-level creativity. *Journal of Creativity, 33*(3), 100066.
9. McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality, 60*(2), 175-215.
10. Jaschek, C., von Thienen, J., Borchart, K. P., & Meinel, C. (2023). The CollaboUse Test for automated creativity measurement in individuals and teams: a construct validation study. *Creativity Research Journal, 35*(4), 677-697.
11. Alabbasi, A. M. A., Paek, S. H., Kim, D., & Cramond, B. (2022). What do educators need to know about the Torrance Tests of Creative Thinking: A comprehensive review. *Frontiers in psychology, 13,* 1000385.
12. Malaie, S., Spivey, M. J., & Marghetis, T. (2024). Divergent and convergent creativity are different kinds of foraging. *Psychological Science*, 09567976241245695.
13. Wu, C. L., Huang, S. Y., Chen, P. Z., & Chen, H. C. (2020). A systematic review of creativity-related studies applying the remote associates test from 2000 to 2019. *Frontiers in psychology, 11*, 573432.
14. Organisciak, P., Acar, S., Dumas, D., & Berthiaume, K. (2023). Beyond semantic distance: Automated scoring of divergent thinking greatly improves with large language models. *Thinking Skills and Creativity, 49*, 101356.
15. Gao, Y., Zhang, D., Ma, H., & Du, X. (2020). Exploring creative entrepreneurs' IEO: Extraversion, neuroticism and creativity. *Frontiers in Psychology, 11*, 2170.
16. Wang, H., Zou, J., Mozer, M., Goyal, A., Lamb, A., Zhang, L., ... & Kawaguchi, K. (2024). Can AI be as creative as humans?. *arXiv preprint arXiv:2401.01623.*
17. Koivisto, M., & Grassini, S. (2023). Best humans still outperform artificial intelligence in a creative divergent thinking task. *Scientific reports, 13*(1), 13601.
18. Bellaiche, L., Shahi, R., Turpin, M. H., Ragnhildstveit, A., Sprockett, S., Barr, N., ... & Seli, P. (2023). Humans versus AI: whether and why we prefer human-created compared to AI-created artwork. *Cognitive Research: Principles and Implications, 8*(1), 42.
19. Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems, 3*, 121-154.
20. Kaufman, J. C., & Beghetto, R. A. (2009). Beyond big and little: The four c model of creativity. *Review of general psychology, 13*(1), 1-12.
21. A Construct Validation Study. Creativity Research Journal, 35(4), 677–697.

22. Ross, S. D., Lachmann, T., Jaarsveld, S., Riedel-Heller, S. G., & Rodriguez, F. S. (2023). Creativity across the lifespan: changes with age and with dementia. *BMC geriatrics, 23*(1), 160.
23. McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality, 60*(2), 175-215.
24. Abu Raya, M., Ogunyemi, A. O., Rojas Carstensen, V., Broder, J., Illanes-Manrique, M., & Rankin, K. P. (2023). The reciprocal relationship between openness and creativity: from neurobiology to multicultural environments. *Frontiers in Neurology, 14*, 1235348.