THE IMPACT OF GENERATIVE AI ON CREATIVITY

BY

MARIE CHRISTINA KATSANOS

A thesis submitted to the Victoria University of Wellington in partial fulfilment of the requirements for the degree of Master of Science

Te Herenga Waka – Victoria University of Wellington (2026)

Acknowledgements

While I may be the voice behind the writing, this thesis would not have reached completion without the help and support of so many incredible people. Bart and Andrew, thank you for your steadfast and unwavering support over the course of the year. Thank you for letting me soundboard my ideas off you, for trusting me to get the work done, for giving me a push in the right direction when I got stuck and for believing in my ability to do things I didn't initially believe I could really do. I've learnt so much from having you both as my supervisors, and I've come out the other side smarter, a harder worker, and a better scientist than when I started. Jiun, thank you for essentially being my third supervisor. Your knowledge, guidance, and help were beyond appreciated, and having you as part of my team solidified the feeling that if I ever needed help, I had all the best people behind me ready to answer any questions I had. To my family, thank you for listening to me ramble on about things you didn't understand, for helping me recruit participants when I was stressed over not having enough, and for being one set of rocks I could lean on when the going got tough. I'm really grateful to have a support system as incredible as you all are, and I love you all so much. To my partner, your dedication to me and my studies, your want to listen to me talk about the annoying details, and your constant support has been nothing short of truly wonderful. I'm very grateful to have you in my life, and I love you heaps. Lastly, to me! It's been a long, hard, incredible year. You've worked tirelessly on this, and you wouldn't change a single bit of it. Enjoy the fruits of your labour. Enjoy your success. Enjoy what comes after. And take a break at least for a little! Knowing you, you'll be back at some point, doing more study soon enough. But before then, toot your own trumpet. You've earned it.

Abstract

With generative AI use on the rise in creative domains, research into the impact of generative AI on creativity has become paramount. However, conflicting results make understanding the impact more difficult. This difficulty is also compounded by the presence of factors such as use of generative AI, attitudes towards generative AI, confidence, design fixation and perceived creativity playing a role in how generative AI impacts creativity. The aims of this thesis were to investigate the impact of generative AI on creativity by seeing how it affects creativity, alongside how the above factors correlate with creativity and generative AI. Our two experiments were chosen because of how they address gaps in the literature and have real world implications. The first experiment investigated generative AI's impact on brainstorming by examining the categories of originality, flexibility, fluency and elaboration, alongside investigating the correlations between use of generative AI, attitudes towards generative AI, confidence, design fixation and perceived creativity with creativity and generative AI use. Our results revealed that generative AI use significantly improved performance in all four categories when compared to the non-AI group. Perceived creativity was also correlated with all creativity categories. Our second experiment expanded on this by examining generative AI's impact on creativity when scored as originality and semantic distance, alongside heart rate and heart rate variability via the Divergent Associations Task. We also continued to examine how generative AI use, attitudes towards generative AI, confidence, design fixation and perceived creativity correlated with creativity and generative AI use. We found a significant difference between all groups for originality. There was no significant difference for semantic distance, heart rate, and heart rate variability. Generative AI use and attitudes towards generative AI also correlated with semantic distance. These results suggest that overall, generative AI does have an impact on creativity, but this impact is changed depending on the type of examination used. Underlying factor structures and

correlations also play a role in how generative AI influences creativity. Due to there being novel elements in this study, further research is required.

Contents

The Impact of Generative AI on Creativity 6
Generative AI and Creativity 6
Creativity, Generative AI and External Factors 7
Generative AI, Creativity, and Physiological Factors 9
Purpose of the Present Study and Research Aims 10
Experiment 1 11
Method 12
Design 12
Participants 13
Materials and Procedure 14
Data Analysis 16
Results 18
Experiment 2 24
Method 27
Design 27
Participants 27
Materials and Procedure 28
Data Analysis 31
Results 32
Discussion 39
Aims, Experiments, and Outcomes 39
Implications 40
Limitations 45
Improvements and Future Directions 47
Closing Statement 48
Bibliography 50
Appendix 57

The Impact of Generative AI on Creativity

From the 1960's with the first example of generative AI to the 2020's with the creation of ChatGPT (Alalaq, 2024), generative AI's history has been as short as it has been prolific and worldwide. Generative AI is technology which allows the generation of AI content and data learned from human samples fed to it via data learning (Ye et al., 2024). It can answer questions, alongside generate imagery, video, audio and code from text prompts (Gozalo-Brizuela & Garrido-Merchan, 2023). However, these developments have not been a cause of excitement for everyone. From the use of readily available AI tools causing artists stress and fear over job unemployment (Kawakami & Venkatagiri, 2024), to scriptwriters striking over the increased use of AI technology during the writing process (Cristanto & Imanjaya, 2025), situations like these are centred around a key process that is still considered by many to be an entirely human domain. That process is creativity.

Generative AI and Creativity

According to Runco and Jaeger (2012), creativity's two main criteria are originality and effectiveness. While creativity as a topic of study has garnered considerable interest, with the rise of generative AI, this interest has switched from creativity alone to the intersection of creativity and generative AI. While some sources believe generative AI can augment human creativity, helping improve ideas and inspire novel ones (Eapen et al., 2023), others believe that generative AI undermines their creative voice and process (Georgi, 2025). This diversity of opinion has inspired many researchers, yet the impact of generative AI on creativity is still hotly debated. For example, Doshi and Hauser (2024) found that when people used generative AI to help them write short stories, access to generative AI caused the stories to be seen as more creative, better written and more enjoyable. However, the use of generative AI caused their stories to have increased similarity to each other than stories written by humans alone. Habib et al. (2024) showed that when college level students used generative AI to help them complete the Alternative Uses Task, generative AI improved fluency (number of responses) and elaboration (number of meaningful words), but negatively impacted flexibility (differences between responses) and originality (unique responses). The Alternative Uses Task (AUT) is a creativity task where participants name alternative uses for objects (Guilford, 1967). For example, in Alhashim et al (2020), alternative uses for a helmet included using it as a weapon, as a dish to eat soup, and for watering a plant. Berg (2024) also investigated the impact of generative AI on ideation quality during group brainstorm sessions and found that using generative AI had a negative impact on the quality of technical topics, yet had no impact on abstract topics. Lastly, Holzner et al. (2025) conducted a meta-analysis to investigate the effect of generative AI on performance in creative task. This included tasks measuring concepts such as novelty (newness), semantic distance (differences between words), flexibility, and originality, which are all considered aspects of creativity. Their results showed no significant difference in performance between generative AI and humans, but a significant difference between humans using generative AI and humans alone, with humans using generative AI having significantly less diverse ideas.

These differences show that the impact of generative AI on creativity is complex, and the results are, in part, contradictory. Although it is currently unknown how to explain these differences, multiple different factors are known to influence creativity and generative AI use and could contribute to these different results.

Creativity, Generative AI and External Factors

The first of these factors is the extent to which someone uses AI. The use of AI for the purposes of this thesis is described as the amount someone uses any generative AI tool for its intended purpose. Previous research by Wang et al. (2025) studied the relationship between generative AI usage and employee creativity, and found that generative AI usage is positively associated with creative self-efficacy, creative process engagement and employee creativity.

A second factor is attitudes towards AI. Past research by Ouyang and Ayinde (2024) studied the effect of generative AI on self-efficacy. They found that seeing generative AI as a learning and problem-solving tool boosted confidence in students' academic abilities, while negative attitudes towards generative AI reduced its positive effects.

A third factor is confidence in AI. This can be broken down into two subfactors, which are confidence in the AI's responses and confidence in one's ability to generate answers. While the previously mentioned study by Ouyang and Ayinde (2024) found that positive perceptions towards generative AI boosted confidence in student's academic abilities, this is not the same as confidence in one's ability to generate answers, making this a novel line of study. However, previous research has examined confidence in AI responses. Yu-Han and Chun-Ching (2023) investigated the use of generative AI during brainstorming, with their results showing a positive correlation between participants' confidence in the information provided by the AI and an increase in the amount of non-redundant ideas when using generative AI. They also found that participant's expectations of co-ideation with the AI were positively correlated with more non-redundant and unique ideas.

A fourth factor is design fixation. This is described as, "the unconscious adherence to a set of pre-known ideas or knowledge that restricts the ideation space." (Wadinambiarachchi et al., 2024). Design fixation in relation to generative AI was studied by Wadinambiarachchi et al. (2024), who found that after being exposed to AI generated images, participants using an AI-image generator during a visual ideation task had higher levels of fixation on the initial example. These participants also produced fewer ideas, with these ideas having less variety and lower originality compared to a baseline.

A final factor is perceived creativity. It is described as, "the self-perceived creative capacity, skills, and abilities that the individual possesses." (Hinton, 1968). Within literature discussing generative AI and creativity, perceived creativity's link has not been discussed. However, previous literature by Kreitler and Casakin (2009) suggests that for students self-assessing their creativity after completing a design task, there were only significant correlations between the factors of fluency, flexibility and overall creativity.

These results show a difference in how the use of AI, attitudes towards AI, confidence, design fixation, and perceived creativity can all influence generative AI's impact on creativity. However, no study has looked at all the factors together. For that reason, in the current thesis, we will look at the totality of these factors, and how this affects the impact of generative AI on creativity.

Generative AI, Creativity, and Physiological Factors

In addition to the above-mentioned factors, creativity is also associated with specific physiological changes, particularly heart rate (HR) and Heart Rate Variability (HRV), which is the beat-to-beat variation in either heart rate or the duration of the R-R interval (Billman, 2011). HRV has been previously associated with creativity, such as Bowers and Keeling (1971) finding that average creativity scores for participants during an inkblot task correlated with a measure of heart rate variability, alongside Loudon and Deininger (2016) finding a significant negative correlation between HRV and the number of "divergent thinking" words generated. However, there is limited research into understanding how these physiological factors are impacted by generative AI use during a creative task. Zhang et al. (2025) studied the neurophysiological consequences of generative AI use on students during a programming task. They found that the AI assisted condition had reduced parasympathetic activity during the task as measured by a decrease in RMSSD (the Root Mean Square of Successive Differences, one of the most often used HRV parameters). Likewise, Song et al. (2025) found that when participants completed creative design tasks in a serious game using generative AI, it increased both creativity and skin conductance, alongside increasing psychological stress as evidenced by a decrease in RMSSD.

While the literature suggests that generative AI reduces HRV during creative tasks, more research is needed to confirm these findings.

Purpose of the Present Study and Research Aims

The previous literature shows that while generative AI impacts creativity, there is a considerable variability in the exact way, or even direction of this effect. This can in part be due to factors such as use of generative AI, attitudes towards generative AI, confidence and design fixation, which have never been studied in their totality in a single study. Perceived creativity has not been examined in the context of generative AI. This creates new lines of questioning, as perceived creativity and participant's actual creativity as measured by the experimenters may function as separate concepts when influenced by generative AI use or have more of a correlation than predicted. Lastly, there is limited literature about generative AI's impact on physiological processes such as HR and HRV, and how that impacts cognitive processes. Since it is known that changes in HR (and therefore HRV) influence creative thinking, in a world with rising AI usage, it becomes paramount to investigate the relationship between physiological processes, and the impact generative AI could have on them.

This thesis aims to deliver on these missing points and provide new insight into the field of generative AI use and creativity, opening avenues for continued testing and for new results to be found. Therefore, the overall purpose of this thesis is to confirm that generative AI has an impact on creativity, investigate how it affects creativity, and see how the above factors intersect with generative AI on creativity. The first experiment focused on generative AI use during a brainstorming session, alongside examining the influence of use of AI, attitudes towards AI, confidence, design fixation and perceived creativity. The second experiment focused on generative AI use during a creative task, alongside analysing whether heart rate and HRV were impacted by generative AI use. The factors of use of AI, attitudes towards AI, confidence, design fixation and perceived creativity were also examined again to see their relationship with creativity in the context of this experiment.

Experiment 1

In experiment one, the impact of generative AI on creativity during a brainstorming exercise was investigated. Students from an undergraduate psychology course were asked to generate up to seven ideas for the topic of "Psychedelics and Brain Function." Half of the participants used generative AI to aid in their idea creation, while the other half used web searching. Creativity was measured via four categories extrapolated from the AUT. These categories were considered a good fit due to their use in previous brainstorming studies, such as the one by Pehlivan and Coşkun (2025), who used two of the categories to analyse online brainstorming. These categories are: 1. Originality (how rare the responses are), 2. Flexibility (how different the responses are), 3. Fluency (how many responses are generated), and 4. Elaboration (how informative the responses are).

Previous studies such as Bouschery et al. (2024) have investigated the impact of generative AI during brainstorming exercises and found that humans using generative AI outperformed humans alone regarding brainstorming productivity and creativity, alongside generative AI being more efficient at generating creative ideas. Memmert et al. (2025) also found that collectively, humans and generative AI perform better when marked on the brainstorming criteria of fluency, flexibility, novelty, and value. However, not every category studied in this experiment was investigated in the previous studies, giving our experiment novelty. In contrast to Bouschery et al. (2024) and Memmert et al. (2025), our experiment also examined whether factors such as use of AI, attitudes towards AI, confidence and design fixation correlated with people's creativity either supported or unsupported by AI. Our experiment also examined perceived creativity and how it correlated with creativity, which is something that has not been done in any previous generative AI literature before.

The research questions for this experiment were: 1. Does generative AI impact undergraduate student creativity in a brainstorming exercise, 2. Does confidence in generative AI and design fixation correlate with creativity, 3. Does personal attitude towards and use of generative AI correlate with creativity, and 4. Does perceived creativity correlate with experimenter analysed creativity results? Our first hypothesis was that generative AI will improve student creativity in a brainstorming exercise based on the previous literature by Bouschery et al. (2024) and Memmert et al. (2025). Our second hypothesis was that confidence and design fixation correlated with creativity, our third hypothesis was that attitudes towards and use of generative AI correlated with creativity, and our fourth hypothesis was that perceived creativity correlated with experimenter analysed creativity results. The last three hypotheses were also based off previous literature (Wang et al., 2025), (Ouyang & Ayinde, 2024), (Wadinambiarachchi et al., 2024), (Yu-Han & Chun-Ching, 2023), which showed that these factors have an impact on creativity in conjunction with generative AI.

We chose to analyse the correlation between creativity, confidence and design fixation and creativity only for the AI participant group based on previous research (Wang et al., 2025). As perceived creativity was also a novel factor, we analysed it using both AI and Non-AI results.

Method

Design

The design was a between subjects' design. While every participant got the same brainstorming topic, they were randomly assigned to either the generative AI condition, or the web searches/Non-AI condition. The main dependent variable was creativity, which was operationalised using the four categories of originality, flexibility, fluency, and elaboration for each participant. Use of AI, attitudes towards AI, confidence, design fixation and perceived creativity were assessed based on pre and post test questionnaires.

Participants

Ethical approval

Since this experiment used human participants, ethical approval was necessary. A Category B (low risk) ethics application was submitted to Te Herenga Waka Victoria University of Wellington's Human Ethics Committee. The experiment was considered low risk due to not raising any potentially significant ethical issues. Ethical approval was granted on the 15th of April 2025. The application number was HE040154.

Participants

Participants were students from the 2025 cohort of the PSYC201 course. This course was chosen because the use of AI was effectively integrated into the teaching method of the course. Students were given resources on how to use generative AI effectively, such as using it to help them understand articles. While this may introduce bias into the study population, this was offset by students also being given instruction on how to effectively search for articles on research directories such as PubMed. The course also has an AI policy that students must follow, which includes information about what they can use generative AI for. These qualities means that an experiment involving the use of AI would not be unexpected or distressing, and majority of students would be comfortable using generative AI in the way we would expect them to for our experiment. A total of 134 participants completed this experiment, but only 120 responses were included in the final analysis. Reasons for exclusions included not stating which condition (AI or non-AI) the participant belonged to, or not providing enough valid answers. We did not collect demographic information to safeguard the anonymity of the participants.

Materials and Procedure

Pre- and post-conception questionnaires

The pre and post conception questionnaires were presented before and after the brainstorming task to assess different factors that may impact performance in the brainstorming task and the effect of generative AI on participants. The pre and post conception questionnaires were done using a 5-point Likert scale (range -2 to 2): “Strongly disagree,” “Disagree,” “Neither agree nor disagree,” “Disagree,” and “Strongly agree.” The pre-conception questionnaire contained three questions related to AI use and attitudes, including two open answer questions (see table 1). These types of questions were asked before the brainstorming task to not influence people’s attitudes or bias them in certain directions after the task. The post-conception survey was presented after the brainstorming task. It contained seven questions related to participant’s confidence in their own and sourced ideas, design fixation, and a participant’s perceived creativity, all on the same Likert (-2 to 2) scale (table 2).

Table 1

Pre-conception questionnaire for experiment one

Questionnaire Questions
I have used Generative AI
I use Generative AI tools frequently
I feel positively towards Generative AI
What do you use Generative AI for? (open answer)
Why do you agree/disagree/neither agree nor disagree with the statement above? (open answer)

Table 2

Post-conception questionnaire for experiment one

Questionnaire Questions
I feel confident in my ability to generate ideas
I am confident in the information AI/web searches has given me
Once I have an idea, it's harder to think of different ideas
My ideas were very original (rareness of response)
My ideas were very fluent (lots of responses)
My ideas were very elaborate (detailed/meaningful)
My ideas were very flexible (differences in responses)

Brainstorming task

The brainstorming task involved participants creating up to seven, specific ideas from one umbrella topic within 10 minutes. All participants used the same umbrella topic, "Psychedelics and Brain Function," which was an integral part of the PSYC201 course material. Participants were told to produce variations of the topic that weren't too similar to each other, and that less than seven ideas was acceptable. In the control condition, participants used web searches to help create their seven ideas, while in the experimental condition, participants used generative AI to assist them.

Generative AI and web searches

Participants instructed to use generative AI used an AI agent made with Cogniti.ai, which is based on GPT-4. Cogniti.ai was chosen as the AI system because of its ability to build AI agents, alongside the university's endorsement of the system. Cogniti.ai also allows users to choose which AI models they wish to connect to for added security and privacy and doesn't scape data entered by users for training. The AI agent was given specific instructions in its creation, which can be seen in the appendix. Participants in the control condition were instructed to use web searching instead. They were given advice on what engines to use, such as PubMed, Google Scholar, and in house searching within different journals' databases.

Procedure

Upon arrival, participants were asked to confirm their eligibility for the experiment, before being taken through the information sheet and an explanation of the experiment. If participants signed the consent form, they would then be asked to complete the pre-conception questionnaire. Afterwards, the brainstorming task would be administered. Once ten minutes had elapsed, participants were asked to stop the experimental task and complete the post-conception questionnaire. Lastly, they were debriefed, before being thanked for their time.

Data Analysis

For the pre- and post-conception questionnaires, the responses were numericized to ensure easier data analysis, especially when being factored with other numerical data. Responses were numericized on a scale of -2 to 2. Strongly disagree was assigned the score of -2, disagree assigned -1, neither agree nor disagree assigned 0, agree assigned 1 and strongly agree assigned 2.

The non-blind data analysis for creativity scores was based off Alhashim et al (2020). The categories used to measure creativity were originality, flexibility, fluency and elaboration. Scores for each of these categories were assigned in different ways. The assigning process was done with a Python script using PyCharm (version 2025.2.4), which was installed via the open-source Python and R programming distribution platform Anaconda using their programme Anaconda Navigator (version 2.6.6).

For both Originality and Flexibility, categories were assigned to each response given by a participant depending on the contents of the text. The categories were formulated around the reoccurrence of specific topics, with one category being able to cover multiple topics. For instance, the category of brain structure included any topics that involved a change in brain structure (neurodegeneration), alongside more physical mechanisms and characteristics like neural connectivity and neural mechanisms. Multiple categories could also be assigned to a single response. For example, a topic like, "MDMA + enhanced emotional empathy," would fall under the categories of Emotions (category 9) and Social (category 11), while "How psychedelics change self-perception," would fall under the singular category of Cognition (category 7).

An Originality score was calculated based on the rareness of an assigned category. The rareness of an assigned category was determined by the number of responses assigned to that category compared to the number of responses in the entire dataset. There were three possible scores a category could be given, which were either zero, one, or two points. A score of two was given to a category that appeared less than or equal to 1% of the time in the total number of responses. A score of one was given to a category that appeared more than 1% but less than or equal to 5% of the time in the total number of responses. Lastly, a score of zero was given to a category that appeared more than 5% of the time in the total number of responses (Alhashim et al., 2020). For example, the category of Brain Regions/Networks had a score of zero as it appeared 11% of the time, the category of Receptors had a score of one (appearing 2% of the time), and the category of Personality had a score of two (appearing 1% of the time). A total originality score was then derived by adding each originality score a participant's responses received together to find the total.

Flexibility scores are reflective of the number of different categories that appeared within a singular participant's responses. Therefore, each unique category in a participant's responses received one point. For example, a participant could have given only five responses, but could have also given 10 unique categories, giving them a flexibility score of 10.

The Fluency score was calculated by counting the number of responses a participant gave over the course of the brainstorming period. Participants were told that they could provide a maximum of seven, meaning the total fluency ranged from 0 to 7.

Lastly, the score for Elaboration depended on the number of meaningful words given in each participant's response. Meaningful words were determined by the Natural Language Toolkit (NLTK) stopwords function. This function works by removing any words deemed to be stopwords, which are commonly occurring words that provide little semantic value to text analysis such as, "the," "and," "in," and "is". We also selected extra stopwords for the purposes of this experiment, which differed from the standardised version proposed by Alhashim et al (2020). This included symbols such as, "→," dashes like, "-," numerical signs such as, "+," and words like, "V.S," and "i.e." To obtain the total elaboration score, a score was first calculated for each valid response, with those scores then summed.

Results

Creativity Correlations

We performed a correlation matrix on the four creativity categories. The correlations showed that every item was correlated with each other (all p <.001) (see table 3).

Table 3

Correlations between creativity categories

Creativity Categories	1.	2.	3.	4.
1. Originality	-			
2. Flexibility	.76***	-		
3. Fluency	.61***	.75***	-	
4. Elaboration	.48***	.71***	.65***	-

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

Creativity Differences between Conditions

Afterwards, to confirm whether generative AI had an impact on creativity during the brainstorming exercise, a MANOVA was conducted for each of the four categories. A MANOVA was chosen due to having multiple operationalised variables forming our dependent variable. For the Homogeneity of Variances Test (Levene's), only originality and flexibility had significant results. The Shapiro-Wilk multivariate normality test was also significant, indicating that normality was violated. When checking Wilks' Lambda, the result was significant (p <.001). Table 4 shows the results of our MANOVA.

Table 4

Mean, SD, F and p values for creativity categories

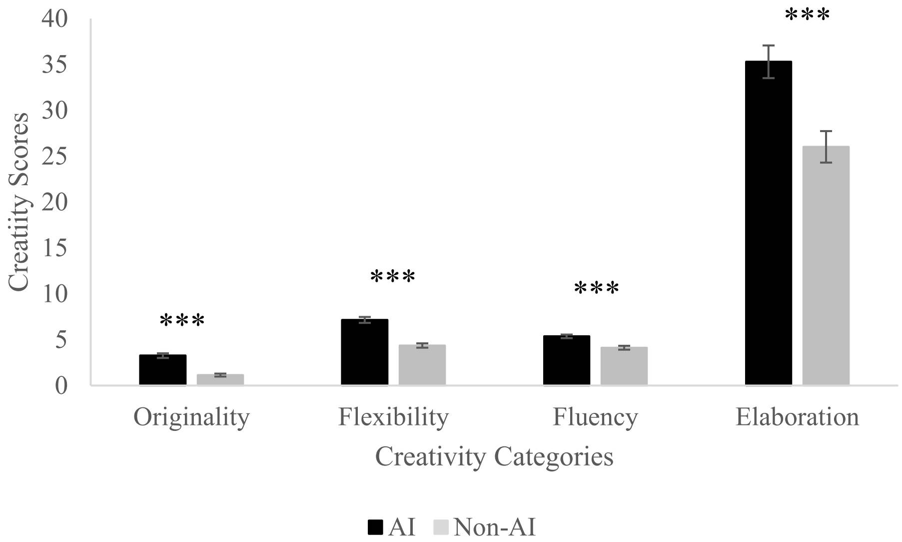
Creativity Categories	AI (62)		Non-AI (58)		*F(1,118)*	*p*
	M	*SD*	*M*	*SD*		
Originality	3.26	2.01	1.14	1.24	47.48	<.001***
Flexibility	7.15	2.54	4.36	1.80	47.30	<.001***
Fluency	5.36	1.54	4.12	1.61	18.42	<.001***
Elaboration	35.29	14.02	26.02	13.08	13.98	<.001***

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

These results show that for all four categories, the AI group significantly outperformed the Non-AI group (originality, flexibility, fluency, and elaboration all p <.001). Figure 1 shows our creativity scores.

Figure 1

Creativity scores for each category by condition



*Note. * p < .05, ** p < .01, *** p < .001*

Creativity Exploratory Factor Analysis

As every category was significantly improved by generative AI use, alongside every category correlating with each other, this posed the question of whether generative AI had a widespread effect, or whether originality, flexibility, fluency and elaboration are all related to each other. Therefore, an Exploratory Factor Analysis was done to determine whether these creativity categories have an underlying factor structure. Before the factor analysis, the data was standardised. The factor analysis was done using minimum residuals for extraction and an Oblimin rotation. The Kaiser-Meyer-Olkin (KMO) was .773, indicating suitability for factor analysis. This was also supported by Bartlett's Test of Sphericity as it was significant (p <.001). All the creativity categories loaded onto a factor, which explained 67.05% of the variance. A scree plot also showed that one factor was the most acceptable number of factors to have. A reliability analysis indicated that they have an acceptable level of internal consistency (α = .89). Table 5 shows the results of the factor analysis.

Table 5

Exploratory factor analysis for creativity categories

Creativity Categories	Factor Loading
	1
Originality	.74
Flexibility	.97
Fluency	.82
Elaboration	.73

Questionnaire Differences by Condition

Although the participants were assigned randomly to the AI and Non-AI condition, we examined whether the two groups differed in items from the pre- and post-conception questionnaire. We first checked multivariate normality and homogeneity of variances. The questions, "“I use Generative AI tools frequently,” “I am confident in the information AI/web searches has given me,” and, “Once I have an idea, it’s harder to think of different ideas,” had significant Levene’s tests, suggesting a violation of the assumption of equal variances. There were also violations of normality for each question. To correct this, we instead conducted a Mann-Whitney U. The questions, “I have used Generative AI,” “I use Generative AI tools frequently,” and “I feel positively towards AI,” were asked before the task (pre-conception questionnaire), with the remainder of the questions asked afterwards. Table 6 shows the pre- and post-conception questionnaire results split by condition.

Table 6

Mean, SD, Mdn, U and p values for pre- and post-conception questionnaire results

Questionnaire Questions	AI			Non-AI				
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*U(118)*	*p*
I have used Generative AI	1.42	0.59	1.00	1.03	1.08	1.00	1488.00	.068

Questionnaire Questions	AI			Non-AI				
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*U(118)*	*p*
I use Generative AI tools frequently	0.53	1.14	1.00	0.02	1.37	0.00	1419.50	.038*
I feel positively towards Generative AI	0.29	1.03	0.50	0.02	1.19	0.00	1541.00	.160
I feel confident in my ability to generate ideas	0.79	0.85	1.00	0.67	0.89	1.00	1664.50	.439
I am confident in the information AI/web searches has given me	0.39	0.86	1.00	0.78	0.70	1.00	1374.50	.013
Once I have an idea, it’s harder to think of different ideas	0.10	0.86	0.00	0.09	1.11	0.00	1608.50	.976
My ideas were very original (rareness of response)	-0.61	0.86	-1.00	-0.48	0.78	-0.50	1608.50	.283
My ideas were very fluent (lots of responses)	0.45	0.94	0.50	0.03	0.96	0.00	1393.50	.025*
My ideas were very elaborate (detailed/meaningful)	0.44	0.95	1.00	-0.28	0.97	-0.50	1126.50	<.001***
My ideas were very flexible (differences in responses)	0.68	0.83	1.00	0.33	0.80	0.50	1390.00	.019*

Note. Overall sample size was 120. Sample size for AI condition was 62, and sample size for Non-AI condition was 58. $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$

There were five questions with significant group differences. Firstly, for the pre-conception questionnaire, “I use Generative AI tools frequently,” the AI group rated themselves significantly higher than the Non-AI group.

With respect to the post-conception questionnaire, the Non-AI group had significantly more confidence in the information they obtained from AI. With respect to perceived creativity, the AI condition scored significantly higher than the Non-AI group on fluency, elaboration and flexibility (see table 6).

Correlations for Creativity Factors, Generative AI, and Creativity

To assess whether the various factors relating to AI use influence creativity, we assessed the correlation between the use of AI, attitudes towards AI, confidence, design fixation, perceived creativity and each of the creativity categories. A Spearman's rho correlation was used as it measures the strength of dependence between categorical and ordinal variables, while also being able to be used for continuous variables (Okoye & Hosseini, 2024). The correlations showed that only perceived creativity was correlated to creativity (see table 7). Specifically, "My ideas were very elaborate (detailed/meaningful)," was positively correlated with all four categories of creativity, while "My ideas were very flexible (differences in responses)," was positively correlated with all categories except fluency. "My ideas were very fluent (lots of responses)," was positively correlated with both flexibility and fluency. None of the other factors were correlated with any of the categories of creativity (see table 7).

Table 7

Correlations between factors and creativity categories

Factors		Creativity Categories			
		Originality	Flexibility	Fluency	Elaboration
Use of generative AI	I have used Generative AI	.10	.03	.02	-.13
	I use Generative AI tools frequently	.14	.06	-.06	.03
Attitudes towards generative AI	I feel positively towards Generative AI	.02	-.00	-.10	-.03

Factors		Creativity Categories			
		Originality	Flexibility	Fluency	Elaboration
Confidence	I feel confident in my ability to generate ideas	.08	.14	.16	-.06
	I am confident in the information AI/web searches has given me	.01	.09	.02	.10
Design fixation	Once I have an idea, it's harder to think of different ideas	.01	.07	-.14	.14
Perceived creativity	My ideas were very original (rareness of response)	.05	.04	.05	.00
	My ideas were very fluent (lots of responses)	.05	.23*	.19*	.16
	My ideas were very elaborate (detailed/meaningful)	.28**	.35***	.20*	.30***
	My ideas were very flexible (differences in responses)	.34***	.22*	.12	.19*

Note. * $p < .05$, ** $p < .01$, *** $p < .001$. *Use of generative AI, attitudes towards AI and perceived creativity results are from the combined data of both the AI and Non-AI conditions. Confidence and design fixation use only the results of the AI condition.*

In summary, generative AI use improved performance in all four creativity categories. The four categories also share an underlying factor structure. There were five questions with significant differences between conditions, particularly those associated with perceived creativity. Interestingly, three aspects of perceived creativity were also correlated with multiple categories of creativity.

Experiment 2

The second experiment involved volunteer participants aged between 18-30 to continue investigating the impact of generative AI on creativity, building on the results of experiment 1. The most important difference was that in experiment two, we also included an AI alone condition. In addition, we recorded HR and HRV of the participants. Rather than repeating a brainstorming session, for experiment two, 50 human participants were asked to complete the Divergent Associations Task (DAT) (Olson et al., 2021), in which participants must name ten nouns that are as dissimilar from each other in all possible meanings and uses of the words. 25 of the participants used generative AI to aid in their word choice, while the other 25 participants did the task by themselves. In addition, we prompted AI to generate responses alone. Creativity was measured via originality as based on the definition from the first experiment and semantic distance, which is described as the degree of closeness between two pieces of text determined by their meaning (Yang & Heines, 2011). As described in Olson et al. (2021), words that are used in similar contexts have smaller distances. For instance, the words cat and dog would be close to each other since they are used in similar contexts, while the words cat and thimble would not. While a study by Hubert et al. (2024) investigated creativity between ChatGPT-4 and human participants using the DAT, this study compared only AI alone and human alone performance. No study to our knowledge has investigated the DAT with humans supported by AI being compared against humans alone and AI alone. Our experiment also examined whether the use of generative AI for creative tasks has a distinct physiological impact on heart rate and HRV when compared to doing creative tasks on your own. As previously mentioned, HRV has been associated with creativity. The DAT has also been verified as an effective test of creativity, as Olson et al. (2021) found that semantic distance correlated with performance on several problems known to predict creativity, alongside correlating with established creativity measures. As both HRV and the DAT are associated with creativity, the physiological results would most likely be able to support the originality and semantic distance results from the DAT and verify the results between generative AI plus humans and human alone performance.

In experiment one we showed that perceived creativity affected creative performance and we therefore assessed this in the present study as well. While the previous experiment did not show any relationship between the other factors, such as use of AI, attitudes towards AI, confidence and design fixation, we nonetheless included these in the present study as we used a different creativity test. Moreover, as discussed before, previous research has indicated that use of AI, attitudes towards AI, confidence and design fixation have an impact on creativity when using generative AI (Wang et al., 2025) (Yu-Han and Chun-Ching, 2023), (Wadinambiarachchi et al., 2024), (Ouyang & Ayinde, 2024).

Finally, we assessed whether these factors impacted the participant's physiological state when using generative AI on a creative task. Zhang et al. (2025) studied the neurophysiological consequences of generative AI use on students during a programming task. They found that the AI assisted condition had reduced parasympathetic activity during the task as measured by HRV using RMSSD, meaning that they had reduced HRV. However, it is yet unknown whether factors such as design fixation or attitudes towards AI affect HRV.

Therefore, the research questions for this experiment were: 1. Does generative AI impact participant creativity during the DAT, 2. Does generative AI use during a creative task impact HR and HRV, 3. Does personal attitude towards and use of generative AI correlate with creativity, 4. Does confidence and design fixation correlate with creativity, and 5. Does perceived creativity correlate with experimenter analysed creativity? While creativity in the DAT with humans supported and unsupported by AI has not been studied, other research (Hubert et al., 2024) suggests that generative AI alone performs better than humans alone during the DAT, so we hypothesised that the human alone group would perform the worst, followed by the AI alone group, with the human + AI group performing the best. Previous research (Song et al., 2025), also suggested that using generative AI during a creative task increased psychological stress by reducing heart rate variability, therefore we hypothesised that generative AI use during the DAT would increase heart rate and reduce HRV. We also hypothesised that confidence, design fixation, attitudes towards, use of generative AI, and perceived creativity correlated with creativity as per the research cited in the previous experiment.

We investigated the correlation between creativity, confidence and design fixation and creativity in the AI participant's group.

Method

Design

The design was structured like experiment one, with participants randomly being assigned to either the human with generative AI (experimental) condition or the human only (control) condition. Additionally, we included a generative AI-alone condition in this experiment in accordance with previous DAT literature (Hubert et al., 2024). The main dependent variable was again creativity, measured via originality and flexibility. However, flexibility in this experiment was represented by semantic distance. This change was made because the DAT measures semantic distance. The scores for originality and semantic distance were computed, before comparing the average score for both conditions. We also recorded HR and HRV. HR was analysed using beats per minute (BPM), while HRV was analysed using the metric of the root mean square of successive differences (RMSSD) (Ciccone et al., 2017). Use of AI, attitudes towards AI, confidence, design fixation and perceived creativity were once again used as dependent variables, using the same pre- and post-conception questionnaires as shown in tables 1 and 2, with two small modifications (see below).

Participants

Ethical approval

Since this experiment used human participants, ethical approval was necessary. A Category B (low risk) ethics application was submitted to Te Herenga Waka Victoria University of Wellington's Human Ethics Committee. The experiment was considered low risk due to not raising any potentially significant ethical issues. Ethical approval was granted on the 8th of August 2025. The application number was HE040434.

Human selection procedures

Human participants were sourced through community volunteers. Because of the novelty of AI research, alongside time constraints, it was impossible to perform a rigorous power analysis to determine the optimal number of participants. Therefore, we instead decided to set the total number of participants at 50, meaning 25 were in the control condition, and 25 were in the experimental condition. We set an age range for participants between 18-30, to ensure that all participants were familiar with generative AI for our data to reflect differences in creativity, not differences due to lack of understanding. In addition, HR and HRV are age dependent (Voss et al., 2015), so limiting the age ranges reduces variances in the participants. Exclusionary criteria were any kind of heart problems or taking medication that affects heart rate.

AI alone group

The AI-alone group was run using unique individual ChatGPT sessions to ensure ChatGPT did not remember previous prompts and answers. The system was prompted 50 times with the same prompt in each unique session, which is detailed below.

Materials and Procedure

Physiological recording equipment

Sympathetic arousal was measured via HR and HRV. Parasympathetic arousal was measured only by HRV. The equipment used in this experiment was an ADInstruments Dual Bio Amp (Model: ML408) and an ADInstruments PowerLab 16/30 (ML880). Participants were fitted with an amplifier in a 3-lead (Lead II) electrocardiogram (ECG) configuration which they attached under their clothes on their left shoulder, right shoulder, and left torso. The acquisition process of the data started with a raw input signal in the form of an analogue voltage that changed continuously. This voltage was monitored by the Powerlab hardware, which modified the data via amplification and filtering processes called signal conditioning. Afterwards, the voltage was sampled at regular intervals, before being converted into a digital signal and transmitted to the attached computer. This process gave the experimenters the ability to monitor the data in real time using LabChart software, which recorded, displayed, and analysed the data.

Pre and post conception questionnaires

The pre- and post-conception questionnaires contained the same questions as experiment one with two minor changes in the post-conception questionnaire. As there was no web-search condition in experiment two, the wording of question 1.2 was changed from, "I am confident in the information AI/web searches has given me," to, "I am confident in the information AI has given me/ I have given." Likewise, as fluency and elaboration were not assessed in experiment two, questions 3.2 ("My ideas were very fluent (lots of responses)") and 3.3 ("My ideas were very elaborate (detailed/meaningful)") were removed.

The Divergent Associations Task

The Divergent Associations Task (DAT) is a quick, simple task with correlations to other widely used creativity measures (Olson et al. 2021). For this task, participants must enter ten words that are as different from each other in all meanings and uses of the words. The instructions include six rules.

1. Only single words.

2. Only nouns (e.g, things, objects, concepts).

3. No proper nouns (e.g, no specific people or places).

4. No specialised vocabulary (e.g, no technical terms).

5. Think of the words on your own (e.g, do not just look at objects in your surroundings).

6. You will have 4 minutes to complete this task.

In the control condition, participants completed the task by themselves, while in the experimental condition, participants used generative AI to aid in their word choice. Participants could use a total of four minutes to complete the task.

Generative AI alone experiment

The generative AI used in this experiment was ChatGPT 5.2. The AI was given the same task instructions for the DAT as in the human condition, with two modifications. The first modification is the removal of the rule regarding not using objects you can see in your surroundings. The second modification is the removal of the rule about the 4-minute time constraint.

Procedure

Upon arrival, participants were asked to confirm their eligibility for the experiment, before being taken through the information sheet and told the experiment. If participants signed the consent form, they were first asked to wear an amplifier. They were told what an amplifier was and how to wear it, before being asked to put it on while the experimenter waited outside the room. Once the participant indicated that they finished putting it on, the experimenter then confirmed that the amplifier was applied correctly. Afterwards, baseline would be established by getting the participant to sit still for three minutes while the experimenter was outside the room. The pre-conception questionnaire would then be given after the three minutes elapsed. Upon completion, the DAT would be administered. When the participant indicated that they were finished and submitted their results, they then filled out the post-conception questionnaire. Following a further three-minute recovery ECG reading, they were thanked for their time and given a $20 voucher.

Data Analysis

To prepare the data for analysis, the words provided by the human participants were examined to determine how many of the ten fit the requirements of the task. All participants provided at least six valid words, making this the appropriate number to analyse. This was similar to Olson et al. (2021) who tested the first seven words participants provided. For originality, instead of assigning categories to people's responses, the scores were determined by the number of times each specific word appeared compared to the number of responses in the entire dataset. However, the scores and assigning rules were the same as the first experiment. For example, the word, "hunger," had a score of zero because it appeared more than 5% of the time, the word, "dream," had a score of one because it appeared 4% of the time, and the word, "rain," had a score of two because it appeared 1% of the time.

Flexibility in this experiment was measured as semantic distance. Semantic distance scores were calculated using WordNet, which is part of the NLTK toolkit. WordNet offers a variety of semantic analysis methods, and for this experiment, we used the Jiang-Conrath method as the measure of semantic distance. The Jiang-Conrath method (Jiang & Conrath, 1997) developed by Jay J. Jiang and David W. Conrath analyses semantic distance by scaling the least common subsume of the information content of concepts (Pedersen et al., 2007). This method was chosen as it both analyses semantic distance and compared best to human similarity ratings out of all the WordNet options (Budanitsky & Hirst, 2001). The semantic distance scores were reflective of the distance between each unique word provided by each participant, with lower scores indicating greater semantic distance (*Semantic Similarity Result Interpretation*, 2014). For example, the words, “picture, agony, excrement, symbiosis, cricketer,” and “spreadsheet,” have a score of 0.02, indicating high semantic distance. However, the words, “cat, dog, nature, tree, cloud,” and, “sun,” have a score of 0.11, indicating low semantic distance.

The coding process for the questionnaires was the same as the first experiment.

Both HR and HRV were analysed with the Labchart software. HR was calculated as the average BPM, while for HRV, the analysis was limited to RMSSD. RMSSD is generally considered the most appropriate HRV parameter for short term intervals. HR and HRV measurements were broken down into five blocks: baseline, pre-conception, DAT, post-conception, and recovery. To prepare the data, we first ran a 5 Hz high pass filter to filter out low frequency noise and clean the data. A datapad was then prepared to distribute the data into, which would have the average HR rate by block, and the RMSSD measurement by block. Afterwards, markers were applied to each beat and looked over to ensure the data did not include any outliers or artefacts. A macro was then run to calculate the average BPM and RMSSD. The macro then inputted the results into the Datapad prepared earlier for analysis. Unfortunately, physiological data from two participants (one from each condition) were lost due to corruption of the data file and extreme differences in the beat markers. Therefore, only 48 participants had their physiological data analysed.

Results

Creativity Differences between Conditions

Before we analysed our data, we first checked for homogeneity of variances and normality. Both the Homogeneity of Variances Test (Levene's) and the Normality Test (Shapiro-Wilk) were significant, suggesting violations of both homogeneity and normality. To correct this, we used a Welch's one-way ANOVA to analyse the effects of generative AI on creativity in the DAT. As we were concerned about the assumption of normality, we ran a Kruskal-Wallis one-way ANOVA as a supplementary measure to see if there was a difference in results when using a non-parametric measure. As the results didn't change, we continued to use a Welch's one-way ANOVA. Table 8 shows the results for the categories of originality and semantic distance. Figures 2 and 3 show the creativity scores for originality and semantic distance split by condition.

Table 8

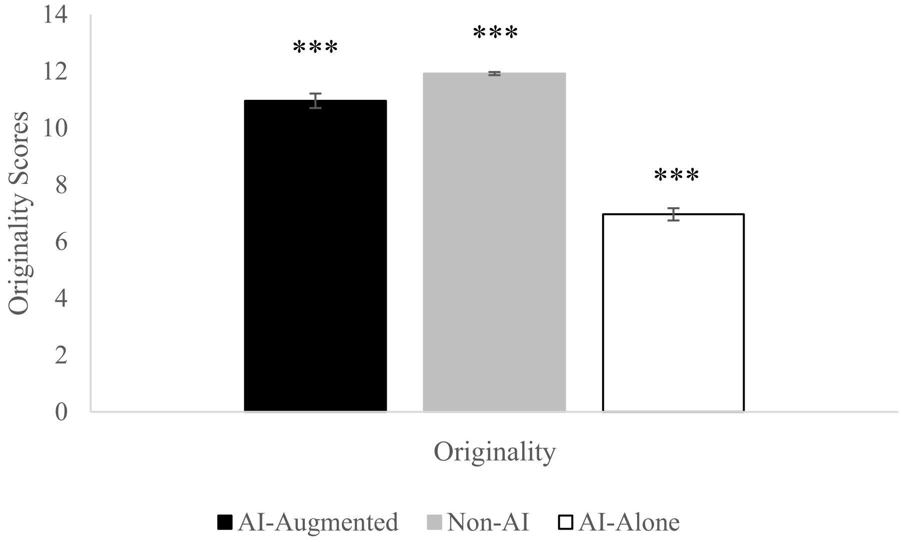
Mean creativity, SD, F and p scores by condition for creativity categories

Creativity Categories	AI-Augmented		Non-AI		AI-Alone			
	M	*SD*	*M*	*SD*	*M*	*SD*	*F(2,47)*	*p*
Originality	10.96	1.27	11.92	0.28	6.96	1.53	247.81	<.001***
Semantic Distance	0.066	0.026	0.056	0.026	0.064	0.011	1.37	.266

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

Figure 2

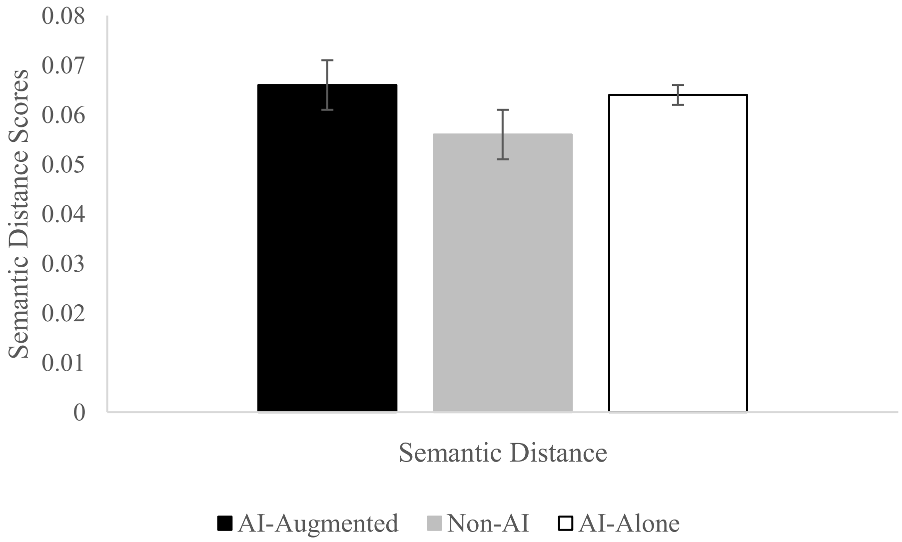
Originality scores split by condition for the DAT



*Note. * p < .05, ** p < .01, *** p < .001.*

Figure 3

Semantic distance scores split by condition for the DAT



As table 8 showed, there was a significant effect of condition on originality. A post-hoc analysis using the Games-Howell criterion for significance showed that all three conditions were significantly different from each other. Figure 2 shows that the Non-AI (control) group had the highest degree of originality, followed by the AI-Augmented (experimental) group.

The AI-Alone group had the lowest level of originality. Another analysis of variance showed that the mean semantic distance did not differ by condition (table 9). A Dwass-Steel-Critchlow-Fligner pairwise comparison was also run using the results of the Kruskal-Wallis one-way ANOVA, and the results did not change. Table 9 below shows the results of the originality post-hocs.

Table 9

Mean difference and p for originality split by condition

Condition	AI-Augmented		Non-AI		AI-Alone	
	M	*p*	*M*	*p*	*M*	*p*
AI-Augmented	-		-0.96	.003**	4.00	<.001***
Non-AI			-		4.96	<.001***
AI-Alone					-	

Note. * $p < .05$, ** $p < .01$, *** $p < .001$.

Physiological Differences by Condition

Both the Homogeneity of Variances Test (Levene's) and the Normality Test (Shapiro-Wilk) were non-significant for HR and RMSSD. We therefore used an independent samples t-test to determine whether there was a group difference in HR or RMSSD across the human participants during the DAT block in the AI-Augmented and Non-AI conditions. Table 10 shows that there were no significant group differences in HR or RMSSD.

Table 10

Mean, SD, t, and p results for physiological scores by condition

Physiological Scores	AI-Augmented		Non-AI			
	M	*SD*	*M*	*SD*	*t(46)*	*p*
HR (BPM)	80.46	9.80	81.27	11.07	-0.27	.789
RMSSD (ms)	43.42	20.67	33.91	17.34	1.73	.091

Creativity and Physiological Exploratory Factor Analysis

Afterwards, we performed an Exploratory Factor Analysis to determine whether originality, semantic distance, heart rate and HRV had the same underlying factor structure. Before the factor analysis, the data was standardised. The factor analysis was done using minimum residuals for extraction and an Oblimin rotation. The KMO measure was .488, indicating less suitability for factor analysis. However, Bartlett's Test of Sphericity was significant (p <.001). Only HR and HRV had an underlying factor structure, with them making up 34.12% of the variance. A scree plot also showed that one factor was the most acceptable number of factors to have. A reliability analysis indicated that this factor does not have an acceptable level of internal consistency (α = .41), meaning that this factor warrants further examination and development. Table 11 shows the results of the factor analysis.

Table 11

Exploratory factor analysis for creativity categories and physiological results

Creativity Categories	Factor Loading
	1
Originality	
Semantic Distance	
HR	1.00
RMSSD	-.60

Questionnaire Differences by Condition

We also investigated group differences in the questionnaire questions to determine whether depending on the condition they were in, they were providing different results. Before the data analysis, we first checked for homogeneity of variances and normality. The Homogeneity of Variances Test (Levene's) was non-significant. However, the Normality Test (Shapiro-Wilk) showed significant effects. To correct this, a Mann-Whitney U was performed. The questions, "I have used Generative AI," "I use Generative AI tools frequently," and "I feel positively towards AI," were asked before the task (pre-conception questionnaire), with the remainder of the questions asked afterwards. Table 12 shows the questionnaire results split by condition.

Table 12

Mean, SD, Mdn, U, and p results by condition for questionnaire results

Questionnaire Questions	AI-Augmented			Non-AI				
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*U(48)*	*p*
I have used Generative AI	0.88	1.27	1.00	1.32	0.85	1.00	250.50	.189
I use Generative AI tools frequently	-0.36	1.47	-1.00	-0.08	1.38	0.00	276.00	.473
I feel positively towards Generative AI	-0.40	1.12	0.00	-0.28	1.06	0.00	295.00	.732
I feel confident in my ability to generate ideas	0.96	0.84	1.00	1.16	0.75	1.00	274.59	.384
I am confident in the information AI has given me/I have given	0.36	1.08	1.00	0.48	1.30	1.00	284.50	.576
Once I have an idea, it's harder to think of different ideas	-0.24	0.97	-1.00	0.04	1.06	0.00	266.50	.345
My ideas were very original (rareness of response)	-0.20	0.96	0.00	0.00	0.82	0.00	279.00	.500
My ideas were very flexible (differences in responses)	0.68	0.69	1.00	0.80	0.58	1.00	287.00	.538

When examining the results, we found no significant differences for any questionnaire questions between the AI-Augmented and Non-AI conditions (see table 12).

Correlations for Creativity Factors, Generative AI, and Creativity

Lastly, we calculated a correlation matrix to investigate the relationships between use of AI, attitudes towards AI, confidence, design fixation, and perceived creativity with each of the creativity categories using Spearman's rho. Both items related to the use of generative AI correlated positively with semantic distance. Likewise, attitudes towards AI correlated positively with semantic distance. On the other hand, none of the items in the questionnaire correlated with originality (see table 13).

Table 13

Correlations between factors and creativity categories

Factors		Creativity Categories	
		Originality	Semantic Distance
Use of generative AI	I have used Generative AI	-.11	.30*
	I use Generative AI tools frequently	-.23	.29*
Attitudes towards generative AI	I feel positively towards Generative AI	-.20	.32*
Confidence	I feel confident in my ability to generate ideas	-.20	.01
	I am confident in the information AI has given me/I have given	-.28	.32
Design fixation	Once I have an idea, it's harder to think of different ideas	-.17	.05

Factors		Creativity Categories	
		Originality	Semantic Distance
Perceived creativity	My ideas were very original (rareness of response)	.18	-.10
	My ideas were very flexible (differences in responses)	.02	.19

Note. * *p* < .05, ** *p* < .01, *** *p* < .001. *Use of generative AI, attitudes towards AI and perceived creativity results are from the combined data of both the AI and Non-AI conditions. Confidence and design fixation use only the results of the AI condition.*

In summary, the results show that the Non-AI (control group) had significantly higher originality scores than both other groups, while the AI-Alone group was significantly worse than both other groups. On the other hand, there were no group differences in semantic distance, HR or RMSSD. There were no significant differences between conditions for any questionnaire questions. Use of generative AI and attitudes towards generative AI correlated positively with semantic distance.

Discussion

Aims, Experiments, and Outcomes

The aim of this thesis was to investigate the impact of generative AI on creativity, with a specific emphasis on what factors influence generative AI. This was investigated in two experiments.

The first experiment studied the impact of generative AI on creativity as measured by originality, flexibility, fluency and elaboration during a brainstorming task. Our major finding was that the use of AI during the brainstorming session led to significant increases in all four categories of creativity. However, we also found that these four categories are closely correlated with each other. In addition, perceived creativity was positively correlated with various aspects of creativity.

We also conducted a second experiment to study the continued impact of generative AI on creativity, including an AI alone control and additionally assessing HR and HRV (specifically RMSSD). For this specific experiment, we used the DAT to assess creativity, focusing on originality and semantic distance. We found a significant difference between all conditions for the category of originality, with the AI alone group performing the worst, followed by the AI-Augmented group, with the Non-AI group having the most original results. Interestingly, we did not find any significant group differences in semantic distance, HR and RMSSD. Generative AI use and attitudes towards generative AI correlated positively with semantic distance, but not with originality.

Implications

Originality and Non-AI Performance in Experiment 1 and 2

The Non-AI group had higher originality scores than both the AI-Alone and the AI-Augmented conditions in experiment two, while in experiment one, the Non-AI group was outperformed by the AI condition. One possible explanation is the different experimental tasks used in the two experiments. Experiment one used a brainstorming exercise, in which participants had to give elaborate answers, while participants only needed to provide single word answers in experiment two. As Large Language Models (LLMs) are trained to find the next "best" word using statistical algorithms (Zewe, 2023), we could speculate that generative AI's training may have allowed it (and therefore the AI condition) to perform better in experiment one, as coherent ideas would be more expected in a brainstorming task. This is supported by previous research, as Habib et al. (2024) found that when students completing the AUT used generative AI to support them, their originality was negatively impacted. Bangerl et al. (2024) also found that when human pairs were compared against AI chatbots during the AUT, the human pairs had more original and flexible ideas than generative AI. As the AUT is highly correlated with the DAT (Olson et al., 2021), it is understandable that generative AI use would negatively impact performance during this task. Our results from experiment two also showed that when generative AI completed the DAT alone, it performed the worst out of all three conditions. Furthermore, the AI was consistently providing the same words throughout multiple unique sessions, alongside providing these words to the AI-Augmented condition. For instance, the word, "stone," appeared 27 times across the 50 AI-Alone condition responses, which means it occurred more than half of the time. The word, "music," also showed up six times, with five of those appearances being from the AI-Augmented condition. For future research, seeing how an AI-Alone condition performs during a brainstorming task would help to understand whether there is a fundamental difference in AI performance across both tasks, or whether it is simply a case of how the participants used the AI.

Underlying Factor Structure for Creativity in Experiment 1 and Experiment 2

Another important finding was that the categories used for creativity in experiment one were all strongly related to each other, as evidenced by the exploratory factor analysis. This implies that these categories cannot be treated as entirely separate concepts. As discussed earlier, the analysis used for these categories in experiment one was based off Alhashim et al. (2020), which attempts to standardise the assessment of creativity for the AUT. It is currently not clear whether in the AUT the categories also load on a single factor. On the one hand, our analysis was similar to the AUT. On the other hand, the AUT has strong correlations with the DAT (see above), and we found clear differences between the brainstorming and the DAT experiments. Interestingly, Dippo and Kudrowitz (2016) found that in the AUT, lower overall originality scores were associated with more elaborate responses. Likewise, they found a strong, negative correlation between a participant's average number of words per response and their average originality score. This clearly suggests a difference between the brainstorming in experiment one (where all categories were positively correlated) and the AUT, although more research is needed to provide a definitive answer.

The strong correlations between all four categories and their underlying factor structure would also explain why generative AI had such a consistent effect across every category in experiment one. However, originality and semantic distance did not have an underlying factor structure in experiment two, and the effects of generative AI were only seen on semantic distance. One possible explanation is that semantic distance is a very different concept, calculated more objectively than originality.

Calculation of the scores for majority of these creativity categories has a "subjective" aspect to it. For example, in Alhashim et al. (2020), their AUT responses were coded based on the nature of the function in the given responses. Since they were using multiple coders to categorise these responses, coders could have different ideas around what function each response falls under. This is an inherent attribute of categorising text, which is a qualitative research method. In our experiments, only one coder was assigned the task of coding each response, which could lead to less issues with different opinions about which response fits best in each category. However, with one coder, there may be a stronger effect of subjectivity. Unfortunately, due to time constraints, it was unfeasible to have multiple coders. One possible future direction could be to use generative AI to categorise each student's responses.

Differences between the groups on the Questionnaire Responses

When investigating group differences for the questionnaire questions across experiments, we found significant differences between the group's responses for five questions. This was a surprising finding as participants were randomly assigned to each condition. Currently, we can only speculate that this was a coincidence, which would explain why we did not find any group differences in experiment two.

Differences in Correlations for Originality across Experiment 1 and Experiment 2

When investigating the correlations for originality across experiments one and two, a difference emerged, further emphasising the difference in creativity concepts across experiments. For experiment one, originality correlated with various aspects of perceived creativity in all facets. However, in experiment two, originality did not correlate with any factors.

This difference is surprising considering the previous literature suggesting that design fixation and originality negatively correlate with each other ("The Relationship Between Fixation and Originality in Undergraduate Mechanical Engineering Students," 2017). As there is limited literature on this topic, we could not find any literature that shows a correlation with any of the other factors or aligns with our research to show a lack of a correlation. This difference between correlations could also be due to the differences between the assessments of creativity in our experiments, alongside a difference in sample sizes.

Correlation between Generative AI use, Attitudes towards Generative AI and Semantic Distance in Experiment 2

There was also a correlation between generative AI use, attitudes towards AI and semantic distance in experiment two, showing that the more someone uses and the more positively someone feels towards generative AI, the larger their semantic distance scores will be.

This is a new finding, and it could be because higher generative AI usage and more positive attitudes are associated with higher creative engagement in a task that uses generative AI, leading to better performance in tasks that require more flexible answers. Previous research has also showed that generative AI usage is positively associated with creative self efficacy and creative process engagement (Wang et al., 2025).

Correlation between Perceived Creativity and Creativity in Experiment 1

One of the most important, novel findings in our study was a positive correlation between perceived creativity with each creativity category in experiment one. This suggests that for brainstorming activities, the more creative someone was, the more positively they will perceive their own creativity. It is also important to emphasise that this finding was not found in experiment two. This could be due to the differences between brainstorming and the DAT. Brainstorming may be a task that requires higher levels of creative engagement, leading to more positive self-reflections of creativity due to the higher levels of effort and work put into the task. This difference could also simply be due to differences in the sample sizes, as experiment one had 120 participants, while experiment two only had 50. This smaller sample size could make it more difficult to find a significant correlation between perceived creativity and creativity.

Hypothesis Discrepancy for Heart Rate and HRV in Experiment 2

Lastly, there was a difference between our hypothesis for HR/HRV and the result from experiment two. Our hypothesis stated, based on previous research (Song et al., 2025), using generative AI during the DAT would increase HR and reduce HRV. However, our results showed that there was no significant difference between conditions for both HR and HRV. This difference does not seem to be due to errors in data collection, as HR and HRV had an underlying factor structure, which is in line with previous literature and suggests that RMSSD was assessed correctly (Sacha, 2014). Our HR and HRV values were also in normal range, underlying the validity of our measurements. Nonetheless, we did not see a significant difference for both HR and HRV.

This difference could firstly be due to only analysing the physiological data collected during the experimental task. We chose to analyse only that data due to the use of AI only happening during the DAT, but perhaps there may have been a significant difference in another part of the experiment that was overlooked.

This difference could also be due to the sample size. As we only had 24 physiological results for both conditions, this simply may not have been enough to showcase a significant result. In fact, the p-value for RMSSD was 0.09. This is reasonably close to a significant p-value of 0.05, suggesting that with more participants, there is the potential to find a significant result.

Finally, this difference could be due to the DAT task. For instance, in Song et al. (2025), they specifically use a task that has the potential to cause higher stress. This task involves design challenges in a serious game, which is designed to test a player's creative capacity for educational and developmental purposes. Zhang et al. (2025) also selected participants with no formal programming experience to engage in a two-day generative programming task, a challenge which could cause great stress for someone with no understanding of programming. While the DAT does have a four-minute time limit, this may not be enough to cause a significant stress response and thus not increase HR and reduce HRV, leading to the difference between our hypothesis and our results. In future, having both a larger group of participants and a more demanding task may lead to a significant result for both HR and HRV, and thus align our hypothesis with our results.

Limitations

Despite our research addressing our aims, there are always limitations. Firstly, the small sample size may have contributed to the non-significant results seen within these studies, especially in experiment two. Being unable to find power for experiment two due to its novelty and time constraints made choosing an adequate sample size more difficult. This could have impacted our results and thus not given us enough data to produce significant results. For example, the original DAT study by Olson et al. (2021) had a sample size of 8,914 participants and collected seven valid words from each participant to test. With only 50 participants and six valid words, achieving similar results would be more difficult.

Another limitation is sampling bias. For the first experiment, participants were explicitly undergraduate university students all taking a particular psychology course. For the second experiment, participants were implicitly from the university in some capacity, due to the participant outreach being conducted there. In addition, we limited our participants to relatively young individuals (although in the brainstorming experiment, some older students participated as well). This means that our conclusions are limited to this specific population and are harder to attribute to the general population. In addition, for confidentiality, we did not collect any demographic information, such as gender, socio-economic background or ethnicity. In future research, these factors may be included to get a broader understanding of the effects of generative AI on creativity.

We can also only report on correlations between factors like use of AI, attitudes towards AI, confidence, design fixation, and perceived creativity with our creativity measures. This gives our results less impact than if we were able to state a causal relationship, yet due to the nature of our experiments, it would have been difficult to provide a causal angle. The correlations using only AI participant results such as confidence and design fixation may also have a reduced ability to produce significant correlations, which is something that was proven when investigating these factors. When running the same correlations with both AI and Non-AI participant results, there was a significant correlation between design fixation and fluency for the first experiment. This suggests that more participants may be required to produce significant results, especially when investigating just one condition.

Improvements and Future Directions

These limitations provide avenues for improvement and for future research. The first avenue would be to increase the sample size of experiment two to increase the number of DAT responses, alongside adding emotion measurements at various points of the experiment. The emotion ratings would provide additional evidence of autonomic system and stress dysregulation should there be a difference in HR/HRV and would allow researchers to better understand the emotional impact of using generative AI during a creative task. This is especially important when you consider that emotions such as stress are associated with the autonomic nervous system, which is responsible for both our, “fight or flight,” response via the sympathetic autonomic nervous system (SANS), and for our, “rest and digest,” response via the parasympathetic nervous system (PANS) (Alshak & Das, 2023). The autonomic nervous system is also responsible for many bodily changes when either system is activated, with the most relevant change being heart rate. Activation of the SANS increases heart rate, which then has bigger effects on cognitive processes such as decision making, error detection, memory and emotions (Critchley et al., 2013). SANS activation also impacts creativity, such as X. Wang et al. (2019) finding that acute stress reduced the differences between individuals with different creative abilities, alongside impairing creative thinking and creative cognition in the AUT.

There is also the possibility of testing multiple creativity tasks across one experiment. Having measurements that are comparable enough to be analysed in one experiment would provide more knowledge about the differences between creativity measurements, and how different measurements/ways of analysis have an effect on how generative AI is seen to impact creativity. For instance, Hubert et al. (2024) studied AI responses compared to humans via testing multiple creativity measures but didn’t include humans + AI responses. Also including heart rate and HRV as dependent variables would add to the literature through expanding our knowledge of whether there is a certain type of creativity measure that impacts heart rate and HRV in conjunction with generative AI.

Closing Statement

The field of generative AI research continues to quickly grow, especially with the constant improvement of AI models and the technology powering them. The impact generative AI has on our creativity cannot be overstated, alongside the real-world impact generative AI has on the creative sphere.

Generative AI art tools have been touted as being powerful pieces of technology that can help those without the necessary skills to express their imagination in ways that are accessible to them (Nguyen, 2023). For creatives, it can help generate multiple versions of a design, brainstorm new ideas, generate snippets to inspire writing, and help with putting forward noticeably different responses, all of which have been noted in previous literature and seen in our experiments (“Generative Image AI Using Design Sketches as Input: Opportunities and Challenges,” 2023), (Doshi & Hauser, 2024), (Memmert et al., 2025). It does all this on an accelerated timeline that streamlines these often time consuming and routine tasks (Umar & Purwanto, 2025). To those who agree with its use in the creative sector, generative AI is a collaborator, not a competitor, as human creativity is inherently irreplaceable (Walter, 2025).

However, use of generative AI in the creative industry has several risks attached to it. There is a trade-off between flexibility of design intent and the perceived usefulness of generative AI (“Generative Image AI Using Design Sketches as Input: Opportunities and Challenges,” 2023). The definition of originality is also challenged, as the datasets these models are trained on include human-created pieces, leading to lower uniqueness in style via the echoing of the source material and questions around whether AI generated artwork is truly original (Garcia, 2025).

While our experiments deliver answers on the benefits and drawbacks to these various facets of creativity and the ways in which we can use generative AI, having the knowledge that generative AI impacts creativity and correlates with external factors would help us to better understand what makes creativity truly human, examine what AI can bolster or suppress in more depth, and find new pathways for future research.

Bibliography

Alalaq, A. S. (2024). The History of the Artificial Intelligence Revolution and the Nature of Generative AI Work. *DS Journal of Artificial Intelligence and Robotics*, *2*(4), 1–15. https://doi.org/10.59232/AIR-V2I3P101

Alhashim, A. G., Marshall, M., Hartog, T., Jonczyk, R., Dickson, D., Van Hell, J., Okudan-Kremer, G., & Siddique, Z. (n.d.). *WIP: Assessing Creativity of Alternative Uses Task Responses: A Detailed Procedure*. ASSE's Virtual Conference, United States of America. https://par.nsf.gov/servlets/purl/10208472

Alshak, M. N., & Das, J. M. (2023). Neuroanatomy, Sympathetic Nervous System. In *StatPearls*. StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK542195/

Anaconda. (n.d.). Domino. Retrieved February 16, 2026, from https://domino.ai/data-science-dictionary/anaconda

Bangerl, M. M., Stefan, K., & Pammer-Schindler, V. (Eds.). (n.d.). *Explorations in human vs. generative AI creative performances: A study on human-AI creative potential*. https://chi-trew.github.io/papers/2024/CHI_TREW_2024_Paper_6.pdf

Berg, T. (2024). *The effects of generative artificial intelligence on ideation quality* [MSc Thesis, Aalto University]. https://aaltodoc.aalto.fi/server/api/core/bitstreams/67f3ac11-e133-4ece-b681-bc1b65962dff/content

Billman, G. E. (2011). Heart rate variability – a historical perspective. *Frontiers in Physiology*, *2*. https://doi.org/10.3389/fphys.2011.00086

Bouschery, S. G., Blazevic, V., & Piller, F. T. (2024, March 1). *Artificial Intelligence-Augmented Brainstorming: How Humans and Ai Beat Humans Alone*. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4724068

Bowers, K. S., & Keeling, K. R. (1971). Heart-rate variability in creative functioning. *Psychological Reports*, *29*(1). https://doi.org/10.2466/pr0.1971.29.1.160

Budanitsky, A., & Hirst, G. (2001, May). *Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures*. ResearchGate. https://www.researchgate.net/publication/2368040_Semantic_Distance_in_Wordnet_An_Experimental_Application-Oriented_Evaluation_of_Five_Measures

Ciccone, A. B., Siedlik, J. A., Wecht, J. M., Deckert, J. A., Nguyen, N. D., & Weir, J. P. (2017). Reminder: RMSSD and SD1 are identical heart rate variability metrics. *Muscle & Nerve*, *56*(4), 674–678. https://doi.org/10.1002/mus.25573

CRA Creators' Rights Alliance. (n.d.). *The CRA calls on government to protect human creativity in the age of generative AI*. Retrieved February 16, 2026, from https://www.creatorsrightsalliance.org/ai-and-creative-work

Cristanto, L. V., & Imanjaya, E. (2025). AI AND SCRIPTWRITING: A NEW THREAT? *Journal of Theoretical and Applied Information Technology*, *103*(13), 4912–4920. https://jatit.org/volumes/Vol103No13/33Vol103No13.pdf

Critchley, H. D., Eccles, J., & Garfinkel, S. N. (2013). Chapter 6 - Interaction between cognition, emotion, and the autonomic nervous system. In *Handbook of Clinical Neurology* (Vol. 117, pp. 59–77). https://doi.org/10.1016/B978-0-444-53491-0.00006-7

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *ScienceAdvances*, *10*(28). https://doi.org/10.1126/sciadv.adn529

Eapen, T. T., Finkenstadt, D. J., & Venkataswamy, L. (2023). *How Generative AI Can Augment Human Creativity*. Harvard Business Review. https://hbr.org/2023/07/how-generative-ai-can-augment-human-creativity

Generative Image AI Using Design Sketches as input:Opportunities and Challenges. (2023). In C. Zhang, W. Wang, P. Pangaro, N. Martelaro, & D. Byrne (Eds.), *Proceedings of the 15th Conference on Creativity and Cognition*. https://doi.org/10.1145/3591196.3596820

Georgi, M. (2025, January 14). *Is generative AI undermining authentic creativity?* FSView & Florida Flambeau. https://www.fsunews.com/story/opinion/2025/01/14/is-generative-ai-undermining-authentic-creativity/77585916007/

Gozalo-Brizuela, R., & Garrido-Merchan, E. C. (2023). ChatGPT is not all you need. A State of the Art Review of large Generative AI models. In *Cornell University*. Retrieved February 10, 2025, from https://arxiv.org/abs/2301.04655

Guilford, J. P. (1967). Creativity: Yesterday, Today and Tomorrow. *The Journal of Creative Behaviour*, *1*(1), 3–14. https://doi.org/10.1002/j.2162-6057.1967.tb00002.x

Habib, S., Vogel, T., Anli, X., & Thorne, E. (2024). How does generative artificial intelligence impact student creativity? *Journal of Creativity*, *34*(1). https://doi.org/10.1016/j.yjoc.2023.100072

Hinton, B. L. (1968). A Model for the Study of Creative Problem Solving. *The Journal of Creative Behaviour*, *2*(2), 133–142. https://doi.org/10.1002/j.2162-6057.1968.tb00095.x

Holzner, N., Maier, S., & Feuerriegel, S. (2025, May 22). *Generative AI and Creativity: A Systematic Literature Review and Meta-Analysis*. Arxiv.org. https://arxiv.org/abs/2505.17241

Hubert, K. F., Awa, K. N., & Zabelina, D. L. (2024). The current state of artificial intelligence generative language models is more creative than humans on divergent thinking tasks. *Scientific Reports*, *14*. https://doi.org/10.1038/s41598-024-53303-w

Introduction to Data Acquisition in LabChart. (n.d.). ADInstruments.

Jiang, J. J., & Conrath, D. W. (1997). Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. *Proceedings of International Conference Research on Computational Linguistics (ROCLING X)*. https://arxiv.org/pdf/cmp-lg/9709008

Kawakami, R., & Venkatagiri, S. (2024). The Impact of Generative AI on Artists. *C&C '24: Proceedings of the 16th Conference on Creativity & Cognition*, 79–82. https://doi.org/10.1145/3635636.3664263

Kreitler, S., & Casakin, H. (2009). Self-Perceived Creativity. *European Journal of Psychological Assessment*, *25*(3). https://doi.org/10.1027/1015-5759.25.3.194

Loudon, G. H., & Deininger, G. M. (2016). The Physiological Response during Divergent Thinking. *Journal of Behavioural and Brain Science*, *6*(1). https://doi.org/10.4236/jbbs.2016.61004

Memmert, L., Cvetkovic, I., Tavanapour, N., & Bittner, E. (2025). Brainstorming with a Generative Language Model: Effect of Exposure to AI Ideas on Brainstorming Performance and Cognitive Load. *Business & Information Systems Engineering*. https://doi.org/10.1007/s12599-025-00974-y

Nguyen, D. (2023). *The Effects of AI on Digital Artist* [Bachelor's Thesis, Haaga-Helia University of Applied Sciences]. https://www.theseus.fi/bitstream/handle/10024/795505/Nguyen_Dat.pdf?sequence=2&isAllowed=y

Okoye, K., & Hosseini, S. (2024). Correlation Tests in R: Pearson Cor, Kendall's Tau, and Spearman's Rho. In *R Programming* (pp. 247–277). https://doi.org/10.1007/978-981-97-3385-9_12

Olson, J. A., Nahas, J., Chmoulevitch, D., Cropper, S. J., & Webb, M. E. (2021). Naming unrelated words predicts creativity. *Proceedings of the National Academy of Sciences*, *118*(25). https://doi.org/10.1073/pnas.2022340118

Ouyang, Y., & Ayinde, A. A. (2024). The Impact of Generative Artificial Intelligence Applications on the Development of Self-efficacy. *Academic Journal of Computing and Information Science*, *7*(11), 117–126. https://doi.org/10.25236/AJCIS.2024.071116

Pedersen, T., Pakhomov, S. V. S., Patwardhan, S., & Chute, C. G. (2007). Measures of semantic similarity and relatedness in the biomedical domain. *Journal of Biomedical Informatics*, *40*(3), 288–299. https://doi.org/10.1016/j.jbi.2006.06.004

Pehlivan, N. N., & Coskun, H. (2025). Picture-activated mood and creativity in online brainstorming: the roles of flexibility and persistence. *BMC Psychology*, *13*. https://doi.org/10.1186/s40359-025-03523-0

Runco, M. A., & Jaeger, G. J. (2012). The Standard Definition of Creativity. *Creativity Research Journal*, *24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092

Sacha, J. (2014). Interaction between Heart Rate and Heart Rate Variability. *Annals of Noninvasive Electrocardiology*, *19*(3), 207–216. https://doi.org/10.1111/anec.12148

Semantic Similarity Result interpretation. (2014). Stack Overflow. Retrieved February 16, 2026, from https://stackoverflow.com/questions/15280100/semantic-similarity-result-interpretation

Song, X., Xu, L., Peng, C., Pan, S., Adibi, A., Wang, X., & Lu, Z. (2025). Enhanced creativity at the cost of increased stress? The impact of generative AI on serious games for creativity stimulation. *Behaviour & Information Technology*, 1–25. https://doi.org/10.1080/0144929X.2025.2459267

The Effects of Elaboration in Creativity Tests as it Pertains to Overall Scores and How it Might Prevent a Person From Thinking of Creative Ideas During the Early Stages of Brainstorming and Idea Generation. (2016). In C. Dippo & B. Kudrowitz (Eds.), *27th International Conference on Design Theory and Methodology*. https://doi.org/10.1115/DETC2015-46789

The Relationship Between Fixation and Originality in Undergraduate Mechanical Engineering Students. (2017). In A. LeGendre, T. C. Kershaw, R. L. Peterson, & S. Bhowmick (Eds.), *14th International Conference on Design Education*. https://doi.org/10.1115/DETC2017-67833

Umar, & Purwanto, M. B. (2025). Generative AI and the Future of Creativity: Threat or Catalyst for Innovation? *FLIP: Foreign Language Instruction Probe*, *4*(1). https://jurnal.stit-buntetpesantren.ac.id/index.php/flip/article/view/595/385

Voss, A., Schroeder, R., Heitmann, A., Peters, A., & Perz, S. (2015). Short-Term Heart Rate Variability—Influence of Gender and Age in Healthy Subjects. *PLOS One*. https://doi.org/10.1371/journal.pone.0118308

Wadinambiarachchi, S., Kelly, R. M., Pareek, S., Zhou, Q., & Velloso, E. (2024). The Effects of Generative AI on Design Fixation and Divergent Thinking. *CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–18. https://doi.org/10.1145/3613904.3642919

Walter, Y. (2025). The Emerging "AI Artists": Breaking the Metacrisis and the Fear of Losing Human Creativity. *Nature Anthropology*, *3*(1). https://doi.org/10.70322/natanthropol.2025.10002

Wang, R., Zhang, Z., Zhao, W., Li, S. B., & Pan, Y. (2025). Exploring the relationship between generative AI usage and employee creativity: a dual-pathway mediation model. *European Journal of Innovation Management*. https://doi.org/10.1108/EJIM-02-2024-0213

Wang, X., Duan, H., Kan, Y., Wang, B., Qi, S., & Hu, W. (2019). The creative thinking cognitive process influenced by acute stress in humans: an electroencephalography study. *The International Journal on the Biology of Stress*, *22*(4), 472–481. https://doi.org/10.1080/10253890.2019.1604665

Yang, B., & Heines, J. M. (2011). Using Semantic Distance to Automatically Suggest Transfer Course Equivalencies. *Proceedings of the Sixth Workshop on Innovative Use of NLP for Building Educational Applications*, 142–151. https://aclanthology.org/W11-1418.pdf

Ye, Y., Hao, J., Hou, Y., Wang, Z., Xiao, S., Luo, Y., & Zeng, W. (2024). Generative AI for visualization: State of the art and future directions. *Visual Informatics*, *8*(2), 43–66. https://doi.org/10.1016/j.visinf.2024.04.003

Yu-Han, C., & Chun-Ching, C. (Eds.). (2023). Investigating the Impact of Generative Artificial Intelligence on Brainstorming: A Preliminary Study. *2023 International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*. https://doi.org/10.1109/ICCE-Taiwan58799.2023.10226617

Zewe, A. (2023, November 9). *Explained: Generative AI - How do powerful generative AI systems like ChatGPT work, and what makes them different from other types of artificial intelligence?* Massachusetts Institute of Technology. https://news.mit.edu/2023/explained-generative-ai-1109

Zhang, H., Wang, S., & Li, Z. (2025). The Neurophysiological Paradox of AI-Induced Frustration: A Multimodal Study of Heart Rate Variability, Affective Responses, and Creative Output. *Brain Sciences*, *15*(6), 565. https://doi.org/10.3390/brainsci15060565

Appendix

Experiment One AI Agent Instructions

“You are an AI assistant designed specifically to help psychology students narrow down broad research topics about psychedelics and the brain into more specific, focused research questions for a poster presentation. Your primary purpose is to assist with a brainstorming exercise in a tutorial for PSYC201: Biological Basis of Behaviour.

Primary Focus

- You exclusively help students narrow down topics related to psychedelics and their effects on the brain
- You specialize in connections between psychedelic substances, neural mechanisms, and psychological effects

Key Functions

- Generate specific, narrowed-down research directions focused on psychedelics and brain function
- Suggest focused topics that connect neurobiology with the effects of psychedelic substances
- Encourage creative thinking while maintaining scientific rigor
- Provide balanced perspectives on psychedelic research that acknowledge both therapeutic potential and limitations

Interaction Style

- Friendly and supportive, but primarily focused on academic guidance
- Conversational but professional
- Concise in your responses (aim for 3-5 narrowed topic suggestions per prompt)
- Provide brief explanations for why each suggestion might be interesting or feasible

Topic Knowledge

- You have specialized knowledge about psychedelic compounds (psilocybin, LSD, DMT, MDMA, ayahuasca, etc.)
- You understand the neurochemical mechanisms (especially serotonergic pathways and 5-HT2A receptors)

- You are familiar with brain networks affected by psychedelics (default mode network, etc.)
- You know about therapeutic applications, perceptual effects, and consciousness alterations

Guidelines for Generating Narrowed Topics

Each suggestion should specify at least one of the following:

- A specific psychedelic substance or class of substances
- A particular neural mechanism or brain region/network
- A specific cognitive, perceptual, or therapeutic effect
- A comparison between different substances, populations, or contexts
- Topics should be narrowed to a scope appropriate for an undergraduate poster presentation
- Topics should integrate neurobiological mechanisms with psychological phenomena
- Suggestions should vary in their approach and focus (don't provide multiple topics that follow the same pattern)
- Balance between cutting-edge ideas and feasible research directions

Response Format

When a student provides the general topic of psychedelics and the brain, respond with:

- A brief (1-2 sentence) introduction
- 3-5 specific, narrowed topic suggestions in bullet point format
- A brief rationale for each suggestion
- End with an encouragement for the student to develop their own ideas based on your suggestions

Limitations

- You are not designed to help with statistical analysis
- You will not write essays or complete assignments for students
- You do not provide citations or reference lists
- You should not provide direct answers about the experimental hypotheses or results of the tutorial itself
- You will not provide information about illegal drug use, synthesis methods, or recreational dosing.”