Computers in Human Behavior: Artificial Humans 4 (2025) 100140

Contents lists available at ScienceDirect

Computers in Human Behavior: Artificial Humans

journal homepage: www.journals.elsevier.com/computers-in-human-behavior-artificial-humans

# If ChatGPT can do it, where is my creativity? generative AI boosts performance but diminishes experience in creative writing

Peidong Mei [a,b,c,*], Deborah N. Brewis [d], Fortune Nwaiwu [e], Deshan Sumanathilaka [f], Fernando Alva-Manchego [g], Joanna Demaree-Cotton [b]

[a] School of Psychology, University of Aberdeen, King's College, AB24 3FX1, Aberdeen, UK
[b] Uehiro Oxford Institute, University of Oxford, Saint Ebbe's St, Oxford, OX11PT, UK
[c] Department of Engineering, University of Exeter, The Queen's Dr, Exeter, EX4 4QH, UK
[d] School of Management, University of Bath, Claverton Down, Bath, BA2 7AY, UK
[e] Department of Mechanical Engineering, University of Bath, Claverton Down, Bath, BA2 7AY, UK
[f] School of Mathematics and Computer Science, Swansea University, Swansea, SA1 8EN, UK
[g] School for Computer Science and Informatics, Cardiff University, Abacws, Senghennydd Road, Cardiff, CF24 4AG, UK

ARTICLE INFO

*Keywords:*
Human-AI interaction
Creativity
Cognition
Perception
GenAI
ChatGPT
Education

ABSTRACT

As generative AI (GenAI) becomes more sophisticated, it is increasingly being used as a tool to enhance creative expression and innovation. Along with its potential benefits, it is imperative that we examine pitfalls in how generative AI may affect the quality of creative thinking and possibly lead to a narrowing of diversity both in representation and thought. In this study, we employed an experimental design with 225 university students who completed a creative writing task with pre- and post-task surveys to assess ChatGPT's impact on their performance and experiences compared to a control group who did not use ChatGPT. Results show that using ChatGPT enhanced creativity of output and reduced the difficulty and effort required for the task, particularly for non-native English speakers. However, it also diminished the value and enjoyment of the task and raised moral concerns. We contribute to the nascent literature on GenAI by showing how ChatGPT assistance could potentially bolster human creativity by facilitating content delivery or providing useful counterpoint ideas. We also significantly advance scholarship on understanding experience of GenAI, demonstrating that bypassing the cognitive effort required for creativity by using ChatGPT could be harmful to the creative process and experience of creative tasks, especially when steps are not taken to address the use of AI in a transparent manner. Finally, our novel mixed-method study design offers a contribution to the methodological frameworks for the study of the effects and experience of GenAI. We discuss the study results in relation to implications for educational practices and social policy and argue that our results support recommending an integration of generative AI into higher education alongside practices that help to mitigate the negative impacts of AI use on student experience.

## 1. Introduction

This study seeks to increase our understanding of how the use of ChatGPT affects the cognitive processes involved in creative thinking. Since its release on November 30, 2022, ChatGPT, a generative artificial intelligence (GenAI), has become a major platform, acquiring users at an unprecedented rate. It reached 1 million users within the first five days (Marr, 2023) and grew to 100 million users in the following months (TIME, 2024). As of October 2024, it boasts over 200 million weekly active users (DemandSage, 2024). This GenAI tool leverages a large pre-trained language model and can draw on it to generate novel text within seconds. According to the Higher Education Policy Institute (HEPI), more than half of students have used generative AI to assist with assessments (HEPI, 2024). The rapid rise of this platform has led to a surge in concern over its impact on learning outcomes, and around critical and creative thinking.

News coverage of generative AI and its effects in education reflected an initial 'chaos' experienced by universities, and assessment systems were quickly reviewed for suitability considering the availability of GenAI (The Russell Group, 2023). Sullivan et al. (2023) analysed 100 media articles from Australia, New Zealand, the United States, and the United Kingdom, and found that 88 % of them raised concerns about

* Corresponding author. School of Psychology, University of Aberdeen, King's College, AB24 3FX1, Aberdeen, UK.
*E-mail address:* peidong.mei@abdn.ac.uk (P. Mei).

https://doi.org/10.1016/j.chbah.2025.100140
Received 30 December 2024; Received in revised form 12 March 2025; Accepted 19 March 2025
Available online 19 March 2025

academic integrity. While important, debates about assessment and academic integrity have arguably overshadowed other important questions to do with the impact that the use of GenAI has on the learning process itself. As Lee et al. (2025) rightly note, GenAI represents just one more tool in a long history of technologies (including *writing*, *printing*, *calculators*, and the *Internet*) that humans have developed to assist them in thinking tasks. The way in which higher education is to adapt to the availability of GenAI must be informed by an evidence-based understanding of how it affects learning. Central to this question is how exactly using ChatGPT, and similar generative AI tools, impacts how students engage with and produce academic work. According to Marshall McLuhan's media theory—specifically his concept that "*the medium is the message*"— it is essential to consider not only the information that AI-generated tools produce but also the medium through which this information is consumed and processed. McLuhan's theory suggests that the specific operation of and engagement with GenAI tools, such as ChatGPT, influence cognition, shaping how individuals process information, engage with knowledge, and develop thinking skills (McLuhan, 1994, 2017).

On the one hand, there are several reasons for thinking that the use of GenAI may enhance productivity and creative outputs. Indeed, as GenAI becomes more sophisticated, it is increasingly framed as a productivity-enhancing tool, with growing acceptance among academic stakeholders (Kasneci et al., 2023). Research has demonstrated that ChatGPT can produce results that are in line with human efforts in certain academic fields. For example, Terwiesch (2023) reported that ChatGPT could achieve a B+ in an MBA course, suggesting that AI can perform well in structured tasks where information is synthesised based on clear rules. However, the extent to which GenAI aids productivity may depend on the type of academic task in question. Early findings indicate that AI tools like ChatGPT perform particularly well in technical disciplines such as coding and mathematics (Frieder et al., 2023) but struggle with more abstract or interpretive tasks in fields like the humanities and social sciences (Terwiesch, 2023). These variations suggest that AI's impact on outputs will differ across disciplines, depending on the complexity and nature of the tasks involved. In addition, GenAI may help to promote inclusion in learning through expanding readily available access to knowledge (Elbanna & Armstrong, 2024; Rasul et al., 2023), regardless of geographical, socio-economic, or educational barriers. In a study conducted with students from countries such as Indonesia, Malaysia, and Nigeria, Veza et al., 2023 found that 75.6 % of respondents believed ChatGPT would enhance accessibility and inclusivity in higher education. ChatGPT, for instance, has made access to knowledge easier through its capacity to synthesise vast amounts of information on a wide range of topics. This capability may particularly promote inclusion for domain-specific tasks that would otherwise require specialised expertise or resources (Spero, 2024; Watson & Romic, 2025). However, the extent to which GenAI facilitates creativity and in-depth learning while using it to complete educational tasks is less clear. It seems that the use of GenAI tools leads to 'cognitive offloading' (Stadler, Bannert, & Sailer, 2024), but the concern is that this may diminish the user's skills in relation to the tasks that have been outsourced to GenAI. For instance, if the use of AI does lead to some degree of cognitive disengagement from the learning process, this may pose a risk to the development of higher-order thinking skills. If so, this could potentially lead to 'de-skilling' in the population at large.

In this study, we sought to explore what is happening to the cognitive processes involved in learning as students use ChatGPT to assist them. By exploring both the performance and experience changes that occurred while engaging with ChatGPT in a creative writing task, we contribute significantly to the existing literature in two ways. First, we extend our understanding of the impacts of GenAI in the context of higher education, in which creativity has a fundamental value for learning. Second, we also contribute a novel methodology by collecting data in a more naturalistic setting, where participants not only engaged with ChatGPT to complete a task, but also reported their experience before and after that task. The mixed-method combination of behavioural and survey data collected throughout the process of engaging with ChatGPT provides richer and more ecologically valid insights of ChatGPT's effects, compared to self-reported survey data that have been used in other studies on this topic. Additionally, we utilised ChatGPT to automate the assessment of task performance. The robustness of this approach as reposted in the validation (section 2.3) later, highlights a new possibility for further methodological innovation.

### 1.1. Literature review: how is the use of generative AI reshaping learning?

There is a vibrant interest in research around ChatGPT, including some studies thus far on its effects in Higher Education. In this section we synthesise the literature by identifying the gaps to highlight areas our study contributes.

#### 1.1.1. Generative AI and creative thinking

Emerging research on how the use of GenAI tools is reshaping cognitive processes has primarily focused on 'critical thinking'. Critical thinking is a complex set of cognitive processes involving recall of knowledge, comprehension, application, analysis, evaluation and creation (Bloom et al., 1964; see also Hyder & Bhamani, 2016; Dabić, 2016). Much of the extant research has, however, looked at the problem through interrogating adjacent processes, such as investigating whether framing GenAI outputs as questions (Danry et al., 2023) or forcing a 'pause' (Buçinca et al., 2021) can improve elements of critical thinking, whether LLMs can be used to promote reflection (Du et al., 2024), or whether LLM's can be used as part of pedagogical tools (Lee et al., 2023; Tanprasert et al., 2024; Yuan et al., 2023).

A recent study, conducted in collaboration with Microsoft, investigated the effects of GenAI on critical thinking among knowledge workers—those whose work consists of trading knowledge itself rather than a product (see Blackler, 1995). The study specifically explored self-evaluation of critical thinking while using GenAI. The researchers found that the user's own confidence in the capacities of GenAI predicted whether critical thinking was perceived to have been enacted on the part of the user – specifically, higher confidence in GenAI was associated with less perceived critical thinking, and lower confidence with more (Lee et al., 2025). The authors concluded that when GenAI was used the nature of critical thinking was reshaped and shifted toward information verification, response integration and task stewardship. The study design took the entire process of critical thinking as its target of study. We build on the insights of this study by taking an alternative approach and focus specifically on the 'top' level of Bloom's taxonomy: creation.

Our study focuses on creativity, specifically through the medium of generating written content, for two interrelated reasons. Firstly, while creative generation of written content is one of the most commonly used methods for assessing learning outcomes in universities, content generation constitutes the signature function of GenAI like ChatGPT, raising the question of how GenAI impacts learning and outputs when students are asked to 'create' and use material generated by GenAI in the process.

Second, according to Bloom's taxonomy—one of the most influential pedagogical frameworks—creativity is theorised to involve the pinnacle of higher-order skills in the learning process. Bloom's Taxonomy categorises cognitive skills into six hierarchical levels: remembering, understanding, applying, analysing, evaluating, and creating, with each level representing increasingly complex mental processes, from basic recall of information to the synthesis of new ideas. Creation, according to this framework, is the production of something *new*. While mere writing need only involve lower-order thinking skills, creativity, according to this framework, is considered to involve higher-order thinking skills, including design, conjecture, and synthesis (Bloom et al., 1964; Hyder & Bhamani, 2016; Nikolić & Dabić, 2016).

Whilst we know that GenAI is able to reproduce some of the lower-order processes in Bloom's taxonomy, its capacity to support or

replicate these higher-order skills remains to be fully understood. On the one hand, there are some reasons to predict that GenAI is unable to support the higher-order skills involved in creativity, both in terms of the output it generates and in terms of the cognitive engagement of its users. The quality of what is generated by GenAI has been found variable (Buruk, 2023, Ironsi & Solomon Ironsi, 2025). In terms of professional output, GenAI has been found to produce good quality academic texts but can also produce 'hallucinations' of entirely fictitious ideas presented as facts (Babl & Babl, 2023 who asked the GenAI to write academic abstracts), although the quality of the output has been shown to be enhanced through an iterative interaction with the user (Yang et al., 2025, Cheng et al., 2025).

Even if GenAI can be prompted to generate high-quality creative outputs, the fear is that students may de-prioritise critical engagement in their learning, ultimately diminishing their motivation and even ability to think independently and creatively and to develop higher-order cognitive skills that have classically between viewed as central to creative writing and research (Kasneci et al., 2023; Malinka et al., 2023; Wild, 2023). It is not yet clear how much of this could be mitigated through the interactive use of GenAI, although dyadicinteraction with GenAI may enhance deeper understanding of the learning material and lead to more tailored learning experiences compared to a non-interactive use (see Lecler et al., 2023; Ray, 2023). We might also be concerned about mechanised convergence in the increased use of GenAI; that AI generation tends to result in less diverse textual output (Lee et al., 2025).

On the other hand, Lee et al. (2025) argue that the use of GenAI need not lead to disengagement from higher-order creative thinking processes. They point out that even users who use GenAI output without editing it *'may have nonetheless performed a critical, reflective judgement in forming the decision not to edit it'*. Instead of leading to a lack of critical thinking on the part of the user, the authors found that critical thinking while using GenAI takes a different shape. Knowledge workers in their study engaged critically in setting clear goals, refining prompts, and assessing AI-generated content to meet specific criteria and standards; and they engaged reflectively in verifying outputs against external sources and their own expertise. This is supported by earlier work by Sarkar (2023) suggesting that AI shifts knowledge work from material production to critical integration of ideas. If this is right, using GenAI could in fact lead to greater engagement in higher-order creative processes on the part of the user, leaving the 'lower-order' processes of textual production to the GenAI.

Levels of creativity have widely been assessed in terms of divergent thinking. Creativity is a multifaceted construct often associated with the generation of novel and useful ideas (Runco & Jaeger, 2012). Creativity is thus enabled in part by divergent thinking — the ability to explore multiple potential solutions — and this has become widely used as a measure of creativity (Guilford, 1967; Runco & Jaeger, 2012). Two important components of divergent thinking are considered in the present study: fluency and flexibility. Fluency describes the quantity of ideas generated, whilst flexibility considers the quality of those ideas by counting the number of distinct conceptual domains or categories represented in the responses (see Reiter-Palmon et al., 2019). Considering divergent thinking through a dual measurement framework provides us with a structured approach to evaluating creativity, allowing for insights into both the breadth and depth of creative thought. In constructing this framework, we contribute a novel measure of creativity in output facilitated by GenAI.

A related question is *for whom* GenAI might facilitate or impede creativity, and therefore its impact on social inequalities in education. There is vibrant research being conducted on the potential uses of GenAI in teaching, such as in automated systems for improving writing skills (Cummings, Monroe, & Watkins, 2024; Ironsi & Solomon Ironsi, 2025) reflective writing (Chang et al., 2025) or argumentation skills (Wambsganss et al., 2021). If GenAI can enhance pedagogical capability, we might also investigate whether GenAI has potential to facilitate greater access for students from marginalised communities in higher education. On one hand, we must be mindful that several studies have underscored the ways in which the outputs of GenAI tools like ChatGPT can inadvertently reinforce biases in areas such as gender and politics (Gross, 2023; Motoki et al., 2024; Rutinowski et al., 2024). Watson & Romic (2025) emphasise that ChatGPT, like other technologies, is shaped by societal influences. Thus, its responses can reflect the subjective leanings originated from its data and embedded in its training, potentially skewing responses in subtle yet significant ways (Chen et al., 2024). These biases can be problematic as they may reinforce existing inequalities or propagate inaccurate representations of marginalised groups – shaping student perceptions in a way that reflects these biases rather than promoting critical thinking or inclusivity. Consequently, the algorithms that drive AI tools could lead to a homogenization of ideas and learning experiences by amplifying certain patterns and trends present in the training data, while excluding outliers or unconventional ideas. On the other hand, it has the potential to equalise students with a varied levels of command in English language, which could be beneficial in learning exercises and assessments where English language accuracy is not a stipulated learning outcome but a medium to convey other skills (see for example Prather et al., 2024 looking at programming).

In the present study therefore, we examined the performance and experiences of diverse groups of students, varied by both their English proficiency and prior experience with ChatGPT. Only by including a rich and varied sample can we truly understand ChatGPT's potential role in promoting inclusion or reinforcing inequalities. This diversity is essential not only for achieving a comprehensive understanding of ChatGPT's potential influences on inclusion, but also for uncovering nuanced interactions between student backgrounds and AI-assisted learning. For instance, students who are native English speakers may leverage ChatGPT differently than those who rely on it for language support, potentially affecting their creative and writing processes. Similarly, those with prior experience using ChatGPT may have developed strategies to maximise its benefits or avoid its pitfalls, leading to different outcomes in terms of performance and experiences.

#### 1.1.2. Experiences of generative AI

An important facet in exploring the potential and limitations of GenAI in higher education, is to explore users' experiences of it. The use of any new technology is set to change not only the outputs of its users, but also users' perceptions of learning itself as it comes to be modified by the availability of the new tool. In the context of higher education, this might include values in relation to learning, perceptions of effort required for educational tasks, and perceptions of the acceptability of the technology. We have some evidence in the extant literature on the attitudes of students. Students in computer science were asked about the potential and dangers that could come from the use of GenAI in programming. The students were concerned about resulting programmer laziness, occupational anxiety, and incorrect information; with some participants (7 of 41) raising the potential for a negative effect of GenAI on the development of thinking skills. However, a higher number of them (13 out of 41) perceived no disadvantages at all in using GenAI (Yilmaz & Yilmaz, 2023a). Investigating the degree of acceptance of GenAI among students, Yilmaz et al. (2023) demonstrated that the use of GenAI tools can enhance not only the student's programming outputs but also their self-efficacy and motivation (Yilmaz & Yilmaz, 2023b). In a more recent study, Gasaymeh et al. (2024) found that students had moderate concerns regarding data security and misinformation but also were optimistic about its potential to foster creativity and innovation. The authors concluded that greater engagement with students to increase their technical familiarity of the technology, along with discussions about the ethical implications of using it, were important.

Through a creative writing task, the present study examined not only the performances and attitudes of individuals regarding these two skills but also their experience of engaging with ChatGPT. Understanding these experiences is crucial, as it provides insights into how students

interact with AI-generated content and how it influences their perceptions of creativity. While AI tools can produce creative outputs, there is a growing need to investigate how students perceive AI-generated creativity in order to understand the relationship between users' intrinsic motivation to engage with complex problems, the value they place on their own creative processes, the perceived moral acceptability of using the technology in the context of higher education, and the impact of AI on enjoyment of the learning process.

### 1.2. Research hypotheses

Based on the review of the literature, as presented above, we adopted the following research hypotheses.

(1) the use of ChatGPT will enhance performance in creative writing tasks, potentially leading to increased levels of divergent thinking and improved writing quality. Furthermore, it is hypothesised that,
(2) the use of ChatGPT may also alter the perception, possibly diminishing the perceived value, enjoyment, and intellectual difficulty of creating skills during the task, while simultaneously reducing the effort required for creating.

Moreover, this research will probe into the moral concerns surrounding the use of such generative tools, questioning their moral acceptability. Ultimately, this study (see Fig. 1) aims to investigate the ways in which ChatGPT represents a tool of empowerment, and/or a threat, within higher education.

## 2. Methods

### 2.1. Participants

Methods, hypotheses and analyses of the study were pre-registered online at Open Science Frame (see link to the pre-registration in the acknowledge section). This study was conducted using the survey platform Qualtrics and participants were recruited via Prolific, a commonly used crowdsourcing platform for research. A total of 266 participants completed the study, all current students studying at a university in the United Kingdom. To maximise data quality, we used attention check and explicitly asked participants to carefully read the instructions and commit to the task. Forty-three people who failed the attention check were excluded from further analysis, leaving a final sample of 225.

The sample had an average age of 29.07 years (SD = 8.59 years), with 51 % male participants, 48 % female and 1 % identifying as other genders. The ethnic composition included 46 % White, 29 % Black, 18 % Asian, and 8 % from mixed or other racial backgrounds. Participants self-rated their socio-economic status on a 10-point scale, where 1 represents "worst off" and 10 represents "best off", with a mean rating of 5.74 (SD = 1.53).

Regarding education levels, 60 % of participants were undergraduates, while 31 % were postgraduates (including 13 PhD students), and 4 % were studying other degrees. Their academic interests spanned a wide range of disciplines: 48 % were studying Social Sciences (e.g., Psychology, History, Economics), 26 % were in Natural Sciences (e.g., Biology, Physics, Chemistry), 18 % were in Formal Sciences (e.g., Mathematics, Computer Science), and 19 % were studying Applied Sciences and professional degrees (e.g., Medicine, Engineering, Law, Architecture). An additional, 4 % of participants were studying other subjects.

### 2.2. Design

To explore potential differences in results between native and non-native English speakers, as well as between individuals familiar with ChatGPT and those new to the technology, participants were randomly selected based on their English proficiency and prior experience with ChatGPT. This selection ensured balanced representation across both control and experimental conditions, resulting in four distinct groups.

- Group 1 (Non-Native, Inexperienced): Non-native English speakers with no prior use of ChatGPT, n = 58 (26 %);
- Group 2 (Native, Inexperienced): Native English speakers with no prior experience with ChatGPT, n = 60 (27 %);
- Group 3 (Non-Native, Experienced): Non-native English speakers with prior experience using ChatGPT, n = 54 (24 %);
- Group 4 (Native, Experienced): Native English speakers with prior experience using ChatGPT, n = 53 (24 %).

Among the 107 participants (47 % of the total sample) with prior experience with ChatGPT: 50 % used it occasionally, 48 % used it regularly, and 2 % work in a relevant field and have explored many of its advanced capabilities. A detailed breakdown of their demographic information is provided in Table 1 in the *Supplementary Information* (SI).

All participants completed pre-task and post-task surveys, along with a creative writing task. 'Creative writing' can in common parlance refer to specific genres of writing including fiction and poetry. However, for our purposes, we define creative writing more generally as a task that requires the participant to use the cognitive skills involved in being creative. In a pre-task survey, participants first rated their perceived value, enjoyment, intellectual difficulty and required effort for the creating and writing part of a "typical creative writing task". They also

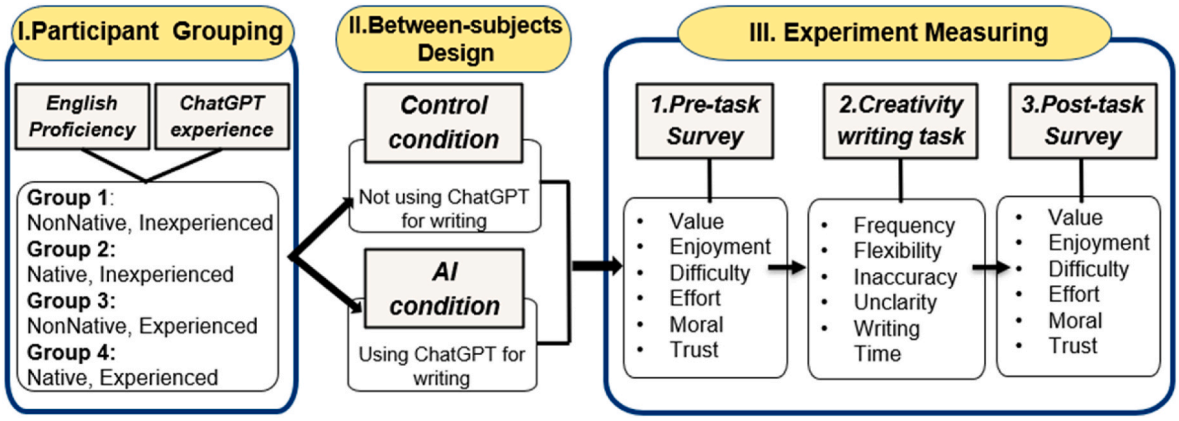


**Fig. 1. Study design overview.** Participants were first labelled into four sub-groups based on their English proficiency and prior ChatGPT experience. Then, we used a between-subjects design to randomly assign all participants to one of the two experimental conditions (Control vs. AI). Subsequently, participants completed the same survey to report their experiences, both before and after taking part in the creativity writing task.

rated their trust in, and moral acceptance of, using generative AI for creative writing tasks. These perceptions were measured using a 5-point Likert scale (1 'not at all' - 5 'very much').

After the pre-task survey, the experiment used a between-subjects design to randomly assign participants into one of two treatment conditions: the control condition (these participants were forbidden from using ChatGPT for the task) and the experimental condition or 'AI group' (these participants were explicitly asked to make sure they used ChatGPT for the task). Participants were required to report their usage of any tools (in the control condition) or ChatGPT (in the experimental condition) after completing the writing task.

The creative writing task involved describing uses for a hypothetical "*printer that can print anything in 2097*" within a 10-min time limit by writing a short piece (approx. 100 words), "to best demonstrate their creativity and writing quality". Creativity was measured by the number of unique uses proposed, corresponding to the instruction, "*describe as many unique uses as possible for this printer, AND do so as quickly as possible in 10 minutes*". The writing quality requirement instructed participants to "*minimise grammar errors (e.g., spelling, grammar, and punctuation), AND maximise clarity of the writing (e.g., word choice, sentence structure, tone adjustments)*". These assessment criteria were further emphasised with comprehension checks. Participants who did not pass the comprehension checks were reminded to review the task instructions and were required to answer correctly before proceeding to the next step.

After completing the task, participants took a post-task survey to re-evaluate their perceptions (value, enjoyment, difficulty, and effort of creating and writing, and trust and moral acceptance of using AI), based on this creative writing experience. These questions were identical to those in the pre-task survey but focused specifically on participants' experience with this task compared to a 'typical creative writing task'. Additionally, they also provided opinions on the impact of transparency on moral judgement by rating the moral acceptance when '*people were transparent about their use of generative AI tools, e.g., acknowledging using ChatGPT to write a piece*'. Demographic information, including age, gender, ethnicity, socioeconomic status, degree, and subject, was collected.

### 2.3. Assessment

In this study, participants demonstrated intellectual creativity by generating different uses for a printer. Their performance was measured by the two key elements of divergent thinking.

1. **Fluency** (*quantity* of ideas) indicated by the total number of different uses of the printer, and
2. **Flexibility** (*variety/quality* of ideas) indicated by the total number of unique conceptual domains represented by these uses.

For example, *"printer can be used to print houses, print villages, and print buildings"* would count as three different uses (houses, villages and buildings) and one context (constructions), so the frequency score would be 3 and flexibility would be 1. *"Printing super-fast vehicles for travel"* and *"Printing airplanes to transport goods"* refer to two different uses (vehicles and airplanes) and distinct contexts (travel and transport goods), so the frequency score would be 2 and flexibility would be 2.

Since manually identifying use cases for all submissions can be time consuming and subject to individual differences, we automated this process using large language models (LLM). We crafted a prompt that provides the LLM with a sequence of instructions for identifying use cases in each submission, as well as a few examples to further guide identification. This corresponds to what is called a few-shot prompting approach (Lee et al., 2024). As LLMs, we used GPT4o and GPT4 Turbo, maintaining a temperature of 0 for consistency in response generation.

In order to validate that the automatic counts were accurate, we manually annotated a sample of responses and compared them to the LLM's outputs. More specifically, we selected a random sample of 40 responses (5 from each group and condition) to be annotated by three human raters from the research team. The raters, who were blinded to both group and condition, conducted their evaluations independently. The raters were provided with clear guidelines on evaluating flexibility and frequency, aligned with the few-shot examples used in the LLM inference process. To evaluate consistency, we computed the inter-rater agreement using the intraclass correlation coefficient (ICC), specifically ICC (2), which measures pairwise agreement across multiple raters.

Given the three-rater setup, we averaged the pairwise ICC scores to report an overall agreement level. The ICC scores among all four raters is reported in Table 2 in the SI. In order to interpret the agreement score, we use the scale suggested in (Koo & Li, 2016). The average ICC score among the human raters was 0.81 for both frequency and flexibility, indicating good agreement. Furthermore, when comparing the average ratings from the human raters with those from GPT-4o, we observed ICC scores of 0.93 for frequency and 0.83 for flexibility, demonstrating strong alignment between human and LLM assessments. These findings suggest that the LLM's assessment of creativity aligns well with human evaluation, supporting the reliability of AI-based methods in creativity analysis.

Participants' writing quality was assessed using Grammarly, a widely used grammar checking tool. We evaluated the inaccuracy by counting the number of grammar mistakes (including spelling, grammar, and punctuation) and unclarity (the number of suggestions for word choice, sentence structure, tone adjustments) detected by Grammarly.

## 3. Results

### 3.1. Analyses

In line with the study pre-registration, descriptive statistics and a series of regression models were used to test the hypothesised effects of using ChatGPT on creative writing task performance. The treatment condition (*using ChatGPT* vs. *not using ChatGPT during the task*) was used as the independent variable, and performance (*creativity* and *writing quality*) as the dependent variable (see section 3.2). The same analysis was repeated five times within each group as well as across groups. We also conducted a heterogeneity analysis (see section 3.2.1) to validate the performance differences across various demographic factors (e.g., age, gender, ethnicity, SES and educational background). We further explored the effects of different human-AI Interaction (HAI) dyads by conducting the same analysis with HAI type (*Human-only; AI-only; AI-led; Dyadic*) as the independent variable and performance as the dependent variable (see section 3.2.2). Additionally, we calculated the frequency of creativity performance at the group-level to compare it with individual-level performance (see section 3.2.3).

Similarly, descriptive statistics and regression models were used to analyse the effects on perception. As experience was measured both before and after the task, and we were interested in the changes, the treatment condition was again used as the independent variable, while the differential perception ($\Delta Experience = Post. - Pre.$) was treated as the dependent variable (see section 3.3). The same analysis was repeated for all four groups, as well as across groups. Full results of these analyses were reported in the SI. Furthermore, we used descriptive statistics to analyse the moral acceptance of using ChatGPT in creative writing tasks and conducted *t*-test to compare the effects of with and without a transparency measure on moral attribution.

### 3.2. Performance

To understand the impact of ChatGPT on various aspects of performance in the creative writing task, we conducted a series of regression analyses for each group, using *frequency*, *flexibility*, *inaccuracy, unclarity*, and *writing time*, as the dependent variable. Significant performance differences between the AI and control conditions were observed in most of the participant groups, and also when analysed independently of the

groups (see Table 3 in SI).

As shown in Fig. 2, for most groups, participants in the AI condition outperformed those in the control condition on frequency, flexibility, accuracy, and clarity, while they also spent less time completing the task. There were a few exceptions to this overall pattern, where there was no difference between the control and AI conditions. Namely, in Group 2 (Native, Inexperienced), no significant differences were observed for frequency ($b = 0.91$, $SE = 2.30$, $p = .69$) or clarity ($b = -1.30$, $SE = 1.90$, $p = .51$), and in Group 3 (Non-Native, Experienced) ($b = -2.30$, $SE = 1.50$, $p = .14$), clarity differences were also non-significant. However, in no groups did the control condition outperform the AI condition.

In Group 4 (Native, Experienced), the only significant effect of AI usage was on writing time, where participants using AI completed the task faster compared to the control group. No significant differences were found in flexibility, correctness, frequency, or clarity between the AI and control groups.

Overall, we found that participants who used ChatGPT were more likely to perform better across all aspects of the task, including higher frequency and flexibility of creativity, less inaccuracy and unclarity from their writing, and less time required to complete the task. These results suggest that AI provides a strong advantage to people in their performance of the creative writing task, with non-native English speakers benefiting particularly from this advantage.

### 3.2.1. Heterogeneity analysis

Furthermore, to validate the robustness of the results, we grouped all participants based on 6 demographic factors: age, gender, ethnicity, socioeconomic status, degree, and subject. Four additional subgroups (other gender, other ethnicity, other degree and other subject) were excluded from the analysis due to small sample sizes and limited clarity. Across the remaining 21 demographic subgroups, we compared participants' performance in both the AI and control conditions. As shown in Fig. 3, participants in the AI condition consistently outperformed those in the control condition. This advantage was evident in various aspects, including creativity, writing quality, and time efficiency.

### 3.2.2. HAI dyad analysis

As a result of early discussions, we performed an additional analysis to explore whether different forms of interaction with AI would result in different outcomes. The self-reported tool usage for the task reveals that a significant majority (90 %) in the control condition did not use any digital tools during the task. However, participants in the AI condition reported varied interactions with ChatGPT: 27 % relied entirely on ChatGPT to complete the task, 47 % made minor or major changes to the AI-generated output, and 21 % used their own prompts or engaged in a feedback loop with ChatGPT. Additionally, 5 % of the usage was unclassified.

Hence, to dissect how ChatGPT affected performance, we categorised participants based on their engagement with ChatGPT into the following four classes: Human-only (participants who did not use ChatGPT or any other digital tools for the task); AI-only (participants who entirely relied on ChatGPT to complete the task); AI-led (participants who primarily relied on ChatGPT to complete the task); Human-AI-dyadic (participants

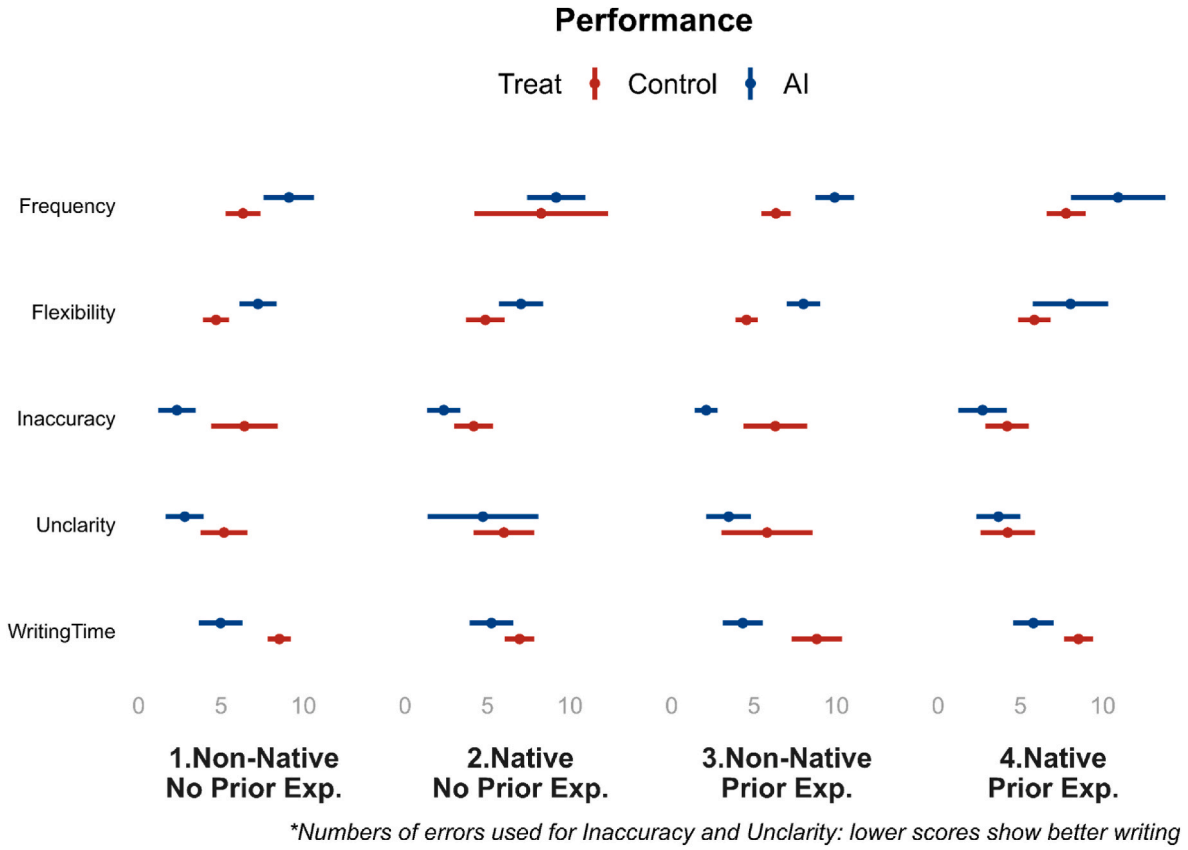


**Fig. 2. Performance compared across 4 subgroups.** The dots on the solid lines show the mean of performance score with the accompanying error bars delineating the 95 % confidence intervals for these means. Mean performance scores under the control condition (did not use ChatGPT for the task) are shown in red, while those in the AI condition (used ChatGPT for the task) appear in blue. *From left to right, each panel shows one group (Group 1: Non-native English speakers with no prior use of ChatGPT; Group 2: Native English speakers with no prior experience with ChatGPT; Group 3: Non-native English speakers with prior experience using ChatGPT; Group 4: Native English speakers with prior experience using ChatGPT). From the top to the bottom, each row indicates one aspect of task performances (frequency of creativity: numbers of different uses, flexibility of creativity: numbers of unique uses;* inaccuracy *of writing: numbers of grammar errors; unclarity of writing: numbers of clarity suggestions; writing time: time spent for completing the writing).*

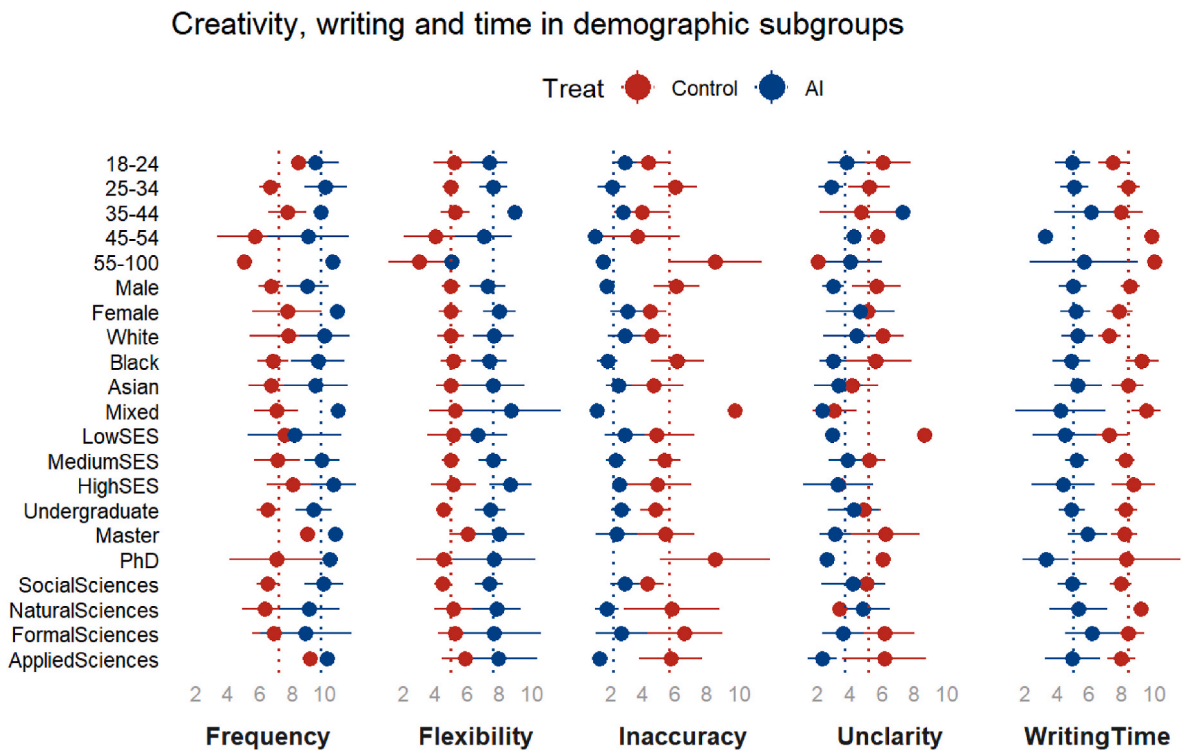


**Fig. 3. Performance compared across demographics.** The dots on the solid lines show the mean of performance score with the accompanying error bars delineating the 95 % confidence intervals for these means. Mean performance scores under the control condition (did not use ChatGPT for the task) are shown in red, while those in the AI condition (used ChatGPT for the task) appear in blue. *From the top to the bottom, each row shows one demographic subgroup. From left to right, each panel shows one aspect of task performances (frequency of creativity: numbers of different uses, flexibility of creativity: numbers of unique uses; inaccuracy of writing: numbers of grammar errors; unclarity of writing: numbers of clarity suggestions; writing time: time spent for completing the writing).*

who interacted actively with ChatGPT during the task).

As shown in Fig. 4, we found that participants demonstrated the lowest levels of creativity, writing quality, and time efficiency in the Human-only dyad condition compared to the other conditions (AI-only, AI-led, Human-only). Moreover, the benefits of using ChatGPT varied depending on the type of interaction. Performance was highest in the AI-only condition, followed by the AI-led and Human-AI-dyadic conditions (See Table 4 in the SI).

### 3.2.3. Group-level analysis

Following from our analysis of individual-level performance discussed in Section 3.2.1 previously (see Fig. 2), we then examined the group-level performance for creativity. This allowed us to assess the extent of creative thinking by collections of participants that share a characteristic or condition.

Fig. 5 suggests that groups using ChatGPT generally demonstrate higher creativity in terms of flexibility across various conditions, while results for frequency are more varied (See Table 5 in the SI). The most pronounced difference is observed in the *Task Treatment* group (*Grp 9 in* Fig. 5), where people using ChatGPT outperform the Control group in both Flexibility (635 vs. 552) and Frequency (442 vs. 313).

When breaking down the groups into different combinations of English proficiency and prior experience with ChatGPT, higher flexibility scores were found in all AI conditions regardless of these characteristics, as shown by the comparisons across *Grps 1–8* in Fig. 5. In terms of frequency, participants in the AI conditions outperformed those in the control condition in most cases, with exceptions in two groups: *Non-Native* and *Inexperienced (Grp 1* in Fig. 5: 140 vs. 193) and *Inexperienced* (*Grp 7 in* Fig. 5: 312 vs. 344). These two groups share the common characteristic of having no prior experience with ChatGPT.

These results indicate that ChatGPT generally enhances creative outputs, particularly for flexibility, across different groups, regardless of English proficiency or prior experience with the tool. However, the results are nuanced: while flexibility scores increase consistently in all AI-supported conditions, the improvement in frequency scores appears more limited for individuals without prior experience with ChatGPT. This suggests that while ChatGPT consistently boosts creative flexibility, its impact on the volume of creative output may be moderated by other factors.

### 3.3. Perception

We assessed the impact of completing the writing task (with or without ChatGPT) on participants' perceptions of the perceived value, enjoyment, intellectual difficulty and effort involved in creative writing, and its impact on moral judgments of using AI for creative writing, by comparing the difference in participant ratings in the pre-task survey to the post-task survey. The results are represented in Fig. 6. Negative values indicate a perceived decrease, meaning participants' ratings (e.g. of the perceived value of creativity) declined after completing the task compared to their initial perception before the task.

The changes in perception for creative writing showed a similar pattern in Groups 1 (Non-native, Inexperienced), 2 (Native, Inexperienced), and 3 (Non-native, Experienced). In the AI condition, both the required effort and perceived difficulty for creative writing decreased significantly after using ChatGPT to complete the task. However, the perceived enjoyment and value of creativity and writing also declined significantly in the AI condition. Interestingly, in the control condition, these perceptions remained relatively stable before and after the task. Changes in effort, difficulty, enjoyment, and value were minimal in the control group. Details of the regression models are provided in Table 6 of the SI.

However, the above perception changes were not found in Group 4 (Native, Experienced). Native English speakers with prior ChatGPT

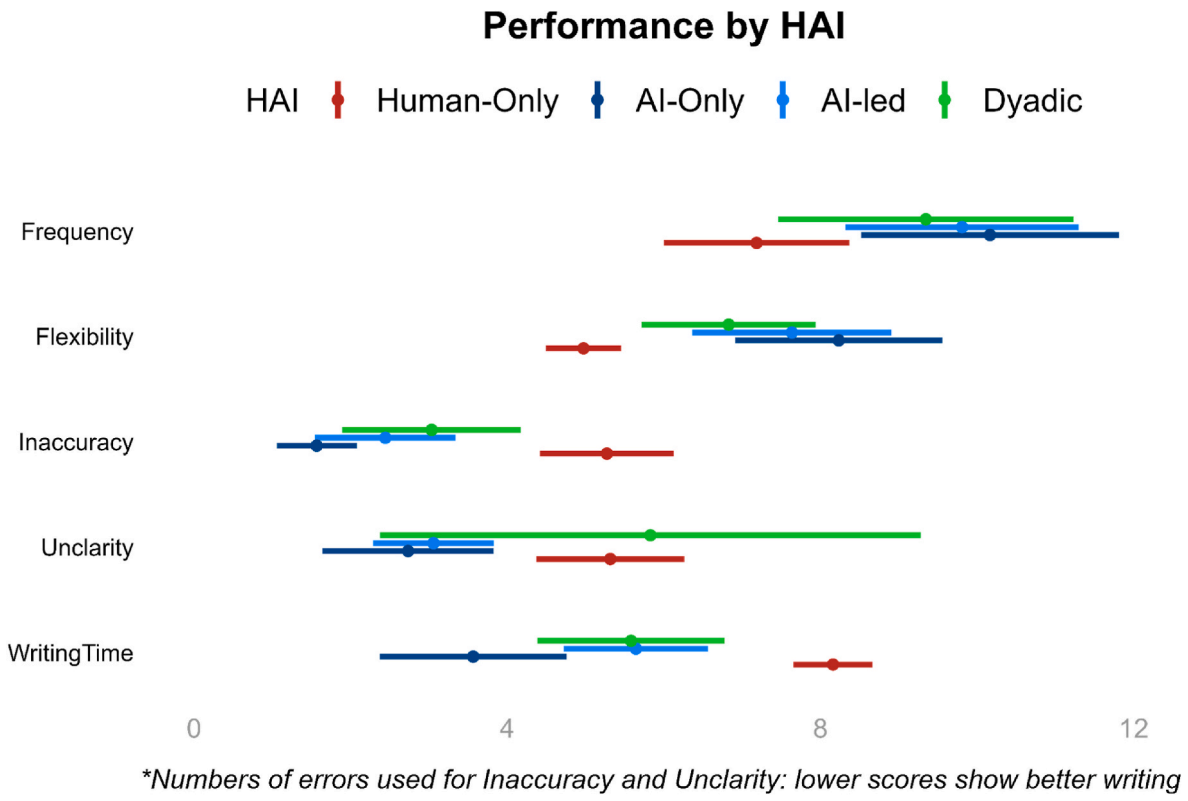


**Fig. 4. Performance compared across Human-AI Dyad (HAI).** The dots on the solid lines show the mean of performance score with the accompanying error bars delineating the 95 % confidence intervals for these means. Mean performance scores for human-only input (where participants completed the task without using ChatGPT) are shown in red, while those in the AI-only dyad (where participants used ChatGPT to complete the task entirely) appear in blue. Lighter blue corresponds to the AI-led dyad (where participants used ChatGPT to complete the task mostly) and green indicates dyadic interaction (where participants engaged in a feedback loop with ChatGPT to complete the task). *From the top to the bottom, each row indicates one aspect of task performances (frequency of creativity: numbers of different uses, flexibility of creativity: numbers of unique uses;* inaccuracy *of writing: numbers of grammar errors; unclarity of writing: numbers of clarity suggestions; writing time: time spent for completing the writing).*

experience reported similar levels of perceived value ($b = -0.06$, $SE = 0.18$, $p = .73$), enjoyment ($b = -0.12$, $SE = 0.25$, $p = .62$), difficulty ($b = -0.05$, $SE = 0.21$, $p = .81$), and effort ($b = -0.16$, $SE = 0.21$, $p = .45$) both before and after completing the task. This suggests that the use of AI had a limited impact on their experience. Three exceptions to the overall trend were found in Groups 2 and 3. Participants in these groups showed no significant change in their enjoyment or effort for creativity and writing after completing the task, regardless of whether they used ChatGPT. Specifically, in Group 2 (Native, Inexperienced), the changes in perceived value ($b = 0.31$, $SE = 0.24$, $p = .20$) and enjoyment ($b = 0.44$, $SE = 0.22$, $p = .06$) were not statistically significant. Similarly, Group 3 (Non-native, Experienced) showed no significant change in enjoyment ($b = -0.10$, $SE = 0.21$, $p = .66$).

Participants' moral judgements changed consistently across all 4 participant groups. Regardless of their English proficiency or prior ChatGPT experience, participants in the control condition showed decreased trust and moral acceptance in using AI for creative writing tasks, after completing the task independently. Specifically, the mean difference (before and after the task) in moral acceptance in the control group decreased by −0.73 ($SE = 0.11$), and the mean difference in trust decreased by −0.80 ($SE = 0.11$). In contrast, participants who used ChatGPT during the task displayed increased trust and moral acceptance of AI afterward. The AI group showed a mean increase of 0.76 ($SE = 0.09$) in trust and 0.63 ($SE = 0.12$) in moral acceptance. Participants' moral acceptance changed significantly in conditions *if people were transparent about their use of generative AI tools*, difference = −1.44, $t(221) = -16.70$, $p < .001$, Cohen's $d = -1.12$.

To summarise, participants in the AI condition experienced more significant shifts in their perceptions, while the perception changes remained relatively stable in the control condition. Specifically, participants who used ChatGPT for the creative writing task found it to be less effortful and intellectually demanding. However, their enjoyment of the task and their perceived value of creativity and writing skills decreased. Non-native English speakers, as well as native English speakers without prior ChatGPT experience, were most affected by these perception changes.

In addition, while ChatGPT assistance appeared to positively influence participants' task performance, there was a demonstrated lack of trust and moral acceptance toward the use of AI in creative writing. However, transparency measures—such as acknowledging the use of ChatGPT— seemed to change their moral acceptance. The mean moral acceptance with transparency ($M = 3.75$, $SE = 0.07$) was significantly higher than without ($M = 2.31$, $SE = 0.07$), $t(221) = 16.51$, $p < .001$, 95 % CI [1.27, 1.61]. The increase in moral acceptance (with vs. without transparency) was significantly larger in the control condition ($M = 1.83$, $SE = 0.12$) compared to the AI condition ($M = 1.05$, $SE = 0.11$), $t(220) = 4.72$, $p < .001$, 95 % CI [0.36, 0.90]. Therefore, when transparency measures were employed, participants in the AI condition who initially showed relatively high acceptance further increased their acceptance, while those who did not use AI during the task became more open to the use of ChatGPT.

## 4. Discussion

### 4.1. Study summary

In this study, we examined students' performance and perception changes through a creative writing task, under two conditions: using

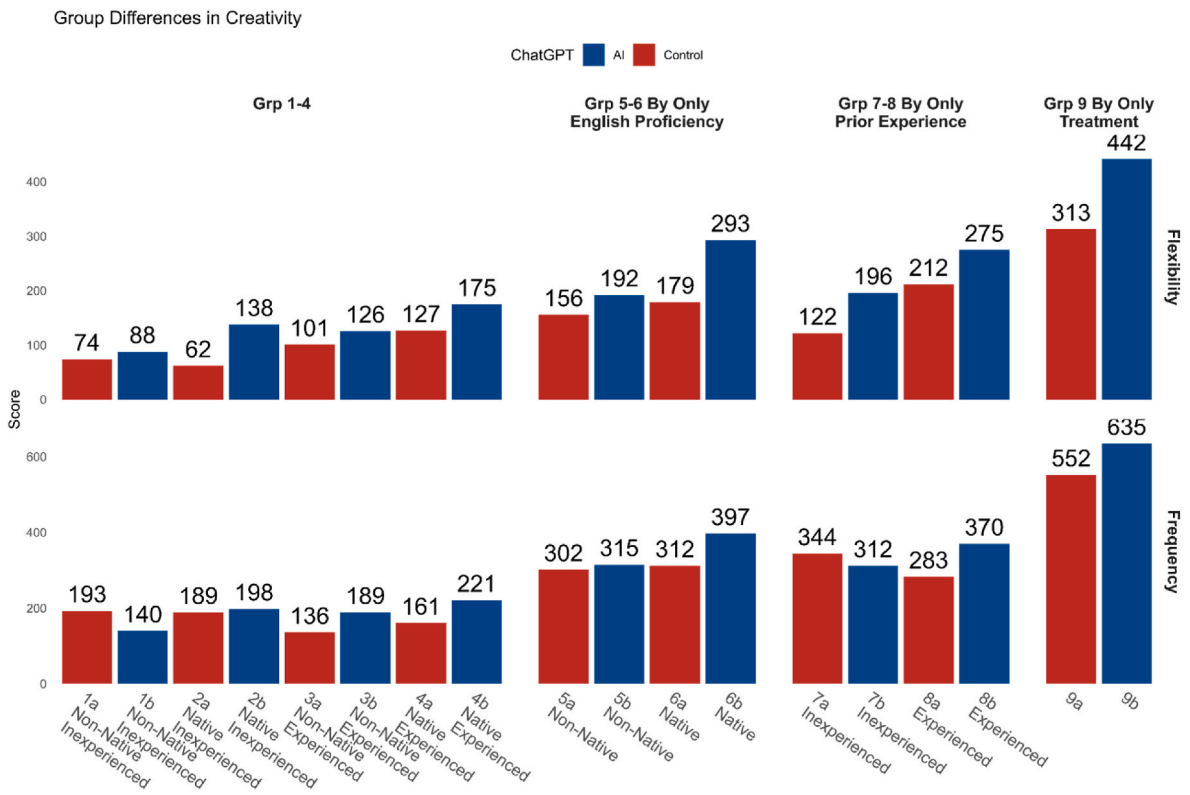


**Fig. 5. Performance compared at group-level for creativity.** The bars show the creativity scores, with the red colour indicating using ChatGPT (the 'Control' condition) for the task and blue for not using ChatGPT (the 'AI' condition). The top panel represents *flexibility* scores *(of creativity: numbers of unique uses)* and the bottom panel shows *frequency* scores (*of creativity: numbers of different uses*).

ChatGPT and not using it. We explored these changes across four groups based on participants' English proficiency and prior experience of ChatGPT. Our findings confirmed our hypotheses and revealed that groups using ChatGPT, especially people who are not native English speakers, demonstrated enhanced creativity and writing performance. They also found the task easier and less effortful. However, we also found that their perceived value and enjoyment of the creative process decreased. These results suggest that while ChatGPT has the potential to empower students by improving productivity when used thoughtfully, its impact on motivation and appreciation for creative thinking may require careful consideration.

The study provides valuable insights to inform the responsible integration of AI in higher education while preserving students' connection to the creative process. More broadly, it prompts reflection on how the rise of "machine culture" (Brinkmann et al., 2023) is reshaping people's perceptions of knowledge generation and transmission. By exploring the role of AI in student thinking processes, especially creativity, we gain a deeper understanding of how these technologies may affect not just academic performance, but also the ways in which learners perceive and value creativity and intellectual engagement in an increasingly digitised future.

### 4.2. Discussion of the findings

#### 4.2.1. Using ChatGPT enhances delivery of a creative writing task

The findings support our first hypothesis and show that ChatGPT enhances students' productivity by boosting their creative performance, a finding that aligns with its known capabilities in text processing (Abdullah et al., 2022; Haleem et al., 2022). This prompts questions about whether the use of AI enhances divergent thinking or merely reduces the time spent on correcting writing accuracy. This reduction in time and effort, however, could be significant. If ChatGPT alleviates the cognitive load required for writing, this may allow students to focus more on the creative aspects of their tasks (for more discussion of these ideas see Kalyuga, 2011; Paas & Sweller, 2014; Stadler et al., 2024). By reducing the mental effort required for drafting and organising ideas, GenAI could free up cognitive resources, enabling students to dedicate more attention to developing creative concepts. This is further supported by the improved writing quality and reduced time spent in the AI-assisted condition, suggesting that ChatGPT's capacity to process words allows students to articulate and deliver their ideas more efficiently. An enhancement of creativity may also result from the range of ideas provided by the AI, and/or from the inspiration that comes from being presented with surprising ideas. Being mindful of the 'hallucination' risk in the output, as ChatGPT draws ideas from a rich resource of knowledge and reassembles them, these unique combinations both contributed to the creative writing task directly and may also have stimulated participants' imaginations (Austin et al., 2012; Haase & Hanel, 2023). With education about the technical functionality of the tool, and its limitations, it could help students in higher education to develop integrative and critical skills.

#### 4.2.2. Using ChatGPT lowers the language barrier for non-native English speakers

Our study shows that this benefit is particularly pronounced for non-native English speakers. For these students, ChatGPT not only assists in formalising ideas but also in effectively 'translating' their thoughts into coherent English (Marrone, Taddeo, & Hill, 2022; Su, 2024; Fitria, 2023; Feng Teng, 2024). Its ability to facilitate clearer communication therefore empowers students from diverse linguistic backgrounds to express their ideas more confidently. This could enhance their learning experiences in higher education and allow for a refocusing of their efforts on the quality of ideas in creative outputs. Ultimately, by bridging language barriers, ChatGPT has the potential to expand knowledge access and foster greater inclusion in higher education, where the language of the host country is a necessary medium for demonstrating other skills (see

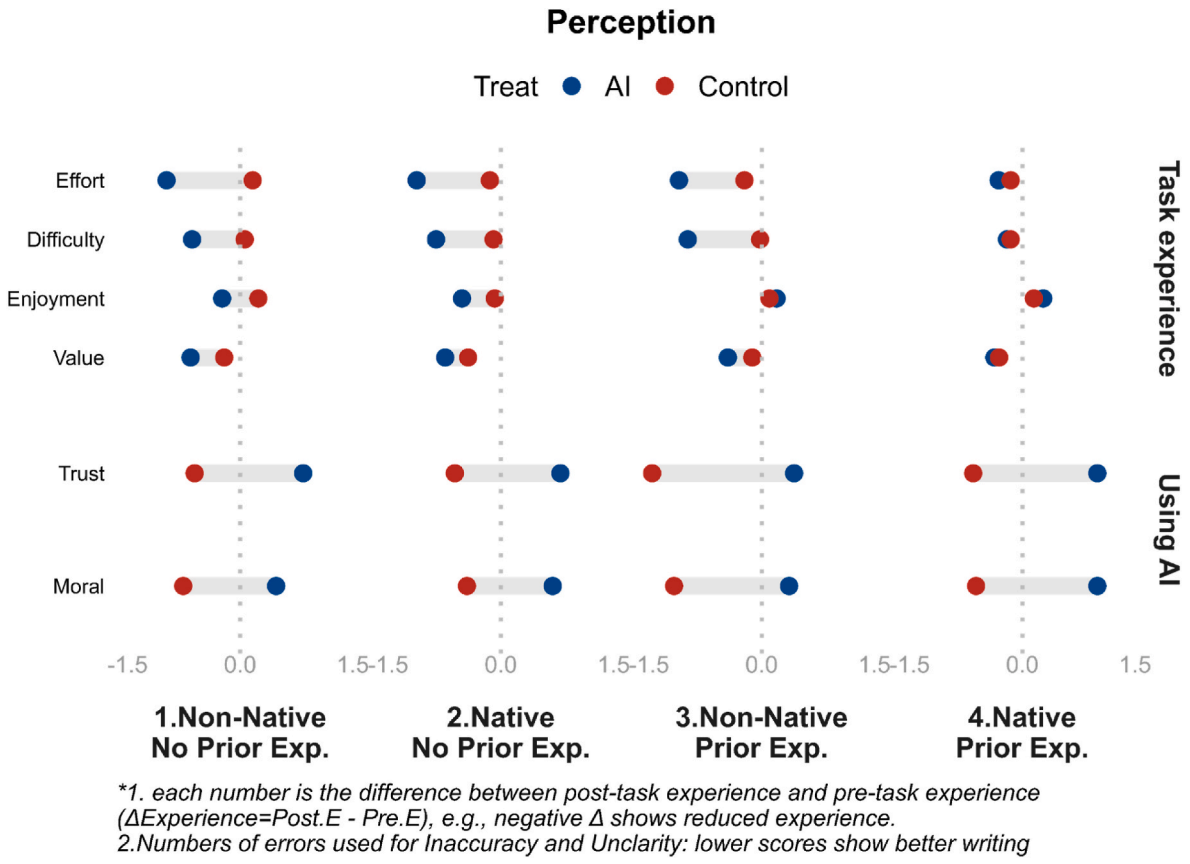


**Fig. 6. Perception (changes before and after task) compared across 4 subgroups.** The dots show the mean of the perception change reported in the pre-task and post-task surveys (*ΔExperience=Post. – Pre.*). The grey bars show the difference in the changes observed between the groups under AI (used ChatGPT for the task) and control (did not use ChatGPT for the task) conditions. Mean perception changes under the control condition are shown in red, while those in the AI condition appear in blue. *From left to right, each panel shows one group (Group 1: Non-native English speakers with no prior use of ChatGPT; Group 2: Native English speakers with no prior experience with ChatGPT; Group 3: Non-native English speakers with prior experience using ChatGPT; Group 4: Native English speakers with prior experience using ChatGPT). From the top to the bottom, each row indicates one aspect of task perceptions (trust and moral acceptance in using AI for creative writing; effort required, intellectual difficulty, enjoyment and value of creative writing).*

also Rawas, 2024).

#### 4.2.3. *Using ChatGPT may support diversity of ideas*

We found that the AI-assisted groups performed well not only at individual levels but also at group levels, demonstrating greater overall flexibility and frequency in their creative writings. Therefore, the fear that generative AI simply recycles ideas within a limited database of existing information, and thus may lead to limited scope for novel thinking, was not supported in this study. Additionally, group-level creativity in the AI condition appears to surpass that of the control conditions. This observation supports the growing body of literature (Cecutti et al., 2021; Sparrow et al., 2011; Chemero, 2023). This somewhat challenges the initial concerns that technological advancements may harm cognition. Our results highlight that AI can complement, and perhaps enhance, human creativity by offering novel perspectives and stimulating concepts that might not emerge in traditional contexts. While the inclusion benefits are evident, the results also provide a more nuanced perspective, showing that AI could also contribute to collective creative efforts without necessarily limiting intellectual diversity.

#### 4.2.4. *Using ChatGPT undermines the value and enjoyment of creativity*

Although ChatGPT has made tasks like creative writing easier and less effortful, our findings support concerns about the potential devaluation and diminished enjoyment of the creative process. Magni et al. (2024) highlight the importance of creativity as an advanced cognitive function that contributes to human satisfaction (see also Fusi et al., 2024). Notably, this perceived playfulness plays a vital role for behavioural intention, even more significant than perceived values and performance expectancy (Gajendragadkar et al., 2024; Moghavvemi et al., 2017; Parveen et al., 2024).

However, the rise of GenAI has inevitably altered this dynamic. When AI performs well in creative tasks, students may feel there is less of a need to engage deeply with the thinking process. This disengagement can undermine the "need for cognition" — the degree to which individuals seek, engage in, and enjoy effortful cognitive activity — which is most critical for academic performance, even beyond cognitive abilities (Lavrijsen, Preckel, & Verschueren, 2023; Lavrijsen et al., 2022). Thus, while the use of GenAI may not have a direct negative impact on students' cognitive skills, and may even help revitalise them by highlighting the qualities that distinguish humans from machines, our findings show that GenAI may undermine students' motivation to engage in the cognitive effort that is necessary for deeper intellectual enjoyment (contra Lee et al., 2025; see also Stadler et al., 2024).

#### 4.2.5. *Transparency mitigates ethical concerns*

Another concern highlighted by this study is the ethics of using GenAI in higher education, especially with the low ratings of moral acceptability and trust of unregulated AI found in relation to our third hypothesis. As Tlili and colleagues (2023) discuss, ChatGPT raises several ethical questions regarding its truthfulness, response quality (fairness and honesty), and the potential for misuse. For example, even sophisticated AI detectors can be easily bypassed with minor edits, such as adding a single word, which can reduce the likelihood of detection.

Moreover, ChatGPT can provide misleading justifications for inaccurate responses, such as claiming oversights or format problems. This kind of manipulation, whether intentional or due to biassed algorithms, poses risks for students who trust ChatGPT for reliable information. However, our results demonstrated that simple measures, e.g. *acknowledging using ChatGPT to write a piece*, can effectively help to mitigate these concerns. Openly acknowledging the use of ChatGPT for the creative writing task enhanced both trust and moral acceptance of the tool among our participants.

We might also consider this finding in relation to a broader framing of higher education as provider of professional skills (Wardat et al., 2023) and that the effective usage of GenAI may become regarded as a key expectation of their academic careers. In this overall climate, the moral acceptability of GenAI may increase in the future.

### 4.3. Limitations and further study

In order to contextualise the conclusions and to guide future studies, we present the limitations in the design of the present study.

#### 4.3.1. Measures of creativity

Although we used divergent thinking as a primary measure of creativity, it is important to recognise that this represents only one facet of a multidimensional construct. Creativity encompasses various cognitive skills, including convergent thinking, which involves narrowing down possibilities to find the most appropriate solution (Ward, 2007). Our assessment primarily focused on the quantity of creative ideas generated, which may not fully capture the quality and impact of creative output. Not all creative outputs are equally valuable or impactful. Additional dimensions of creativity extend beyond mere idea generation, emphasising novelty, originality, value, and appropriateness of the creative output (Glăveanu et al., 2021). Therefore, creative cognition involves not only producing new ideas but also selecting, prioritising, and acting upon them. Future studies should consider integrating additional measures of creativity, such as evaluating both the quality and quantity of creative responses, to provide a more comprehensive understanding of how generative AI influences creative output from ideation to implementation.

#### 4.3.2. Assessment of creativity

We utilised LLMs to automate the process of creativity rating, a practice that has been documented in prior research (Cropley & Marrone, 2025; Marrone, Cropley, & Wang, 2023). While this approach allows for efficiency in evaluating a large dataset of writing samples, it is crucial to recognise that these creativity ratings can be approximate and require further validation. The use of AI for rating creativity introduces the possibility of bias, as the AI system may have inherent limitations in understanding nuanced or highly novel creative outputs. Future studies should consider a hybrid approach, combining AI-assisted rating with human expert evaluation to leverage the strengths of both methods. In this study, we conducted agreement checks among three human raters on a random subset (approx. 20 %) of the entire dataset, and found evidence of high interrater reliability. Still, ideally, a larger portion of the data should have been assessed by multiple human raters, with a more comprehensive inter-rater reliability measurement. This would help validate the AI-generated ratings further and identify any systematic biases in the automated assessment process.

#### 4.3.3. Larger and more diverse sample

Although we recruited hundreds of participants in this study, the sample size was relatively small after assignment to subgroups, especially for the Human-AI dyads. This limitation may restrict the generalisability of the relevant findings. Additionally, the significant benefits observed in non-native English speakers highlight the need to expand our investigation to other potentially disadvantaged groups in higher education. For example, individuals with neurodiversity, specific learning needs and different cultural backgrounds. Future studies should consider larger scale studies with more diverse groups to gain insights into how generative AI impacts creativity across different cognitive profiles, linguistic backgrounds, and cultural contexts.

Finally, our study captured a snapshot of ChatGPT's impact on creativity through a single writing task. While this offers valuable initial insights, it is crucial to conduct longitudinal studies to understand the long-term effects of AI assistance on creative skills. Such studies would help determine whether the observed effects persist, evolve, or change over time. Furthermore, exploring the long-term use of generative AI could reveal how creative processes adapt to reflect the technology advancement.

## 5. Conclusions

Our study firstly contributes to the growing literature on GenAI in education and plays a crucial role in the small pool of studies that directly investigate the effects of GenAI use on cognitive skills. Second, the study advances understandings of the experience of using GenAI in a creative learning task, through a controlled study.

Our research found that using ChatGPT enhanced the delivery of a creative writing task in fluency and flexibility of the output; that using ChatGPT lowered the language barrier of the creative-writing task for non-native English speakers; that using ChatGPT may – with critical use – support the diversity of ideas that is produced; that using ChatGPT undermined the perceived value and enjoyment of a creative task; and that transparency over the use of GenAI may mitigate ethical concerns that are currently held about it. In all, our findings suggest that the critical use of GenAI in higher education has potential to enhance creativity in learning rather than to harm it. In pursuing this, it will be important to engage students in learning about how the technology works, its limitations, and also what it means for the development of 'creative skills' in higher education. Our findings highlight the specific need to address opportunities for students to mobilise complex cognitive skills during creative tasks in order to ensure that they can experience the deeper intellectual satisfaction of learning. Moreover, our findings suggest that clear transparency guidelines should accompany activities that allow the use of GenAI.

In concluding, we outline some of the practical implications of our findings for higher education.

### 5.1. Implications for teaching and learning

Results from the study in general show potential for increase in creativity when students interact with ChatGPT. This supports our recommendation that educators, as gatekeepers of knowledge creation and skill development, should adapt their teaching approach and philosophy to incorporate the up-skilling of students in the usage of generative AI tools and LLMs such as ChatGPT, and to develop assessment mechanisms that would aid students in further honing these valuable skills.

The adoption and integration of AI as a tool that enables and enhances learning outcomes represents a paradigm shift in socio-technological trends that leverages technology to "superhumanise" students' capabilities in ways that can significantly improve overall learning outcomes both for the students and for society which benefits from the knowledge created within higher education institutions. The term "superhumanising" suggests that AI provides support that exceeds human capabilities by potentially transforming traditional learning paradigms. Whilst AI can offer valuable assistance to students and learners, there remains of course a risk of overreliance on technology, which may negatively impact critical thinking and problem-solving skills among students. Especially when there are utilitarian motivations, such as the desire to streamline learning processes through AI, the use of GenAI may inadvertently diminish students' sense of agency and ownership over their learning experiences and even lead to cheating. We

must, however, bear in mind that the enhancement of learning by using tools, including digital technologies, is not itself a new phenomenon. The question that must therefore be asked by educators is one of value - what is the value that we will attribute to the cognitive skills involved in creating without the aid of AI, versus those skills involvd in working *with* it.

While AI can automate certain aspects of learning, it does not adequately address the complex, multifaceted nature of creative expression and critical thinking. Therefore, educators should equip students not just to use AI tools, but to understand their capabilities, limitations, and ethical implications. Rather than competing with AI or training secondary-passive recipients in learning, it is possible to foster complementary skills that position students as primary, active agents. Further, our results from the inexperienced participants in the sample suggest that good understanding of generative AI is necessary in order to really benefit from it (supported by results in Kay, 2023).

In light of the findings from this study, it is also important that the design of assessment is carefully reconsidered. The way in which stakeholders within the educational sector and society at large conceive of plagiarism may need re-evaluation, as the role of AI in content generation blurs the lines of authorship and intellectual property. Moreover, results from the study feed into wider and complex discussions of the suitability of assessing work as individual or as relational (i.e. group work) (see also Dochy, Segers, & Sluijsmans, 1999; Van den Berg, Admiraal, & Pilot, 2006; Pereira, Flores, & Niklasson, 2016; Benvenuti et al., 2023). The reframing of how society perceives thinking about creative work as a relational and collaborative endeavour that may include AI would complement transformations that are already occurring in the world of work, where skills in the effective and critical usage of generative AI are already in demand (Shen & Guo, 2024; Yang & Zhao, 2024; Yuan & Liu, 2025). By fostering active human-AI cooperation and reframing education as a process of developing uniquely human capabilities alongside technological literacy, we can prepare students for a future where success depends on effectively leveraging AI while maintaining human creativity, critical thinking, and ethical judgement (see more discussion of HAI in Jin, Lin, & Lai, 2025). In increasing the attention within higher education to these issues, we bring into focus the complexity of using AI as a tool for learning, and may mitigate the decreased enjoyment of creative tasks as seen in our results.

### 5.2. Implications for industry and policy

Our study also offers insights for business stakeholders, AI programmers, and policymakers. As AI-based digital technologies and large language models like ChatGPT continue to proliferate, we will witness a more widespread occurrence of the human-AI dyad. This trend is particularly evident in content creative industries, which are entering a new phase of transformation where generative AI is changing the nature of job roles, skill requirements and skill value (Cetinic & She, 2022; Coeckelbergh, 2023). Our study suggests that on the one hand, recognition of AI skills will be important in mitigating the devaluing of creative skills (see also Köbis & Mossink, 2021). On the other hand, the government may need to be aware of these potential industry changes and develop appropriate support for those whose livelihoods may be affected in the short-term.

More widely it has been noted that there is a bias aGenAInst AI use, which is often manifested in hesitancy among individuals to fully accept AI-generated intellectual, artistic and creative works due to the perception of it being "superhuman" (Fang et al., 2024; Ragot et al., 2020). This draws attention to an underlying assumption in our value system that creativity is most valuable when it is 'human' and 'independent'. Consequently, the rise in creative products of the human-AI dyad will likely present complex moral, economic, social and cultural dilemmas, prompting discussions about the boundaries between human and AI. Our findings suggest a need to shift how we evaluate products made with the input of generative AI, moving beyond raw performance metrics to consider its usefulness in helping humans perform tasks (see also Choudhury & Shamszare, 2023; Hitsuwari et al., 2023; Shahzad, Xu & Javed, 2024) which may include a consideration of its perceived usefulness beyond higher education, in society at large. This includes assessing not just accuracy and efficiency, but also the perceived usefulness of AI in various contexts.

It is pertinent to state that findings from this study feed into the conversation on the usage of AI tools, especially within higher education, by highlighting the potential increased inclusion that arises from the empowerment of generative AI users, especially non-native speakers. In this sense, the super-humanising potential may be stronger for some who currently find themselves linguistically marginalised. An important part of this conversation, however, must be to bear in mind that access to ChatGPT itself is not universal. ChatGPT and other generative AI platforms operate with both free and paid services, which means that there will be a new dynamic of privilege and marginalisation that may emerge - where some have access to more sophisticated tiers of the tool - as the technology becomes more widely used, something that higher educational institutions would do well to keep in view.

There is a question here regarding the role that legislation and regulation can play to address the consequences of increased reliance on generative AI. One of the priorities is to maintain content quality as AI usage expands, and one route towards this may be the tagging of AI-generated content to increase transparency and facilitate fact-checking. The tendency of generative AI to produce hallucinations – fabricated information presented as fact – is a significant concern. Recent research has shown that leading language models – ChatGPT, Deepseek, Grok, PaLM, Claude, Qwen, and Llama – were on average producing hallucinations in 75 % of the cases (Dahl et al., 2024), with many users unaware of this prevalence (Williamson & Prybutok, 2024). Thus, tagging the source of AI-generated content is urgently needed and this could be something mandated at a regulatory level.

## CRediT authorship contribution statement

**Peidong Mei:** Writing – review, editing & revising, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Deborah N. Brewis:** Writing – review, editing & revising, Writing – original draft, Conceptualization. **Fortune Nwaiwu:** Writing – review, editing & revising, Writing – original draft, Conceptualization. **Deshan Sumanathilaka:** Validation, Software. **Fernando Alva-Manchego:** Writing – review & editing, Validation, Conceptualization. **Joanna Demaree-Cotton:** Writing – review, editing & revising.

## Ethics approval

The project has been approved by the Ethics Committee of the Faculty of Environment, Science, and Economy (FESE) at the University of Exeter, UK, and the Research Ethics Committee at the University of Bath, UK. Information forms were provided to all participants, and consent forms were collected from each participant before any data collection commenced. Additionally, participants were debriefed at the conclusion of the study.

## Appreciation

We also thank Pranjal Jain, a PhD student in Computer Science at Swansea University, and Prasad Shetye, a Master's student in Business and Analytics at Swansea University, for their assistance in the initial recruitment process of the study.

## Pre-registration

This study is pre-registered on Open Science Framework. The pre-

registration can be accessed at: https://osf.io/dt5up?mode=&revisionId=&view_only=

## Funding source

This work is funded by the GW4 [Cru23_03, PROJECT 124760]; 'Reshaping Learning in the Digital Era – How Does ChatGPT Affect Our Learning Performance, Experience and Attitude'.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We thank Pranjal Jain, a PhD student in Computer Science at Swansea University, and Prasad Shetye, a Master's student in Business and Analytics at Swansea University, for their assistance in the initial recruitment process of the study. We also thanks Elif Duman Cogen and Benoit Dillet for their support on this project.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chbah.2025.100140.

## References

Abdullah, M., Madain, A., & Jararweh, Y. (2022). ChatGPT: Fundamentals, applications and social impacts. In *2022 ninth international conference on social networks analysis, management and security (SNAMS)* (pp. 1–8). IEEE.

Austin, R. D., Devin, L., & Sullivan, E. E. (2012). Accidental innovation: Supporting valuable unpredictability in the creative process. *Organization Science, 23*(5), 1505–1522.

Babl, F. E., & Babl, M. P. (2023). Generative artificial intelligence: Can ChatGPT write a quality abstract? *Emergency Medicine Australasia, 35*(5), 809–811.

Benvenuti, M., Cangelosi, A., Weinberger, A., Mazzoni, E., Benassi, M., Barbaresi, M., et al. (2023). Artificial intelligence and human behavioral development: A perspective on new skills and competences acquisition for the educational context. *Computers in Human Behavior, 148*, Article 107903. https://doi.org/10.1016/j.chb.2023.107903

Blackler, F. (1995). Knowledge, knowledge work and organizations: An overview and interpretation. *Organization Studies, 16*(6), 1021–1046.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1964). *Taxonomy of educational objectives* (Vol. 2). New York: Longmans. Green.

Brinkmann, L., Baumann, F., Bonnefon, J. F., Derex, M., Müller, T. F., Nussberger, A. M., … Rahwan, I. (2023). Machine culture. *Nature Human Behaviour, 7*(11), 1855–1868.

Buçinca, Z., Malaya, M. B., & Gajos, K. Z. (2021). To trust or to think: Cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-computer Interaction, 5*(CSCW1), 1–21.

Buruk, O. O. (2023). Academic writing with GPT-3.5: Reflections on practices, efficacy and transparency. *arXiv preprint arXiv:2304.11079*. https://doi.org/10.48550/arXiv.2304.11079.

Cecutti, L., Chemero, A., & Lee, S. W. (2021). Technology may change cognition without necessarily harming it. *Nature Human Behaviour, 5*(8), 973–975.

Cetinic, E., & She, J. (2022). Understanding and creating art with AI: Review and outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications, 18*(2), 1–22.

Chang, C. Y., Lin, H. C., Yin, C., & Yang, K. H. (2025). Generative AI-assisted reflective writing for improving students' higher order thinking. *Educational Technology & Society, 28*(1), 270–285.

Chemero, A. (2023). LLMs differ from human cognition because they are not embodied. *Nature Human Behaviour, 7*(11), 1828–1829.

Chen, F., Wang, L., Hong, J., Jiang, J., & Zhou, L. (2024). Unmasking bias in artificial intelligence: A systematic review of bias detection and mitigation strategies in electronic health record-based models. *Journal of the American Medical Informatics Association, 31*(5), 1172–1183.

Cheng, Y., Fan, Y., Li, X., Chen, G., Gašević, D., & Swiecki, Z. (2025). Asking generative artificial intelligence the right questions improves writing performance. *Computers and Education: Artificial Intelligence.* , Article 100374. https://doi.org/10.1016/j.caeai.2025.100374

Choudhury, A., & Shamszare, H. (2023). Investigating the impact of user trust on the adoption and use of ChatGPT: Survey analysis. *Journal of Medical Internet Research, 25*, Article e47184.

Coeckelbergh, M. (2023). Narrative responsibility and artificial intelligence: How AI challenges human responsibility and sense-making. *AI & Society, 38*(6), 2437–2450.

Cropley, D. H., & Marrone, R. L. (2025). Automated scoring of figural creativity using a convolutional neural network. *Psychology of Aesthetics, Creativity, and the Arts, 19*(1), 77–86. https://doi.org/10.1037/aca0000510.

Cummings, R. E., Monroe, S. M., & Watkins, M. (2024). Generative AI in first-year writing: An early analysis of affordances, limitations, and a framework for the future. *Computers and Composition, 71*, Article 102827.

Dabić, T. (2016). The Bloom's taxonomy revisited in the context of online tools. In *Sinteza 2016-international scientific conference on ICT and E-business related research* (pp. 315–320). Singidunum University.

Dahl, M., Magesh, V., Suzgun, M., & Ho, D. E. (2024). Large legal fictions: Profiling legal hallucinations in large language models. *Journal of Legal Analysis, 16*(1), 64–93.

Danry, V., Pataranutaporn, P., Mao, Y., & Maes, P. (2023). Don't just tell me, ask me: Ai systems that intelligently frame explanations as questions improve human logical discernment accuracy over causal ai explanations. In *Proceedings of the 2023 CHI conference on human factors in computing systems* (pp. 1–13).

DemandSage. (2024). ChatGPT statistics – 200 million active users. https://www.demandsage.com/chatgpt-statistics/.

Dochy, F. J. R. C., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education, 24*(3), 331–350.

Du, W., Zhu, Z., Xu, X., Che, H., & Chen, S. (2024). CareerSim: Gamification design leveraging LLMs for career development reflection. In *Extended abstracts of the CHI conference on human factors in computing systems* (pp. 1–7).

Elbanna, S., & Armstrong, L. (2024). Exploring the integration of ChatGPT in education: Adapting for the future. *Management & Sustainability: An Arab Review, 3*(1), 16–29.

Fang, X., Che, S., Mao, M., Zhang, H., Zhao, M., & Zhao, X. (2024). Bias of AI-generated content: An examination of news produced by large language models. *Scientific Reports, 14*(1), 5224.

Feng Teng, M. (2024). A systematic review of ChatGPT for English as a foreign language writing: Opportunities, challenges, and recommendations. *International Journal of TESOL Studies, 6*(3).

Fitria, T. N. (2023). Artificial intelligence (AI) technology in OpenAI ChatGPT application: A review of ChatGPT in writing English essay. *ELT Forum: Journal of English Language Teaching, 12*(1), 44–58.

Frieder, S., Pinchetti, L., Griffiths, R. R., Salvatori, T., Lukasiewicz, T., Petersen, P., et al. (2023). Mathematical capabilities of chatgpt. *Advances in Neural Information Processing Systems, 36*, 27699–27744.

Fusi, G., Giannì, J., Borsa, V. M., Colautti, L., Crepaldi, M., Palmiero, M., & Rusconi, M. L. (2024). Can creativity and cognitive reserve predict psychological well-being in older adults? The role of divergent thinking in healthy aging. *Healthcare, 12*(3), 303. MDPI.

Gajendragadkar, S., Arora, R., Trivedi, R., & Neelam, N. (2024). From intentions to action: How behavioural intentions shape employee performance through digital learning? *Journal of Workplace Learning, 36*(5), 348–363.

Gasaymeh, A. M. M., Beirat, M. A., & Abu Qbeita, A. A. A. (2024). University students' insights of generative artificial intelligence (AI) writing tools. *Education Sciences, 14*(10), 1062.

Gläveanu, V. P., Ness, I. J., & de Saint Laurent, C. (2021). Creative learning in digital and virtual environments. *Opportunities and challenges of technology–enabled learning and creativity*. New York and London: Routledge.

Gross, N. (2023). What ChatGPT tells us about gender: A cautionary tale about performativity and gender biases in AI. *Social Sciences, 12*(8), 435.

Guilford, J. P. (1967). Creativity: Yesterday, today and tomorrow. *Journal of Creative Behavior, 1*(1), 3–14.

Haase, J., & Hanel, P. H. (2023). Artificial muses: Generative artificial intelligence chatbots have risen to human-level creativity. *Journal of Creativity, 33*(3), Article 100066.

Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations, 2*, Article 100089.

Higher Education Policy Institute. (2024). Provide or punish? Students' views on generative AI in higher education. https://www.hepi.ac.uk/2024/02/01/provide-or-punish-students-views-on-generative-ai-in-higher-education/.

Hitsuwari, J., Ueda, Y., Yun, W., & Nomura, M. (2023). Does human–AI collaboration lead to more creative art? Aesthetic evaluation of human-made and AI-generated haiku poetry. *Computers in Human Behavior, 139*, Article 107502.

Hyder, I., & Bhamani, S. (2016). Bloom's Taxonomy (cognitive domain) in higher education settings: Reflection brief. *Journal of Education and Educational Development, 3*(2), 288.

Ironsi, C. S., & Solomon Ironsi, S. (2025). Experimental evidence for the efficacy of generative AI in improving students' writing skills. *Quality Assurance in Education, 33*(2), 237–252.

Jin, F., Lin, C. H., & Lai, C. (2025). Modeling AI-assisted writing: How self-regulated learning influences writing outcomes. *Computers in Human Behavior, 165*, Article 108538.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review, 23*, 1–19.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274.

Kay, J. (2023). Foundations for Human-AI teaming for self-regulated learning with explainable AI (XAI). *Computers in Human Behavior, 147*, Article 107848.

Köbis, N., & Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry. *Computers in Human Behavior, 114*, Article 106553.
Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine, 15*(2), 155–163.
Lavrijsen, J., Preckel, F., & Verschueren, K. (2023). Seeking, mastering, and enjoying cognitive effort: Scrutinizing the role of need for cognition in academic achievement. *Learning and Individual Differences, 107*, Article 102363.
Lavrijsen, J., Vansteenkiste, M., Boncquet, M., & Verschueren, K. (2022). Does motivation predict changes in academic achievement beyond intelligence and personality? A multitheoretical perspective. *Journal of Educational Psychology, 114*(4), 772.
Lecler, A., Duron, L., & Soyer, P. (2023). Revolutionizing radiology with GPT-based models: Current applications, future possibilities and limitations of ChatGPT. *Diagnostic and Interventional Imaging, 104*(6), 269–274.
Lee, S., Choi, D., Lee, M., Choi, J., & Lee, S. (2023). Fostering youth's critical thinking competency about AI through exhibition. In *Proceedings of the 2023 CHI conference on human factors in computing systems* (pp. 1–22).
Lee, U., Jung, H., Jeon, Y., Sohn, Y., Hwang, W., Moon, J., et al. (2024). Few-shot is enough: Exploring ChatGPT prompt engineering method for automatic question generation in English education. *Education and Information Technologies, 29*(9), 11483–11515.
Lee, H. P. H., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., et al. (2025). *The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers*.
Magni, F., Park, J., & Chao, M. M. (2024). Humans as creativity gatekeepers: Are we biased against AI creativity? *Journal of Business and Psychology, 39*(3), 643–656.
Malinka, K., Peresíni, M., Firc, A., Hujnák, O., & Janus, F. (2023). On the educational impact of chatgpt: Is artificial intelligence ready to obtain a university degree? *Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V., 1*, 47–53.
Marr, B. (2023). *A short history of ChatGPT: How we got to where we are today*. Forbes. https://www.forbes.com/sites/bernardmarr/2023/05/19/a-short-history-of-chatgpt-how-we-got-to-where-we-are-today/.
Marrone, R., Cropley, D. H., & Wang, Z. (2023). Automatic assessment of mathematical creativity using natural language processing. *Creativity Research Journal, 35*(4), 661–676.
Marrone, R., Taddeo, V., & Hill, G. (2022). Creativity and artificial intelligence—a student perspective. *Journal of Intelligence, 10*(3), 65.
McLuhan, M. (1994). *Understanding media: The extensions of man*. MIT Press.
McLuhan, M. (2017). The medium is the message. In *Communication theory* (pp. 390–402). Routledge.
Moghavvemi, S., Paramanathan, T., Rahin, N. M., & Sharabati, M. (2017). Student's perceptions towards using e-learning via Facebook. *Behaviour & Information Technology, 36*(10), 1081–1100.
Motoki, F., Pinho Neto, V., & Rodrigues, V. (2024). More human than human: Measuring ChatGPT political bias. *Public Choice, 198*(1), 3–23.
Nikolić, M., & Dabić, T. (2016). *The Bloom's Taxonomy Revisited in the Context of Online Tools* (pp. 315–320). in Sinteza 2016 - International Scientific Conference on ICT and E-Business Related Research. https://doi.org/10.15308/Sinteza-2016-315-320
Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. *The Cambridge handbook of multimedia learning, 27*, 27–42.
Parveen, K., Phuc, T. Q. B., Alghamdi, A. A., Hajjej, F., Obidallah, W. J., Alduraywish, Y. A., et al. (2024). Unraveling the dynamics of ChatGPT adoption and utilization through structural equation modeling. *Scientific Reports, 14*(1), Article 23469.
Pereira, D., Flores, M. A., & Niklasson, L. (2016). Assessment revisited: A review of research in assessment and evaluation in higher education. *Assessment & Evaluation in Higher Education, 41*(7), 1008–1032.
Prather, J., Reeves, B. N., Denny, P., Leinonen, J., MacNeil, S., Luxton-Reilly, A., … Barbre, G. (2024). Breaking the programming language barrier: Multilingual prompting to empower non-native English learners. *arXiv preprint arXiv:2412. 12800*. https://doi.org/10.48550/arXiv.2412.12800.
Ragot, M., Martin, N., & Cojean, S. (2020). Ai-generated vs. human artworks. a perception bias towards artificial intelligence?. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems* (pp. 1–10).
Rasul, T., et al. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching, 6*, 41–56.
Rawas, S. (2024). ChatGPT: Empowering lifelong learning in the digital age of higher education. *Education and Information Technologies, 29*(6), 6895–6908.
Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems, 3*, 121–154.
Reiter-Palmon, R., Forthmann, B., & Barbot, B. (2019). Scoring divergent thinking tests: A review and systematic framework. *Psychology of Aesthetics, Creativity, and the Arts, 13*(2), 144–152.
Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*(1), 92–96.
Rutinowski, J., et al. (2024). The self-perception and political biases of ChatGPT. *Human Behavior and Emerging Technologies, 2024*, Article 7115633.
Sarkar, A. (2023). Exploring perspectives on the impact of artificial intelligence on the creativity of knowledge work: Beyond mechanised plagiarism and stochastic parrots. In *Proceedings of the 2nd annual meeting of the symposium on human-computer interaction for work* (pp. 1–17).
Shahzad, M. F., Xu, S., & Javed, I. (2024). ChatGPT awareness, acceptance, and adoption in higher education: The role of trust as a cornerstone. *International Journal of Educational Technology in Higher Education, 21*(1), 46.
Shen, Y., & Guo, H. (2024). "I feel AI is neither too good nor too bad": Unveiling Chinese EFL teachers' perceived emotions in generative AI-mediated L2 classes. *Computers in Human Behavior, 161*, Article 108429.
Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science, 333*(6043), 776–778.
Spero, T. A. (2024). Using ChatGPT to improve access, inclusion, and equity in and through English-medium instruction in higher education. In *ChatGPT and global higher education: Using artificial intelligence in teaching and learning* (p. 135).
Stadler, M., Bannert, M., & Sailer, M. (2024). Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior, 160*, Article 108386.
Su, J. F. (2024). Using ChatGPT to enhance writing proficiency in an EFL college English course in taiwan. In *Selected papers from the thirty-third international symposium on English language teaching and learning* (p. 377).
Sullivan, M., Kelly, A., & McLaughlan, P. (2023). *ChatGPT in higher education: Considerations for academic integrity and student learning*. Edith Cowan University. https://ro.ecu.edu.au/ecuworks2022-2026/2501/.
Tanprasert, T., Fels, S. S., Sinnamon, L., & Yoon, D. (2024). Debate chatbots to facilitate critical thinking on youtube: Social identity and conversational style make a difference. In *Proceedings of the 2024 CHI conference on human factors in computing systems* (pp. 1–24).
Terwiesch, C. (2023). Would chat GPT3 get a wharton MBA? A prediction based on its performance in the operations management course. *Mack Institute for Innovation Management*. https://mackinstitute.wharton.upenn.edu/wp-content/uploads/2023/01/Would-ChatGPT-get-a-Wharton-MBA.pdf.
The Russell Group. (2023). New principles on use of AI in education. https://russellgroup.ac.uk/news/new-principles-on-use-of-ai-in-education/.
TIME. (2024). Why ChatGPT is the fastest growing web platform ever. https://time.com/6253615/chatgpt-fastest-growing/.
Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., et al. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart learning environments, 10*(1), 15.
Van den Berg, I., Admiraal, W., & Pilot, A. (2006). Design principles and outcomes of peer assessment in higher education. *Studies in Higher Education, 31*(3), 341–356.
Veza, I., et al. (2023). How ChatGPT affects education landscape: Effects of ChatGPT on higher education accessibility and inclusivity. In *International conference on advancing and redesigning education* (pp. 569–579).
Wambsganss, T., Kueng, T., Soellner, M., & Leimeister, J. M. (2021). ArgueTutor: An adaptive dialog-based learning system for argumentation skills. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–13).
Ward, T. B. (2007). Creative cognition as a window on creativity. *Methods, 42*(1), 28–37.
Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *Eurasia Journal of Mathematics, Science and Technology Education, 19*, Article em2286.
Watson, S., & Romic, J. (2025). ChatGPT and the entangled evolution of society, education, and technology: A systems theory perspective. *European Educational Research Journal, 24*(2), 205–224.
Wild, B. (2023). ChatGPT: Cardiff students admit using AI on essays. *BBC News*. http://www.bbc.co.uk/news/uk-wales-65167321.
Williamson, S. M., & Prybutok, V. (2024). The era of artificial intelligence deception: Unraveling the complexities of false realities and emerging threats of misinformation. *Information, 15*(6), 299. https://doi.org/10.3390/info15060299
Yang, K., Raković, M., Liang, Z., Yan, L., Zeng, Z., Fan, Y., … Chen, G. (2025). Modifying AI, enhancing essays: How active engagement with generative AI boosts writing quality. In *Proceedings of the 15th international learning Analytics and knowledge conference* (pp. 568–578).
Yilmaz, F. G. K., Yilmaz, R., & Ceylan, M. (2023). Generative artificial intelligence acceptance scale: A validity and reliability study. *International Journal of Human-Computer Interaction, 40*(24), 8703–8715. https://doi.org/10.1080/10447318.2023.2288730
Yilmaz, R., & Yilmaz, F. G. K. (2023a). Augmented intelligence in programming learning: Examining student views on the use of ChatGPT for programming learning. *Computers in Human Behavior: Artificial Humans, 1*(2), Article 100005. https://doi.org/10.1016/j.chbah.2023.100005.
Yilmaz, R., & Yilmaz, F. G. K. (2023b). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Computers and Education: Artificial Intelligence, 4*, Article 100147.
Yang, L., & Zhao, S. (2024). AI-induced emotions in L2 education: Exploring EFL students' perceived emotions and regulation strategies. *Computers in Human Behavior, 159*, 108337.
Yuan, K., Lin, H., Cao, S., Peng, Z., Guo, Q., & Ma, X. (2023). CriTrainer: An adaptive training tool for critical paper reading. In *Proceedings of the 36th annual ACM symposium on user interface software and technology* (pp. 1–17).
Yuan, L., & Liu, X. (2025). The effect of artificial intelligence tools on EFL learners' engagement, enjoyment, and motivation. *Computers in Human Behavior, 162*, Article 108474.