Generative AI and Creativity

An enquiry into creative performance, independent thinking and cultural diversity

Prepared for Peter Gates | Literature searched to 26th October 2026

Introduction

On 25 October 2026, I posted the following on BlueSky:

> *I've been using ChatGPT for a while. Today I gave it some notes for a study I'm thinking of doing and asked it to draft an outline. It gave me some 1300 words back. I redrafted this in my own style and asked it to analyse my style. It came back with a 2000 word analysis which was very impressive.*
>
> *It will henceforth provide writing in what it considers my writing style rather than its own rather bland style. I also gave it my own Word stylesheet which it now uses to format writing. It is quite impressive. Though I appreciate there are others who think I might be dining with the devil.*

Marcus Hill (@drmarcushill.bsky.social) promptly replied with

> *Not so much dining with the devil as using a technology which is provably detrimental to your own ability to think creatively which is built on environmentally disastrous and economically non-viable infrastructure and a stolen corpus of training data.*

Unfortunately, Marcus didn't link to the research he was referring to, so I did a quick search myself and found several articles with a range of studies, one of which was a systematic review of 28 studies (Holzner, Maier and Feuerriegel, 2025). A surface skim of the research did not appear to support the claim that generative AI was indeed *"a technology which is provably detrimental to your own ability to think creatively"*. Consequently, I asked ChatGPT

> *Can you produce for me a broad literature review on Generative AI and creativity, in my own writing style along with references and drop it into my Word stylesheet and link directly to articles. Can you also indicate in which countries the research was located.*

What follows is my redraft of what ChatGPT produced. Linking directly to original articles would get over the claim that ChatGPT might invent spurious articles and references.

I do not believe for one moment that producing a review using ChatGPT will in any way satisfy the sceptics who do not believe such an approach is valid or reliable. However, that is not my intention. My intention here is merely to co-produce something that those who see the other side of the coin might find useful.

Is this my work? Is it a creative output? Is it valid? Does it help anyone? These are all valid questions, and challenges which we will need to consider as AI moves us forward (or back depending on one's point of view). I personally still have, and repeatedly refer to, Chris Hart's seminal books on literature reviews and searches (Hart, 1998, 2001)[1] [2]. I also use Google, Google Scholar, NUSearch[3], all of which I believe make literature searches more effective - none of which are controversial. Yet using a LLM is still divisive. However, I personally do not believe we can put that genie back in the bottle. What I do believe is we need to embrace new technologies to begin to control how we learn and make use of them. I did this in 1972 when I was asked to try an early LED electronic 4-function calculator when I worked for Commercial Union instead of using mechanical devices. Similarly, I tried out an early Sinclair ZX80 computer with children in the school I worked

[1] Hart, C. (1998) *Doing a Literature Review*, London: Sage.
[2] Hart, C. (2001) *Doing a Literature Search*, London: Sage.
[3] The University of Nottingham's literature search tool.

in 1980, and extensive (and creative) use of calculators in the early 1980s when I was told they would destroy mathematics teaching. I see unapologetically engaging with ChatGPT in the same light.

For now, onto the original issues. Is generative AI detrimental to creativity?

The question behind the claim

The claim that people who make use of generative artificial intelligence become less creative has some empirical support, but it brings together several quite different propositions. But let me first unpick one phrase in that sentence. The phrase "*make use of"* brings with it a number of assumptions related to how we use the technology. Do we depend/rely on it to do the legwork for us or do we "interact with" it after developing our own skill at constructing "prompts" – something which I believe is a cored part of the creativity needed in using the technology. Do we do some of the work first then ask AI to improve it, extend it and develop it? Wherein is the creativity here?

A person may produce a more highly rated piece of writing "*with AI assistance*" whilst contributing less of the thinking and legwork involved in its production (such as on-line searching). Several people may each individually produce more creative work, yet their work may become more uniform as the LLM applies the same algorithm. Equally, an improvement whilst using a tool may disappear when that tool is removed, without the person's original ability having deteriorated. These distinctions matter because the evidence required to establish each proposition is different.

My concern here is to examine what the research actually permits us to say. The emerging literature provides credible evidence of improvements in some creative products and of restrictions on creative exploration in other circumstances. There is also evidence of convergence between people's outputs. There are many cases where people "claim" to be able to identify where a piece of writing was created by AI. However, what of all the pieces of writing these same people think wasn't created by AI? Confirmation bias cold well be at play here. Much less securely established in the current literature is a general and enduring decline in independent creative capacity caused by prolonged reliance on AI. That uncertainty should neither be used to dismiss the risks, nor quietly replaced by a certainty that the research has not earned. Widespread use of AI has not been around long enough to allow studies of prolonged use.

This is a broad critical narrative review, requested specifically by me, produced by ChatGPT and redrafted by me, rather than an exhaustive systematic review. It draws on targeted searches of scholarly publications, research repositories and authors' institutional records; it is followed by examination of relevant methods and related papers where they are accessible. It covers creative writing, visual design, everyday ideation, workplace innovation and education, alongside research on ownership, cultural expression and cognitive engagement. The main empirical literature dates from 2022 to 2026; an earlier definitional paper supplies conceptual context. Peer-reviewed articles and conference papers are distinguished from preprints. Where only an abstract or publication record could be checked, the discussion remains at that level. The geographical register records uncertainty explicitly rather than treating an author's university address as the location of the research.

What is being measured as creativity

Runco and Jaeger (2012) discuss the familiar definition of creativity as requiring both originality and effectiveness. Something unusual is not necessarily useful, meaningful or appropriate. This apparently straightforward formulation nevertheless leaves substantial room for disagreement about who judges effectiveness and within which social or cultural setting. A novel business proposal, a child's mathematical conjecture and a challenging poem do not acquire their significance through the same forms of judgement.

Many studies operationalise creativity through divergent thinking tasks, such as proposing unusual uses for an everyday object. Researchers may count the number of ideas, their distribution across categories and their originality. Other studies ask independent assessors to rate a story or product proposal. Automated measures may estimate how far apart ideas sit within a model's representation of language. Each approach makes a phenomenon measurable, but measurability is not in itself sufficient evidence of validity. Semantic distance is an indicator of difference; it does not establish artistic significance.

There is a further problem when the assessed output has been jointly produced. A higher score demonstrates something about the performance of the person using the system. It does not automatically demonstrate that the person has acquired a greater capacity to work independently. To investigate learning or deterioration, the research needs comparable assessments without assistance, preferably repeated over time. Otherwise, the instrument measures an accomplishment that depends partly on the continuing availability of the tool.

The same caution applies to confidence. People may feel more creative because a task has become easier or because the output looks more accomplished. Such confidence can support participation, but it should be examined alongside performance rather than substituted for it. These are questions about the validity of the inference, not objections to studying creativity experimentally.

Creative products can improve whilst their range narrows

Doshi and Hauser (2024) randomly allocated 293 UK-based writers to produce short stories with no AI ideas, access to one idea, or access to up to five. The assisted stories were rated more highly for novelty and usefulness, with particularly marked benefits among writers with lower initial creativity scores. However, the stories also became more similar to one another. The experiment therefore supports both an improvement in individual creative output and a reduction in collective diversity. Its brief writing task does not establish whether participants subsequently became more or less capable writers without assistance.

Lee and Chung (2024) investigated everyday and innovation-related problems across five experiments, comparing ChatGPT assistance with unaided work and conventional web searching. AI assistance increased the assessed creativity of ideas. Meincke, Nave and Terwiesch (2025) subsequently reanalysed the original datasets and identified reduced diversity within the resulting pools of ideas. This is an instructive scholarly exchange: the apparent disagreement largely concerns the level at which success is assessed. It is also important that the later paper is a reanalysis, not five additional independent experiments.

For an organisation seeking one workable suggestion, a higher average rating might be useful. For a group trying to explore a wide range of possibilities, repeated variations on a similar suggestion may be a serious limitation. A collection of individually attractive ideas can still leave important possibilities unexplored. The issue becomes more consequential if many organisations consult similar systems and accept similar assumptions about what constitutes a sensible problem or an acceptable solution.

Inspiration and fixation

The Australian laboratory study by Wadinambiarachchi and colleagues (2024) provides direct evidence of a detrimental effect within a particular creative activity. Sixty participants with visual design experience undertook a design task using no support, image search or an AI image generator. Those using generative AI showed greater fixation on an initial example and produced fewer ideas, with less variety and originality than the baseline group. This was a short task with a specific interface and design brief; it cannot by itself establish a lasting loss of creative ability.

Cheng and Zhang (2025) offer a more differentiated account. Their experiments found that collaborating with an unconstrained language model could improve creativity on

simpler tasks whilst reducing it on more complex tasks, where fixation became a problem. Changing the way the model responded mitigated fixation in complex tasks, but could weaken the benefits on simpler ones. The finding makes the interaction itself part of the explanation. It is insufficient simply to ask whether AI was present.

There is also contrary evidence which a balanced review needs to retain. Ashkinaze and colleagues' conference paper (2025), developed from a 2024 preprint, reports a dynamic experiment involving 844 participants from 48 countries. Participants encountered ideas from AI and earlier participants before contributing their own. High exposure to AI ideas increased collective diversity without improving individual creativity. This differs from asking a chatbot to supply one's answer. The circulation and combination of examples can produce different effects from repeated dependence on a common source of suggestions.

Taken together, these studies suggest a conditional process. An example can open a possibility or establish an unnoticed boundary. Subsequent activity depends on the diversity of examples, the problem, the interface and what the person does with the material. Fixation also predates generative AI; what is distinctive here is the ease with which a persuasive example can be supplied, repeatedly, to very large numbers of people.

Workplace creativity and the importance of prior resources

Sun and colleagues (2025) conducted a randomised field experiment with 250 employees in a technology consulting firm in China. AI assistance improved supervisor-rated creativity and externally rated novelty. Benefits were greater among employees who actively monitored and regulated their own thinking. Not every outcome was equally secure: the usefulness effect was not statistically significant in the model containing controls. Nevertheless, this study extends the evidence beyond a brief online exercise into an organisational setting.

It would be tempting to place this alongside the benefits for initially less creative writers and conclude that AI either equalises or amplifies existing differences. The findings do not justify such a simple choice. Initial writing performance and metacognitive strategies are different characteristics, measured in different tasks. An employee who knows when an answer is inadequate may benefit from resources which a less experienced user cannot yet evaluate. Access to the same system does not ensure an equivalent capacity to use it.

This raises a social question as well as a cognitive one. If effective use depends on prior knowledge, confidence and the freedom to reject a suggestion, then the consequences will partly reflect existing educational and occupational inequalities. That is an interpretation of the evidence, rather than a causal result established across the field. It identifies something that future research ought to investigate directly: who can use AI to extend their agency, and who is encouraged merely to accept what it supplies?

Using AI and learning to work with AI

A further distinction concerns the difference between using generative AI and learning how to incorporate it into an existing intellectual or creative practice. These are not equivalent activities. Giving someone access to a system and asking them to complete a task establishes relatively little about what they might achieve after developing a more informed and discriminating way of working with it. Consequently, studies comparing users and non-users need to consider what experience participants bring, what preparation they receive and whether the interaction itself changes as they become more familiar with its possibilities and limitations.

My own experience over several weeks suggests that this learning involves more than discovering how to phrase an instruction or "prompt". It includes recognising which questions are worth asking, supplying relevant context, identifying what an answer has overlooked and deciding when to pursue, challenge or abandon or return to develop and extend a line of enquiry. Sometimes an initially unsatisfactory response helps clarify what I was trying to ask. The question develops through the exchange. This is a personal

observation of my own growth with AI rather than evidence of a general effect, but it identifies a process which a single encounter with an unfamiliar tool may fail to capture.

The emerging literature on prompt literacy provides a useful way of examining this process. Tour and Zadorozhnyy (2025) conceptualise engagement with AI as an iterative cycle of constructing a prompt, interpreting and evaluating the response, and reconsidering the prompt. Their discussion draws on literacy theory to connect these activities with meaning, purpose and critical analysis. Although developed for language education, the framework suggests that effective interaction requires judgement throughout the exchange. It is a conceptual contribution, however, rather than an experimental demonstration that such practices necessarily improve creativity.

More directly relevant empirical evidence comes from Urban and colleagues (2025). Their analysis of 77 university students in Prague found that stronger creative problem-solving performance was associated with supplying elaborated context, specifying evaluation criteria and giving the system particular feedback. Higher-performing students engaged in iterative dialogue, monitoring and regulating the generation of ideas, whereas lower-performing students tended to use ChatGPT as an information resource. The associations do not establish that teaching these practices would produce the same improvement: stronger students may already have possessed relevant knowledge or metacognitive capabilities. Nevertheless, the study makes differences in the organisation of interaction visible.

Oppenlaender, Linder and Silvennoinen (2025, first online 2024) similarly treat prompting for AI-generated art as a skill requiring acquisition. Across three studies, participants could recognise aspects of prompt quality but encountered difficulties in producing effective prompts, including limitations in their vocabulary for visual styles. Familiarity with ordinary language therefore does not ensure competence in directing a generative system. Nor, however, should repeated use automatically be equated with learning. Misiejuk and colleagues (2026), examining prompting across four assignments, report a tendency towards minimally contextualised, low-effort prompts. Practice can consolidate an unproductive habit as well as develop a capability.

What I have described as developing a "rapport" with AI may capture another aspect of this experience. I use the term cautiously, to describe an increasingly workable pattern of exchange rather than to assume reciprocal human understanding. Through explanations, corrections and examples, I make my expectations more explicit whilst learning how to interpret and redirect the responses. Continuity can reduce the need to begin each enquiry afresh. Yet familiarity may also encourage unwarranted trust, and a response which sounds increasingly congenial need not become increasingly reliable. A productive rapport would therefore have to accommodate disagreement and verification.

For research on creativity, the implication is that "AI use" is too coarse a category to carry the explanatory weight sometimes placed upon it. We need comparisons between different histories and forms of engagement, including supported learning over time. We also need to distinguish improvements in directing AI from improvements in independent creative capacity. Learning to work effectively with a system is an achievement worth investigating, but it does not, by itself, settle what happens to the person's thinking when the system is absent.

Education and mathematical creativity

Liu, Zuo, Feng and Lu (2025) compared human–ChatGPT, human–Google and human–human collaboration in a quasi-experiment involving 230 Chinese university students. The ChatGPT group produced the most fluent and flexible mathematical solutions, whereas human pairs produced the most original solutions. The ChatGPT group also overestimated its performance. This is particularly helpful evidence because it separates dimensions that a single overall creativity score might conceal. As a quasi-experiment in one educational context, it warrants more cautious causal interpretation than a well-controlled randomised trial.

For mathematics education, the distinction has practical substance. Producing several solution methods may be valuable, but it does not necessarily show that a student can explain why they work, choose between them or adapt them when the problem changes. A finished solution can conceal the source of the mathematical decisions. Conversely, an initially awkward human exchange may involve substantial conceptual development even if its final written product is less polished.

The educational question therefore concerns the organisation of activity. We need to know whether learners formulate conjectures, challenge assumptions and account for their decisions, and whether they retain those capacities afterwards. AI could be incorporated into such an enquiry, for example by supplying a solution for students to examine. It could also remove the very work through which understanding develops. These are plausible pedagogical differences, not a claim that one prescribed sequence has already been shown to work universally.

Practising writers artists and comedians

Research with established practitioners adds dimensions that short performance tests often overlook. Ippolito and colleagues' Wordcraft white paper (2022) followed 13 published writers over eight weeks through interviews and journals. Participants identified uses for brainstorming and developing details, but also problems with distinctive voice, control and understanding a story. The small qualitative study is not evidence of an average treatment effect. Its value lies in showing what experienced writers actually need from a tool and how those needs differ.

Mirowski and colleagues (2024) worked with 20 professional comedians in workshops at the Edinburgh Festival Fringe and online. Participants often regarded the generated material as bland and found that moderation and dominant cultural assumptions could obstruct satire and minority perspectives. Some nevertheless found details that stimulated further work. Since participants already used AI creatively and were recruited through relevant networks, their accounts should not be generalised to all comedians. They do show that creative usefulness includes the ability to challenge conventions, not simply to produce fluent material.

Zhou and Lee (2024) examined more than four million artworks from over 50,000 users on an art-sharing platform. AI adoption was associated with increased production and favourable audience responses, alongside declining average content novelty and visual novelty. Maximum content novelty could nevertheless increase. The scale and natural setting are strengths, although adoption was not randomly allocated and platform responses are a particular measure of value. More work and more approval need not mean a broader artistic culture.

These studies invite attention to the process of selection. A practitioner may generate many possibilities and reject almost all of them. Another may accept the first coherent result. A count of finished artefacts will not reveal the difference. Nor will it tell us whether a person's distinctive voice has become easier to express or more difficult to maintain.

Whose culture is being reproduced

Agarwal, Naaman and Vashistha (2025) conducted a controlled experiment with 118 participants from India and the United States undertaking culturally grounded writing tasks in English. AI suggestions provided greater efficiency gains for American participants and shifted Indian participants' writing towards Western patterns, affecting both content and expression. This is evidence of cultural convergence within a particular language and tool setting. It should not be treated as a description of all Indian writers or all models.

Qadri and colleagues (2024) approached evaluation through dialogue with three artists and three commentators connected with Persian Gulf art worlds, especially Iranian contexts. Their workshop paper draws on Howard Becker's conception of art worlds to examine creative practice as socially situated. The participants' residences cannot simply

be equated with the cultural region under discussion. Their work suggests that evaluation should include historical knowledge, representation and the purposes artists themselves bring to the technology.

From this perspective, originality cannot be adequately understood as distance between strings of words. An image may look unfamiliar to an external assessor whilst reproducing a familiar stereotype for those represented. A polished description may erase precisely the local distinctions that gave an account its significance. The question of whose judgement becomes authoritative has something in common with Bourdieu's concern with the social organisation of cultural value. That is a useful interpretative connection; these studies do not themselves test a Bourdieusian explanation.

The geographical concentration of research consequently matters twice. It limits the populations to which findings can be generalised, and it can shape the standards by which creativity is recognised. An English-language study with online volunteers offers evidence about those participants under those conditions. It cannot stand without qualification for creative expression across languages, communities and institutions.

Can AI appear creative without making people more creative

Koivisto and Grassini (2023) compared human responses with chatbot responses on an alternative uses task. Chatbots performed well on average, whilst the strongest human responses remained competitive. The article has a published correction, which should be consulted with the original. This comparison concerns performance on a test, rather than the effect of AI use on human development. It is therefore relevant to what a system can produce, but cannot establish either learning or deterioration among users.

Porter and Machery (2024) found that non-expert readers had difficulty identifying AI-generated poetry and rated it favourably compared with poems by established poets. Ratings were also affected by what participants were told about authorship. Such findings demonstrate the importance of audience and framing. They do not settle whether accessibility on a first reading captures the value of a poem, or whether the process that produced it resembles human artistic activity.

The distinction is especially important when a model is used to judge another model's work. Agreement between an automated score and a human rating may be useful, but neither should be treated as an uncontested account of creativity. A method can reliably reward a restricted kind of accomplishment. We still need to examine what it leaves out.

Dependence and what remains when assistance is removed

Liu, Zhou, Huang and Li's preprint (2024) followed 61 university students through a seven-day experiment and a later follow-up. Performance improved during five days with ChatGPT, but the advantage was absent when participants worked without it on day seven and at follow-up. Greater similarity in outputs persisted. The authors describe performance as reverting to baseline; the results do not establish a general fall below participants' original ability. The research team was based in China, but the report does not explicitly identify the university where participants were recruited.

The distinction is not pedantic. Losing an assisted advantage, failing to learn a transferable skill and suffering a deterioration in an existing skill are different findings. Establishing the last requires a credible counterfactual for what would otherwise have happened, comparable assessments and enough follow-up to distinguish an adjustment from a sustained change. A month is informative, but it is not a history of dependence over several years.

Lee and colleagues (2025) surveyed 319 “knowledge workers” (i.e. people who “think for a living”) from several countries, collecting accounts of AI use. Greater confidence in AI was associated with less reported critical thinking, and the work described shifts towards checking and integrating outputs. This is evidence about perceived effort and practice.

Self-report and the observational design limit causal inference; critical thinking is also related to, but not interchangeable with, creativity.

Kosmyna and colleagues' MIT-associated preprint (2025, revised December 2025) compared essay writing with a language model, web search or no tools. Fifty-four participants completed the first three sessions and 18 completed the fourth. The AI group showed weaker measured EEG connectivity, poorer recall and lower reported ownership. The findings merit investigation, but an EEG difference is not a direct measure of creative deterioration. Stankovic and colleagues' preprint commentary (2025/2026) raises concerns about sample size, EEG analysis, reproducibility and reporting. Both the original and the criticism require scrutiny; neither a striking title nor a neuroscience measure removes the need for methodological judgement.

Human collaboration provides a demanding comparison

A recent paper by Lai and colleagues, available online in 2026 and assigned to a 2027 journal issue, directly compares human–human and human–AI collaboration on alternative uses and short story tasks. Its abstract reports 68 human pairs and 72 human–AI pairs; the latter include an AI partner, so these are not equivalent counts of human participants. The authors report poorer human creative responses with AI partners and suggest that verbosity can account for an apparent AI advantage. These claims are included at abstract level here, with the participant country unconfirmed.

This comparison broadens the original question. The alternative to AI need not be an isolated individual. It may be discussion with another person who misunderstands, objects, recalls a different experience or changes the direction of the enquiry. If a study compares AI with unaided solitary work, it answers a different question from a comparison with sustained human collaboration. Both comparisons have value, but they should not be silently substituted for one another.

What the reviews add

Heigl's systematic review, first published online in 2025 and allocated to a 2026 issue, covers 64 studies across creative contexts. It maps adoption, processes, outcomes and perceptions, and identifies long-term and socioeconomic questions requiring further research. Its breadth is useful precisely because the field is larger than the question of whether a chatbot can score highly on a creativity test.

Holzner, Maier and Feuerriegel's meta-analysis preprint (2025) combines 28 studies involving 8,214 participants. It reports a modest average advantage for people collaborating with AI in creative performance, Hedges' $g = 0.27$, alongside reduced diversity, $g = -0.86$. These are standardised differences, not percentage changes, and the full participant total does not apply separately to every outcome. The pooled estimates summarise heterogeneous tasks and should not be read as a universal effect on people.

The peer-reviewed review by de Rooij and Biskjaer (2026) synthesises 19 studies and 61 effect sizes. It finds a small overall homogenisation effect, with stronger convergence in more constrained ideation tasks. Its related conference publication reports $d = 0.334$; that publication and the journal article belong to the same research programme and should not be counted as independent replications. The difference from the earlier meta-analysis makes attention to inclusion criteria, measures and task composition essential before comparing effect sizes.

Caprioli's review (2026) distinguishes consequences for creative outcomes from consequences for the process. It identifies timing, interaction design, task characteristics and user expertise as important conditions. This is a useful organising account, rather than evidence that a single preferred method has been validated across contexts. Reviews also overlap in the studies they include. Adding their sample totals together would give an inflated impression of the independent evidence.

What can be concluded and what still needs investigation

The strongest conclusion from this review is conditional. Generative AI can improve assessed creative products in some circumstances. It can also encourage fixation, reduce originality within a task, or narrow the range of outputs across users. These effects can coexist because they concern different outcomes. Evidence of sustained deterioration in people's independent capacity remains more limited than evidence about immediate performance and convergence.

The methodological weaknesses are not all the same. Random allocation strengthens causal inference about a particular intervention, but cannot make a short task representative of a lifetime of creative practice. A large observational dataset captures activity in a natural setting, yet selection into AI use may remain consequential. Interviews recover intentions and frustrations that a score misses, but cannot establish population-wide frequencies. These however are the standard limitations of quantitative and qualitative research. A useful review needs these forms of evidence in conversation rather than arranged in a simple hierarchy in which one method answers every question.

Future research should compare different patterns of use, include delayed tasks without assistance and follow people for substantially longer periods. It should distinguish novices from experienced practitioners and investigate languages and communities that existing research represents poorly. It should examine what people reject as well as what they accept, and whether improved production is accompanied by greater understanding, ownership or freedom to pursue their own questions. The effects of institutional pressure to produce more quickly also warrant attention; choosing to use a tool is different from being required to organise one's work around it.

A provisional judgement based on this review is that the risk lies partly in allowing a readily available response to decide the shape of the enquiry before the enquiry has properly begun. Developing an initial account, testing it against alternatives and retaining opportunities to work unaided are reasonable precautions. They are not a guarantee, nor should they become a ritual which substitutes for examining the quality of the work. The central issue is how far the person remains able to formulate, pursue and defend a creative judgement of their own.

Geographical register of empirical research

The table below distinguishes participant or data location from the institutional base of the research teams. “Unconfirmed” means that the location was not established from the accessible material, not that the authors necessarily omitted it. Cultural identity, nationality, residence and the venue of a conference are different things. The table should therefore be read as an evidence map rather than as a ranking of national research.

Study	Participant or data setting	Design and geographical qualification
Agarwal et al. 2025	India and USA; 118 participants	Controlled English-language writing experiment; US-based team.
Ashkinaze et al. 2025	48 countries; 844 participants	Largest response shares from USA, Canada, Germany, UK and Australia; US-based team.
Cheng and Zhang 2025	Participant country unconfirmed	China-based team; experiments vary task complexity and model response constraints.
Doshi and Hauser 2024	UK; 293 writers	UK-resident Prolific sample; UK-based team; randomised story-writing experiment.

Study	Participant or data setting	Design and geographical qualification
Ippolito et al. 2022	Multiple countries; breakdown unconfirmed	13 published writers; eight-week commissioned writing study; Google Research.
Koivisto and Grassini 2023	Online Western-country sample; UK and USA included	Human–chatbot performance comparison; team based in Finland and Norway; see published correction.
Kosmyna et al. 2025	US/MIT-associated laboratory study	54 participants in sessions 1–3; 18 in session 4. Exact catchment unconfirmed; EEG and essay outcomes.
Lai et al. online 2026	Participant country unconfirmed; USA institutional record	68 human pairs and 72 human–AI pairs. Abstract-level assessment; 2027 issue assignment.
Lee and Chung 2024	Participant country unconfirmed	Five experiments; US-based team. Do not infer residence from author affiliations.
Lee et al. 2025	Multinational; 319 workers	Largest country groups: UK 37, Canada 25, USA 20, South Africa 18, Poland 17. Team: UK/USA.
Liu Q. et al. 2024	University location unconfirmed; China-based team	61 students, seven-day experiment and 30-day follow-up; preprint.
Liu Z. et al. 2025	China; 230 university students	Quasi-experiment in mathematical creativity; Yangzhou University team.
Meincke et al. 2025	Same data as Lee and Chung	Reanalysis; no new fieldwork. Author affiliations: USA and Germany.
Mirowski et al. 2024	UK venue and online; 20 comedians	Ten at Edinburgh Fringe; ten across online workshops. Venue does not establish nationality.
Porter and Machery 2024	Online sample; residence unconfirmed here	US-based team; experiments on poetry identification and evaluation, not user skill development.
Qadri et al. 2024	Persian Gulf/Iranian cultural contexts; residences unspecified	Three artists and three commentators; US and UK professional links; workshop/track paper.
Sun et al. 2025	China; 250 consulting employees	Randomised field experiment in one firm. Team affiliations: USA, China and Singapore.

Study	Participant or data setting	Design and geographical qualification
Wadinambiarachchi et al. 2024	Australia; 60 participants	Melbourne laboratory context and university recruitment; conference venue was Hawaii, not fieldwork location.
Zhou and Lee 2024	Art-sharing platform; country distribution unconfirmed	Over four million artworks and 50,000 users; US-based team; observational design.

References and direct article links

Author and year citations in the main text link to the relevant article or repository record. The links below provide the full reference and, where available, an accessible manuscript. Journal access may still depend on a subscription. Publication status and geographical notes are included to make the basis of the review open to inspection.

Agarwal, D., Naaman, M. and Vashistha, A. (2025). AI Suggestions Homogenize Writing Toward Western Styles and Diminish Cultural Nuances. Proceedings of CHI 2025, article 1117, pp. 1–21.

Peer-reviewed conference paper. Participant countries: India and USA. Team: USA.

Ashkinaze, J., Mendelsohn, J., Qiwei, L., Budak, C. and Gilbert, E. (2025). How AI Ideas Affect the Creativity, Diversity, and Evolution of Human Ideas: Evidence From a Large, Dynamic Experiment. Proceedings of the Collective Intelligence Conference.

Peer-reviewed conference paper; earlier preprint in 2024. Team: USA. Multinational participants.

Caprioli, S. (2026). Collaborating with GenAI: A blessing or a curse for creativity? Current Opinion in Psychology, 73, 102425.

Published review, online 7 September 2026. Author based in Germany; synthesis rather than a single-country experiment.

Cheng, X. and Zhang, L. (2025). Inspiration booster or creative fixation? The dual mechanisms of LLMs in shaping individual creativity in tasks of different complexity. Humanities and Social Sciences Communications, 12, 1563.

Peer-reviewed journal article. Team: Renmin University of China. Participant country not independently confirmed here.

de Rooij, A. and Biskjaer, M. M. (2026). Does generative AI make us think alike? A systematic review and meta-analysis of homogenisation effects in human–AI co-creation. Behaviour & Information Technology, advance online publication, 31 August.

Peer-reviewed review of 19 studies. Netherlands-linked research team; international evidence. Related conference report: Generative AI Makes Creative Output More Homogeneous, DOI 10.1145/3822301.3822304; not an independent replication.

Doshi, A. R. and Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. Science Advances, 10(28), eadn5290.

Peer-reviewed journal article. Participant and team country: UK.

Heigl, R. (2026; first online 2025). Generative artificial intelligence in creative contexts: a systematic review and future research agenda. Management Review Quarterly, 76, pp. 955–992.

Peer-reviewed systematic review of 64 studies. Author: Goethe University Frankfurt, Germany. International literature; online and issue years differ.

Holzner, N., Maier, S. and Feuerriegel, S. (2025). Generative AI and Creativity: A Systematic Literature Review and Meta-Analysis. arXiv:2505.17241.

Preprint version checked. Team: LMU Munich, Germany. International synthesis of 28 studies; 8,214 participants across included studies.

Ippolito, D., Yuan, A., Coenen, A. and Burnam, S. (2022). Creative Writing with an AI-Powered Writing Assistant: Perspectives from Professional Writers. arXiv:2211.05030.

White paper/preprint version checked. Google Research; writers from multiple countries and backgrounds. Country-by-country residence breakdown not established.

Koivisto, M. and Grassini, S. (2023). Best humans still outperform artificial intelligence in a creative divergent thinking task. Scientific Reports, 13, 13601.

Peer-reviewed journal article; corrected in February 2024, DOI 10.1038/s41598-024-54745-y. Team: Finland and Norway. Online participants rather than a Finnish or Norwegian national sample.

Kosmyna, N., Hauptmann, E., Yuan, Y. T., Situ, J., Liao, X.-H., Beresnitzky, A. V., Braunstein, I. and Maes, P. (2025). Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task. arXiv:2506.08872, version 2, 31 December.

Preprint version checked, not treated as a peer-reviewed creativity trial. US/MIT-associated study; exact recruitment catchment not independently verified here.

Lai, C. K. Y., Luchini, S. A., Lauharatanahirun, N. and Beaty, R. (online 2026; issue 2027). Creative or uncreative partner: Comparing humans and AI in collaborative creative tasks. Computers in Human Behavior, 187, 109172.

Peer-reviewed publication record and abstract checked; issue assigned to February 2027. Penn State institutional record: USA. Participant country not confirmed; included as online-available research, not a completed 2027 study.

Lee, B. C. and Chung, J. (2024). An empirical investigation of the impact of ChatGPT on creativity. Nature Human Behaviour, 8, pp. 1906–1914.

Peer-reviewed journal article. Team: University of Houston and Rice University, USA. Participant country not independently established from accessible material.

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R. and Wilson, N. (2025). The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers. Proceedings of CHI 2025, article 1121.

Peer-reviewed conference paper. Team: UK and USA; multinational respondents.

Liu, Q., Zhou, Y., Huang, J. and Li, G. (2024). When ChatGPT is gone: Creativity reverts and homogeneity persists. arXiv:2401.06816.

Preprint version checked. Team: Peking University and Beijing Yuxin Technology Company, China. University recruitment location is not explicitly identified.

Liu, Z., Zuo, H., Feng, J. and Lu, Y. (2025). Who Is the Best Helper for University Students' Mathematical Creativity? A Quasi-Experimental Study of Human–ChatGPT, Human–Google, and Human–Human Co-Creation. Computer Applications in Engineering Education, 33(6), e70100.

Peer-reviewed journal article; abstract and publication record checked. Participants: China. Team: Yangzhou University, China.

Meincke, L., Nave, G. and Terwiesch, C. (2025). ChatGPT decreases idea diversity in brainstorming. Nature Human Behaviour, 9, pp. 1107–1109.

Peer-reviewed Matters Arising reanalysis of Lee and Chung's datasets. Team: USA and Germany. Not a new participant sample.

Mirowski, P. W., Love, J., Mathewson, K. W. and Mohamed, S. (2024). A Robot Walks into a Bar: Can Language Models Serve as Creativity Support Tools for Comedy? An

Evaluation of LLMs' Humour Alignment with Comedians. Proceedings of ACM FAccT 2024.

Peer-reviewed conference paper. Workshops: Edinburgh, UK, and online. Mixed participant backgrounds; national residence breakdown not established.

Misiejuk, K., López-Pernas, S., Kaliisa, R. and Saqr, M. (2026). Cognitive offloading in student–AI collaboration: A longitudinal analysis of prompting strategies. Computers in Human Behavior Reports, 22, 101130.

A useful qualification to the argument that experience necessarily develops expertise. Participant country not confirmed from the material checked.

Oppenlaender, J., Linder, R. and Silvennoinen, J. (2025; online 2024). Prompting AI Art: An Investigation into the Creative Skill of Prompt Engineering. International Journal of Human–Computer Interaction, 41(16), 10207–10229.

Authors based in Finland and the USA; these affiliations should not be mistaken for the countries of the crowdsourced participants.

Tour, E. and Zadorozhnyy, A. (2025). Conceptualizing and Operationalizing Prompt Literacy for English Language Learners. Journal of Adolescent & Adult Literacy, 69, e70020.

A scholarly discussion article, with authors based in Australia and Hong Kong, rather than a participant study. Particularly relevant to an account of learning through dialogue.

Porter, B. and Machery, E. (2024). AI-generated poetry is indistinguishable from human-written poetry and is rated more favorably. Scientific Reports, 14, 26133.

Peer-reviewed journal article. Team: University of Pittsburgh, USA. Participant residence not independently confirmed in the accessible primary material.

Qadri, R., Mirowski, P., Gabriellan, A., Mehr, F., Gupta, H., Karimi, P. and Denton, R. (2024). Dialogue with the Machine and Dialogue with the Art World: Evaluating Generative AI for Culturally-Situated Creativity. NeurIPS 2024 Creative AI Track; arXiv:2412.14077.

Workshop/track paper; assessed as a small qualitative and methodological contribution. Persian Gulf/Iranian cultural focus; not an Iran-resident survey.

Runco, M. A. and Jaeger, G. J. (2012). The Standard Definition of Creativity. Creativity Research Journal, 24(1), pp. 92–96.

Conceptual article, preceding current generative AI. Authors based at the University of Georgia, USA; no experimental participant country.

Stankovic, M., Hirche, E., Kollatzsch, S. and Doetsch, J. N. (2025/2026). Comment on: Your Brain on ChatGPT: Accumulation of Cognitive Debt When Using an AI Assistant for Essay Writing Tasks. arXiv:2601.00856.

Methodological commentary preprint; submitted 29 December 2025, January 2026 identifier. No new participant sample; author countries not established here.

Sun, S., Li, Z. A., Foo, M.-D., Zhou, J. and Lu, J. G. (2025). How and for whom using generative AI affects creativity: A field experiment. Journal of Applied Psychology, 110(12), pp. 1561–1573.

Peer-reviewed journal article. Fieldwork: China. Authors affiliated with institutions in the USA, China and Singapore.

Urban, M. et al. (2025). Prompting for creative problem-solving: A process-mining study. Learning and Instruction, 99, 102156.

Research with 77 students in Prague, Czech Republic. The closest empirical match to the distinction herein identified.

Wadinambiarachchi, S., Kelly, R. M., Pareek, S., Zhou, Q. and Velloso, E. (2024). The Effects of Generative AI on Design Fixation and Divergent Thinking. Proceedings of CHI 2024.

Peer-reviewed conference paper. Laboratory research and team: Melbourne, Australia.

Zhou, E. and Lee, D. (2024). Generative artificial intelligence, human creativity, and art. PNAS Nexus, 3(3), pgae052.

Peer-reviewed observational platform study. Team: Boston University, USA. Platform users are not a US national sample; country distribution unconfirmed.