

# Assistance, Not Usurpation: LLMs as Catalysts for Human Originality

## Paper type: Position Paper

**Roberto Gallotta**  
Institute of Digital Games  
University of Malta  
Malta  
roberto.gallotta@um.edu.mt

**Antonios Liapis**  
Institute of Digital Games  
University of Malta  
Malta  
antonios.liapis@um.edu.mt

**Georgios N. Yannakakis**  
Institute of Digital Games  
University of Malta  
Malta  
georgios.yannakakis@um.edu.mt

### Abstract

While prior generations of AI served primarily as functional tools, today’s Large Language Models (LLMs) function as generative decision-making assistants across diverse creative fields. This paper examines the critical risks accompanying this advancement, namely the accumulation of cognitive debt, the homogenization of ideas, and workflow friction stemming from prompt sensitivity and ambiguous user requirements. In this position paper we argue that the goal of LLM integration is not full automation, which carries these risks, but enhanced collaboration. To this end, LLMs must be intentionally designed as semi-autonomous partners operating within structured, human-centric workflows, so as to mitigate LLM’s limitations and prioritize the amplification of human originality. This paper outlines design imperatives and research directions for the computational creativity community to realize this vision.

### Introduction

Historically, Artificial Intelligence (AI) tools served primarily utilitarian roles, automating repetitive tasks and optimizing existing constraints (Colton and Wiggins 2014). While useful, these systems lacked the creative capacity and contextual awareness necessary to meaningfully contribute during the ideation phase of creative work (Liu 2025; Kadenhe, Al Musleh, and Lompot 2025). Such systems were often confined to a problem-solving paradigm, whereas the frontier of the computational creativity field has aimed to generate artifacts (Colton and Wiggins 2014). The advent of Large Language Models (LLMs), however, has disrupted this landscape entirely. LLMs, powered by unprecedented scale and training data, now function as generative collaborators across creative domains, from writing and design to multimedia and scientific discovery (Kadenhe, Al Musleh, and Lompot 2025). This shift in the AI landscape forces computational creativity researchers to reconsider fundamental assumptions about human-AI creative partnership, particularly concerning the distribution of autonomy, the locus of creative authority, and the role of human agency in generative collaboration.

The field of computational creativity has historically emphasized understanding creative processes: how humans generate, refine, and judge ideas. Contemporary LLM deployment, often characterized as a shift from predictable,

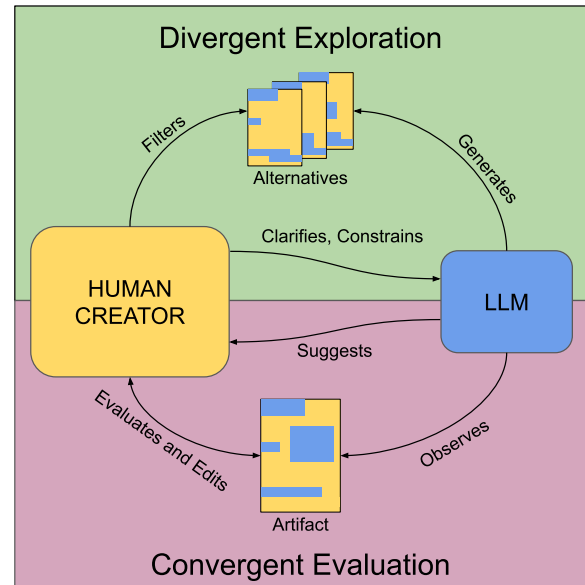


Figure 1: Our vision of LLMs as catalysts for human originality: LLMs generate alternatives that users can select and evaluate, following specifications provided by the user. Users retain authority over final decisions, refining selected ideas through iterative feedback loops with the LLM. All outputs are documented to ensure shared authorship.

rule-based systems to opaque, probabilistic models, from “symbolic caterpillars” to “stochastic butterflies” (Veale 2024), frequently sidesteps these considerations in favor of rapid automation. The urgency of discussing the risks LLMs bring in creative contexts cannot be overstated. Users who passively rely on LLM outputs suffer from cognitive debt, where diminished originality persists even after LLM use ceases (Kosmyna et al. 2025; Kumar et al. 2025). The systems themselves introduce workflow friction: their extreme sensitivity to prompt variations and opaque reasoning processes often create conflicts that undermine collaborative intent (Huang et al. 2025a). These are not merely technical limitations. They instead reflect fundamental tensions between LLMs capabilities and the requirements of authentic creative practices.

In light of these dangers, we argue that LLMs must be reoriented from fully autonomous systems to semi-autonomous collaborators embedded within structured, human-centric workflows. This perspective aligns with computational creativity’s longstanding commitment to move beyond “fire-and-forget” approaches toward mixed-initiative systems that distribute autonomy, authority, and initiative between human and machine. Computational creativity research has formalized multiple modes of human-computer collaboration, from alternating co-creativity (Kantosallo and Toivonen 2016) to task-divided approaches (Tubb and Dixon 2014), each designed to clarify roles and optimize the balance between human and computational contributions (Lin et al. 2023). This paper aims to articulate how this principle can be realized in practice, leveraging insights from both computational creativity research and recent studies on human-AI collaboration. By grounding LLMs within structured creative processes, we can mitigate their inherent limitations while amplifying human originality rather than replacing it. A diagram overview of our vision is shown in Figure 1.

This position paper focuses on text-based LLMs, setting aside the distinct landscape of multimodal large language models (MLLMs) that accept and generate diverse modalities including images, audio, and video. While existing work explores creativity and MLLMs (Huang et al. 2025b), the unique affordances and limitations of multimodal systems warrant separate investigation beyond the scope of this work.

## The Creative Ceiling of LLMs

Creativity has long been defined as the ability to generate artifacts that are novel, surprising, and valuable (Boden 2004). The conceptualization of creativity, however, varies depending on the creative persona or perspective one adopts, from the individualist creator to the collective practitioner (Johnson 2012). These are products of human cognitive processes such as conceptual blending and analogical reasoning, leading to genuinely novel insights. This definition reflects a commitment to understanding the mechanisms underlying creative thought, not merely automating its outputs. Despite their impressive surface capabilities, LLMs exhibit fundamental limitations that prevent them from reliably engaging in these processes. These limitations are not superficial but architectural, stemming directly from how these models encode, process, and represent meaning. Novelty, depth, and subjectivity are three critical limitations directly impacted by these characteristics.

LLMs operate via next-token prediction, selecting the most probable continuation of text. When fine-tuned with Reinforcement Learning from Human Feedback (Ouyang et al. 2022), they are further optimized toward preferred, conventional outputs, to the detriment of creative diversity (Padmakumar and He 2023). This mechanism creates an insurmountable novelty ceiling: LLMs bias toward high frequency patterns in training data, unable to reliably generate the surprising, low-probability combinations that characterize genuine creative work (Peeperkorn, Brown, and Jor-

danous 2023; Špela Vintar and Javoršek 2025). The inability of LLMs to break out of probabilistic modes leads to generation of homogeneous, underdeveloped ideas, less original and more predictable (Yang et al. 2025).

Creativity rooted in subjective experience, such as poetry, personal narratives, and emotionally resonant design requires episodic memory and individuated perspective, which LLMs cannot convincingly emulate (Russell et al. 2025). As LLMs learn statistical regularities between word tokens, and not relationships between language and lived experience, they lack genuine grounding (Chakrabarty et al. 2024; Ismayilzada et al. 2025; Kumar et al. 2025). They consequently adhere to paradigmatic patterns, failing to capture nuance, contradiction, and the idiosyncratic perspectives that distinguish human creativity. In the case of creative writing, LLMs overuse clichés and tropes, often generating dialogue that lacks subtext and is instead direct and expository, which feels unnatural (Chakrabarty et al. 2024). The homogeneity of LLM outputs, often referred to as “AI slop” when mass produced with the goal of driving engagement, is a direct consequence of these architectural limitations, not a tuning problem (Hern and Milmo 2024). Where human creatives draw on lived experience to produce singular, unrepeatably voices, LLMs synthesize patterns across thousands of similar examples, converging toward centrality rather than distinction. This homogeneity stands in sharp contrast to the authenticity and perceived legitimacy that humans attribute to creative works grounded in genuine experience (Colton, Pease, and Saunders 2018).

Computational creativity systems must not only generate but also evaluate creative merit (Colton and Wiggins 2014). Evaluating creativity requires understanding the viewer-artifact interaction: how humans extract novel and useful perspectives from artifacts they encounter (Indurkha 2012). While LLM-generated content may satisfy the average user, it almost never passes expert validation on creativity tests when compared to human professionals. This has been shown by Alharthi (2025) with stories evaluated with the Torrance test of creative thinking (Torrance 1968). Whether this can be generalized to other domains is an open question. More critically, LLMs lack the genuine understanding necessary for authentic critical judgment (Avlonitou and Papadaki 2025; Chakrabarty et al. 2024). They cannot reliably apply the evaluative standards that experienced creators use to refine their work. This dual limitation means that human judgment must remain authoritative, particularly during the convergent phases of creative work where ideas are refined and selected (Ruan et al. 2025).

## Pitfalls of Autonomous LLMs in Creativity

The breadth of LLM applicability is precisely what makes unconstrained automation dangerous in creative contexts. While full automation promises efficiency, it obscures profound cognitive, procedural, and epistemic costs for humans that undermine both immediate creative outcomes and long-term creative capacity (Khurana et al. 2025). We discuss four major pitfalls of deploying LLMs as fully autonomous systems: human cognitive and creative costs

of *over-reliance*, *systemic instability* arising from ambiguous design contexts, the *erosion of user agency* in creative decision-making, and the *limitations of the agentic paradigm*, which relies even less on human involvement and oversight.

### **Over-Reliance**

When users interact with LLMs through unscaffolded conversational interfaces, shallow engagement is the predictable outcome (Kumar et al. 2025). Users passively overly accept LLMs suggestions without critical evaluation (Yang et al. 2025), missing the deliberation and refinement essential to iterative creative practice (Choudhury and Chaudhry 2024; Kim et al. 2025), with the risk of behavioral drift (Lopez-Lopez et al. 2026). This dynamic mirrors the phenomenon of production blocking in group creativity, where the presence of external structure can inhibit individual ideation (Sosa and Gero 2013). Negative effects of passive reliance on LLMs extend beyond the immediate assistance period. Exposure to LLM assistance measurably reduces users' subsequent originality and idea diversity (Kumar et al. 2025). This diminished originality persists even after users discontinue using LLMs, suggesting that the cognitive scaffolding of such systems may produce lasting, detrimental imprints on creative processes. Users appear to internalize the paradigmatic, high-frequency patterns produced by LLMs, losing access to their distinctive creative voices. The "Do It For Me vs. Do It With Me" study (Khurana et al. 2025) confirms this finding, showing that users strongly prefer semi-autonomous systems with explicit control, especially in exploratory, creative tasks. Full automation in early exploratory phases measurably reduces control, learnability, and perceived utility (Noy and Zhang 2023; Shneiderman 2020).

### **Systemic instability**

Beyond cognitive costs, the automation of LLMs in creative workflows may introduce structural vulnerabilities and instability. Creative work involves emergent goals, vague specifications, and constraints that materialize during the process itself (Grace, Gero, and Saunders 2013; Suwa, Purcell, and Gero 1998). While humans naturally update requirements as needed, LLMs may respond unpredictably, as they are extremely sensitive to even minor prompt variations (Huang et al. 2025a). Outputs shift across iterations, resting on contradictory assumptions about the problem space. Designers investing in LLM-generated artifacts often discover too late that fundamental interpretations diverged, requiring extensive revision and leading to frustration (Huang et al. 2025a).

### **Erosion of User Agency**

The misalignment between designer intent and LLM interpretation becomes nearly impossible to diagnose given the opacity of LLMs reasoning, rendering it essentially a very powerful "black box" that erodes trust (Benk et al. 2025; Lynch et al. 2025). LLMs introduce errors through either hallucinations (Hicks, Humphries, and Slater 2024),

which can lead to the generation of non-existent or irrelevant relationships, or simply because they struggle to maintain a comprehensive understanding of all the details of multiple components in complex systems (Ruan et al. 2025). Rather than accelerating work, autonomous LLM deployment can create the illusion of progress while introducing costly, cascading failures (Becker et al. 2025; Neumann and Singh 2025).

### **Limits of the Agentic Paradigm**

Contemporary research into agentic LLMs focuses on systems capable of autonomous goal-setting, planning, and iteration without human intervention. Recently, some emerging platforms have begun experimenting with LLM agents operating in shared social environments where they interact and influence one another (Edwards 2026). This represents a substantial departure from previous paradigms as it invokes systems that could, in principle, generate emergent collective behaviors grounded in social interaction (Gubenko, Lubart, and Houssemand 2022; Andrus and Fulda 2022).

However, current implementations of agentic LLMs remain far removed from realizing this potential. Multi-agent systems designed to simulate collective intelligence or consensus reasoning frequently suffer from significant practical drawbacks, including prohibitive computational cost, high inference latency, and persistent process loss (Pan et al. 2025). More crucially, the LLMs in these systems frequently exhibit conversational stagnation and redundancy, converging repeatedly on overlapping concepts or aligning to the most "intelligent-sounding" agent (Rathi, Jurafsky, and Zhou 2025). They consistently fail to generate an emergent collective performance that exceeds the sum of their individual components. Rather than amplifying LLM intelligence and creativity, fully automated multi-agent approaches often amplify inefficiency and computational waste. Systems such as BILLY (Pai et al. 2025) attempt to overcome these drawbacks by capturing the benefits of collaboration within a single model via activation steering, significantly reducing inference times and costs. Yet even these optimizations remain fundamentally bound by the architectural limitations discussed previously.

### **Regulatory, Economic, and Ethical Challenges**

As LLMs become increasingly integrated into creative practices, their rapid technological advancement raises urgent regulatory, legal, and social questions that demand immediate attention (Independent 2019). These concerns extend beyond technical and cognitive dimensions, implicating issues of governance, intellectual property, equity, and labor.

The rapid adoption of generative AI has created unprecedented urgency for governments and institutions to establish effective regulatory responses (Alharthi 2025). Policymakers face the challenge of developing flexible regulatory frameworks that balance public safety, innovation, and public interest. However, current frameworks such as the EU AI Act (Madiaga 2024) may prove too rigid for the dynamic and rapidly evolving nature of generative AI systems (Alharthi 2025). The pace of AI innovation systematically outpaces

institutional capacity for deliberation and response, raising concerns whether regulations can meaningfully constrain AI systems when technological change is so rapid (Lamb and Brown 2023).

Unlike previous AI systems where authorship of the final product could be linked directly to the user, generative AI blurs the attribution line (McCormack, Gifford, and Hutchings 2019). This epistemic crisis regarding the evidential status of creative works directly challenges copyright systems designed for an era of human authorship (Kantosalo 2020). Legal debates increasingly focus on whether unlicensed use of copyrighted works for training large models qualifies as “fair use”, and whether permitting such training destabilizes existing licensing markets for human creators (Hwang, Shin, and Lee 2025). Policymakers are increasingly urged to introduce regulations mandating transparency regarding AI training data consent, attribution practices, and accountability mechanisms designed to prevent exploitation and digital forgery (Avlonitou and Papadaki 2025; Kyi et al. 2025).

Beyond regulatory and legal concerns lies a deeper critique rooted in political economy. The deployment of foundation models can be framed as an automated continuation of neoliberal austerity politics. These systems function to reduce public expenditure by systematically replacing socially embedded practices with privatized, efficiency-oriented, and market-driven operations (Koch and Zanchetta 2025). This perspective situates LLMs within a longer history of technological systems that reinforce existing power asymmetries and consolidate resources among a small number of technology corporations. The push toward full automation of creative work is not merely a technical matter but a political choice that warrants opposition through labor movements and political activism directed against the underlying austerity politics (Koch and Zanchetta 2025).

## Future of LLM Design

The risk introduced so far are neither inevitable nor insurmountable. Rather, they illuminate the contours of a more principled research and design agenda for LLM-assisted creativity. The field of computational creativity has long understood that effective creative support systems require structured interaction models grounded in understanding of creative processes (Kantosalo et al. 2020; Shneiderman 2009). As explained in the previous sections, LLMs lack genuine epistemic agency and autonomy (Špecián 2026). LLMs alone underperform in creative domains compared to human-guided alternatives (Rondini et al. 2026), revealing that the perceived independence of LLMs is largely illusory. By reorienting LLMs within intentional workflows constrained by human authority, we position them not as autonomous agents but as sophisticated materials through which humans think, explore, and refine their ideas (Rozenal and de Rooij 2026). In this section we outline three interconnected design imperatives that emerge from both LLM constraints and the demonstrated user preferences.

## LLMs as Refiners, not Generators

Currently, LLM outputs are often incorrectly treated as end-point products. To achieve effective LLM augmentation, we must move beyond simple generation in favor of facilitating deep, iterative refinement grounded in critical evaluation. Instead of producing final answers, LLMs should create intermediate artifacts that feed into structured critique cycles. Rather than asking LLMs to continuously generate ideas, more effective systems should establish iterative feedback loops where critique is converted into actionable refinement steps (Hou et al. 2025; Li et al. 2025). This approach mirrors the creative cycle of writing and reading, where generation and interpretation inform one another iteratively (Gervás and León 2014). For example, the *FlexMind* system allows users to ask for trade-offs and mitigations for each idea generated by the LLM, which then become specific refinement requests (Yang et al. 2025). By shifting from one-off generation to a critique-driven workflow, users in this study rated more favorably the perceived quality and depth of the ideas generated through iterative feedback loops.

A critical requirement for future systems supporting a reliable workflow is the employment of reasoning and traceability mechanisms. Approaches such as Chain-of-Thought (?), which decompose requests into justified steps, or other task decomposition frameworks provide crucial mechanisms for mitigating hallucinations while offering users transparent justifications they can critique and refine (Yang et al. 2025; Huang et al. 2025a). Explanations themselves however are still prone to hallucinations, misleading users in decision making processes (Cabitza et al. 2024). Despite this, transparency is essential, as users cannot meaningfully engage with a system’s output if they cannot understand or verify its reasoning. Yet transparency alone is insufficient; systems must also be proactive. When a user’s request is ambiguous or incomplete, the system should ask for clarifications rather than proceeding with unexamined assumptions. Resolving ambiguity before starting to generate prevents the cascading failures that arise when LLMs misinterpret vague requirements (Sami et al. 2024). This represents a fundamental shift from systems that operate passively on user inputs toward systems that actively collaborate through seeking and confirming essential contextual information (Vijayvargiya et al. 2024; Sun et al. 2025).

## Distributed Authority

The ideation process is typically divided into two distinct phases: divergent exploration (generating alternatives) and convergent evaluation (selecting and refining) (Kumar et al. 2025; Paulus, Coursey, and Kenworthy 2018; Ulrich 2018). Research in cognitive psychology demonstrates that convergent thinking (the focused ability to evaluate and refine ideas) serves as a necessary threshold for divergent thinking to produce measurable creative outcomes (Zhu et al. 2019). Current LLMs are deployed almost exclusively to exploration, but are rarely consulted for evaluation and selection of ideas (Li et al. 2025). This asymmetry reflects both user intuitions and empirical findings: humans remain

skeptical of LLM judgment, and research demonstrates that this skepticism is well-founded. LLMs lack genuine understanding and reasoning, as the calculation of word sequence probabilities is not equivalent to human knowledge or a “mind” (Peeperkorn, Brown, and Jordanous 2023). This lack of true reasoning is exacerbated by the LLMs’ grounding problem: they learn relationships between words, not between words and real-world entities as humans do (Peeperkorn, Brown, and Jordanous 2023). These shortcomings lead to the problem of hallucination, or “plausibility trap”, making such models unreliable judges of real-world ideas (Li et al. 2025).

In user-centric systems, LLMs must be part of the workflow while users retain decision-making authority, particularly during convergent phases where critical selection and judgment occur. This does not mean excluding LLMs from evaluation contexts. Rather, it means grounding the models as much as possible in the operational domain, and ensuring human evaluation remains the final arbiter, while LLM input serves an advisory, transparent role. This grounding places a dual onus on developers and users to co-manage the model’s integration within the creative domain, ensuring that AI assistance does not override human agency. Users should understand not only what the LLM recommends, but also why. Providing contrastive explanations, for example, would highlight the difference between the LLM’s recommendation and the user’s likely reasoning, preventing skill erosion and allowing users to identify specific knowledge gaps (Bućinca et al. 2025). Such transparency enables informed disagreement and the ability to override suggestions (Shen et al. 2025), ensuring that the collaboration enhances, rather than replaces, human expertise.

Systems that preserve human authority should also leverage specialized LLM techniques tailored to enhance specific creative dimensions, rather than relying on general-purpose models. One example is models trained with Creative Preference Optimization (Ismayilzada et al. 2025), where different signals from multiple creative dimensions are injected directly into the model’s training objective. While outputs from these models are judged by humans to be more novel, diverse, and surprising, they often sacrifice overall quality. Another technique is BILLY (Pai et al. 2025), where different persona vectors are integrated in the model’s activation space, representing shifts in activation states when responding with specific personas (e.g., creative, environmentalist) compared to neutral responses. This avoids expensive multi-agent communication while enabling efficient multi-perspective generation. Users consistently prefer systems that enable autotelic creativity (the intrinsically pleasurable exploration of a possibility space) rather than systems designed primarily for efficiency (Compton and Mateas 2015). While such specializations transform LLMs from generic tools into focused augmentation systems that amplify specific creative dimensions, they cannot resolve the inherent shortcomings of LLMs identified previously. Rather than allowing open-ended, autonomous LLM generation, LLMs must be deployed within workflows that acknowledge these limitations rather than pretend they don’t exist. Roles, decision points, and interaction patterns must be explicitly de-

signed to preserve human authority, preventing LLMs from operating autonomously.

## Shared Authorship

The question of who deserves credit for creative work becomes increasingly fraught in human-AI contexts (Bown 2015). Contemporary human-AI co-creation often results in “fuzzy hybrid authorship”, where the boundary between human intent and machine generation is so ambiguous that assigning credit, accountability, or tracing the creative history (provenance) of the artifact becomes nearly impossible (Alharthi 2025; Hwang, Shin, and Lee 2025). This ambiguity is further complicated by the fact that narrative context and framing are often absent from AI-generated artifacts (Charnley, Pease, and Colton 2012). This “shadowed agency” (Hwang, Shin, and Lee 2025) complicates the creative process, making it difficult to delineate where a user’s contribution ends and the AI’s influence begins. To address this, future design systems must embed mechanisms for tracking and documenting contribution provenance, measuring what portions of a creative artifact originated from human input, LLM generation, or iterative human-AI refinement (Kadenhe, Al Musleh, and Lompot 2025). This kind of transparency would serve both ethical and practical functions: it enables fair attribution while helping users understand the nature of their collaborative process.

Beyond the practical need for attribution, embedding ethical frameworks directly into generative systems can serve as a powerful catalyst for innovation. Treating ethics itself as an aesthetic dimension of creativity opens novel pathways for creative exploration (Ventura and Gates 2018). Rather than acting as a constraint, injecting ethical instructions into the model’s prompt for creative processes disrupts its tendency to fall back on high-frequency, predictable patterns (Rabonato and Berton 2025). This intentional disruption forces the AI away from its default outputs and toward less probable, more imaginative, and diverse narrative combinations (Rabonato and Berton 2025). Altering the model’s prompt, however, does not solve the fundamental limitations of LLMs discussed previously. Therefore, future design systems must drive LLM generation under ethical constraints, either by embedding ethical principles into the model’s architecture or by integrating human oversight throughout the creative process to ensure alignment with ethical standards.

## Directions for Computational Creativity Research

Establishing actionable, realistic goals for LLM integration in creative practice requires grounding research in the actual creative process. Rather than treating LLMs as monolithic tools, we must understand the distinct phases of creative work and where LLMs can meaningfully contribute. Small-scale, transparent systems have long been recognized as valuable for exploring specific creative techniques in isolation (Montfort and Fedorova 2012), and may provide a model for domain-specialized LLM deployment. To this end, we propose specific open questions, aiming to inspire

further investigation within the computational creativity and machine learning communities.

**How can we employ LLMs in specific and personalized creative domains?** The creative process typically unfolds across distinct phases: divergent exploration (idea generation), iterative refinement (development and articulation), and convergent evaluation (judgment and selection) (Kumar et al. 2025; Paulus, Coursey, and Kenworthy 2018; Ulrich 2018). The source of creativity in collective environments involves complex dynamics of ideation and interaction (Maher 2012), dynamics that specialized Small Language Models (SLMs) might better support through domain-specific expertise. Rather than assuming LLMs are universally beneficial across all phases, we must empirically and theoretically map where they genuinely excel and where they introduce friction. Current approaches have largely focused on massive, generalized models driven by the economic incentives of large corporations, often trained on data of unclear provenance with significant ethical concerns. These training corpora frequently perpetuate data colonialism, reproducing Western aesthetics, linguistic hierarchies, and cultural stereotypes that narrow the creative possibilities available to users globally (Kamran 2023; Liu 2023). Instead, future research should prioritize specialized SLMs tailored to specific creative domains and individual users. Unlike general-purpose LLMs trained on diverse internet-scale data, SLMs are fine-tuned on curated, domain-specific corpora, encoding expert knowledge directly into their parameters. Such models often outperform larger models and even humans on domain-specific tasks (Marco, Rello, and Gonzalo 2025), and offer greater accessibility, as they can be deployed locally on consumer-level hardware (Xu et al. 2025). For example, a model explicitly trained on the Blender wiki could be integrated directly into the application to assist designers. This approach minimizes reliance on Retrieval-Augmented Generation and lowers the probability of hallucinations, as relevant domain knowledge is embedded within the model’s parameters. Crucially, such a dedicated assistant removes the need for the model to be aware of unrelated concepts, thereby streamlining the inference process and reducing computational overhead. Rather than blindly embedding LLMs into every phase of creative work, research should systematically investigate where specialized SLMs provide genuine value to the creative process. How could SLMs be leveraged during exploratory ideation, generating diverse alternatives? Or during iterative refinement, articulating and developing nascent ideas? Can they reliably assist during convergent evaluation, providing transparent, justifiable critiques that inform human judgment? Which approaches best preserve individual creative identity while still providing meaningful assistance? How do these approaches scale across diverse creative domains? Answering these questions requires empirical studies, mapping LLM utility to specific workflow phases.

**How do we measure human-LLM contribution?** When humans and LLMs collaborate, the boundary between human and machine contribution becomes ambiguous. Establishing rigorous methodologies for measuring this contribution serves dual purposes: enabling fair attribution and deep-

ening our understanding of effective collaboration patterns. In text-based domains, metrics such as text perplexity (how random the text is, where less predictable text is more likely to be human-authored) or “burstiness” (the length of sentences, where sentences of uniform length are more likely to be LLM-generated) have been used to estimate LLM involvement in the writing process (Gralha and Pimentel 2024). However, these metrics fail to account for intellectual contributions preceding textual generation, such as conceptual framing, thematic direction, and critical feedback that shapes LLM behavior. A writer may contribute only 20% of final text but 80% of the creative vision; conversely, an LLM may generate most text while the human merely accepts suggestions uncritically. Understanding how humans perceive and value computational agents’ contributions in collaborative scenarios remains an open question (Jordanous 2017), one that attribution methodologies must address. Accurately assessing authorship requires complex judgment of how much final ideas were defined by the user versus how much they were influenced by interaction with the machine (Epstein et al. 2023). The challenge of attribution intensifies in non-textual domains. In 3D modeling, how do we measure the extent to which geometric modifications reflect human intent versus LLM suggestion? In music composition, how do we distinguish melodic ideas originating from user input versus algorithmic elaboration? In visual design, how do we quantify the contribution of color palettes, compositional guidance, and iterative refinement? These questions highlight that in many creative contexts an objective contribution metric may not only be undefinable but conceptually void of significance, as the emergent output is often greater than the sum of its discrete parts. Consequently, these metrics must take into account the unique affordances and creative practices for each medium. Advances in this direction would support both ethical attribution and research into effective collaborative design patterns. From a research perspective, contribution measurement illuminates which collaborative patterns produce superior creative outcomes, informing the design of future systems.

**How should interfaces transform for creative LLM partnership?** Current human-LLM interaction relies predominantly on conversational interfaces based on chat windows where users formulate requests in natural language and await responses. While intuitive for information retrieval, conversational interaction proves suboptimal for creative work requiring tight feedback loops, contextual awareness, and seamless integration with domain-specific tools. Moving beyond standard chat windows, future GUIs should prioritize operation-specific, contextualized interactions, drawing on existing examples of non-linear interfaces (Choi et al. 2025; Vanderdonck et al. 2005). For instance, in LLMaker (Gallotta, Liapis, and Yannakakis 2024), a designer should be able to initiate a conversation regarding a specific game enemy simply by right-clicking its sprite. A writer could highlight a passage to solicit critique or refinement suggestions. A composer could select a melodic segment to request variations maintaining specific tonal qualities. This interaction is more intuitive for designers, as it mimics direct manipulation, and is computationally efficient for the LLM,

as the selection provides immediate, relevant context for subsequent operations. This workflow also supports the refinement requirements highlighted previously, enabling the LLM to request clarifications based on specific local context. Future interface designs might draw on computational models of flow and social creativity (Webster, Zachos, and Maiden 2013), ensuring that LLM assistance operates within the user’s zone of creative engagement. What other modalities can be employed to capture users’ implicit context? How can we design interfaces that seamlessly integrate LLM assistance into existing creative tools without disrupting established workflows? How can we balance transparency and usability, ensuring users understand LLM reasoning without overwhelming them with technical details?

**How proactive should LLM assistants be in creative contexts?** To enhance proactiveness, operations carried out solely by the user could be monitored by the LLM, which could then intervene only when relevant. This capability introduces fundamental questions about the appropriate level of LLM agency in creative work. At what point does an LLM’s proactive intervention transition from being a “helpful partner” to a “distracting agent”? How can systems determine when intervention is opportune versus intrusive? What signals should trigger LLM engagement without disrupting the user’s creative flow? How can we design monitoring mechanisms that respect user autonomy while enabling timely assistance? Addressing these questions will require interdisciplinary research combining insights from human-computer interaction, cognitive psychology, and design studies.

These future design directions aim to mitigate existing integration challenges and promote the adoption of LLMs in more cautious, user-centric applications. By addressing these issues and proposing clearer architectural and interaction standards, we hope to establish a safer, more productive research trajectory for both the computational creativity community and LLM researchers at large.

## Conclusions

In this paper we argued for the fundamental reorientation of how computational creativity researchers and AI practitioners must conceptualize and deploy LLMs in creative contexts. Rather than embracing full automation, LLMs must be intentionally designed as semi-autonomous collaborators embedded within structured, human-centric workflows. This approach aligns with computational creativity’s decades-long commitment to understanding and supporting creative practice. It offers a principled pathway to realize the genuine augmentative potential of LLMs while mitigating the cognitive, procedural, and epistemic risks that unconstrained automation poses.

This position rests on multiple observations. Firstly, we have established that LLMs suffer fundamental, architectural limitations preventing them from reliably automating high-level creative work. They struggle with novelty and originality because their training paradigm biases toward high-frequency patterns. They fail at genuinely subjective creative tasks because they lack episodic grounding and

lived experience. Most critically, they cannot serve as reliable judges of creative merit. This dual failure in both generation and evaluation demands that human authority remains paramount. These are not superficial limitations remediable through incremental improvements. Rather, they are inherent to how LLMs process and represent meaning. Recognizing these limitations, however, is not pessimistic. Rather, it invites us to orchestrate the complementary strengths of humans and LLMs.

Secondly, we have documented tangible harms arising when humans overly rely on systems operating with excessive autonomy in creative contexts. Unstructured automation produces shallow engagement and cognitive offloading that persists and degrades subsequent creative performance. It introduces systemic failures when design contexts involve genuine ambiguity. These costs are neither abstract nor speculative: they have been empirically demonstrated across multiple studies and domains, and they are concerning for the creative community at large. Yet they are also avoidable through deliberate design choices that preserve human agency and embed critique throughout the creative process.

Finally, we have outlined a coherent research and design agenda. Effective LLM-assisted creative systems must involve iterative critique and refinement, employ transparent reasoning mechanisms, proactively resolve ambiguity, and most importantly, retain human decision-making authority especially in evaluative contexts. These systems must leverage specialized LLM techniques tailored to enhance specific creative dimensions rather than relying on general-purpose models. They must address the question of authorship, attribution, and creative provenance that human-AI collaboration raises. Concrete research directions include developing domain-specific small language models that reduce hallucination and computational overhead, formalizing methodologies for measuring human-AI contribution, and designing context-aware user interfaces that support rather than replace human judgment.

Recognizing the limitations of LLM creativity unlocks their genuine value. It shifts the focus from viewing LLMs as replacements for human creativity to intellectual catalysts, sparking new imaginative pathways that neither humans nor machines could traverse alone. LLMs can synthesize patterns, rapidly generate diverse alternatives, and articulate novel possibilities. But these capacities are most valuable not as endpoints but as scaffolding for human refinement, critique, and judgment. This perspective mandates that LLMs assist human creators by amplifying their reach, accelerating iteration, and expanding conceptual horizons, rather than usurping the creative authority that remains fundamentally human.

Realizing this vision necessitates sustained interdisciplinary research combining insights from computational creativity, human-computer interaction, machine learning, and the creative professions themselves. It requires validation of design frameworks that enhance human originality while preserving creative agency. It demands investment in specialized LLM techniques and evaluation frameworks that optimize for creative dimensions such as novelty over generic language quality. It calls for establishing practi-

cal and ethical standards for documenting hybrid authorship and creative provenance. Most fundamentally, it requires the creative and research communities to resist the fallacious allure of full automation and instead embrace the more challenging, more rewarding path of authentic human-AI partnership. This requires a shift from focusing on the final product to the creative process itself.

The conversation about LLMs as tools for creativity is only beginning. What is clear is that the path forward lies neither in unbridled over-reliance nor in reflexive rejection of these powerful models. It lies instead in thoughtful, deliberate integration: embedding LLMs within workflows and applications designed to preserve human agency, to amplify human judgment, and to ensure that creative practice remains fundamentally human even as it becomes increasingly, inevitably augmented by artificial intelligence.

### Author Contributions

R.G. and A.L. defined the position. R.G. was in charge of writing the manuscript. A.L. and G.N.Y. were involved in revising the manuscript.

### Acknowledgments

R.G. thanks his fellow PhD students at the Institute of Digital Games for the fruitful discussions over lunch.

### References

- Alharthi, S. A. 2025. Generative AI in game design: Enhancing creativity or constraining innovation? *Journal of Intelligence* 13(6).
- Andrus, B., and Fulda, N. 2022. A data-driven architecture for social behavior in creator networks. In *Proceedings of the International Conference on Computational Creativity*.
- Avlonitou, C., and Papadaki, E. 2025. AI: An active and innovative tool for artistic creation. *Arts* 14(3).
- Becker, J.; Rush, N.; Barnes, E.; and Rein, D. 2025. Measuring the impact of early-2025 AI on experienced open-source developer productivity. *arXiv preprint arXiv:2507.09089*.
- Benk, M.; Wettstein, L.; Schlicker, N.; von Wangenheim, F.; and Scharowski, N. 2025. Bridging the knowledge gap: understanding user expectations for trustworthy LLM standards. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence*. AAAI Press.
- Boden, M. A. 2004. *The Creative Mind: Myths and Mechanisms*. Taylor & Francis Group.
- Bown, O. 2015. Attributing creative agency: Are we doing it right? In *Proceedings of the International Conference on Computational Creativity*.
- Buçinca, Z.; Swaroop, S.; Paluch, A. E.; Doshi-Velez, F.; and Gajos, K. Z. 2025. Contrastive explanations that anticipate human misconceptions can improve human decision-making skills. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery.
- Cabitza, F.; Fregosi, C.; Campagner, A.; and Natali, C. 2024. Explanations considered harmful: The impact of misleading explanations on accuracy in hybrid human-AI decision making. In Longo, L.; Lapuschkin, S.; and Seifert, C., eds., *Explainable Artificial Intelligence*. Springer Nature Switzerland.
- Chakrabarty, T.; Laban, P.; Agarwal, D.; Muresan, S.; and Wu, C.-S. 2024. Art or artifice? Large language models and the false promise of creativity. *arXiv preprint arXiv:2309.14556*.
- Charnley, J.; Pease, A.; and Colton, S. 2012. On the notion of framing in computational creativity. In *Proceedings of the International Conference on Computational Creativity*.
- Choi, D.; Son, K.; Jung, H.; and Kim, J. 2025. Expandora: Broadening design exploration with text-to-image model. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery.
- Choudhury, A., and Chaudhry, Z. 2024. Large language models and user trust: Consequence of self-referential learning loop and the deskilling of health care professionals. *Journal of Medical Internet Research*.
- Colton, S., and Wiggins, G. A. 2014. Computational creativity: The final frontier? In *Proceedings of the International Conference on Computational Creativity*.
- Colton, S.; Pease, A.; and Saunders, R. 2018. Issues of authenticity in autonomously creative systems. In *Proceedings of the International Conference on Computational Creativity*.
- Compton, K., and Mateas, M. 2015. Casual creators. In *Proceedings of the International Conference on Computational Creativity*.
- Edwards, B. 2026. AI agents now have their own Reddit-style social network, and it's getting weird fast. Available online at <https://arstechnica.com/information-technology/2026/01/ai-agents-now-have-their-own-reddit-style-social-network-and-its-getting-weird-fast/>.
- Epstein, Z.; Hertzmann, A.; Akten, M.; Farid, H.; Fjeld, J.; Frank, M. R.; Groh, M.; Herman, L.; Leach, N.; Mahari, R.; Pentland, A.; Russakovsky, O.; Schroeder, H.; and Smith, A. 2023. Art and the science of generative AI. *arXiv preprint arXiv:2306.04141*.
- Gallotta, R.; Liapis, A.; and Yannakakis, G. 2024. LL-Maker: A game level design interface using (only) natural language. In *Proceedings of the 2024 IEEE Conference on Games (CoG)*.
- Gervás, P., and León, C. 2014. Reading and writing as a creative cycle: the need for a computational model. In *Proceedings of the International Conference on Computational Creativity*.
- Grace, K.; Gero, J.; and Saunders, R. 2013. Learning how to reinterpret creative problems. In *Proceedings of the International Conference on Computational Creativity*.

- Gralha, J. G., and Pimentel, A. S. 2024. Gotcha GPT: Ensuring the integrity in academic writing. *Journal of Chemical Information and Modeling*.
- Gubenko, A.; Lubart, T.; and Houssemand, C. 2022. From social robots to creative humans and back. In *Proceedings of the International Conference on Computational Creativity*.
- Hern, A., and Milmo, D. 2024. Spam, junk ... slop? the latest wave of AI behind the 'zombie internet'. Available online at <https://www.theguardian.com/technology/article/2024/may/19/spam-junk-slop-the-latest-wave-of-ai-behind-the-zombie-internet>.
- Hicks, M. T.; Humphries, J.; and Slater, J. 2024. ChatGPT is bullshit. *Ethics and Information Technology*.
- Hou, M.; Wu, L.; Liao, Y.; Yang, Y.; Zhang, Z.; Zheng, C.; Wu, H.; and Hong, R. 2025. A survey on generative recommendation: Data, model, and tasks. *arXiv preprint arXiv:2510.27157*.
- Huang, R.; Yang, T.; Feng, S.; and Li, F. 2025a. Design coco-pilot: QFD as decision logic for conflict-aware human-LLM collaborative product conceptual design. *Journal of Engineering Design*.
- Huang, Z.; Zhong, S.; Zhou, P.; Gao, S.; Zitnik, M.; and Lin, L. 2025b. A causality-aware paradigm for evaluating creativity of multimodal large language models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Hwang, Y.; Shin, D.; and Lee, J. H. 2025. Who owns AI-generated artwork? Revisiting the work of generative AI based on human-AI co-creation. *Telematics and Informatics* 98.
- Independent. 2019. Ethics guidelines for trustworthy AI. Available online at <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.
- Indurkha, B. 2012. Whence is creativity? In *Proceedings of the International Conference on Computational Creativity*.
- Ismayilzada, M.; Laverghetta, A.; Luchini, S. A.; Patel, R.; Bosselut, A.; Plas, L. v. d.; and Beaty, R. 2025. Creative preference optimization. *arXiv preprint arXiv:2505.14442*.
- Johnson, C. G. 2012. The creative computer as romantic hero? computational creativity systems and creative personae. In *Proceedings of the International Conference on Computational Creativity*.
- Jordanous, A. 2017. Co-creativity and perceptions of computational agents in co-creativity. In *Proceedings of the International Conference on Computational Creativity*.
- Kadenhe, N.; Al Musleh, M.; and Lompot, A. 2025. Human-AI co-design and co-creation: A review of emerging approaches, challenges, and future directions. In *Proceedings of the AAAI Symposium Series*.
- Kamran, A. 2023. Decolonizing artificial intelligence: Unveiling biases, power dynamics, and colonial continuities in AI systems. *RMS journal* 2023.
- Kantosalo, A., and Toivonen, H. 2016. Modes for creative human-computer collaboration: Alternating and task-divided co-creativity. In *Proceedings of the International Conference on Automated Planning and Scheduling*.
- Kantosalo, A.; Ravikumar, P. T.; Grace, K.; and Takala, T. 2020. Modalities, styles and strategies: An interaction framework for human-computer co-creativity. In *International Conference on Innovative Computing and Cloud Computing*.
- Kantosalo, A. 2020. Five C's for human-computer co-creativity - An update on classical creativity perspectives. In *Proceedings of the International Conference on Computational Creativity*.
- Khurana, A.; Su, X.; Wang, A. Y.; and Chilana, P. K. 2025. Do it for me vs. do it with me: Investigating user perceptions of different paradigms of automation in copilots for feature-rich software. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.
- Kim, S. S. Y.; Vaughan, J. W.; Liao, Q. V.; Lombrozo, T.; and Russakovsky, O. 2025. Fostering appropriate reliance on large language models: The role of explanations, sources, and inconsistencies. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery.
- Koch, A., and Zanchetta, G. 2025. Foundation models as agents of austerity. In *Proceedings of the International Conference on Computational Creativity*.
- Kosmyrna, 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 preprint arXiv:2506.08872*.
- Kumar, H.; Vincentius, J.; Jordan, E.; and Anderson, A. 2025. Human creativity in the age of LLMs: Randomized experiments on divergent and convergent thinking. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.
- Kyi, L.; Mahuli, A.; Silberman, M. S.; Binns, R.; Zhao, J.; and Biega, A. J. 2025. Governance of generative AI in creative work: Consent, credit, compensation, and beyond. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery.
- Lamb, C. E., and Brown, D. G. 2023. Should we have seen the coming storm? Transformers, society, and CC. In *Proceedings of the International Conference on Computational Creativity*.
- Li, S.; Padilla, S.; Bras, P. L.; Dong, J.; and Chantler, M. 2025. A review of LLM-assisted ideation. *arXiv preprint arXiv:2503.00946*.
- Lin, Z.; Ehsan, U.; Agarwal, R.; Dani, S.; Vashishth, V.; and Riedl, M. 2023. Beyond prompts: Exploring the design space of mixed-initiative co-creativity systems. In *Proceedings of the International Conference on Computational Creativity*.
- Liu, Z. 2023. Cultural bias in large language models: A comprehensive analysis and mitigation strategies. *Journal of Transcultural Communication*.

- Liu, Z. 2025. Human-AI co-creation: A framework for collaborative design in intelligent systems. *arXiv preprint arXiv:2507.17774*.
- Lopez-Lopez, E.; Abels, C. M.; Lorenz-Spreen, P.; Lewandowsky, S.; and Herzog, S. M. 2026. Boosting metacognition in entangled human-AI interaction to navigate cognitive-behavioral drift. *arXiv preprint arXiv:2602.01959*.
- Lynch, A.; Wright, B.; Larson, C.; Ritchie, S. J.; Mindermann, S.; Hubinger, E.; Perez, E.; and Troy, K. 2025. Agentic misalignment: How LLMs could be insider threats. *arXiv preprint arXiv:2510.05179*.
- Madiega, T. A. 2024. Artificial intelligence act. Available online at [https://www.europarl.europa.eu/thinktank/en/document/EPRS\\_BRI\(2021\)698792](https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698792).
- Maher, M. L. 2012. Computational and collective creativity: Who's being creative? In *Proceedings of the International Conference on Computational Creativity*.
- Marco, G.; Rello, L.; and Gonzalo, J. 2025. Small language models can outperform humans in short creative writing: A study comparing SLMs with humans and LLMs. In Rambow, O.; Wanner, L.; Apidianaki, M.; Al-Khalifa, H.; Eugenio, B. D.; and Schockaert, S., eds., *Proceedings of the 31st International Conference on Computational Linguistics*. Association for Computational Linguistics.
- McCormack, J.; Gifford, T.; and Hutchings, P. 2019. Autonomy, authenticity, authorship and intention in computer generated art. *arXiv preprint arXiv:1903.02166*.
- Montfort, N., and Fedorova, N. 2012. Small-scale systems and computational creativity. In *Proceedings of the International Conference on Computational Creativity*.
- Neumann, A., and Singh, J. 2025. Cascading effects: A multifaceted governance challenge in AI systems. In *Proceedings of the EurIPS 2025 Workshop on Private AI Governance*.
- Noy, S., and Zhang, W. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Science*.
- Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C. L.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; Schulman, J.; Hilton, J.; Kelton, F.; Miller, L.; Simens, M.; Askell, A.; Welinder, P.; Christiano, P.; Leike, J.; and Lowe, R. 2022. Training language models to follow instructions with human feedback. In *Proceedings of the International Conference on Neural Information Processing Systems*.
- Padmakumar, V., and He, H. 2023. Does writing with language models reduce content diversity? *arXiv preprint arXiv:2309.05196*.
- Pai, T.-M.; Wang, J.-I.; Lu, L.-C.; Sun, S.-H.; Lee, H.-Y.; and Chang, K.-W. 2025. BILLY: Steering large language models via merging persona vectors for creative generation. *arXiv preprint arXiv:2510.010157*.
- Pan, M. Z.; Cemri, M.; Agrawal, L. A.; Yang, S.; Chopra, B.; Tiwari, R.; Keutzer, K.; Parameswaran, A.; Ramchandran, K.; Klein, D.; Gonzalez, J. E.; Zaharia, M.; and Stolica, I. 2025. Why do multiagent systems fail? In *ICLR 2025 Workshop on Building Trust in Language Models and Applications*.
- Paulus, P.; Coursey, L.; and Kenworthy, J. 2018. *Divergent and Convergent Collaborative Creativity*. Palgrave.
- Peeperkorn, M.; Brown, D. G.; and Jordanous, A. 2023. On characterizations of large language models and creativity evaluation. In *Proceedings of the International Conference on Computational Creativity*.
- Rabonato, R. T., and Berton, L. 2025. Fairness as a creative resource: Challenges and opportunities in creative computing. In *Proceedings of the International Conference on Computational Creativity*.
- Rathi, N.; Jurafsky, D.; and Zhou, K. 2025. Humans over-rely on overconfident language models, across languages. *arXiv preprint arXiv:2507.06306*.
- Rondini, S.; Alvarez-Martin, C.; Angermair-Barkai, P.; Penacchio, O.; Paz, M.; Pelowski, M.; Dediu, D.; Rodriguez-Fornells, A.; and Cerda-Company, X. 2026. Stable diffusion models reveal a persisting human-AI gap in visual creativity. *Advanced Science*.
- Rozental, S., and de Rooij, A. 2026. Neither tool nor collaborator: Rethinking human-AI co-creativity in artistic practice with material engagement theory. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*.
- Ruan, S.; Sheng, R.; Wen, X.; Wang, J.; Zhang, T.; Wang, Y.; Dwyer, T.; and Li, J. 2025. Qualitative study for LLM-assisted design study process: Strategies, challenges, and roles. *arXiv preprint arXiv:2507.10024*.
- Russell, M.; Shah, A.; Blaney, G.; Amores, J.; Czerwinski, M.; and Jacob, R. J. K. 2025. Neural and cognitive impacts of AI: The influence of task subjectivity on human-LLM collaboration. *arXiv preprint arXiv:2506.04167*.
- Sami, M. A.; Waseem, M.; Zhang, Z.; Rasheed, Z.; Systä, K.; and Abrahamsson, P. 2024. AI based multiagent approach for requirements elicitation and analysis. *arXiv preprint arXiv:2409.00038*.
- Shen, H.; Knearem, T.; Ghosh, R.; Alkiek, K.; Krishna, K.; Liu, Y.; Ma, Z.; Petridis, S.; Peng, Y.-H.; Qiwei, L.; Rakshit, S.; Si, C.; Xie, Y.; Bigham, J. P.; Bentley, F.; Chai, J.; Lipton, Z.; Mei, Q.; Mihalcea, R.; Terry, M.; Yang, D.; Morris, M. R.; Resnick, P.; and Jurgens, D. 2025. Position: Towards bidirectional human-AI alignment. *arXiv preprint arXiv:2406.09264*.
- Shneiderman, B. 2009. *Creativity Support Tools: A Grand Challenge for HCI Researchers*. Springer London.
- Shneiderman, B. 2020. Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction*.
- Sosa, R., and Gero, J. S. 2013. Brainstorming in solitude and teams: A computational study of group influence. In *Proceedings of the International Conference on Computational Creativity*.
- Sun, W.; Zhou, X.; Du, W.; Wang, X.; Welleck, S.; Neubig,

G.; Sap, M.; and Yang, Y. 2025. Training proactive and personalized LLM agents. *arXiv preprint arXiv:2511.02208*.

Suwa, M.; Purcell, T.; and Gero, J. 1998. Macroscopic analysis of design processes based on a scheme for coding designers' cognitive actions. *Design Studies*.

Torrance, E. P. 1968. A longitudinal examination of the fourth grade slump in creativity. *Gifted Child Quarterly*.

Tubb, R., and Dixon, S. 2014. A four strategy model of creative parameter space interaction. In *Proceedings of the International Conference on Computational Creativity*.

Ulrich, F. 2018. Exploring divergent and convergent production in idea evaluation: Implications for designing group creativity support systems. *Communications of the Association for Information Systems* 43.

Vanderdonckt, J.; Grolaux, D.; Van Roy, P.; Limbourg, Q.; Macq, B.; and Michel, B. 2005. A design space for context-sensitive user interfaces. In *Proceedings of the ISCA 14th International Conference on Intelligent and Adaptive Systems and Software Engineering*.

Veale, T. 2024. From symbolic caterpillars to stochastic butterflies: Case studies in re-implementing creative systems with LLM. In *Proceedings of the International Conference on Computational Creativity*.

Ventura, D., and Gates, D. 2018. Ethics as aesthetic: A computational creativity approach to ethical behavior. In *Proceedings of the International Conference on Computational Creativity*.

Vijayvargiya, S.; Zhou, X.; Yerukola, A.; Sap, M.; and Neuhbig, G. 2024. Interactive agents to overcome ambiguity in software engineering. *arXiv preprint arXiv:2409.00038*.

Webster, S.; Zachos, K.; and Maiden, N. 2013. An emerging computational model of flow spaces to support social creativity. In *Proceedings of the International Conference on Computational Creativity*.

Xu, D.; Zhang, H.; Yang, L.; Liu, R.; Huang, G.; Xu, M.; and Liu, X. 2025. Fast on-device LLM inference with NPUs. In *Proceedings of the 30th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 1*. Association for Computing Machinery.

Yang, Y.; Mohanty, V.; Chen, Y.-Y.; Hong, M. K.; Martelaro, N.; and Kittur, A. 2025. FlexMind: Supporting deeper creative thinking with LLMs. *arXiv preprint arXiv:2509.21685*.

Zhang, J.; Hu, S.; Lu, C.; Lange, R.; and Clune, J. 2025. Darwin Godel machine: Open-ended evolution of self-improving agents. *arXiv preprint arXiv:2505.22954*.

Zhu, W.; Shang, S.; Jiang, W.; Pei, M.; and Su, Y. 2019. Convergent thinking moderates the relationship between divergent thinking and scientific creativity. *Creativity Research Journal*.

Špecián, P. 2026. Machine advisors: Integrating large language models into democratic assemblies. *Social Epistemology*.

Špela Vintar, and Javoršek, J. J. 2025. The truth is no diaper: Human and AI-generated associations to emotional words.

In *Proceedings of the International Conference on Computational Creativity*.