Searching Together: A Theory of Human-AI Co-Creativity

Vivianna Fang He  
University of St. Gallen, viviannafang.he@unisg.ch

Yash Raj Shrestha  
University of Lausanne, yashraj.shrestha@unil.ch

Phanish Puranam  
INSEAD, phanish.puranam@insead.edu

Ella Miron-Spektor  
INSEAD, ella.miron-spektor@insead.edu

The recent developments have enabled generative AI (GAI) to produce content across various modalities that is often indistinguishable from human-generated content, challenging the long-held belief that creativity is a human prerogative. In the light of this transformative technological evolution, we propose a novel theory of human-AI co-creativity (HACC), conceptualizing creativity as a joint search process. This creative search consists of three main tasks—construction of a search space, movement through this space, and evaluation of points in it. In developing a framework of multi-actor joint search, we systematically analyze the space of possible configurations of human-AI collaboration for creative tasks, along the dimensions of specialization of actors and sequencing of tasks. By delving into the intersections of human creativity and advancements in GAI, we explore the implications of the human-AI collaborative creative process within organizational contexts and spur future discussions around developing human and AI co-creativity while sustaining human skills and agency.

Keywords: AI; Creativity; Collaboration; Search; Human-Centric

Electronic copy available at: https://ssrn.com/abstract=4603650
Deep learning algorithms involving training multi-layered neural networks have allowed the identification of subtle and complex connections between inputs and outputs in a wide variety of contexts (LeCun, Bengio & Hilton 2015). Yet, creativity—the generation of novel and useful ideas or solutions (Amabile et al., 1996)—seemed beyond the reach of Artificial Intelligence (AI), and till recently theorists routinely took it as a hallmark of human ingenuity (Argote et al. 2021). However, the assumption that creativity is a uniquely human prerogative is being challenged by the latest product of deep learning algorithms—generative AI (GAI). GAI, such as the Large Language Models, can generate novel content indistinguishable from the products of humans across modalities (e.g., text, images, and music, Jo 2023). The introduction of these technological advancements in organizations has not only brought to the fore older discussions but also sparked new scholarly debates (Amabile, 2020; Dasborough, 2023) on the implications of AI for human creativity.

Creativity has been at the center of inquiries in both computer science and organizational science, but neither discipline has thus far offered a theory about how the creative process could be transformed when humans and AI work together. Computer scientists are primarily interested in the automation of creativity, which involves the production of novel, yet useful, ideas, problem solutions, or other outputs (Jordanous, 2012). Extant work in this discipline consists of identifying patterns in human creative content and using those patterns to replicate human creativity (Boden, 2016). In contrast, scholars in organization and management focus on personal and organizational factors that impact the creativity of individuals or small groups of individuals working together (see Anderson, Potocnik & Zhou, 2014 for a review). Studies on human AI interaction have explored whether humans understand and trust the information produced by AI (Lebovitz, Lifshitz-Assaf & Levina, 2022), suffer from algorithm aversion (Dietvorst, Simmons, & Massey, 2015),
and how they evaluate the creative content produced by AI (Raj, Berg, & Seamans, 2023) and its authenticity (Jago & Carroll, 2023). These streams of work have evolved largely independently with little cross-fertilization. Most importantly, none have considered the full range of possible divisions of labor between humans and AI in the creative process. As a result, despite much insight into how to leverage AI for generating creative outputs and how humans evaluate or resist these outputs, less is known about possible configurations of how humans and AI can collaborate to co-create.

We propose a theory about human-AI collaboration in the creative process. Drawing on insights from organizational science and computer science, we develop a “creativity as search” framework, conceptualizing how humans and AI collaborate in the search for creative ideas. Considering this joint search for novel and useful ideas as human-AI co-creativity (HACC), we explore various HACC configurations and their implications. Our framework integrates two important streams of work: first, an established body of literature on idea generation and evaluation by individuals and teams (e.g., Perry-Smith & Mannucci, 2017) and second, a nascent stream of research on the application of AI for generating creative outcomes alongside humans (Shrestha et al., 2019).

We make three main contributions. First, integrating research on generative AI (Bommasani et al. 2021, Foster 2022) and human unstructured problem solving (Mednick, 1962; Simonton, 2003), we theorize creativity as a search process, consisting of the construction of a search space, movement through this space, and evaluation of points in it. Some of the constructs within this process are well understood by human creativity scholars (Amabile, 1996; Perry-Smith & Mannucci, 2017) but the process is not sufficiently compared with or contrasted by AI creativity. Second, we develop a framework covering a broad range of possible human-AI co-creativity
configurations by explicitly considering the different roles human and AI could play across different phases of the creative process (Puranam, 2021). Our framework extends computer scientists’ focus on automating creativity and organizational scientists’ investigation of algorithm aversion or AI apprehension. Third, we offer theoretically founded conjectures about the implications of different human-AI co-creativity configurations for not only the creativity of the team, but more importantly, for its human members. Understanding this distinction, we believe, has significant theoretical, practical, and ethical impact.

**Differences between Humans and AI in Creative Search Processes**

Search, the pursuit of solutions to foreseen or existing problems, is a common denominator of both humans and AI creativity, serving as a foundational principle in many discoveries and innovations (Argote et al., 2021). We argue that creativity can be modeled as a process of search through a conceptual space, structured by a “fitness landscape” (Levinthal, 1997), in which creative ideas and solutions can be generated and evaluated for their appropriateness or utility. This formulation enables a synthesis of creativity studies across computer science and organizational science domains, promoting understanding of various creativity types, including combinatory, exploratory, and transformative creativity, each reflecting different mechanisms of idea generation and evaluation within the conceptual space.

Formally, we consider creativity as a search process consisting of three distinct tasks: A) construction of the search space through a selection of dimensions, B) movement through this space so that new points are considered and C) evaluation of the fitness of each point reached, possibly based on noisy signals. Within this process, components A and B underlie idea generation. Because individuals differ in their dimensions used for constructing the search space and their speed of traversal through it, idea generation varies across individuals. Component C corresponds
precisely to idea evaluations. Since estimates of fitness for the same point also likely vary across individuals, so will their idea evaluations.

We further argue that there are marked differences between human and AI creative processes in all three components. First, humans often possess a diverse yet shallow array of accessible data, encompassing life experiences, cultural, and social contexts, many of which remain elusive to AI due to their uncodified nature or regulatory constraints. In contrast, AI can access vast, but narrower data sets and represent them in high dimensional spaces, overcoming the limitations inherent to human cognitive capacity and memory, utilizing advancements in computing infrastructures and deep learning algorithms. The development in representation learning and the emergence of generative AI models (e.g., GPT-4) exemplify these capabilities. Second, bounded by cognitive biases and societal influences (Kahneman, 2003), humans often exhibit localized movements within conceptual spaces, limiting their creative scope. In contrast, algorithms can navigate extensive conceptual distances, exploring diverse, novel territories with fewer constraints of cognition and attention. Third, humans assess creative outcomes based on personal and collective preferences, acquired through social interactions and experiential data. Humans’ intuitive evaluations are prone to overfitting, overgeneralization, and various biases. While AI lacks such intuitive understanding and contextual evaluation of creative products, algorithmic evaluations can potentially offer more impartial, practical, and unbiased assessments.

**Configurations of Human-AI Co-creativity**

Our comparative assessment of the relative strengths and weakness of humans and AI underscores possibilities for humans and AI to collaborate in creative tasks. Depending on how tasks are distributed, division of labor varies along two dimensions—specialization among actors and sequencing of tasks (Puranam, 2021). Specialization refers to whether different tasks are
assigned to different actors (high specialization), or alternatively, all actors participate in most or all tasks (low, or no specialization). Sequencing of tasks refers to the temporal order in which tasks are executed. When a task produces outputs that become inputs to another task, these two tasks are sequentially dependent and need to be executed in sequence. Alternatively, when two tasks are not sequentially dependent, they can be executed simultaneously, in parallel (Puranam, Raveendran and Knudsen, 2012).

Specialization and sequence, combined with the component tasks in the search process, namely a. constructing and b. traversing the search space, and c. evaluating the fitness of a resulting point) give rise to various HACC configurations. For example, without specialization, humans and AI could both perform generation or evaluation, and they could work in sequence or parallel. With specialization, given the inherent sequential structure (i.e., the generation of ideas precedes the evaluation of them), there is no possibility of parallel work. Therefore, in the case with specialization, the human or the AI must play the first or second mover. Put differently, when only the AI generates ideas and only humans evaluate, humans necessarily act after the AI. We outline the possible configurations (Appendix I) and discuss the challenges and opportunities accordingly. Here we focus on the key trade-offs associated with the two main aspects of possible divisions of labor, namely the extent of specialization between human and AI actors and the sequence in which tasks are accomplished.

Collaboration with Specialization between Humans and AI

The primary advantage of specialization lies in leveraging the distinctive strengths of both humans and AI in constructing and navigating creative spaces. This advantage was clearly evidenced by Doshi and Hauser (2023), who found that leveraging OpenAI’s GPT-4 led to an 8% to 9% increase in writers' creativity, with notable enhancements of up to 22% to 26% among less
creative writers, thereby underscoring the transformative potential of specialized human-AI partnerships in enhancing creativity beyond individual capacities.

However, specialization between humans and AI in creative tasks also poses substantial disadvantages, predominantly due to the challenges in integrating efforts and the consequential risk of human skill decay. The integration of efforts is impeded by the inherent incommensurability of search spaces between human and AI-generated ideas, with AI's multifaceted dimensions often being intricate for humans to assess and vice versa (Fügener et al., 2022). This complexity is further exacerbated by a prevailing human distrust in AI, rooted in transparency deficiencies, control loss fears, and bias concerns (Rahman, 2021). This mistrust skews human evaluations and acceptance of AI contributions to creative processes. Further, even if integration of efforts may eventually become seamless, human skill may decay as humans concede the ideation process to AI, leading to a diminished human ability in creativity over time, potentially validating the belief in machine superiority in creativity. In addition, legal and practical predicaments ensue, raising unresolved issues about the originality and legality of AI-generated content (Smits & Borghuis, 2022). Current litigations like those against OpenAI, Midjourney, and Stable Diffusion show a prevailing concern about source attribution and rights infringement, reflecting deeper philosophical contemplations on the essence of creativity and originality in both AI and human creations.

**Collaboration without specialization between Human and AI**

We argue that configurations without specialization (i.e., ensembling, Choudhury et al., 2023) can capitalize on the “gains from diversity” while avoiding some of the challenges pertaining to integration of effort and skill decay. Yet, making search spaces used by humans and AI commensurable—such that one’s dimensions can be mapped into another’s dimensions—
remains difficult in these configurations. Ensembling better utilizes humans’ and AI’s different training data and resulting models—in the case of humans, life experiences and world views, respectively (Choudhury et al., 2023). Leveraging this diversity enables (1) production of more search spaces and moves through search spaces than either humans or AI alone could; (2) cancellation of errors in evaluation. Since both human and AI can engage in idea generation and evaluation, the disadvantages created by separating evaluation from generation we noted above as well as the potential skill loss for humans are both mitigated with ensembling. Humans may enhance their skills by observing and reverse engineering AI’s ideas. Consider the game of Go, by observing the novel moves of Alpha Go, humans can be inspired to reverse engineer them and gain a deeper understanding of the game itself (e.g., Shin et al. 2021).

The most significant challenge facing ensembling is the aggregation of humans and AI outputs. Rather than static error cancellation when averaging numbers or counting votes, combining the creative outputs (e.g., music) of humans with that of AI requires operators that are close in spirit to genetic algorithms. In Appendix II, we describe the “re-composition” operator that allows novelty to be generated from existing ideas that can be represented in a common space. These operators require that the search spaces within which human and AI have generated their alternatives are commensurable. However arriving at a common search space is by no means easy. Studying musicians co-producing songs with GAI, Huang et al. (2020) observed a structural difference preventing musicians work effectively with GAI: Human musicians are trained and used to produce songs comprising multiple modular sections, including vocal and instrumental parts that work in parallel. Conversely, GAI produced sound features in hierarchical structures that were not meaningful for human musicians to integrate.
Leveraging the power of HACC

Given that each configuration has its unique advantages and disadvantages, how should we best utilize the power of HACC? Were the goal of HACC simply producing the best combined creative outputs, we would be agnostic between specialization and ensembling (either in serial or parallel) and care only about output quality. However, when we care about preserving human creativity, we must a) prioritize ensembling in general and b) specialization that retains human creativity. This might entail either tolerating the inefficiency associated with the incommensurability problem and human distrust in AI, or the opportunity cost of not utilizing AI where it is superior. Crucially, if the challenges associated with ensembling could be addressed, human creativity may be better preserved or even improved. Two streams of work are currently being developed toward mitigating these challenges.

The first stream concerns translating human dimensions to AI readable spaces. Humans effortlessly navigate complex cognitive tasks such as interpreting strings of characters in poetry, utilizing evolved, specialized cognitive processing mechanisms. Recent advancements in neuroscience and machine learning leverage data from brain activities and gaze patterns, and cognitive processing signals, to enhance neural networks' language understanding and improve performance across various tasks like sentiment analysis and named entity recognition, surpassing the capabilities of purely text-based models. The second one concerns translating AI dimensions into human comprehensible spaces. The ever-increasing complexity prevents human from understanding the inner working of many AI models, necessitating advancements in AI explainability, particularly in design and creative domains. To address this, developments in intuitive interfaces and explainable designs have come to enable more effective human-AI co-creative systems where both entities proactively contribute to problem-solving (Yannakakis &
Togelius, 2014). Other tools like Sentient Sketchbook offer visual feedback on autonomous design suggestions, allowing designers to make informed decisions on adapting or discarding AI-generated designs based on the elucidated desirability or undesirability of the suggestions.

In conclusion, we develop a framework to systematically analyze the entire space of HACC configurations and their trade-offs, from a human-centric perspective. By conceptualizing creativity as a search process, we provide a robust framework for dissecting the what (construction, and transversal through the search space; evaluation of the resulting points), how (with or without specialization) and when (in sequence or in parallel) for humans and AI to collaborate in this process. Our framework unites the relevant conceptual elements from computer and organization science, thereby allowing cross-fertilization between disciplines that are most pivotal for furthering the understanding of human and AI creativity. We hope our framework and perspective stimulate further exploration into human and AI co-creativity while sustaining human skills and agency.
References


**Appendix I: HACC Configurations (in the case of multi-actor brainstorming)**

<table>
<thead>
<tr>
<th>Specialization</th>
<th>Involved Actor</th>
<th>Creativity phases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Idea generation</strong> (construction and traversal of search space)</td>
</tr>
<tr>
<td>Yes</td>
<td>H only</td>
<td>Humans specialize in producing candidate solutions</td>
</tr>
<tr>
<td></td>
<td>AI only</td>
<td>AI (e.g., GPT4) specializes in producing candidate solutions</td>
</tr>
<tr>
<td>No</td>
<td>Both H and AI, in parallel</td>
<td>A “super team” of humans and AI engages in producing candidate solutions; each actor works independently</td>
</tr>
<tr>
<td></td>
<td>Both H and AI, in sequence</td>
<td>A “super team” of humans and AI engages in producing candidate solutions; each actor builds on the ideas of others (e.g., elaboration)</td>
</tr>
</tbody>
</table>
Appendix II: The “Re-Composition” operator for aggregating creative outputs

Research on genetic algorithms shows that aggregation of creative outputs is possible in a common search space through concatenation, re-composition (via crossover and including with random mutations), inter- and extrapolation (Cantu-Paz, 2000). Drawing on this literature, we propose three basic operators that can be used to aggregate generated ideas by different actors to create further novelty.

During the idea generation phase, given humans’ and AI’s independent search results (e.g., as an example consider a set of multiple points \{(x1,y1),(x2,y2)\} in a 2 dimensional conceptual space), there are three ways to aggregate these ideas—simple concatenation (i.e. presenting the output in sequence one after the other), stacking (i.e., presenting the outputs together to be evaluated and consumed in parallel), and recomposition (i.e., blending elements of different output). For instance, when producing music, one could create a new tune by putting human produced drum beats immediately after AI produced keyboard sound in a sequence (i.e., concatenation) or make the human and AI play the keyboard and drums simultaneously (i.e., stacking).

Alternatively, and more ambitiously, one could take the keyboard and drum beats and blend them seamlessly - where sections of drum and keyboard may play iteratively in sequence and or/parallel, and new sections might involve new sounds that are a hybrid of the sounds created by drums and keyboard (i.e., recomposition). We describe this in more detail.

Recomposition entails first decomposing each creative output into sub-components. Sub-components derived from different outputs can potentially be functional equivalents. For initial ideas \((x1,y1)\) and \((x2,y2)\), if \(x1\) and \(x2\) (and the 2 y coordinates) are perfect functional equivalents, the ideas produce the same fitness. They need not be for our arguments, though and even if they produce comparable fitness, our arguments below go through. Starting with two ideas \({x1, y1}\) and \({x2, y2}\), new points in this space can be identified through convex combinations of the original points.

To illustrate, consider recomposition in poetry. We can begin by defining a common conceptual space in which poems exist, and the components of a poem are its coordinates in this space. Given certain components, we then consider how components of poems produced by multiple poets (one of whom may be an AI) can be pulled apart and shuffled. Components can be the lines in fixed-length poems (e.g., haiku), or themes in free verse poems. Since every haiku
must have three lines, and the second lines in two haikus have a degree of functional equivalence, lines produced in parallel by humans and AI can be swapped. Recomposition thus involves breaking down a current concept into coordinates (its components), take another concept that shares none of the components, and recompose the two into a new object that shares many but not all components of the first. The resulting “remixed” haiku would of course still need to be evaluated for whether it has any aesthetic merit.

More sophisticated ways include, for example, interpolation, which builds new outputs by finding points in the concept space that lie between two original points (e.g., imagine a tricycle as an intermediate between a bicycle and a four-wheeled carriage). To be sure, the resulting output of aggregation may not always be useful or appealing, which is why we need a separate evaluation function for assessing the outcomes of aggregation.

Reference: